Random Graphs and Complex Networks

Remco van der Hofstad

Department of Mathematics and Computer Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven, The Netherlands
rhofstad@win.tue.nl

October 12, 2011

Preface

These lecture notes are intended to be used for master courses, where the students have a limited prior knowledge of special topics in probability. Therefore, we have included many of the preliminaries, such as convergence of random variables, probabilistic bounds, coupling, martingales and branching processes. These notes are aimed to be self-contained, and to give the readers an insight in the history of the field of random graphs.

The field of random graphs was started in 1959-1960 by Erdős and Rényi, see [96, 97, 98, 99]. At first, the study of random graphs was used to prove deterministic properties of graphs. For example, if we can show that a random graph has with a positive probability a certain property, then a graph must exist with this property. The method of proving deterministic statements using probabilistic arguments is called the *probabilistic method*, and goes back a long way. See among others the preface of a standard work in random graphs by Bollobás [44], or the work devoted to it *The Probabilistic Method* [12]. Erdős was one of the first to use this method, see e.g., [94], where it was shown that the Ramsey number $R(k)$ is at least $2^{k/2}$. The Ramsey number $R(k)$ is the value n for which any graph of size at least n is such that either itself or its complement contains a complete graph of size at least k. Erdős in [94] shows that for $n \leq 2^{k/2}$, the fraction of graphs for which the graph or its complement contains a complete graph of size k is bounded by $1/2$, so that there must be graphs of size $n \leq 2^{k/2}$ for which the graph nor its complement contains a complete graph of size k.

The initial work by Erdős and Rényi on random graphs has incited a great amount of work on the field. See the standard references on the subject [44] and [134] for the state of the art. Particularly [97] is a highly impressive paper. In it, a rather complete picture is given of the various phase transitions that occur on the Erdős-Rényi random graph. An interesting quote appearing in [97, Page 2-3] is the following:

> "*It seems to us worthwhile to consider besides graphs also more complex structures from the same point of view, i.e. to investigate the laws governing their evolution in a similar spirit. This may be interesting not only from a purely mathematical point of view. In fact, the evolution of graphs can be seen as a rather simplified model of the evolution of certain communication nets...*"

This was an excellent prediction indeed! Later, interest in random graphs of a different nature arose. Due to the increase of computer power, it has become possible to study so-called *real networks*. Many of these networks turned out to share similar properties, such as the fact that they are *small worlds*, and are *scale-free*, which means that they have degrees obeying *power laws*. The Erdős-Rényi random graph does not obey these properties, and, therefore, new graph models were invented. In fact, already in [97], Erdős and Rényi remark that

> *"Of course, if one aims at describing such a real situation, one should replace the hypothesis of equiprobability of all connection by some more realistic hypothesis."*

See [178] and [5] for two reviews of real networks and their properties to see what 'more realistic' could mean. These other models are also partly covered in the classical works [44] and [134], but up to today, there is no text treating random graphs and random graph models for complex networks in a relatively elementary way. See [88] for the most recent book on random graph, and, particularly, dynamical processes living on them. Durrett covers part of the material in this book, and much more, but the intended audience is

rather different. The goal of these notes is to provide a source for a 'Random graphs' course at master level.

In these notes, we treat both results for the Erdős-Rényi random graph, as well as for the random graph models for complex networks. The aim is to give the simplest possible proofs for classical results, such as the phase transition for the largest connected component in the Erdős-Rényi random graph. Some proofs are more technical and difficult, and the sections containing these proofs will be indicated with a star *. These sections can be omitted without losing the logic in the results. We also try to give many exercises that help the reader to obtain a deeper understanding of the material by working at their solutions.

These notes contain two main parts. In the first part consisting of Chapters 1–8, we give complete and self-contained proofs for some of the main results on random graphs. In the second part consisting of Chapters 9–11, we state further results giving many, but not all, proofs. We sometimes add heuristic arguments for the validity of the results and the logic behind the proof, without giving all details. In the first part, we start by investigating the Erdős-Rényi random graph, and describe some of its key aspects, such as the phase transition for the largest connected component, the connectivity transition, as well as degrees. On the basis of these results, we shall see that the Erdős-Rényi random graph is inappropriate as a model for many real networks, primarily due to the fact that the degrees in the Erdős-Rényi random graph have too thin tails. Therefore, we investigate several other models for real networks, focussing on their degree sequences. In the second part, we investigate related properties of the random graph models, focussing on their connectivity and largest connected components, as well as on the distances in such graphs. The investigation of degree sequences allows us to verify the scale-free nature of the random graph models for complex networks, while the distance results allow us to quantify the small-world behavior in such graphs. These two properties together give insight into what a realistic model for a real network should behave like.

These notes would not have been possible without the help of many people. I thank Gerard Hooghiemstra for the encouragement to write these notes, and for using them at Delft University of Technology almost simultaneously while I used these notes at Eindhoven University of Technology in the spring of 2006 and again in the fall of 2008. Together with Piet Van Mieghem, we entered the world of random graphs in 2001, and I have tremendously enjoyed exploring this field together with you, as well as with Henri van den Esker, Dmitri Znamenski, Mia Deijfen and Shankar Bhamidi. I particularly wish to thank Gerard for many useful comments on these notes, solutions to exercises and suggestions for improvements of the presentation of particularly Chapters 2–5.

I thank Christian Borgs, Jennifer Chayes, Gordon Slade and Joel Spencer for joint work on random graphs which are like the Erdős-Rényi random graph, but do have geometry. This work has deepened my understanding of the basic properties of random graphs, and many of the proofs presented here have been inspired by our work in [51, 52, 53]. Special thanks go to Gordon Slade, who has introduced me to the world of percolation, which is a close neighbor of random graphs (see also [110]). It is peculiar to see that two communities work on two so related topics with quite different methods and even terminology, while it has taken such a long time to build bridges between the subjects.

Further I wish to thank Finbar Bogerd, Mia Deijfen, Michel Dekking, Henri van den Esker, Markus Heydenreich, Martin van Jole, Willemien Kets, Xiaotin Yu for remarks, corrections of typos and ideas that have improved the content and presentation of these notes substantially. Wouter Kager has entirely read the February 2007 version of the notes, giving many ideas for improvements of the arguments and of the methodology. I especially wish to thank Dennis Timmers and Eefje van den Dungen, who, as my student assistants, have been a great help in the development of these notes. Dennis has been responsible for the references, many figures and proofs, as well as first versions of the solutions to many of the exercises. Eefje continued in this direction with great enthusiasm, updated the solutions and created many more of them, and helped in typing out parts of the text. Thanks a lot to both of you!

I have tried to give as many references to the literature as possible. However, the number of papers on random graphs is currently exploding. In MathSciNet, see

`http://www.ams.org/mathscinet/`,

there are, on December 21, 2006, a total of 1,428 papers that contain the phrase 'random graphs' in the review text, on September 29, 2008, this number increased to 1614. These are merely the papers on the topic in the math community. What is special about random graph theory is that it is extremely multidisciplinary, and many papers are currently written in economics, biology, theoretical physics and computer science, using random graph models. For example, in Scopus (see `http://www.scopus.com/scopus/home.url`), again on December 21, 2006, there are 5,403 papers that contain the phrase 'random graph' in the title, abstract or keywords, and on September 29, 2008, this has increased to 7,928. It can be expected that these numbers will increase even faster in the coming period, rendering it impossible to review most of the literature.

Contents

Chapter 1
Introduction

In this first chapter, we give an introduction to random graphs and complex networks. The advent of the computer age has incited an increasing interest in the fundamental properties of real networks. Due to the increased computational power, large data sets can now easily be stored and investigated, and this has had a profound impact in the empirical studies on large networks. A striking conclusion from this empirical work is that many real networks share fascinating features. Many are *small worlds*, in the sense that most vertices are separated by relatively short chains of edges. From an efficiency point of view, this general property could perhaps be expected. More surprisingly, many networks are *scale free*, which means that their degrees are size independent, in the sense that the empirical degree distribution is almost independent of the size of the graph, and the proportion of vertices with degree k is close to proportional to $k^{-\tau}$ for some $\tau > 1$, i.e., many real networks appear to have *power-law degree sequences*. These realisations have had fundamental implications for scientific research on networks. This research is aimed to both understand *why* many networks share these fascinating features, and also what the *properties* of these networks are.

The study of complex networks plays an increasingly important role in science. Examples of such networks are electrical power grids and telephony networks, social relations, the World-Wide Web and Internet, collaboration and citation networks of scientists, etc. The structure of such networks affects their performance. For instance, the topology of social networks affects the spread of information and disease (see e.g., [212]). The rapid evolution in, and the success of, the Internet have incited fundamental research on the topology of networks. See [19] and [220] for expository accounts of the discovery of network properties by Barabási, Watts and co-authors. In [181], you can find some of the original papers on network modeling, as well as on the empirical findings on them.

One main feature of complex networks is that they are *large*. As a result, their complete description is utterly impossible, and researchers, both in the applications and in mathematics, have turned to their *local description*: how many vertices do they have, and by which local rules are vertices connected to one another? These local rules are *probabilistic*, which leads us to consider *random graphs*. The simplest imaginable random graph is the Erdős-Rényi random graph, which arises by taking n vertices, and placing an edge between any pair of distinct vertices with some fixed probability p. We give an introduction to the classical Erdős-Rényi random graph and informally describe the scaling behavior when the size of the graph is large in Section 1.5. As it turns out, the Erdős-Rényi random graph is *not* a good model for a complex network, and in these notes, we shall also study extensions that take the above two key features of real networks into account. These will be introduced and discussed informally in Section 1.6.

1.1 Complex networks

Complex networks have received a tremendous amount of attention in the past decade. In this section, we use the Internet as an example of a real network, and illustrate the properties of real networks using the Internet as a key example. For an artist's impression of the Internet, see Figure 1.1.

Measurements have shown that many real networks share two fundamental properties. The first fundamental network property is the fact that typical distances between vertices are small. This is called the 'small-world' phenomenon (see [219]). For example, in Internet, IP-packets cannot use more than a threshold of physical links, and if distances in the

Figure 1.1: The Internet topology in 2001 taken from `http://www.fractalus.com/steve/stuff/ipmap/`.

Internet would be larger than this threshold, e-mail service would simply break down. Thus, the graph of the Internet has evolved in such a way that typical distances are relatively small, even though the Internet is rather large. For example, as seen in Figure 1.2, the AS count, which is the number of Autonomous Systems (AS) which are traversed by an e-mail data set, is most often bounded by 7. In Figure 1.3, the hopcount, which is the number of routers traversed by an e-mail message between two uniformly chosen routers, is depicted.

The second, maybe more surprising, fundamental property of many real networks is that the number of vertices with degree k falls off as an inverse power of k. This is called a 'power-law degree sequence', and resulting graphs often go under the name 'scale-free graphs', which refers to the fact that the asymptotics of the degree sequence is independent of its size. We refer to [5, 86, 178] and the references therein for an introduction to complex networks and many examples where the above two properties hold. The second fundamental property is visualized in Figure 1.4, where the degree distribution is plotted on log-log scale. Thus, we see a plot of $\log k \mapsto \log N_k$, where N_k is the number of vertices with degree k. When N_k is proportional to an inverse power of k, i.e., when, for some normalizing constant c_n and some exponent τ,

$$N_k \sim c_n k^{-\tau}, \tag{1.1.1}$$

then

$$\log N_k \sim \log c_n - \tau \log k, \tag{1.1.2}$$

so that the plot of $\log k \mapsto \log N_k$ is close to a straight line. Here, and in the remainder of this section, we write $\sim$ to denote an uncontrolled approximation. Also, the power exponent τ can be estimated by the slope of the line, and, for the AS-data, this given as estimate of $\tau \approx 2.15 - 2.20$. Naturally, we must have that

$$\sum_k N_k = n, \tag{1.1.3}$$

so that it is reasonable to assume that $\tau > 1$.

Interestingly, in the AS-count, various different data sets (which focus on different parts of the Internet) show roughly the same picture for the AS-count. This shows that the AS-count is somewhat robust, and it hints at the fact that the AS graph is relatively

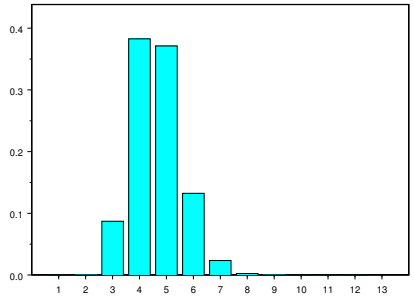

Figure 1.2: Number of AS traversed in hopcount data. Data courtesy of Hongsuda Tangmunarunkit.

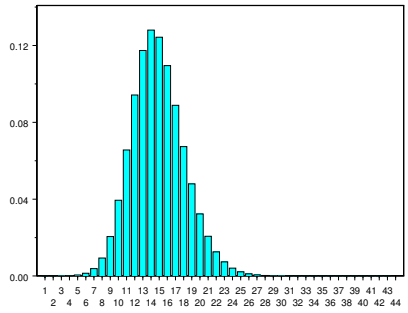

Figure 1.3: Internet hopcount data. Data courtesy of H. Tangmunarunkit.

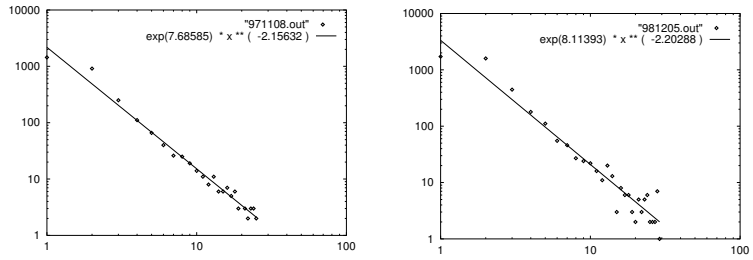

Figure 1.4: Degree sequences AS domains on 11-97 and 12-98 on log-log scale [104]: Power-law degrees with exponent $\approx 2.15 - 2.20$.

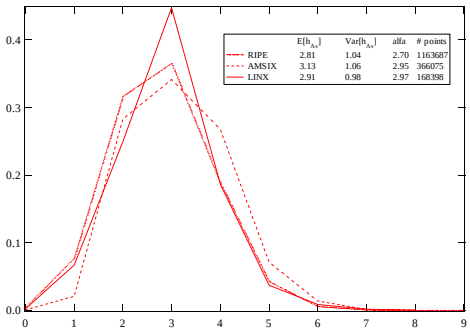

Figure 1.5: Number of AS traversed in various data sets. Data courtesy of Piet Van Mieghem.

homogenous. See also Figure 1.5. For example, the AS-count between AS's in North-America on the one hand, and between AS's in Europe, are quite close to the one of the entire AS. This implies that the dependence on geometry of the AS-count is rather weak, even though one would expect geometry to play a role. As a result, most of the models for the Internet, as well as for the AS graph, ignore geometry altogether.

The observation that many real networks have the above properties have incited a burst of activity in network modeling. Most of the models use random graphs as a way to model the uncertainty and the lack of regularity in real networks. In these notes, we survey some of the proposals for network models. These models can be divided into two distinct types: 'static' models, where we model a graph of a given size as a time snap of a real network, and 'dynamic' models, where we model the growth of the network. Static models aim to describe real networks and their topology at a given time instant, and to share properties with the networks under consideration. Dynamic models aim to *explain* how the networks came to be as they are. Such explanations often focus on the growth of the network as a way to explain the power law degree sequences by means of 'preferential attachment' growth rules, where added vertices and links are more likely to be attached to vertices that already have large degrees.

When we would like to model a power-law relationship between the number of vertices with degree k and k, the question is how to appropriately do so. In Chapters 6, 7 and 8, we discuss a number of models which have been proposed for graphs with a given degree sequence. For this, we let F_X be the distribution function of an integer random variable X, and we denote its probability mass function by $\{f_k\}_{k=1}^{\infty}$, so that

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{k \leq x} f_k. \tag{1.1.4}$$

We wish to obtain a random graph model where N_k, the number of vertices with degree k, is roughly equal to nf_k, where we recall that n is the size of the network. For a power-law relationship as in (1.1.1), we should have that

$$N_k \sim nf_k, \tag{1.1.5}$$

so that

$$f_k \propto k^{-\tau}, \tag{1.1.6}$$

where, to make $f = \{f_k\}_{k=1}^{\infty}$ a probability measure, we take $\tau > 1$, and $\propto$ in (1.1.6) denotes that the left-hand side is proportional to the right-hand side. Now, often (1.1.6)

is too restrictive, and we wish to formulate a power-law relationship in a weaker sense. A different formulation could be to require that

$$1 - F_X(x) = \sum_{k>x} f_k \propto x^{1-\tau}, \tag{1.1.7}$$

for some power-law exponent $\tau > 1$. Indeed, (1.1.7) is strictly weaker than (1.1.6), as indicated in the following exercise:

Exercise 1.1. *Show that when* (1.1.6) *holds with equality, then* (1.1.7) *holds. Find an example where* (1.1.7) *holds in the form that there exists a constant C such that*

$$1 - F_X(x) = Cx^{1-\tau}(1 + o(1)), \tag{1.1.8}$$

but that (1.1.6) *fails.*

An even weaker form of a power-law relation is to require that

$$1 - F_X(x) = L_X(x)x^{1-\tau}, \tag{1.1.9}$$

where the function $x \mapsto L_X(x)$ is a so-called *slowly varying function.* Here, a function $x \mapsto \ell(x)$ is called *slowly varying* when, for all constants $c > 0$,

$$\lim_{x\to\infty} \frac{\ell(cx)}{\ell(x)} = 1. \tag{1.1.10}$$

Exercise 1.2. *Show that $x \mapsto \log x$ and, for $\gamma \in (0,1)$, $x \mapsto e^{(\log x)^\gamma}$ are slowly varying, but that when $\gamma = 1$, $x \mapsto e^{(\log x)^\gamma}$ is not slowly varying.*

The above discussion on real networks has been illustrated by using the Internet as a prime example. We close the discussion by giving references to the literature on the empirical properties of the Internet:

1. Siganos, Faloutsos, Faloutsos and Faloutsos [202] take up where [104] have left, and further study power laws arising in Internet.

2. In [136], Jin and Bestavros summarize various Internet measurements and study how the small-world properties of the AS graph can be obtained from the degree properties and a suitable way of connecting vertices.

3. In [227], Yook, Jeong and Barabási find that the Internet topology depends on geometry, and find that the fractal dimension is equal to $D_f = 1.5$. They continue to propose a model for the Internet growth that predicts this behavior using *preferential attachment* including geometry. We shall discuss this in more detail in Chapter 8.

4. A critical look at the proposed models for the Internet, and particularly the suggestion of preferential attachment in Internet was given by Willinger, Govindan, Paxson and Shenker in [224]. Preferential attachment models shall be described informally in Section 1.1, and are investigated in more detail in Chapters 8 and 11. The authors conclude that the Barabási-Albert model does not model the growth of the AS graph appropriately, particularly since the degrees of the receiving vertices in the AS graph is even larger than for the Barabási-Albert model. This might also explain why the power-law exponent, which is around 2.2 for the AS-graph, is smaller than the power-law exponent in the Barabási-Albert model, which is 3 (see Chapter 8 for this result).

5. An interesting topic of research receiving substantial attention is how the Internet behaves under malicious attacks or random breakdown [74, 75]. The conclusion is that the topology is critical for the vulnerability under intentional attacks. When vertices with high degrees are taken out, then the connectivity properties of random graph models for the Internet cease to have the necessary connectivity properties.

In the remainder of this section, we shall describe a number of other examples of real networks where the small-world phenomenon and the power-law degree sequence phenomenon are observed:

1. 'Six Degrees of Separation' and social networks.
2. Kevin Bacon Game and the movie actor network.
3. Erdős numbers and collaboration networks.
4. The World-Wide Web.

In this section, we shall discuss some of the empirical findings in the above applications, and discuss the key publications on their empirical properties. Needless to say, one could easily write a whole book on each of these examples separately, so we cannot dive into the details too much.

1.1.1 Six degrees of separation and social networks

In 1967, Stanley Milgram performed an interesting experiment. See

`http://www.stanleymilgram.com/milgram.php`

for more background on the psychologist Milgram, whose main topic of study was the obedience of people, for which he used a very controversial 'shock machine'.

In his experiment, Milgram sent 60 letters to various recruits in Wichita, Kansas, U.S.A., who were asked to forward the letter to the wife of a divinity student living at a specified location in Cambridge, Massachusetts. The participants could only pass the letters (by hand) to personal acquaintances who they thought might be able to reach the target, either directly, or via a "friend of a friend". While fifty people responded to the challenge, only three letters (or roughly 5%) eventually reached their destination. In later experiments, Milgram managed to increase the success rate to 35% and even 95%, by pretending that the value of the package was high, and by adding more clues about the recipient, such as his/her occupation. See [167, 215] for more details.

The main conclusion from the work of Milgram was that most people in the world are connected by a chain of at most 6 "friends of friends", and this phrase was dubbed "Six Degrees of Separation". The idea was first proposed in 1929 by the Hungarian writer Frigyes Karinthy in a short story called 'Chains' [139], see also [181] where a translation of the story is reproduced. Playwright John Guare popularized the phrase when he chose it as the title for his 1990 play. In it, Ousa, one of the main characters says:

> "*Everybody on this planet is separated only by six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United states. A gondolier in Venice... It's not just the big names. It's anyone. A native in the rain forest. (...) An Eskimo. I am bound to everyone on this planet by a trail of six people. It is a profound thought.*".

The fact that any number of people can be reached by a chain of at most 6 intermediaries is rather striking. It would imply that two people in as remote areas as Greenland and the

Amazone could be linked by a sequence of at most 6 "friends of friends". This makes the phrase "It's a small world!" very appropriate indeed! Another key reference in the small-world work in social sciences is the paper by Pool and Kochen [194], which was written in 1958, and has been circulating around the social sciences ever since, before it was finally published in 1978.

The idea of Milgram was taken up afresh in 2001, with the added possibilities of the computer era. In 2001, Duncan Watts, a professor at Columbia University, recreated Milgram's experiment using an e-mail message as the"package" that needed to be delivered. Surprisingly, after reviewing the data collected by 48,000 senders and 19 targets in 157 different countries, Watts found that again the average number of intermediaries was six. Watts' research, and the advent of the computer age, has opened up new areas of inquiry related to six degrees of separation in diverse areas of network theory such as power grid analysis, disease transmission, graph theory, corporate communication, and computer circuitry. See the web site

`http://smallworld.columbia.edu/project.html`

for more information on the Small-World Project conducted by Watts. See [219] for a popular account of the small-world phenomenon. Related examples of the small-world phenomenon can be found in [5] and [178].

To put the idea of a small-world into a network language, we define the vertices of the social graph to be the inhabitants of the world (so that $n \approx 6$ billion), and we draw an edge between two people when they know each other. Needless to say, we should make it more precise what it means to "know each other". Possibilities here are various. We could mean that the two people involved have shaken hands at some point, or that they know each other on a first name basis.

One of the main difficulties of social networks is that they are notoriously hard to measure. Indeed, questionaires can not be trusted easily, since people have a different idea what a certain social relation is. Also, questionaires are quite physical, and they take time to collect. As a result, researchers are quite interested in examples of social networks that *can* be measured, for example due to the fact that they are *electronic*. Examples are e-mail networks or social networks such as Hyves. Below, I shall give a number of references to the literature for studies of social networks.

1. Amaral, Scala, Bartélémy and Stanley [13] calculated degree distributions of several networks, among others a friendship network of 417 junior high school students and a social network of friendships between Mormons in Provo, Utah. For these examples, the degree distributions turn out to be closer to a normal distribution than to a power law.

2. In [92], Ebel, Mielsch and Bornholdt investigate the topology of an e-mail network of an e-mail server at the University of Kiel over a period of 112 days. The authors conclude that the degree sequence obeys a power law, with an exponential cut-off for degrees larger than 200. The estimated degree exponent is 1.81. The authors note that since this data set is gathered at a server, the observed degree of the *external* vertices is an underestimation of their true degree. When only the *internal* vertices are taken into account, the estimate for the power-law exponent decreases to 1.32. When taking into account that the network is in fact *directed*, the power-law exponent of the in-degree is estimated at 1.49, while the out-degrees have an exponent of 2.03. The reported errors in the estimation of the exponents are between 0.10 and 0.18.

3. There are many references to the social science literature on social networks in the book by Watts [220], who now has a position in social sciences. In [180], Newman,

Kevin Bacon Number	# of actors
0	1
1	1902
2	160463
3	457231
4	111310
5	8168
6	810
7	81
8	14

Table 1.1: Kevin Bacon Numbers.

Watts and Strogatz survey various models for social networks that have appeared in their papers. Many of the original references can also be found in the collection in [181], along with an introduction explaining their relevance.

4. Liljeros, Edling, Amaral and Stanley [158] investigated sexual networks in Sweden, where two people are connected when they have had a sexual relation in the previous year, finding that the degree distributions of males and females obey power laws, with estimated exponents of $\tau_{\text{fem}} \approx 2.5$ and $\tau_{\text{mal}} \approx 2.3$. When extending to the entire lifetime of the Swedish population, the estimated exponents decrease to $\tau_{\text{fem}} \approx 2.1$ and $\tau_{\text{mal}} \approx 1.6$. The latter only holds in the range between 20 and 400 contacts, after which it is truncated. Clearly, this has important implications for the transmittal of sexual diseases.

1.1.2 Kevin Bacon Game and movie actor network

A second example of a large network in the movie actor network. In this example, the vertices are movie actors, and two actors share an edge when they have played in the same movie. This network has attracted some attention in connection to Kevin Bacon, who appears to be reasonably central in this network. The Computer Science Department at Virginia University has an interesting web site on this example, see The Oracle of Bacon at Virginia web site on

`http://www.cs.virginia.edu/oracle/`.

See Table 1.1 for a table of the Kevin Bacon Numbers of all the actors in this network. Thus, there is one actor at distance 0 from Kevin Bacon (namely, Kevin Bacon himself), 1902 actors have played in a movie starring Kevin Bacon, and 160463 actors have played in a movie in which another movie star played who had played in a movie starring Kevin Bacon. In total, the number of linkable actors is equal to 739980, and the Average Kevin Bacon number is 2.954. In search for "Six Degrees of Separation", one could say that most pairs of actors are related by a chain of co-actors of length at most 6.

It turns out that Kevin Bacon is not the most central vertex in the graph. A more central actor is Sean Connery. See See Table 1.2 for a table of the Sean Connery Numbers. By computing the average of these numbers we see that the average Connery Number is about 2.731, so that Connery a better center than Bacon. Mr. Bacon himself is the 1049th best center out of nearly 800,000 movie actors, which makes Bacon a better center than 99% of the people who have ever appeared in a feature film.

On the web site `http://www.cs.virginia.edu/oracle/`, one can also try out one's own favorite actors to see what Bacon number they have, or what the distance is between them.

We now list further studies of the movie actor network.

Sean Connery Number	# of actors
0	1
1	2272
2	218560
3	380721
4	40263
5	3537
6	535
7	66
8	2

Table 1.2: Sean Connery Numbers

1. Watts and Strogatz [221] investigate the small-world nature of the movie-actor network, finding that it has more clustering and shorter distances than a random graph with equal edge density. Amaral et al. looked closer at the degree distribution to conclude that the power-law in fact has an exponential cut-off.

2. Albert and Barabási [20] use the movie actor network as a prime example of a network showing power-law degrees. The estimated power-law exponent is 2.3.

1.1.3 Erdős numbers and collaboration networks

A further example of a complex network that has drawn substantial attention is the collaboration graph in mathematics. This is popularized under the name "Erdős number project". In this network, the vertices are mathematicians, and there is an edge between two mathematicians when they have co-authored a paper. See

`http://www.ams.org/msnmain/cgd/index.html`

for more information. The Erdős number of a mathematician is how many papers that mathematician is away from the legendary mathematician Paul Erdős, who was extremely prolific with around 1500 papers and 509 collaborators. Of those that are connected by a trail of collaborators to Erdős, the maximal Erdős number is claimed to be 15.

On the above web site, one can see how far one's own professors are from Erdős. Also, it is possible to see the distance between any two mathematicians.

The Erdős numbers has also attracted attention in the literature. In [80, 81], the authors investigate the Erdős numbers of Nobel prize laureates, as well as Fields medal winners, to come to the conclusion that Nobel prize laureates have Erdős numbers of at most 8 and averaging 4-5, while Fields medal winners have Erdős numbers of at most 5 and averaging 3-4. See also

`http://www.oakland.edu/enp`

for more information on the web, where we also found the following summary of the collaboration graph. This summary dates back to July, 2004. An update is expected somewhere in 2006.

In July, 2004, the collaboration graph consisted of about 1.9 million authored papers in the Math Reviews database, by a total of about 401,000 different authors. Approximately 62.4% of these items are by a single author, 27.4% by two authors, 8.0% by three authors, 1.7% by four authors, 0.4% by five authors, and 0.1% by six or more authors. The largest number of authors shown for a single item is in the 20s. Sometimes the author list includes "et al." so that in fact, the number of co-authors is not always precisely known.

Erdős Number	# of Mathematicians
0	1
1	504
2	6593
3	33605
4	83642
5	87760
6	40014
7	11591
8	3146
9	819
10	244
11	68
12	23
13	5

Table 1.3: Erdős Numbers

The fraction of items authored by just one person has steadily decreased over time, starting out above 90% in the 1940s and currently standing at under 50%. The entire graph has about 676,000 edges, so that the average number of collaborators per person is 3.36. In the collaboration graph, there is one large component consisting of about 268,000 vertices. Of the remaining 133,000 authors, 84,000 of them have written no joint papers, and these authors correspond to isolated vertices. The average number of collaborators for people who have collaborated is 4.25. The average number of collaborators for people in the large component is 4.73. Finally, the average number of collaborators for people who have collaborated but are not in the large component is 1.65. There are only 5 mathematicians with degree at least 200, the largest degree is for Erdős, who has 509 co-authors. The diameter of the largest connected component is 23.

The clustering coefficient of a graph is equal to the fraction of ordered triples of vertices a, b, c in which edges ab and bc are present that have edge ac present. In other words, the clustering coefficient describes how often are two neighbors of a vertex adjacent to each other. The clustering coefficient of the collaboration graph of the first kind is $1308045/9125801 = 0.14$. The high value of this figure, together with the fact that average path lengths are small, indicates that this graph is a *small world* graph.

For the Erdős numbers, we refer to Table 1.3. The median Erdős number is 5, the mean is 4.65, and the standard deviation is 1.21. We note that the Erdős number is finite if and only if the corresponding mathematician is in the largest connected component of the collaboration graph.

See Figure 1.6 for an artistic impression of the collaboration graph in mathematics taken from

`http://www.orgnet.com/Erdos.html`

and Figure 1.7 for the degree sequence in the collaboration graph.

We close this section by listing interesting papers on collaboration graphs.

1. In [24], Batagelj and Mrvar use techniques for the analysis of large networks, such as techniques to identify interesting subgroups and hierarchical clustering techniques, to visualize further aspects of the Erdős collaboration graph.

2. Newman has studied several collaboration graphs in a sequence of papers that we shall discuss now. In [177], he finds that several of these data bases are such that

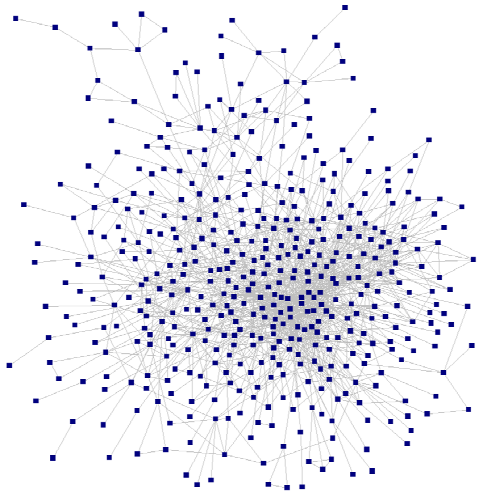

Figure 1.6: An artist impression of the collaboration graph in mathematics.

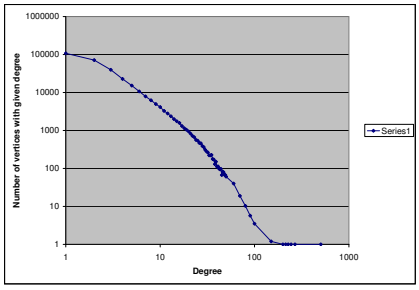

Figure 1.7: The degree sequence in the collaboration graph.

the degrees have power-laws with exponential cut-offs. The data bases are various arXiv data bases in mathematics and theoretical physics, the MEDLINE data base in medicine, and the ones in high-energy physics and theoretical computer science. Also, the average distance between scientists is shown to be rather small, which is a sign of the small-world nature of these networks. Finally, the average distance is compared to $\log n / \log z$, where n is the size of the collaboration graph and z is the average degree. The fit shows that these are quite close. Further results are given in [176].

3. In Barabási et al. [22], the evolution of scientific collaboration graphs is investigated. The main conclusion is that scientists are more likely to write papers with other scientists who have written many papers, i.e., there is a tendency to write papers with others who have already written many. This preferential attachment is shown to be a possible explanation for the existence of power laws in collaboration networks (see Chapter 8).

1.1.4 The World-Wide Web

A final example of a complex network that has attracted enormous attention is the World-Wide Web (WWW). The elements of the WWW are web pages, and there is a directed connection between two web pages when the first links to the second. Thus, while the WWW is virtual, the Internet is physical. With the world becoming ever more virtual, and the WWW growing at tremendous speed, the study of properties of the WWW has grown as well. It is of great practical importance to know what the structure of the WWW is, for example, in order for search engines to be able to explore it. A notorious, but rather interesting, problem is the Page-Rank problem, which is the problem to rank web pages on related topics such that the most important pages come first. Page-Rank is claimed to be the main reason of the success of Google, and the inventors of Page-Rank were also the founders of Google (see [57] for the original reference).

In [6], the authors Albert, Jeong and Barabási study the degree distribution to find that the in-degrees obey a power law with exponent $\tau_{\text{in}} \approx 2.1$ and the out-degrees obey a power law with exponent $\tau_{\text{out}} \approx 2.45$, on the basis of several Web domains, such as `nd.edu`, `mit.edu` and `whitehouse.gov`, respectively the Web domain of the home university of Barabási at Notre Dame, the Web domain of MIT and of the White House. Further, they investigated the distances between vertices in these domains, to find that distances within domains grown linearly with the log of the size of the domains, with an estimated dependence of $d = 0.35 + 2.06 \log n$, where d is the average distance between elements in the part of the WWW under consideration, and n is the size of the subset of the WWW. Extrapolating this relation to the estimated size of the WWW at the time, $n = 8 \cdot 10^8$, Albert, Jeong and Barabási [6] concluded that the diameter of the WWW was 19 at the time, which prompted the authors to the following quote:

> *"Fortunately, the surprisingly small diameter of the web means that all information is just a few clicks away."*

In [154], it was first observed that the WWW also has power-law degree sequences. In fact, the WWW is a *directed* graph, and in [154] it was shown that the in-degree follows a power-law with power-law exponent quite close to 2. See also Figure 1.8.

The most substantial analysis of the WWW was performed by Broder et al. [59], following up on earlier work in [154, 153] in which the authors divide the WWW into several distinct parts. See Figure 1.9 for the details. The division is roughly into four parts:

(a) The central core or Strongly Connected Component (SCC), consisting of those web pages that can reach each other along the directed links (28% of the pages);

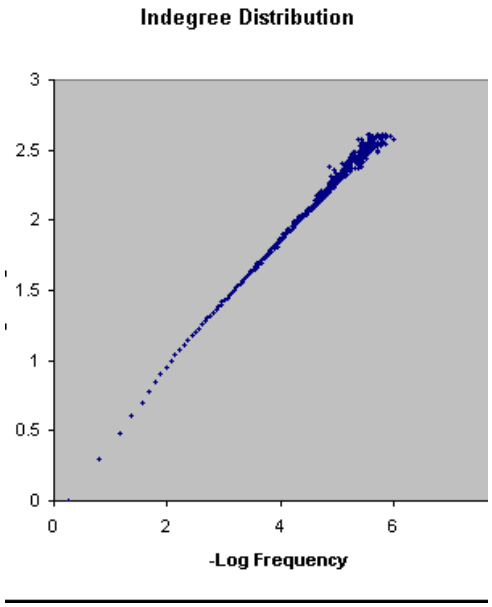

Figure 1.8: The in-degree sequence in the WWW taken from [154].

(b) The IN part, consisting of pages that can reach the SCC, but cannot be reached from it (21% of the pages);

(c) The OUT part, consisting of pages that can be reached from the SCC, but do not link back into it (21% of the pages);

(d) The TENDRILS and other components, consisting of pages that can neither reach the SCC, nor be reached from it (30% of the pages).

Broder et al. [59] also investigate the diameter of the WWW, finding that the SCC has diameter at least 28, but the WWW as a whole has diameter at least 500. This is partly due to the fact that the graph is *directed*. When the WWW is considered to be an *undirected* graph, the average distance between vertices decreases to around 7. Further, it was shown that both the in- and out-degrees in the WWW follow a power-law, with power-law exponents estimated as $\tau_{\text{in}} \approx 2.1, \tau_{\text{out}} \approx 2.5$.

In [2], distances in the WWW are discussed even further. When considering the WWW as a directed graph, it is seen that the distances between most pairs of vertices within the SCC is quite small. See Figure 1.10 for a histogram of pairwise distances in the sample. Distances between pairs of vertices in the SCC tend to be at most 7: Six Degrees of Separation.

We close this section by discussing further literature on the WWW:

1. In [20], it is argued that new web pages are more likely to attach to web pages that already have a high degree, giving a bias towards popular web pages. This is proposed as an explanation for the occurrences of power laws. We shall expand this explanation in Section 1.6, and make the discussion rigorous in Chapter 8.

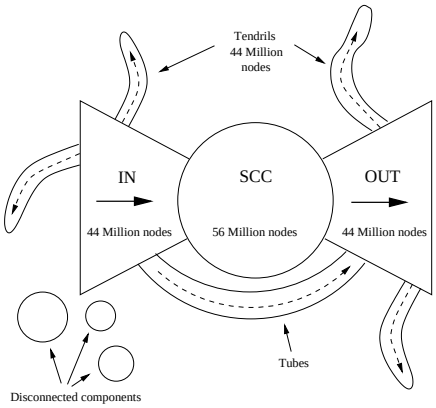

Figure 1.9: The WWW according to Broder et al [59].

2. In [153], models for the WWW are introduced, by adding vertices which copy the links of older vertices in the graph. This is called an *evolving copying model*. In some cases, depending on the precise copying rules, the model is shown to lead to power-law degree sequences. The paper [149] is a nice survey of measurements, models and methods for obtaining information on the WWW, by analyzing how Web crawling works.

3. Barábasi, Albert and Jeong [21] investigate the scale-free nature of the WWW, and propose a preferential attachment model for it. In the proposed model for the WWW in [20, 21], older vertices tend to have the highest degrees. On the WWW this is not necessarily the case, as Adamic and Huberman [3] demonstrate. For example, Google is a late arrival on the WWW, but has yet managed to become one of the most popular web sites. A possible fix for this problem is given in [36] through a notion of *fitness* of vertices, which enhance or decrease their preferential power.

4. The works by Kleinberg [146, 147, 148] investigate the WWW and other networks from a computer science point of view. In [146, 147], the problem is addressed how hard it is to find short paths in small-world networks on the d-dimensional lattice, finding that navigation sensitively depends upon how likely it is for large edges to be present. Indeed, the delivery time of any local algorithm can be bounded below by a positive power of the width of the box, except for one special value of the parameters, in which case it is of the order of the square of the log of the width of the box. Naturally, this has important implications for the WWW, even though the WWW may depend less sensitively on geometry. In Milgram's work discussed in Section 1.1.1, on the one hand, it is striking that there exist short paths between most pairs of individuals, but, on the other hand, it may be even more striking that people actually succeed in *finding* them. In [146], the problem is addressed how "authoritative sources" for the search on the WWW can be quantified. These authoritative sources can be found in an algorithmic way by relating them to the hubs in the network. Clearly, this problem is intimately connected to the Page-Rank problem.

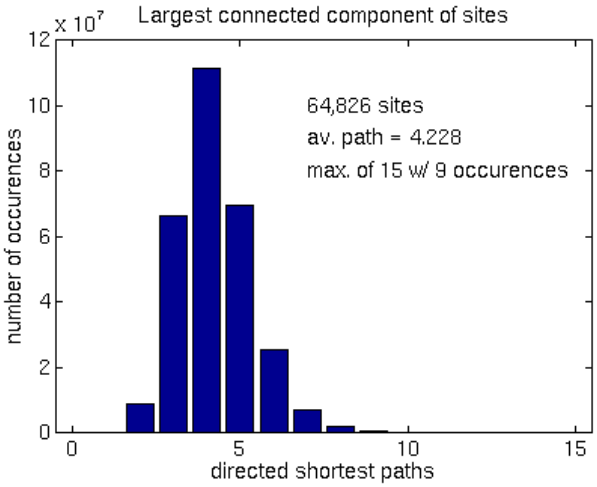

Figure 1.10: Average distances in the Strongly Connected Component of the WWW taken from [2].

1.2 Scale-free, small-world and highly-clustered random graph processes

As described in Section 1.1, many real-world complex networks are large. They share similar features, in the sense that they have a relatively low degree compared to the maximal degree $n-1$ in a graph of size n, i.e., they are 'sparse'. Further, many real networks are 'small worlds', 'scale free' and 'highly clustered'. These notions are empirical, and, hence, inherently not very mathematically precise. In this section, we describe what it means for a model of a real network to satisfy these properties.

Many of real-world networks as considered in Section 1.1, such as the World-Wide Web and collaboration networks, grow in size as time proceeds. Therefore, it is reasonable to consider graphs of growing size, and to define the notions of scale-free, small-world and highly-clustered random graphs as a limiting statement when the size of the random graphs tend to infinity. This naturally leads us to study graph sequences. In this section, we shall denote a graph sequence by $\{G_n\}_{n=1}^{\infty}$, where n denotes the size of the graph G_n, i.e., the number of vertices in G_n.

Denote the proportion of vertices with degree k in G_n by $P_k^{(n)}$, i.e.,

$$P_k^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{D_i^{(n)}=k\}}, \tag{1.2.1}$$

where $D_i^{(n)}$ denotes the degree of vertex $i \in \{1,\ldots,n\}$ in the graph G_n, and recall that the degree sequence of G_n is given by $\{P_k^{(n)}\}_{k=0}^{\infty}$. We use capital letters in our notation to indicate that we are dealing with *random variables*, due to the fact that G_n is a random graph. Now we are ready to define what it means for a random graph process $\{G_n\}_{n=1}^{\infty}$ to be scale free.

We first call a random graph process $\{G_n\}_{n=1}^{\infty}$ *sparse* when

$$\lim_{n\to\infty} P_k^{(n)} = p_k, \tag{1.2.2}$$

for some *deterministic* limiting probability distribution $\{p_k\}_{k=0}^{\infty}$. Since the limit p_k in (1.2.2) is deterministic, the convergence in (1.2.2) can be taken as convergence in probability or in distribution. Also, since $\{p_k\}_{k=0}^{\infty}$ sums up to one, for large n, most of the vertices have a bounded degree, which explains the phrase sparse random graphs.

We further call a random graph process $\{G_n\}_{n=1}^{\infty}$ *scale free with exponent* τ when it is sparse and when

$$\lim_{k\to\infty} \frac{\log p_k}{\log(1/k)} = \tau \tag{1.2.3}$$

exists. Thus, for a scale-free random graph process its degree sequence converges to a limiting probability distribution as in (1.2.2), and the limiting distribution has asymptotic power-law tails described in (1.2.3). This gives a precise mathematical meaning to a random graph process being scale free. In some cases, the definition in (1.2.3) is a bit too restrictive, particularly when the probability mass function $k \mapsto p_k$ is not very smooth. Instead, we can also replace it by

$$\lim_{k\to\infty} \frac{\log[1-F(k)]}{\log(1/k)} = \tau - 1, \tag{1.2.4}$$

where $F(x) = \sum_{y\le x} p_y$ denotes the distribution function corresponding to the probability mass function $\{p_k\}_{k=0}^{\infty}$. In particular models below, we shall use the version that is most appropriate in the setting under consideration. See Section 1.3 below for a more extensive discussion of power laws.

We say that a graph process $\{G_n\}_{n=1}^{\infty}$ is *highly clustered* when

$$\lim_{n\to\infty} C_{\mathbb{G}_n} = C_{\mathbb{G}_\infty} > 0. \tag{1.2.5}$$

We finally define what it means for a graph process $\{G_n\}_{n=1}^{\infty}$ to be a small world. Intuitively, a small world should have distances that are much smaller than those in a lattice or torus. When we consider the nearest-neighbor torus $\mathbb{T}_{r,d}$, then, and when we draw two uniform vertices at random, their distance will be of the order r. Denote the size of the torus by $n = (2r+1)^d$, then the typical distance between two uniform vertices is of the order $n^{1/d}$, so that it grows as a positive power of n.

Let H_n denote the distance between two uniformly chosen *connected* vertices, i.e., we pick a pair of vertices uniformly at random from all pairs of connected vertices, and we let H_n denote the graph distance between these two vertices. Here we use the term graph distance between the vertices v_1, v_2 to denote the minimal number of edges in the graph on a path connecting v_1 and v_2. Below, we shall be dealing with random graph processes $\{G_n\}_{n=1}^{\infty}$ for which G_n is not necessarily connected, which explains why we condition on the two vertices being connected.

We shall call H_n the *typical distance of* G_n. Then, we say that a random graph process $\{G_n\}_{n=1}^{\infty}$ is a *small world* when there exists a constant K such that

$$\lim_{n\to\infty} \mathbb{P}(H_n \le K\log n) = 1. \tag{1.2.6}$$

Note that, for a graph with a *bounded* degree $d_{\max}$, the typical distance is *at least* $(1-\varepsilon)\log n/\log d_{\max}$ with high probability, so that a random graph process with bounded degree is a small world *precisely* when the order of the typical distance is optimal.

For a graph G_n, let $\mathrm{diam}(G_n)$ denote the *diameter* of G_n, i.e., the maximal graph distance between any pair of connected vertices. Then, we could also have chosen to replace H_n in (1.2.6) by $\mathrm{diam}(G_n)$. However, the diameter of a graph is a rather sensitive object which can easily be changed by making small changes to a graph in such a way that the scale-free nature and the typical distance H_n do not change. For example, by adding a sequence of m vertices in a line, which are not connected to any other vertex, the diameter

of the graph becomes at least m, whereas, if m is much smaller than n, H_n is not changed very much. This explain why we have a preference to work with the typical distance H_n rather than with the diameter $\mathrm{diam}(G_n)$.

In some models, we shall see that typical distances can be even much smaller than $\log n$, and this is sometimes called an ultra-small world. More precisely, we say that a random graph process $\{G_n\}_{n=1}^{\infty}$ is an *ultra-small world* when there exists a constant K such that

$$\lim_{n\to\infty} \mathbb{P}(H_n \leq K \log\log n) = 1. \tag{1.2.7}$$

There are many models for which (1.2.7) is satisfied, but $\mathrm{diam}(G_n)/\log n$ converges in probability to a positive limit. This once more explain our preference for the typical graph distance H_n.

We have given precise mathematical definitions for the notions of random graphs being highly clustered, small worlds and scale free. This has not been done in the literature so far, and our definitions are based upon a summary of the relevant results proved for random graph models. We believe it to be a good step forward to make the connection between the theory of random graphs and the empirical findings on real-life networks.

1.3 Tales of tails

In this section, we discuss the occurrence of power laws. In Section 1.3.1, we discuss the literature on this topic, which dates back to the twenties of the last century. In Section 1.3.2, we describe the new view points on power laws in real networks.

1.3.1 Old tales of tails

Mathematicians are always drawn to simple relations, believing that they explain the rules that gave rise to them. Thus, finding such relations uncovers the hidden structure behind the often chaotic appearance. A power-law relationship is such a simple relation. We say that there is a power-law relationship between two variables when one is proportional to a power of the other. Or, in more mathematical language, the variable k and the characteristic $f(k)$ are in a power-law relation when $f(k)$ is proportional to a power of k, that is, for some number τ,

$$f(k) = Ck^{-\tau}. \tag{1.3.1}$$

Power laws are intimately connected to so-called 80/20 rules. For example, when studying the wealth in populations, already Pareto observed a huge variability [187]. Most individuals do not earn so much, but there are these rare individuals that earn a substantial part of the total income. Pareto's principle was best known under the name '80/20 rule', indicating, for example, that 20 percent of the people earn 80 percent of the total income. This law appears to be true much more generally. For example, 20 percent of the people own 80 percent of the land, 20 percent of the employees earn 80 percent of the profit of large companies, and maybe even 20 percent of the scientists write 80 percent of the papers. In each of these cases, no typical size exists due to the high variability present, which explains why these properties are called 'scale-free'.

Intuitively, when a 80/20 rule holds, a power law must be hidden in the background! Power laws play a crucial role in mathematics, as well as in many applications. Power laws have a long history. Zipf [229] was one of the first to find one in the study of the frequencies of occurrence of words in large pieces of text. He found that the relative frequency of words is roughly inversely proportional to its rank in the frequency table of all words. Thus, the most frequent word is about twice as frequent as the second most frequent word, and about three times as frequent as the third most frequent word, etc. In short, with k the rank of the word and $f(k)$ the relative frequency of k^{th} most frequent word, $f(k) \propto k^{-\tau}$ where τ is close to 1. This is called *Zipf's law*.

Already in the twenties, several other examples of power laws were found. Lotka [160] investigated papers that were referred to in the Chemical Abstracts in the period from 1901-1916, and found that the number of scientists appearing with 2 entries is close to $1/2^2 = 1/4$ of the number of scientists with just one entry. The number of scientists appearing with 3 entries is close to $1/3^2 = 1/9$ times the number of scientists appearing with 1 entry, etc. Again, with $f(k)$ denoting the number of scientists appearing in k entries, $f(k) \propto k^{-\tau}$, where τ now is close to 2. This is dubbed *Lotka's Law*. Recently, effort has been put into explaining power-laws using 'self-organization'. Per Bak, one of the central figures in this area, called his book on the topic "How nature works" [18].

Power-law relations are one-step extensions of linear relations. Conveniently, even when one does not understand the mathematical definition of a power law too well, one can still *observe* them in a simple way: in a log-log plot, power laws are turned into straight lines! Indeed, taking the log of the power-law relationship (1.3.1) yields

$$\log f(k) = \log C - \tau \log k, \tag{1.3.2}$$

so that $\log f(k)$ is in a linear relationship with $\log k$, with coefficient equal to $-\tau$. Thus, not only does this allow us to visually inspect whether $f(k)$ is in a power-law relationship to k, it also allows us to estimate the exponent τ! Naturally, this is precisely what has been done in order to obtain the power-law exponents in the examples in Section 1.1. An interesting account of the history of power-laws can be found in [168], where possible explanations why power laws arise so frequently are also discussed.

1.3.2 New tales of tails

In this section, we discuss the occurrence of power-law degree sequences in real networks. We start by giving a heuristic explanation for the occurrence of power law degree sequences, in the setting of exponentially growing graphs. This heuristic is based on some assumptions that we formulate now.

We assume that

(1) the number of vertices $V(t)$ is growing *exponentially* at some rate $\rho > 0$, i.e., $V(t) \approx e^{\rho t}$;

(2) the number $N(t)$ of links *into* a vertex at some time t after its creation is $N(t) \approx e^{\beta t}$. (Note that we then must have that $\beta \leq \rho$, since the number of links into a vertex must be bounded above by the number of vertices.) Thus, also the number of links into a vertex grows exponentially with time.

We note that assumption (1) is equivalent to the assumption that

(1') the lifetime T since its creation of a random vertex has law

$$\mathbb{P}(T > t) = e^{-\rho t}, \tag{1.3.3}$$

so that the density of the lifetime of a random vertex is equal to

$$f_T(t) = \rho e^{-\rho t}. \tag{1.3.4}$$

Then, using the above assumptions, the number of links *into* a random vertex X equals

$$\mathbb{P}(X > i) \approx i^{-\rho/\beta}, \tag{1.3.5}$$

since it is equal to

$$\begin{aligned}\mathbb{P}(X > i) &= \int_0^\infty f_T(t)\mathbb{P}(X > i|T = t)dt \\ &= \int_0^\infty \rho e^{-t\rho}\mathbb{P}(X > i|T = t)dt \\ &= \rho\int_0^\infty e^{-t\rho}\mathbb{1}_{\{e^{\beta t}>i\}}dt \\ &\sim \rho\int_{(\log i)/\beta}^\infty e^{-t\rho}dt \sim e^{-(\log i)\rho/\beta} \sim i^{-\rho/\beta},\end{aligned}$$

where $\mathbb{1}_{\mathcal{E}}$ denotes the indicator of the event $\mathcal{E}$. Stretching the above heuristic a bit further yields

$$\mathbb{P}(X = i) = \mathbb{P}(X > i - 1) - \mathbb{P}(X > i) \sim i^{-(\rho/\beta+1)}. \tag{1.3.6}$$

This heuristic suggests a power law for the in-degrees of the graph, with power-law exponent $\tau = \rho/\beta + 1 \geq 2$. Peculiarly, this heuristic does not only explain the occurrence of power laws, but even of power laws with exponents that are at least 2.

The above heuristic only explains why the *in-degree* of a vertex has a power law. An alternative reason why power laws occur so generally will be given in Chapter 8. Interestingly, so far, also in this explanation only power laws that are at least 2 are obtained.

While power-law degree sequences are claimed to occur quite generally in real networks, there are also some critical observations, particularly about he measurements that produce power laws in Internet. In [155], it is argued that `traceroute`-measurements, by which the Internet-topology is uncovered, could be partially responsible for the fact that power-law degree sequences are observed in Internet. Indeed, it was shown that applying similar methods as `traceroute`-measurements to certain subgraphs of the Erdős-Rényi random graph exhibit power-law degree sequences. Clearly, the Erdős-Rényi random graph does *not* have power-law degree sequences, so that this observation is an *artefact* of the way the measurements are performed. The point is that in Internet measurements, subgraphs are typically obtained by exploring the paths between sets of pairs of vertices. Indeed, we obtain a subgraph of the Internet by only taking that part of the network that appear along a path between the various starting points and destinations, and this is the way how `traceroute` is used in Internet. Assuming that paths are all *shortest-paths*, i.e., there is shortest-path routing, vertices with a high degree are far more likely to appear in one of the shortest paths between our initial set of pairs of vertices. Therefore, such data sets tend to *overestimate* the degrees in the complete network. This bias in `traceroute` data was further studied in [1, 73], in which both for Erdős-Rényi random graphs and for random regular graphs, it was shown that subgraphs appear to obey a power-law.

While the above criticism may be serious for the Internet, and possibly for the World-Wide Web, where degree distributions are investigated using web-crawling, there are many networks which are completely available that also show power-law degree sequences. When the network is completely described, the observed power-laws can not be so easily dismissed. However, one needs to be careful in using and analyzing data confirming power-law degree sequences. Particularly, it could be that many estimates of the power-law degree exponent τ are *biased*, and that the true values of τ are substantially larger. Possibly, this criticism may give an argument why so often power laws are observed with exponents in the interval $(2, 3)$.

1.4 Notation

In these notes, we frequently make use of certain notation, and we strive to be as consistent as possible. We shall denote events by calligraphic letters, such as $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and

$\mathcal{E}$. We shall use $\mathbb{1}_{\mathcal{E}}$ to denote the indicator function of the event $\mathcal{E}$. We shall use capital letters, such as X, Y, Z, U, V, W, to denote random variables. There are some exceptions, for example, F_X and M_X denote the distribution function and moment generating function of a random variable X, and we emphasize this by writing the subscript X explicitly. We say that a sequence of events $\{\mathcal{E}_n\}_{n=0}^{\infty}$ occurs *with high probability* when $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1$. We often abbreviate this as **whp**. We call a sequence of random variables $\{X_i\}_{i=1}^{n}$ i.i.d. when they are independent, and X_i has the same distribution as X_1 for every $i = 2, \ldots, n$.

We shall use special notion for certain random variables, and write $X \sim \mathrm{Be}(p)$ when X has a Bernoulli distribution with success probability p, i.e., $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$. We write $X \sim \mathrm{Bin}(n, p)$ when the random variable X has a binomial distribution with parameters n and p, and we write $X \sim \mathrm{Poi}(\lambda)$ when X has a Poisson distribution with parameter λ.

Furthermore, we write $f(n) = o(g(n))$ as $n \to \infty$ when $g(n) > 0$ and $\lim_{n\to\infty} |f(n)|/g(n) = 0$. We write $f(n) = O(g(n))$ as $n \to \infty$ when $g(n) > 0$ and $\limsup_{n\to\infty} |f(n)|/g(n) < \infty$. Finally, we write $f(n) = \Theta(g(n))$ as $n \to \infty$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

1.5 The Erdős-Rényi random graph: introduction of the model

In the previous sections, we have described properties of real networks. These networks are quite large, and in most cases, it is utterly impossible to describe them explicitly. To circumvent this problem, random graph models have been considered as network models. These random graphs describe by which *local* and *probabilistic* rules vertices are connected to one another. The use of probabilistic rules is to be able to describe the *complexity* of the networks. In deterministic models, often too much structure is present, making the arising networks unsuitable to describe real networks. This approach introduces *randomness* in network theory, and leads us to consider random graphs as network models. However, it does not tell us *what* these random graph models should look like.

The field of random graphs was established in the late fifties and early sixties of the last century. While there were a few papers appearing around (and even before) that time, one paper is generally considered to have founded the field [97]. The authors Erdős and Rényi studied the simplest imaginable random graph, which is now named after them. Their graph has n elements, and each pair of elements is independently connected with a fixed probability. When we think of this graph as describing a social network, then the elements denote the individuals, while two individuals are connected when they know one another. The probability for elements to be connected is sometimes called the *edge probability*. Let $\mathrm{ER}_n(p)$ denote the resulting random graph. This random graph is named after its inventors Erdős and Rényi who introduced a version of it in [97] in 1960. Note that the precise model above is introduced by Gilbert in [108], and in [97] a model was formulated with a fixed number of edges (rather than a binomial number of edges). It is not hard to see that the two models are intimately related (see e.g., Section 4.6, where the history is explained in a bit more detail). The Erdős-Rényi random graph was named after Erdős and Rényi due to the deep and striking results proved in [97], which opened up an entirely new field. Earlier papers investigating random graphs are [94], using the probabilistic method to prove graph properties, and [209], where the model is introduced as a model for neurons.

Despite the fact that $\mathrm{ER}_n(p)$ is the simplest imaginable model of a random network, it has a fascinating *phase transition* when p varies. Phase transitions are well known in physics. The paradigm example is the solid-fluid transition of water, which occurs when we move the temperature from below 0° to above 0°. Similar phase transitions occur in various real phenomena, such as magnetism or the conductance properties of porous materials. Many models have been invented that describe and explain such phase transitions, and we shall see some examples in these notes. As we will see, the Erdős-Rényi random graph exhibits a phase transition in the size of the maximal component, as well as

in the connectivity of the arising random graph.

Indeed, if $p = \lambda/n$ with $\lambda < 1$, then $\mathrm{ER}_n(p)$ consists of many small components having at most size $\Theta(\log n)$. If, otherwise, $\lambda > 1$ the graph consists of one giant component of $\Theta(n)$ and some small components which have size $\Theta(\log n)$. (Recall the notation in Section 1.4.) These properties shall be explained and proved in full detail in Chapter 4. In Chapter 5, we shall also investigate the size for the largest connected component when $\lambda = 1$, and for which λ the Erdős-Rényi random graph is connected.

A rough outline of the ideas behind the proof in Chapters 4–5 is given below. The necessary probabilistic ingredients are described in Chapter 2, for example, stochastic orderings, convergence of random variables, and couplings. In Chapter 3, we describe *branching processes*, which prove to be extremely useful in the analysis of the Erdős-Rényi random graph and many other related random graph models.

To describe these preliminaries, let us investigate the cluster of a vertex in an Erdős-Rényi random graph. We say that $u, v \in \{1, \ldots, n\}$ are *connected* when there exists a path of occupied bonds connecting the two vertices u and v, and we write this as $u \longleftrightarrow v$. We let the cluster of v, i.e., the connected component containing v, be equal to

$$\mathcal{C}(v) = \{y : v \longleftrightarrow y\}, \tag{1.5.1}$$

where, by convention, v is connected to v, so that $v \in \mathcal{C}(v)$. Let $|\mathcal{C}(v)|$ denote the number of vertices in $\mathcal{C}(v)$. Then, the size of the largest connected component of $\mathrm{ER}_n(p)$ is equal to

$$|\mathcal{C}_{\max}| = \max_{v \in \{1,\ldots,n\}} |\mathcal{C}(v)|. \tag{1.5.2}$$

Naturally, the law of $\mathcal{C}(x)$, and, therefore also of $|\mathcal{C}_{\max}|$, depends sensitively on the value of p.

To describe the largest connected component, we explore the different clusters one by one. We start with vertex 1, and explore all the edges that are incident to 1. The endpoints of these edges are clearly elements of the cluster $\mathcal{C}(1)$. Therefore, the exploration of the edges starting from 1 gives rise to a subset of vertices that are in $\mathcal{C}(1)$, namely, precisely the vertices that are at distance 1 in the random graph $\mathrm{ER}_n(p)$ from the vertex 1, i.e., the direct neighbors. Denote the number of different neighbors by X_1. Note that the distribution of the number of direct neighbors X_1 is equal to a binomial random variable with parameters $n-1$ and p, i.e., $X_1 \sim \mathrm{Bin}(n-1, p)$.

When $X_1 = 0$, then $\mathcal{C}(1) = \{1\}$, and we have explored the entire cluster of vertex 1. However, when $X_1 \geq 1$, then there is at least one direct neighbor of 1, and we next explore its direct neighbors. We denote $i_1, \ldots, i_{X_1}$ the vertices that are direct neighbors of 1, where we order these such that $i_1 < i_2 < \ldots < i_{X_1}$.

We now explore the neighbors of i_1. Naturally, when we wish to explore the elements of $\mathcal{C}(1)$, we are only interested in those neighbors of 1 for which we do not yet know that they are part of $\mathcal{C}(1)$. When we fix the number of direct neighbors X_1, then this number of neighbors of i_1 again has a binomial distribution, now with parameters $n-1-X_1$ and probability of success p. Denote the number of vertices by X_2. We emphasize here that the *conditional distribution* of X_2 given X_1 is $\mathrm{Bin}(n-1-X_1, p)$, but the marginal distribution of X_2 is *not* binomial.

When $X_1 \geq 2$, we can also explore the direct neighbors of i_2 that are not yet part of $\mathcal{C}(1)$, and this number, which we denote by X_3, has, conditionally on X_1 and X_2,distribution $\mathrm{Bin}(n-1-X_1-X_2, p)$. This is called *breadth-first search*. In general, when we explore the $(i+1)^{\mathrm{st}}$ vertex of the cluster of vertex 1, we obtain a random number of newly added vertices, denoted by X_{i+1}, which are part of $\mathcal{C}(1)$, and of which the law is $\mathrm{Bin}(N_i, p)$, where

$$N_i = n - 1 - X_1 - \cdots - X_i, \qquad i = 1, 2, \ldots. \tag{1.5.3}$$

Before exploring the i^{th} vertex, the number of vertices whose neighbors we have not yet investigated is equal to

$$1 + X_1 + \ldots + X_i - i, \tag{1.5.4}$$

that is, the number of vertices of which we have decided that they are part of the cluster of vertex 1 minus the number of vertices which have been fully explored. This process continues as long as there are unexplored or active vertices, i.e., it continues as long as

$$1 + X_1 + \ldots + X_i - i \geq 1. \tag{1.5.5}$$

Since finally we explore all vertices in the cluster, we obtain that

$$|\mathcal{C}(1)| = \min\{i : X_1 + \ldots + X_i = i - 1\}. \tag{1.5.6}$$

Similarly, we can explore the clusters of the other vertices that are not elements of $\mathcal{C}(1)$. Say that $j \in \{1, \ldots, n\}$ is the smallest element that does not belong to $\mathcal{C}(1)$. Then, in a similar way as above, we can explore $\mathcal{C}(j)$, the cluster of j. Note, however, that, since $j \notin \mathcal{C}(1)$, the vertices in $\mathcal{C}(1)$ should now be removed from the procedure. Therefore, the number of available vertices decreases. This phenomenon is sometimes called the *depletion of points* effect.

It is well known that when n is large, then the binomial distribution with parameters n and $p = \lambda/n$ is close to the Poisson distribution with parameter λ. More precisely, we have that

$$\mathbb{P}\Big(\mathrm{Bin}(n, \lambda/n) = k\Big) = e^{-\lambda}\frac{\lambda^k}{k!} + o(1), \qquad k = 0, 1, \ldots. \tag{1.5.7}$$

The probability mass function $f_k = e^{-\lambda}\frac{\lambda^k}{k!}$ is the probability mass function of the Poisson distribution with parameter λ. In fact, this result can be strengthened to saying that the proportion of vertices with degree k converges in probability to the Poisson probability mass function f_k, i.e., $\mathrm{ER}(n, \lambda/n)$ is a *sparse random graph process*. In particular, for every fixed i, if we were to know that $X_1, \ldots, X_i$ are not too large (which is true if X_j were Poisson random variables with parameter λ), then

$$N_i = n - 1 - X_1 - \cdots - X_i \approx n. \tag{1.5.8}$$

Thus, we have that a binomial random variable with parameters N_i and success probability $p = \lambda/n$ is approximately Poisson distributed with parameter λ. With this approximation, the random variables $\{X_j\}_{j=1}^{\infty}$ are *independent and identically distributed,* which is often abbreviated by i.i.d. in these notes. In this approximation, we see that the number of unexplored vertices satisfies a recurrence relation given by

$$S_i^* \sim 1 + X_1^* + \ldots + X_i^* - i, \tag{1.5.9}$$

up to the point where $S_i^* = 0$, and where $\{X_i^*\}_{i=1}^{\infty}$ are i.i.d. Poisson random variables with parameter λ. We write

$$T^* = \min\{i : S_i^* = 0\} = \min\{i : X_1^* + \ldots + X_i^* = i - 1\} \tag{1.5.10}$$

for the first time at which $S_i = 0$. In the above simplified model, the random variable T^* could be *infinite*, while in (1.5.6) this is clearly impossible. In (1.5.10), we explore vertices in a tree, and the i^{th} explored individual gives rise to X_i^* children, where $\{X_j^*\}_{j=1}^{\infty}$ are i.i.d. Poisson random variables with parameter λ. The above process is called a *branching process* with a Poisson offspring distribution with parameter or mean λ.

Branching processes are simple models for the evolution of a population, and have received considerable attention in the mathematical literature. See [16, 115, 123] for introductions to the subject. Branching processes have a *phase transition* when the expected offspring varies. When the expected offspring exceeds 1, then there is a positive probability of survival forever, while if the expected offspring is at most 1, then the population dies out with probability one. This phase transition for branching processes is intimately connected to the phase transition on the random graph.

We describe the phase transition on the random graph in Chapter 4. In that chapter, the exploration description of connected components described above will be crucial. In order to make the above steps rigorous, we need some preliminaries. In Chapter 2, we describe the probabilistic preliminaries, such as stochastic ordering, convergence of random variables, coupling theory and martingales. For example, stochastic domination allows us to make the intuition that $X_{i+1} \sim \mathrm{Bin}(N_i, p)$ when $N_i \leq n$ is smaller than a binomial random variable with parameters n and p precise. Convergence of random variables is the right notion to show that a binomial distribution with parameters n and $p = \lambda/n$ is close to the Poisson distribution with parameter λ. A coupling of these two random variables allows us to give a bound on their difference. In Chapter 3, we describe branching processes. We prove the phase transition, and relate super critical branching processes conditioned to die out with subcritical branching processes. We pay particular attention to branching processes with a Poisson offspring distribution.

While the Erdős-Rényi random graph is a beautiful model displaying fascinating scaling behavior for large graphs and varying edge probabilities, its degrees are *not* scale-free, rendering it unrealistic as a network model. Indeed, its typical degree size is the average degree, and there is little variability in it. In particular, no hubs exist. More precisely, the degree of any vertex in an Erdős-Rényi random graph with edge probability $p = \lambda/n$ is precisely equal to a binomial random variable with parameters $n-1$ and success probability $p = \lambda/n$. As a result, the limiting degree of any vertex is equal to a Poisson random variable with mean λ. It is well known that Poisson random variables have thinner tails than power laws. In fact, Poisson random variables have *exponential tails*. See the discussion below (1.5.7), and see Section 5.3 for a proof of the fact that the Erdős-Rényi random graph with edge probability $p = \lambda/n$ is sparse.

Therefore, to model networks more appropriately, we are on the hunt for scale-free random graph models! Remarkably, the fact that the Erdős-Rényi random graph is not a suitable network model was already foreseen by the masters themselves [97]:

> "*Of course, if one aims at describing such a real situation, one should replace the hypothesis of equiprobability of all connections by some more realistic hypothesis.*".

How do power laws arise then in networks, and what can we learn from that? In the next section, we shall describe three models for scale-free networks.

1.6 Random graph models for complex networks

As explained in Section 1.5, Erdős-Rényi random graphs are *not* scale free, whereas, as explained in Section 1.1, many real networks are scale free. In Chapters 6, 7 and 8, we describe three scale-free random graph models. In Chapter 6, we describe the *generalized random graph*. The philosophy of this model is simple: we adapt the random graph in such a way that it becomes scale free. For this, we note that the degrees of the Erdős-Rényi random graph with edge probability $p = \lambda/n$ are close to to a Poisson random variable with mean λ. As mentioned before, these are *not* scale free. However, we can make these degrees scale free by taking the parameter λ to be a random variable with a power law. Thus, to each vertex i, we associate a random variable W_i, and, conditionally on W_i, the edges emanating from i will be occupied with a probability depending on i. There are many ways in which this can be done. For example, in the generalized random graph [58], the probability that edge between vertices s and t, which we denote by st, is occupied, conditionally on the weights $\{W_i\}_{i=1}^n$, is equal to

$$p_{st} = \frac{W_s W_t}{W_s W_t + L_n}, \tag{1.6.1}$$

where $L_n = \sum_{i=1}^n W_i$ is the total weight of the graph, and different edges are conditionally independent given $\{W_i\}_{i=1}^n$. In Chapter 6, we shall prove that this further randomization of the Erdős-Rényi random graph does, in the case when the W_i are i.i.d. and satisfy a power law, lead to scale-free graphs. There are various other possibilities to generalize the Erdős-Rényi random graph, some of which will also be discussed. See [67, 182] for two specific examples, and [47] for the most general set-up of generalized random graphs.

In the second scale-free random graph model, the idea is that we should take the degrees as a start for the model. Thus, to each vertex i, we associate a degree D_i, and in some way connect up the different edges. Clearly, we need that the sum of the degrees $L_n = \sum_{i=1}^n D_i$ is *even*, and we shall assume this from now on. Then we think of placing D_i half-edges or stubs incident to vertex i, and connecting all the stubs in a certain way to yield a graph. One way to do this is to attach all the stub *uniformly*, and this leads to the *configuration model*. Naturally, it is possible that the above procedure does not lead to a simple graph, since self-loops and multiple edges can occur. As it turns out, when the degrees are not too large, more precisely, when they have *finite variance*, then the graph is with positive probability simple. By conditioning on the graph being simple, we end up with a uniform graph with the specified degrees. Sometimes this is also referred to as the *repeated configuration model*, since we can think of conditioning as repeatedly forming the graph until it is simple, which happens with strictly positive probability. A second approach to dealing with self-loops and multiple edges is simply to remove them, leading to the so-called *erased configuration model*. In Chapter 7, we investigate these two models, and show that the degrees are given by the degree distribution, when the graph size tends to infinity. Thus, the erasing and the conditioning do not alter the degrees too much.

The generalized random graph and configuration models *describe* networks, in some sense, quite satisfactorily. Indeed, they give rise to models with degrees that can be matched to degree distributions found in real networks. However, they do not *explain* how the networks came to be as they are. A possible explanation for the occurrence of scale-free behavior was given by Albert and Barabási [20], by a feature called *preferential attachment*. Most real networks grow. For example, the WWW has increased from a few web pages in 1990 to an estimated size of a few billion now. *Growth* is an aspect that is not taken into account in Erdős-Rényi random graphs, but it would not be hard to reformulate them as a growth process where elements are successively added, and connections are added and removed. Thus, growth by itself is not enough to explain the occurrence of power laws. However, viewing real networks as evolving in time does give us the possibility to investigate just *how* they grow.

So, how do real networks grow? Think of a social network describing a certain population in which a newcomer arrives, increasing it by one element. He/She will start to socialize with people in the population, and this process is responsible for the connections to the newly arrived person. In an Erdős-Rényi random graph, the connections to the newcomer will be spread uniformly over the population. Is this realistic? Is the newcomer not more likely to get to know people who are socially active, and, therefore, already have a larger degree? Probably so! We do not live in a perfectly egalitarian world. Rather, we live in a self-reinforcing world, where people who are successful are more likely to become even more successful! Therefore, rather than equal probabilities for our newcomer to acquaint him-/herself to other individuals in the population, there is a bias towards individuals who already know many people. When we think of the degree of elements as describing the *wealth* of the individuals in the population, we live in a world where the rich get richer!

Phrased in a more mathematical way, preferential attachment models are such that new elements are more likely to attach to elements with high degree compared to elements with small degree. For example, suppose that new elements are born with a fixed amount of edges to the older elements. Each edge is connected to a specific older element with a probability which is proportional to the degree of that older element. This phenomenon is now mostly called *preferential attachment*, and was first described informally by Albert and Barabási [20]. See also the book [19] for a highly readable and enthusiastic personal

account by Barabási. Albert and Barabási have been two of the major players in the investigation of the similarities of real networks, and their papers have proved to be very influential. See [5, 6, 7, 20]. The notion of preferential attachment in networks has lead the theoretical physics and the mathematics communities to study the structure of preferential attachment models in numerous papers. For some of the references, see Chapter 8.

While the above explanation is for social networks, also in other examples some form of preferential attachment is likely to be present. For example, in the WWW, when a new web page is created, it is more likely to link to an already popular site, such as Google, than to my personal web page. For the Internet, it may be profitable for new routers to be connected to highly connected routers, since these give rise to short distances. Even in biological networks, a more subtle form of preferential attachment exists.

In Chapter 8, we shall introduce and study preferential attachment models, and show that preferential attachment leads to scale-free random graphs. The power-law exponent of the degrees depends sensitively on the precise parameters of the model, such as the number of added edges and how dominant the preferential attachment effect is, in a similar way as the suggested power law exponent in the heuristic derivation in (1.3.6) depends on the parameters of that model.

In Chapters 6, 7 and 8, we investigate the *degrees* of the proposed random graph models. This explains the scale-free nature of the models. In Chapters 9, 10 and 11, we investigate further properties of these models, focussing on the *connected components* and the *distances* in the graphs. As observed in Section 1.1, most real networks are *small worlds*. As a result, one would hope that random graph models for real networks are such that distances between their elements are small. In Chapters 9, 10 and 11, we shall quantify this, and relate graph distances to the properties of the degrees. A further property we shall investigate is the *phase transition* of the largest connected component, as described in detail for the Erdős-Rényi random graph in Chapter 4.

1.7 Notes and discussion

Chapter 2
Probabilistic methods

In this chapter, we describe basic results in probability theory that we shall rely on in these notes. We describe convergence of random variables in Section 2.1, coupling in Section 2.2 and stochastic domination in Section 2.3. In Section 2.4 we describe bounds on random variables, namely the Markov inequality, the Chebychev inequality and the Chernoff bound. Particular attention will be given to binomial random variables, as they play a crucial role throughout these notes. In Section 2.5, we describe a few results on martingales. Finally, in Section 2.6, we describe some extreme value theory of random variables. In this chapter, not all proofs are given.

2.1 Convergence of random variables

In the random graph with $p = \lambda/n$, for some $\lambda > 0$, we note that the degree of a vertex is distributed as a $\mathrm{Bin}(n-1,p)$ random variable. When n is large, and $np = \lambda$ is fixed, then it is well known that a $\mathrm{Bin}(n-1,p)$ is close to a Poisson random variable with mean λ. In Chapter 4, we make heavy use of this convergence result, and a version of it is stated in Theorem 2.9 below.

In order to formalize that

$$\mathrm{Bin}(n,p) \approx \mathrm{Poi}(np), \tag{2.1.1}$$

we need to introduce the notions of convergence of random variables. For this, we note that random variables are defined to be functions on a sample space. It is well known that there are several possible notions for convergence of functions on function spaces. In a similar fashion, there are several notions of convergence of random variables, three of which we state in the following definition. For more background on the convergence of random variables, we refer the reader to [37].

Definition 2.1 (Convergence of random variables)**.**

(a) *A sequence X_n of random variables converges in distribution to a limiting random variable X when*

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x), \tag{2.1.2}$$

for every x for which $F(x) = \mathbb{P}(X \le x)$ is continuous. We write this as $X_n \xrightarrow{d} X$.

(b) *A sequence X_n of random variables converges in probability to a limiting random variable X when, for every $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0. \tag{2.1.3}$$

We write this as $X_n \xrightarrow{\mathbb{P}} X$.

(c) *A sequence X_n of random variables converges almost surely to a limiting random variable X when*

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1. \tag{2.1.4}$$

We write this as $X_n \xrightarrow{a.s.} X$.

It is not hard to see that convergence in probability implies convergence in distribution. The notion of convergence almost surely is clearly the most difficult to grasp. It turns out that convergence almost surely implies convergence in probability, making it the strongest version of convergence to be discussed in these notes. We shall mainly work with convergence in distribution and convergence in probability.

There are also further forms of convergence that we do not discuss, such as convergence in L^1 or L^2. We again refer to [37], or to introductory books in probability, such as [38, 105, 106, 111].

There are examples where convergence in distribution holds, but convergence in probability fails:

Exercise 2.1. *Find an example of a sequence of random variables where convergence in distribution occurs, but convergence in probability does not.*

Exercise 2.2. *Show that the sequence of random variables $\{X_n\}_{n=1}^{\infty}$, where X_n takes the value n with probability $\frac{1}{n}$ and 0 with probability $1-\frac{1}{n}$ converges both in distribution and in probability to 0.*

We next state some theorems that give convenient criterions by which we can conclude that random variables converge in distribution. In their statement, we make use of a number of functions of random variables that we introduce now.

Definition 2.2 (Generating functions of random variables)**.** *Let X be a random variable. Then*

(a) *The* characteristic function *of X is the function*

$$\phi_X(t) = \mathbb{E}[e^{itX}], \qquad t \in \mathbb{R}. \tag{2.1.5}$$

(b) *The* probability generating function *of X is the function*

$$G_X(t) = \mathbb{E}[t^X], \qquad t \in \mathbb{R}. \tag{2.1.6}$$

(c) *The* moment generating function *of X is the function*

$$M_X(t) = \mathbb{E}[e^{tX}], \qquad t \in \mathbb{R}. \tag{2.1.7}$$

We note that the characteristic function exists for every random variable X, since $|e^{itX}| = 1$ for every t. The moment generating function, however, does not always exist.

Exercise 2.3. *Find a random variable for which the moment generating function is equal to $+\infty$ for every $t \neq 0$.*

Theorem 2.3 (Criteria for convergence in distribution)**.** *The sequence of random variables $\{X_n\}_{n=1}^{\infty}$ converges in distribution to a random variable X*

(a) if and only if the characteristic functions $\phi_n(t)$ of X_n converge to the characteristic function $\phi_X(t)$ of X for all $t \in \mathbb{R}$.

(b) when, for some $\varepsilon > 0$, the moment generating functions $M_n(t)$ of X_n converge to the moment generating function $M_X(t)$ of X for all $|t| < \varepsilon$.

(c) when, for some $\varepsilon > 0$, the probability generating functions $G_n(t)$ of X_n converge to the probability generating function $G_X(t)$ of X for all $|t| < 1 + \varepsilon$ for some $\varepsilon > 0$.

(d) when the X_n are non-negative and integer-valued, and the moments $\mathbb{E}[X_n^r]$ converge to the moments $\mathbb{E}[X^r]$ of X for each $r = 1, 2, \ldots$, provided the moments of X satisfy

$$\lim_{r\to\infty} \frac{\mathbb{E}[X^r]r^m}{r!} = 0 \qquad \forall m = 0, 1, \ldots \tag{2.1.8}$$

(e) when the moments $\mathbb{E}[X_n^r]$ converge to the moments $\mathbb{E}[X^r]$ of X for each $r = 1, 2, \ldots$, and $M_X(t)$, the moment generating function of X, is finite for t in some neighborhood of the origin.

Exercise 2.4. *Show that a Poisson random variable satisfies the moment condition in* (2.1.8).

Exercise 2.5. *Prove that when X is a Poisson random variable with mean λ, then*

$$\mathbb{E}[(X)_r] = \lambda^r. \tag{2.1.9}$$

Exercise 2.6. *Show that the moments of a Poisson random variable X with mean λ satisfy the recursion*

$$\mathbb{E}[X^m] = \lambda\mathbb{E}[(X+1)^{m-1}]. \tag{2.1.10}$$

We finally discuss a special case of convergence in distribution, namely, when we deal with a sum of indicators, and the limit is a Poisson random variable. We write $(X)_r = X(X-1)\cdots(X-r+1)$, so that $\mathbb{E}[(X)_r]$ is the r^{th} *factorial moment of* X.

For a random variable X taking values in $\{0, 1, \ldots, n\}$, the factorial moments of X uniquely determine the probability mass function, since

$$\mathbb{P}(X = k) = \sum_{r=k}^{n} (-1)^{k+r} \frac{\mathbb{E}[(X)_r]}{(r-k)!k!}, \tag{2.1.11}$$

see e.g. [44, Corollary 1.11]. To see (2.1.11), we write

$$\mathbb{1}_{\{X=k\}} = \binom{X}{k}\Big(1 \ -1\Big)^{X-k}, \tag{2.1.12}$$

using the convention that $0^0 = 1$. Then, by Newton's binomial, we obtain

$$\mathbb{1}_{\{X=k\}} = \binom{X}{k}\sum_{i=0}^{X-k}(-1)^i\binom{X-k}{i} = \sum_{i=0}^{\infty}(-1)^i\binom{X}{k}\binom{X-k}{i}, \tag{2.1.13}$$

where, by convention, we take that $\binom{n}{k} = 0$ when $k < 0$ or $k > n$. Rearranging the binomials, we arrive at

$$\mathbb{1}_{\{X=k\}} = \sum_{r=k}^{\infty}(-1)^{k+r}\frac{(X)_r}{(r-k)!k!}, \tag{2.1.14}$$

where $r = k + i$, and taking expectations yields

$$\mathbb{P}(X = k) = \sum_{r=k}^{\infty}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!}, \tag{2.1.15}$$

which is (2.1.11). Similar results also hold for unbounded random variables, since the sum

$$\sum_{r=k}^{n}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!} \tag{2.1.16}$$

is alternatingly smaller than $\mathbb{P}(X=k)$ (for n even) and larger than $\mathbb{P}(X=k)$ (for n odd). This implies the following result:

Theorem 2.4 (Convergence to a Poisson random variable)**.** *A sequence of integer-valued random variables $\{X_n\}_{n=1}^{\infty}$ converges in distribution to a Poisson random variable with parameter λ when, for all $r=1,2,\ldots,$*

$$\lim_{n\to\infty}\mathbb{E}[(X_n)_r]=\lambda^r. \tag{2.1.17}$$

Exercise 2.7. *Show that if*

$$\lim_{n\to\infty}\sum_{r\geq n}\frac{\mathbb{E}[(X)_r]}{(r-k)!}=0, \tag{2.1.18}$$

then also

$$\mathbb{P}(X=k)=\sum_{r=k}^{\infty}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!}, \tag{2.1.19}$$

and use this to conclude that when $\lim_{n\to\infty}\mathbb{E}[(X_n)_r]=\mathbb{E}[(X)_r]$ for all $r\geq 1$, where X_n and X are all integer-valued non-negative random variables, then also $X_n\xrightarrow{d}X$.

Theorem 2.4 is particularly convenient when dealing with sums of indicators, i.e., when

$$X_n=\sum_{i\in\mathcal{I}_n}I_{i,n}, \tag{2.1.20}$$

where $I_{i,n}$ takes the values 0 and 1 only, as the following result shows:

Theorem 2.5 (Factorial moments of sums of indicators)**.** *When $X=\sum_{i\in\mathcal{I}}I_i$ is a sum of indicators, then*

$$\mathbb{E}[(X)_r]=\sum_{i_1,\ldots,i_r\in\mathcal{I}}^{*}\mathbb{E}[\prod_{l=1}^{r}I_{i_l}]=\sum_{i_1,\ldots,i_r\in\mathcal{I}}^{*}\mathbb{P}\big(I_{i_1}=\cdots=I_{i_r}=1\big), \tag{2.1.21}$$

where $\sum^{*}_{i_1,\ldots,i_r\in\mathcal{I}}$ denotes a sum over distinct indices.

Exercise 2.8. *Prove* (2.1.21) *for $r=2$.*

Exercise 2.9. *Compute the factorial moments of a binomial random variable with parameters n and $p=\lambda/n$ and the ones of a Poisson random variable with mean λ, and use this to conclude that a binomial random variable with parameters n and $p=\lambda/n$ converges in distribution to a Poisson random variable with mean λ.*

Proof of Theorem 2.5. We prove (2.1.21) by induction on $r \geq 1$ and for all probability measures $\mathbb{P}$ and corresponding expectations $\mathbb{E}$. For $r = 1$, we have that $(X)_1 = X$, and (2.1.21) follows from the fact that the expectation of a sum of random variables is the sum of expectations. This initializes the induction hypothesis.

In order to advance the induction hypothesis, we first note that it suffices to prove the statement for indicators I_i for which $\mathbb{P}(I_i = 1) > 0$. Then, for $r \geq 2$, we write out

$$\mathbb{E}[(X)_r] = \sum_{i_1 \in \mathcal{I}} \mathbb{E}\big[I_{i_1}(X-1)\cdots(X-r+1)\big]. \tag{2.1.22}$$

Denote by $\mathbb{P}_{i_1}$ the conditional distribution given that $I_{i_1} = 1$, i.e., for any event E, we have

$$\mathbb{P}_{i_1}(E) = \frac{\mathbb{P}(E \cap \{I_{i_1} = 1\})}{\mathbb{P}(I_{i_1} = 1)}. \tag{2.1.23}$$

Then we can rewrite

$$\mathbb{E}\big[I_{i_1}(X-1)\cdots(X-r+1)\big] = \mathbb{P}(I_{i_1} = 1)\mathbb{E}_{i_1}\big[(X-1)\cdots(X-r+1)\big]. \tag{2.1.24}$$

We define

$$Y = X - I_{i_1} = \sum_{j \in \mathcal{I}\setminus\{i_1\}} I_j, \tag{2.1.25}$$

and note that, conditionally on $I_{i_1} = 1$, we have that $X = Y + 1$. As a result, we obtain that

$$\mathbb{E}_{i_1}\big[(X-1)\cdots(X-r+1)\big] = \mathbb{E}_{i_1}\big[Y\cdots(Y-r+2)\big] = \mathbb{E}_{i_1}\big[(Y)_{r-1}\big]. \tag{2.1.26}$$

We now apply the induction hypothesis to $\mathbb{E}_{i_1}\big[(Y)_{r-1}\big]$, to obtain

$$\mathbb{E}_{i_1}\big[(Y)_{r-1}\big] = \sum_{i_2,\ldots,i_r \in \mathcal{I}\setminus\{i_1\}}^{*} \mathbb{P}_{i_1}\big(I_{i_2} = \cdots = I_{i_r} = 1\big). \tag{2.1.27}$$

As a result, we arrive at

$$\mathbb{E}[(X)_r] = \sum_{i_1 \in \mathcal{I}} \mathbb{P}(I_{i_1} = 1) \sum_{i_2,\ldots,i_r \in \mathcal{I}\setminus\{i_1\}}^{*} \mathbb{P}_{i_1}\big(I_{i_2} = \cdots = I_{i_r} = 1\big). \tag{2.1.28}$$

We complete the proof by noting that

$$\mathbb{P}(I_{i_1} = 1)\mathbb{P}_{i_1}\big(I_{i_2} = \cdots = I_{i_r} = 1\big) = \mathbb{P}\big(I_{i_1} = I_{i_2} = \cdots = I_{i_r} = 1\big), \tag{2.1.29}$$

and that

$$\sum_{i_1 \in \mathcal{I}} \sum_{i_2,\ldots,i_r \in \mathcal{I}\setminus\{i_1\}}^{*} = \sum_{i_1,\ldots,i_r \in \mathcal{I}}^{*}. \tag{2.1.30}$$

□

There also exist multidimensional versions of Theorems 2.4 and 2.5:

Theorem 2.6 (Convergence to independent Poisson random variables)**.** *A vector of integer-valued random variables $\{(X_{1,n},\ldots,X_{d,n})\}_{n=1}^{\infty}$ converges in distribution to a vector of independent Poisson random variable with parameters $\lambda_1,\ldots,\lambda_d$ when, for all $r_1,\ldots,r_d \in \mathbb{N}$,*

$$\lim_{n\to\infty} \mathbb{E}[(X_{1,n})_{r_1}\cdots(X_{d,n})_{r_d}] = \lambda_1^{r_1}\cdots\lambda_d^{r_d}. \tag{2.1.31}$$

Theorem 2.7 (Factorial moments of sums of indicators)**.** *When $X_l = \sum_{i \in \mathcal{I}_l} I_{i,l}$ for all $l = 1, \dots, d$ are sums of indicators, then*

$$\mathbb{E}[(X_{1,n})_{r_1} \cdots (X_{d,n})_{r_d}] = \sum_{i_1^{(1)}, \dots, i_{r_1}^{(1)} \in \mathcal{I}_1}^{*} \cdots \sum_{i_1^{(d)}, \dots, i_{r_d}^{(d)} \in \mathcal{I}_d}^{*} \mathbb{P}\big(I_{i_s}^{(l)} = 1 \forall l = 1, \dots, d \& s = 1, \dots, r_l\big). \tag{2.1.32}$$

Exercise 2.10. *Prove Theorem 2.7 using Theorem 2.5.*

The fact that the convergence of moments as in Theorems 2.3, 2.4 and 2.6 yields convergence in distribution is sometimes called the *method of moments*, and is a good way of proving convergence results.

2.2 Coupling

For any λ fixed, it is well known that, when $n \to \infty$,

$$\mathrm{Bin}(n, \lambda/n) \xrightarrow{\mathbb{P}} \mathrm{Poi}(\lambda). \tag{2.2.1}$$

In general, convergence in probability implies convergence in distribution, so that also convergence in distribution follows. To prove this convergence, we will use a *coupling* proof. Couplings will be quite useful in what follows, so we will discuss couplings, as well as the related topic of *stochastic orderings*, in detail. An excellent treatment of coupling theory is given in [214], to which we refer for more details.

In general, two random variables X and Y are *coupled* when they are defined on the same probability space. This means that there is one probability law $\mathbb{P}$ such that $\mathbb{P}(X \in E, Y \in F)$ are defined for all events E and F. This is formalized in the following definition, where it is also generalized to more than one random variable:

Definition 2.8 (Coupling of random variables)**.** *The random variables $(\hat{X}_1, \dots, \hat{X}_n)$ are a coupling of the random variables $X_1, \dots, X_n$ when $(\hat{X}_1, \dots, \hat{X}_n)$ are defined on the same probability space, and are such that the marginal distribution of $\hat{X}_i$ is the same as the distribution of X_i for all $i = 1, \dots, n$, i.e., for all measurable subsets E of $\mathbb{R}$, we have*

$$\mathbb{P}(\hat{X}_i \in E) = \mathbb{P}(X_i \in E). \tag{2.2.2}$$

The key point of Definition 2.8 is that while the random variables $X_1, \dots, X_n$ may be defined on *different* probability spaces, the coupled random variables $(\hat{X}_1, \dots, \hat{X}_n)$ are defined on the *same* probability space. The coupled random variables $(\hat{X}_1, \dots, \hat{X}_n)$ are related to the original random variables $X_1, \dots, X_n$ by the fact that the marginal distributions of $(\hat{X}_1, \dots, \hat{X}_n)$ are equal to the random variables $X_1, \dots, X_n$. Note that one coupling arises by taking $(\hat{X}_1, \dots, \hat{X}_n)$ to be independent, with $\hat{X}_i$ having the same distribution as X_i. However, in our proofs, we shall often make use of more elaborate couplings, which give rise to stronger results.

Couplings are very useful to prove that random variables are somehow related. We now describe a general coupling between two random variables which makes two random variables be with high probability *equal*. We let X and Y be two random variables with

$$\mathbb{P}(X = x) = p_x, \qquad \mathbb{P}(Y = y) = q_y, \qquad x \in \mathcal{X}, y \in \mathcal{Y} \tag{2.2.3}$$

where $\{p_x\}_{x\in\mathcal{X}}$ and $\{q_y\}_{y\in\mathcal{Y}}$ are any two probability mass functions on two subsets $\mathcal{X}$ and $\mathcal{Y}$ of the same space. Then, we define the random vector $(\hat{X},\hat{Y})$ by

$$\mathbb{P}(\hat{X}=\hat{Y}=x)=\min\{p_x,q_x\}, \tag{2.2.4}$$

$$\mathbb{P}(\hat{X}=x,\hat{Y}=y)=\frac{(p_x-\min\{p_x,q_x\})(q_y-\min\{p_y,q_y\})}{\frac{1}{2}\sum_z|p_z-q_z|}, \qquad x\neq y. \tag{2.2.5}$$

First of all, this is a probability distribution, since

$$\sum_x(p_x-\min\{p_x,q_x\})=\sum_x(q_x-\min\{p_x,q_x\})=\frac{1}{2}\sum_x|p_x-q_x|. \tag{2.2.6}$$

Exercise 2.11 (Coupling and total variation distance)**.** *Prove* (2.2.6)*.*

The distance between discrete probability distributions $\{p_x\}_{x\in\mathcal{X}}$ and $\{q_x\}_{x\in\mathcal{X}}$ in (2.2.6) is called the *total variation distance* between the discrete probability mass functions $\{p_x\}_{x\in\mathcal{X}}$ and $\{q_x\}_{x\in\mathcal{X}}$. In general, for two probability measures μ and ν, the total variation distance is given by

$$d_{\mathrm{TV}}(\mu,\nu)=\max_A|\mu(A)-\nu(A)|, \tag{2.2.7}$$

where $\mu(A)$ and $\nu(A)$ are the probabilities of the event A under the measures μ and ν.

When μ and ν are the distribution functions corresponding to two discrete probability mass functions $p=\{p_x\}_{x\in\mathcal{X}}$ and $q=\{q_x\}_{x\in\mathcal{X}}$, so that, for every measurable A with $A\subset\mathcal{X}$, we have

$$\mu(A)=\sum_{x\in A}p_x, \qquad \nu(A)=\sum_{x\in A}q_x, \tag{2.2.8}$$

then it is not hard to see that

$$d_{\mathrm{TV}}(p,q)=\frac{1}{2}\sum_x|p_x-q_x|. \tag{2.2.9}$$

When F and G are the distribution functions corresponding to two continuous densities $f=\{f(x)\}_{x\in\mathbb{R}}$ and $g=\{g(x)\}_{x\in\mathbb{R}}$, so that for every measurable $A\subseteq\mathbb{R}$,

$$\mu(A)=\int_A f(x)dx, \qquad \nu(A)=\int_A g(x)dx, \tag{2.2.10}$$

then we obtain

$$d_{\mathrm{TV}}(f,g)=\frac{1}{2}\int_{-\infty}^{\infty}|f(x)-g(x)|dx. \tag{2.2.11}$$

Exercise 2.12 (Total variation and L^1-distances)**.** *Prove* (2.2.9) *and* (2.2.11)*.*

We now continue investigating the coupling in (2.2.4) for two discrete random variables. By construction,

$$\mathbb{P}(\hat{X}=x)=p_x, \qquad \mathbb{P}(\hat{Y}=y)=q_y, \tag{2.2.12}$$

so that $\hat{X}$ and $\hat{Y}$ have the right marginal distributions as required in Definition 2.8. Moreover, we have that, by (2.2.6),

$$\begin{aligned}\mathbb{P}(\hat{X}\neq\hat{Y})&=\sum_{x,y}\frac{(p_x-\min\{p_x,q_x\})(q_y-\min\{p_y,q_y\})}{\frac{1}{2}\sum_z|p_z-q_z|}\\&=\frac{1}{2}\sum_x|p_x-q_x|=d_{\mathrm{TV}}(p,q).\end{aligned} \tag{2.2.13}$$

It turns out that this is an optimal or maximal coupling. See [214] for details. Indeed, we have that for all x,

$$\mathbb{P}(\hat{X} = \hat{Y} = x) \leq \mathbb{P}(\hat{X} = x) = \mathbb{P}(X = x) = p_x, \tag{2.2.14}$$

and also

$$\mathbb{P}(\hat{X} = \hat{Y} = x) \leq \mathbb{P}(\hat{Y} = x) = \mathbb{P}(Y = x) = q_x, \tag{2.2.15}$$

so that for any coupling we must have that

$$\mathbb{P}(\hat{X} = \hat{Y} = x) \leq \min\{p_x, q_x\}. \tag{2.2.16}$$

Therefore, any coupling must be such that

$$\mathbb{P}(\hat{X} = \hat{Y}) = \sum_x \mathbb{P}(\hat{X} = \hat{Y} = x) \leq \sum_x \min\{p_x, q_x\}. \tag{2.2.17}$$

As a result, we have that, for *any* coupling,

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \geq 1 - \sum_x \min\{p_x, q_x\} = \frac{1}{2} \sum_x |p_x - q_x|. \tag{2.2.18}$$

The coupling in (2.2.4) attains this equality, which makes it the best coupling possible, in the sense that it maximizes $\mathbb{P}(\hat{X} = \hat{Y})$.

In these notes, we will often work with binomial random variables which we wish to compare to Poisson random variables. We will make use of the following theorem, which will be proved using a coupling argument:

Theorem 2.9 (Poisson limit for binomial random variables)**.** *let $\{I_i\}_{i=1}^n$ be independent with $I_i \sim \mathrm{Be}(p_i)$, and let $\lambda = \sum_{i=1}^n p_i$. Let $X = \sum_{i=1}^n I_i$ and let Y be a Poisson random variable with parameter λ. Then, there exists a coupling $(\hat{X}, \hat{Y})$ of (X, Y) such that*

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \sum_{i=1}^n p_i^2. \tag{2.2.19}$$

Consequently, for every $\lambda \geq 0$ and $n \in \mathbb{N}$, there exists a coupling $(\hat{X}, \hat{Y})$, where $\hat{X} \sim \mathrm{Bin}(n, \lambda/n)$ and $\hat{Y} \sim \mathrm{Poi}(\lambda)$ such that

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \frac{\lambda^2}{n}. \tag{2.2.20}$$

Exercise 2.13. *Let $X \sim \mathrm{Bin}(n, \lambda/n)$ and $Y \sim \mathrm{Poi}(\lambda)$. Write $f_i = \mathbb{P}(X = i)$ and $g_i = \mathbb{P}(Y = i)$. Prove that Theorem 2.9 implies that $d_{\mathrm{TV}}(f, g) \leq \lambda^2/n$. Conclude also that, for every $i \in \mathbb{N}$,*

$$\big|\mathbb{P}(X = i) - \mathbb{P}(Y = i)\big| \leq \lambda^2/n. \tag{2.2.21}$$

Proof of Theorem 2.9. Throughout the proof, we let $I_i \sim \mathrm{Be}(p_i)$ and assume that $\{I_i\}_{i=1}^n$ are independent, and we let $J_i \sim \mathrm{Poi}(p_i)$ and assume that $\{J_i\}_{i=1}^n$ are independent. In the proof, we write

$$p_{i,x} = \mathbb{P}(I_i = x) = p_i \mathbb{1}_{\{x=1\}} + (1 - p_i)\mathbb{1}_{\{x=0\}}, \qquad q_{i,x} = \mathbb{P}(J_i = x) = e^{-p_i} \frac{p_i^x}{x!} \tag{2.2.22}$$

for the Bernoulli and Poisson probability mass functions.

For each pair I_i, J_i, the maximal coupling $(\hat{I}_i, \hat{J}_i)$ described above satisfies

$$\mathbb{P}(\hat{I}_i = \hat{J}_i = x) \quad = \quad \min\{p_{1,x}, q_{1,x}\} = \begin{cases} 1 - p_i & \text{for } x = 0, \\ p_i e^{-p_i} & \text{for } x = 1, \\ 0 & \text{for } x \geq 2, \end{cases} \tag{2.2.23}$$

where we have used the inequality $1 - p_i \leq e^{-p_i}$ for $x = 0$. Thus, now using that $1 - e^{-p_i} \leq p_i$,

$$\mathbb{P}(\hat{I}_i \neq \hat{J}_i) = 1 - \mathbb{P}(\hat{I}_i = \hat{J}_i) = 1 - (1 - p_i) - p_i e^{-p_i} = p_i(1 - e^{-p_i}) \leq p_i^2. \tag{2.2.24}$$

Next, let $\hat{X} = \sum_{i=1}^n \hat{I}_i$ and $\hat{Y} = \sum_{i=1}^n \hat{J}_i$. Then, $\hat{X}$ has the same distribution as $X = \sum_{i=1}^n I_i$, and $\hat{Y}$ has the same distribution as $Y = \sum_{i=1}^n J_i \sim \mathrm{Poi}(p_1 + \cdots + p_n)$. Finally, by Boole's inequality and (2.2.24),

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \mathbb{P}\Big(\bigcup_{i=1}^n \{\hat{I}_i \neq \hat{J}_i\} \Big) \leq \sum_{i=1}^n \mathbb{P}(\hat{I}_i \neq \hat{J}_i) \leq \sum_{i=1}^n p_i^2. \tag{2.2.25}$$

This completes the proof of Theorem 2.9. □

For $p = \{p_x\}$ and $q = \{q_x\}$, the total variation distance $d_{\mathrm{TV}}(p, q)$ is obviously larger than $\frac{1}{2}|p_x - g_x|$, so that convergence in total variation distance of $p(n) = \{p_x(n)\}$ to a probability mass function $p = \{p_x\}$ implies pointwise convergence of the probability mass functions $\lim_{n\to\infty} p_x(n) = p_x$ for every x. Interestingly, it turns out that these notions are equivalent:

Exercise 2.14. *Show that if* $\lim_{n\to\infty} p_x(n) = p_x$ *for every* x*, and* $p = \{p_x\}$ *is a probability mass function, then also* $\lim_{n\to\infty} d_{\mathrm{TV}}(p(n), p) = 0$.

2.3 Stochastic ordering

To compare random variables, we will rely on the notion of *stochastic ordering*, which is defined as follows:

Definition 2.10 (Stochastic domination)**.** *Let* X *and* Y *be two random variables, not necessarily living on the same probability space. The random variable* X *is stochastically smaller than the random variable* Y *when, for every* $x \in \mathbb{R}$*, the inequality*

$$\mathbb{P}(X \leq x) \geq \mathbb{P}(Y \leq x) \tag{2.3.1}$$

holds. We denote this by $X \preceq Y$.

A nice coupling reformulation of stochastic ordering is presented in the following lemma:

Lemma 2.11 (Coupling definition of stochastic domination)**.** *The random variable* X *is stochastically smaller than the random variable* Y *if and only if there exists a coupling* $(\hat{X}, \hat{Y})$ *of* X, Y *such that*

$$\mathbb{P}(\hat{X} \leq \hat{Y}) = 1. \tag{2.3.2}$$

Proof. When $\mathbb{P}(\hat{X} \leq \hat{Y}) = 1$, then

$$\begin{aligned}\mathbb{P}(Y \leq x) = \mathbb{P}(\hat{Y} \leq x) &= \mathbb{P}(\hat{X} \leq \hat{Y} \leq x) \\ &\leq \mathbb{P}(\hat{X} \leq x) = \mathbb{P}(X \leq x),\end{aligned} \tag{2.3.3}$$

so that X is stochastically smaller than Y.

For the other direction, suppose that X is stochastically smaller than Y. We define the generalized inverse of a distribution function F by

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \tag{2.3.4}$$

where $u \in [0, 1]$. If U is a uniform random variable on $[0, 1]$, then it is well-known that the random variable $F^{-1}(U)$ has distribution function F. This follows since the function F^{-1} is such that

$$F^{-1}(u) > x \qquad \text{precisely when} \qquad u > F(x). \tag{2.3.5}$$

Denote by F_X and F_Y the marginal distribution functions of X and Y. Then (2.3.1) is equivalent to

$$F_X(x) \geq F_Y(x) \tag{2.3.6}$$

for all x. It follows that, for all $u \in [0, 1]$,

$$F_X^{-1}(u) \leq F_Y^{-1}(u). \tag{2.3.7}$$

Therefore, since $\hat{X} = F_X^{-1}(U)$ and $\hat{Y} = F_Y^{-1}(U)$ have the same marginal distributions as X and Y, respectively, it follows that

$$\mathbb{P}(\hat{X} \leq \hat{Y}) = \mathbb{P}(F_X^{-1}(U) \leq F_Y^{-1}(U)) = 1. \tag{2.3.8}$$

□

There are many examples of pairs of random variables which are stochastically ordered, and we will now describe a few.

Binomial random variables. A simple example of random variables which are stochastically ordered is as follows. Let $m, n \in \mathbb{N}$ be integers such that $m \leq n$. Let $X \sim \mathrm{Bin}(m, p)$ and $Y \sim \mathrm{Bin}(n, p)$. Then, we claim that $X \preceq Y$. To see this, let $\hat{X} = \sum_{i=1}^{m} I_i$ and $\hat{Y} = \sum_{i=1}^{n} I_i$, where $\{I_i\}_{i=1}^{\infty}$ is an i.i.d. sequence of Bernoulli random variables, i.e.,

$$\mathbb{P}(I_i = 1) = 1 - \mathbb{P}(I_i = 0) = p, \qquad i = 1, \ldots, n, \tag{2.3.9}$$

and $I_1, I_2, \ldots, I_n$ are mutually independent. Then, since $I_i \geq 0$ for each i, we have that

$$\mathbb{P}(\hat{X} \leq \hat{Y}) = 1. \tag{2.3.10}$$

Therefore, $X \preceq Y$.

The stochastic domination above also holds when $X = \mathrm{Bin}(n - Z, p)$ and $Y = \mathrm{Bin}(n, p)$, when Z is any random variable that takes non-negative integer values. This domination result will prove to be useful in the investigation of the Erdős-Rényi random graph.

Poisson random variables. Another example of random variables which are stochastically ordered is as follows. Let $\lambda, \mu \in \mathbb{R}$ be such that $\lambda \leq \mu$. Let $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$. Then, $X \preceq Y$. To see this, let $\hat{X} \sim \text{Poi}(\lambda), \hat{Z} \sim \text{Poi}(\mu - \lambda)$, where $\hat{X}$ and $\hat{Z}$ are independent, and let $\hat{Y} = \hat{X} + \hat{Z}$. Then, $\hat{Y} \sim \text{Poi}(\mu)$. Moreover, since $\hat{Z} \geq 0$ for each i, we have that

$$\mathbb{P}(\hat{X} \leq \hat{Y}) = 1. \tag{2.3.11}$$

Therefore, $X \preceq Y$.

Exercise 2.15. *Let X and Y be normal distributions with equal variances σ^2 and means $\mu_X \leq \mu_Y$. Is $X \preceq Y$?*

Exercise 2.16. *Let X and Y be normal distributions with variances $\sigma_X^2 < \sigma_Y^2$ and equal means μ. Is $X \preceq Y$?*

2.3.1 Consequences of stochastic domination

In this section, we discuss a number of consequences of stochastic domination, such as the fact that the means of a stochastically ordered pair of random variables is ordered as well.

Theorem 2.12 (Ordering of means for stochastically ordered random variables)**.** *Suppose $X \preceq Y$. Then*

$$\mathbb{E}[X] \leq \mathbb{E}[Y]. \tag{2.3.12}$$

Proof. We apply Lemma 2.11. Let $\hat{X}$ and $\hat{Y}$ have the same law as X and Y, and be such that $\hat{X} \leq \hat{Y}$ with probability 1. Then

$$\mathbb{E}[X] = \mathbb{E}[\hat{X}] \leq \mathbb{E}[\hat{Y}] = \mathbb{E}[Y]. \tag{2.3.13}$$

□

Theorem 2.13 (Preservation of ordering under monotone functions)**.** *Suppose $X \preceq Y$, and $g \colon \mathbb{R} \to \mathbb{R}$ is non-decreasing. Then $g(X) \preceq g(Y)$.*

Proof. Let $\hat{X}$ and $\hat{Y}$ have the same laws as X and Y and be such that $\hat{X} \leq \hat{Y}$ (see Lemma 2.11). Then, $g(\hat{X})$ and $g(\hat{Y})$ have the same distributions as $g(X)$ and $g(Y)$, and $g(\hat{X}) \leq g(\hat{Y})$ with probability one, by the fact that g is non-decreasing. Therefore, by Lemma 2.11, the claim follows. □

2.4 Probabilistic bounds

We will often make use of a number of probabilistic bounds, which we will summarise and prove in this section.

Theorem 2.14 (Markov inequality)**.** *Let X be a non-negative random variable with $\mathbb{E}[X] < \infty$. Then,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{2.4.1}$$

In particular, when X is integer valued with $\mathbb{E}[X] \leq m$, then

$$\mathbb{P}(X = 0) \geq 1 - m. \tag{2.4.2}$$

By (2.4.2), if the integer random variable has a small mean, then it must be equal to 0 with high probability. This is called the *first moment method*, and is a powerful tool to prove results.

Proof. Equation (2.4.1) follows by

$$a\mathbb{P}(X \geq a) \leq \mathbb{E}[X\mathbb{1}_{\{X\geq a\}}] \leq \mathbb{E}[X]. \tag{2.4.3}$$

□

Theorem 2.15 (Chebychev inequality)**.** *Assume that X is integer valued with $\mathrm{Var}(X) = \sigma^2$. Then,*

$$\mathbb{P}\Big(\big|X - \mathbb{E}[X]\big| \geq a\Big) \leq \frac{\sigma^2}{a^2}. \tag{2.4.4}$$

In particular, when X is integer valued with $\mathbb{E}[X] \geq m$ and $\mathrm{Var}(X) = \sigma^2$, then

$$\mathbb{P}(X = 0) \leq \frac{\sigma^2}{m^2}. \tag{2.4.5}$$

By (2.4.5), if the integer random variable has a large mean, and a variance which is small compared to the square of the mean, then it must be positive with high probability. This is called the *second moment method*.

Proof. For (2.4.4), we note that

$$\mathbb{P}\Big(\big|X - \mathbb{E}[X]\big| \geq a\Big) = \mathbb{P}\Big((X - \mathbb{E}[X])^2 \geq a^2\Big), \tag{2.4.6}$$

and apply the Markov inequality. For (2.4.5), we note that

$$\mathbb{P}(X = 0) \leq \mathbb{P}\big(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]\big) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \leq \frac{\sigma^2}{m^2}. \tag{2.4.7}$$

□

We will often rely on bounds on the probability that a sum of independent random variables is larger than its expectation. For such probabilities, *large deviation theory* gives good bounds. We will describe these bounds here. For more background on large deviations, we refer the reader to [83, 122, 185].

Theorem 2.16 (Cramér's upper bound, Chernoff bound)**.** *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables. Then, for all $a \geq \mathbb{E}[X_1]$,*

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i \geq na\Big) \leq e^{-nI(a)}, \tag{2.4.8}$$

while, for all $a \leq \mathbb{E}[X_1]$,

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i \leq na\Big) \leq e^{-nI(a)}, \tag{2.4.9}$$

where, for $a \geq \mathbb{E}[X_1]$,

$$I(a) = \sup_{t\geq 0}\Big(ta - \log \mathbb{E}[e^{tX_1}]\Big), \tag{2.4.10}$$

while, for $a \leq \mathbb{E}[X_1]$,

$$I(a) = \sup_{t\leq 0}\Big(ta - \log \mathbb{E}[e^{tX_1}]\Big). \tag{2.4.11}$$

Note that the function $t \mapsto ta - \log \mathbb{E}[e^{tX_1}]$ is concave, and the derivative in 0 is $a - \mathbb{E}[X_1] \geq 0$ for $a \geq \mathbb{E}[X_1]$. Therefore, for $a \geq \mathbb{E}[X_1]$, the supremum of $t \mapsto (ta - \log \mathbb{E}[e^{tX_1}])$ will be attained for a $t \geq 0$ when $\mathbb{E}[e^{tX_1}]$ exists in a neighborhood of $t = 0$. As a result, (2.4.10)–(2.4.11) can be combined as

$$I(a) = \sup_t \Big(ta - \log \mathbb{E}[e^{tX_1}] \Big). \tag{2.4.12}$$

Proof. We only prove (2.4.8), the proof of (2.4.9) is identical when we replace X_i by $-X_i$. We rewrite, for every $t \geq 0$,

$$\mathbb{P}\Big(\sum_{i=1}^n X_i \geq na \Big) = \mathbb{P}\Big(e^{t \sum_{i=1}^n X_i} \geq e^{tna} \Big) \leq e^{-nta} \mathbb{E}\Big[e^{t \sum_{i=1}^n X_i} \Big], \tag{2.4.13}$$

where we have used Markov's inequality in Theorem 2.14. Since $\{X_i\}_{i=1}^\infty$ is a sequence of i.i.d. random variables, we have

$$\mathbb{E}\Big[e^{t \sum_{i=1}^n X_i} \Big] = \mathbb{E}[e^{tX_1}]^n, \tag{2.4.14}$$

so that, for every $t \geq 0$,

$$\mathbb{P}\Big(\sum_{i=1}^n X_i \geq na \Big) \leq \Big(e^{-ta} \mathbb{E}[e^{tX_1}] \Big)^n. \tag{2.4.15}$$

Minimizing the right-hand side over $t \geq 0$ gives that

$$\mathbb{P}\Big(\sum_{i=1}^n X_i \geq na \Big) \leq e^{-n \sup_{t \geq 0} \left(ta - \log \mathbb{E}[e^{tX_1}] \right)}. \tag{2.4.16}$$

This proves (2.4.8). □

Exercise 2.17. *Compute $I(a)$ for $\{X_i\}_{i=1}^\infty$ being independent Poisson random variables with mean λ. Show that, for $a > \lambda$,*

$$\mathbb{P}\Big(\sum_{i=1}^n X_i \geq na \Big) \leq e^{-nI_\lambda(a)}, \tag{2.4.17}$$

where $I_\lambda(a) = a(\log(a/\lambda) - 1) + \lambda$. Show also that, for $a < \lambda$

$$\mathbb{P}\Big(\sum_{i=1}^n X_i \leq na \Big) \leq e^{-nI_\lambda(a)}. \tag{2.4.18}$$

Prove that $I_\lambda(a) > 0$ for all $a \neq \lambda$.

2.4.1 Bounds on binomial random variables

In this section, we investigate the tails of the binomial distribution. We start by a corollary of Theorem 2.16:

Corollary 2.17 (Large deviations for binomial distribution)**.** *Let X_n be a binomial random variable with parameters p and n. Then, for all $a \in (p, 1]$,*

$$\mathbb{P}\Big(X_n \geq na\Big) \leq e^{-nI(a)}, \tag{2.4.19}$$

where

$$I(a) = a \log\Big(\frac{a}{p}\Big) + (1-a)\log\Big(\frac{1-a}{1-p}\Big). \tag{2.4.20}$$

Moreover,

$$I(a) \geq I_p(a) \tag{2.4.21}$$

where

$$I_\lambda(a) = \lambda - a - a\log(\lambda/a) \tag{2.4.22}$$

is the rate function of a Poisson random variable with mean λ.

We can recognize (2.4.22) as the rate function of a Poisson random variable with mean λ (recall Exercise 2.17). Thus, Corollary 2.17 suggests that the upper tail of a binomial random variable is thinner than the one of a Poisson random variable.

Proof. We start by proving (2.4.19), using (2.4.8). We note that, by (2.4.10), we obtain a bound with $I(a)$ instead of I_p, where, with $X_1 \sim \mathrm{Be}(p)$,

$$I(a) = \sup_{t\geq 0}\big(ta - \log \mathbb{E}[e^{tX_1}]\big) = \sup_t \Big(ta - \log\big(pe^t + (1-p)\big)\Big) = a\log\Big(\frac{a}{p}\Big) + (1-a)\log\Big(\frac{1-a}{1-p}\Big). \tag{2.4.23}$$

We note that, for $t \geq 0$,

$$pe^t + (1-p) = 1 + p(e^t - 1) \leq e^{p(e^t-1)}, \tag{2.4.24}$$

so that

$$I(a) \geq \sup_t \big(ta - p(e^t - 1)\big) = p - a - a\log\big(p/a\big) = I_p(a). \tag{2.4.25}$$

□

We continue to study tails of the binomial distribution, following [134]. The main bound is the following:

Theorem 2.18. *Let $X_i \sim \mathrm{Be}(p_i)$, $i = 1, 2, \ldots, n$, be independent Bernoulli distributed random variables, and write $X = \sum_{i=1}^n X_i$ and $\lambda = \mathbb{E}[X] = \sum_{i=1}^n p_i$. Then*

$$\mathbb{P}(X \geq \mathbb{E}[X] + t) \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right), \qquad t \geq 0; \tag{2.4.26}$$

$$\mathbb{P}(X \leq \mathbb{E}[X] - t) \leq \exp\left(-\frac{t^2}{2\lambda}\right), \qquad t \geq 0. \tag{2.4.27}$$

Further similar bounds under the same conditions and, even more generally, for independent random variables X_i such that $0 \leq X_i \leq 1$, are given, for example, in [26, 116] and [12, Appendix A].

Exercise 2.18. *Prove that Theorem 2.18 also holds for the Poisson distribution by a suitable limiting argument.*

Proof. Let $Y \sim \mathrm{Bin}(n, \lambda/n)$ where we recall that $\lambda = \sum_{i=1}^n p_i$. Since $x \mapsto \log x$ is concave, we have that for every $x_1, \dots, x_n \in \mathbb{R}$,

$$\sum_{i=1}^n \frac{1}{n} \log(x_i) \le \log\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big). \tag{2.4.28}$$

As a result, for every real u, upon taking the logarithm,

$$\begin{aligned}\mathbb{E}[e^{uX}] = \prod_{i=1}^n (1 + (e^u - 1)p_i) &= e^{n\sum_{i=1}^n \frac{1}{n}\log(1+(e^u-1)p_i)} \\ &\le e^{n\log(1+(e^u-1)\lambda/n)} = \Big(1 + (e^u-1)\lambda/n\Big)^n = \mathbb{E}[e^{uY}].\end{aligned} \tag{2.4.29}$$

Then we compute that, for all $u \ge 0$, by the Markov inequality,

$$\mathbb{P}(X \ge \mathbb{E}[X] + t) \le e^{-u(\mathbb{E}[X]+t)}\mathbb{E}[e^{uX}] \le e^{-u(\mathbb{E}[X]+t)}\mathbb{E}[e^{uY}] = e^{-u(\lambda+t)}(1 - p + pe^u)^n, \tag{2.4.30}$$

where $p = \lambda/n$ and using that $\mathbb{E}[X] = \lambda$.

When $t > n - \lambda$, the left-hand side of (2.4.30) equals 0, and there is nothing to prove. For $\lambda + t < n$, the right-hand side of (2.4.30) attains its minimum for the u satisfying $e^u = (\lambda + t)(1 - p)/(n - \lambda - t)p$. This yields, for $0 \le t \le n - \lambda$,

$$\mathbb{P}(X \ge \lambda + t) \le \left(\frac{\lambda}{\lambda + t}\right)^{\lambda+t} \left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t}. \tag{2.4.31}$$

The bound is implicit in [63] and is often called the *Chernoff bound*, appearing for the first time explicitly in [183].

For $0 \le t \le n - \lambda$,, we can rewrite (2.4.31) as

$$\mathbb{P}(X \ge \lambda + t) \le \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right) - (n-\lambda)\varphi\left(\frac{-t}{n-\lambda}\right)\right), \tag{2.4.32}$$

where $\varphi(x) = (1 + x)\log(1 + x) - x$ for $x \ge -1$ (and $\varphi(x) = \infty$ for $x < -1$). Replacing X by $n - X$, we also obtain, again for $0 \le t \le n - \lambda$,

$$\mathbb{P}(X \le \lambda - t) \le \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right) - (n-\lambda)\varphi\left(\frac{t}{n-\lambda}\right)\right). \tag{2.4.33}$$

Since $\varphi(x) \ge 0$ for every x we can ignore the second term in the exponent. Furthermore, $\varphi(0) = 0$ and $\varphi'(x) = \log(1 + x) \le x$, so that $\varphi(x) \ge x^2/2$ for $x \in [-1, 0]$, which proves (2.4.27). Similarly, $\varphi(0) = \varphi'(0) = 0$ and, for $x \in [0, 1]$,

$$\varphi''(x) = \frac{1}{1+x} \ge \frac{1}{(1+x/3)^3} = \left(\frac{x^2}{2(1+x/3)}\right)'', \tag{2.4.34}$$

so that $\varphi(x) \ge x^2/(2(1 + x/3))$, which proves (2.4.26). □

2.5 Martingales

In this section, we state and prove some results concerning martingales. These results will be used in the remainder of the text. For more details on martingales, we refer the reader to [111, 223].

We assume some familiarity with conditional expectations. For the readers who are unfamiliar with filtrations and conditional expectations given a σ-algebra, we start by giving the simplest case of a martingale:

Definition 2.19 (Martingale)**.** *A stochastic process* $\{M_n\}_{n=0}^{\infty}$ *is a martingale process if*

$$\mathbb{E}[|M_n|] < \infty \qquad \text{for all } n \geq 0, \tag{2.5.1}$$

and

$$\mathbb{E}[M_{n+1}|M_0, M_1, \ldots, M_n] = M_n \qquad \text{for all } n \geq 0. \tag{2.5.2}$$

As can be seen from (2.5.2), a martingale can be interpreted as a fair game. Indeed, when M_n denotes the profit after the n^{th} game has been played, then (2.5.2) tells us that the expected profit at time $n+1$ given the profits up to time n is equal to the profit at time n.

Exercise 2.19. *Show that when* $\{M_n\}_{n=0}^{\infty}$ *is a martingale process, then* $\mu = \mathbb{E}[M_n]$ *is independent of* n*.*

We now give a second definition, which we will need in Chapter 8, where a martingale is defined with respect to a more general filtration.

Definition 2.20 (Martingale definition general)**.** *A stochastic process* $\{M_n\}_{n=0}^{\infty}$ *is a martingale process with respect to* $\{X_n\}_{n=0}^{\infty}$ *if*

$$\mathbb{E}[|M_n|] < \infty \qquad \text{for all } n \geq 0, \tag{2.5.3}$$

M_n *is measurable with respect to the* σ*-algebra generated by* $(X_0, \ldots, X_n)$*, and*

$$\mathbb{E}[M_{n+1}|X_0, \ldots, X_n] = M_n \qquad \text{for all } n \geq 0. \tag{2.5.4}$$

For $X_n = M_n$, the definitions in (2.5.2) and (2.5.4) coincide.

Exercise 2.20. *Let* $\{X_i\}_{0=1}^{\infty}$ *be an independent sequence of random variables with* $\mathbb{E}[|X_i|] < \infty$ *and* $\mathbb{E}[X_i] = 1$*. Show that, for* $n \geq 0$*,*

$$M_n = \prod_{i=0}^{n} X_i \tag{2.5.5}$$

is a martingale.

Exercise 2.21. *Let* $\{X_i\}_{i=0}^{\infty}$ *be an independent sequence of random variables with* $\mathbb{E}[|X_i|] < \infty$ *and* $\mathbb{E}[X_i] = 0$*. Show that, for* $n \geq 0$*,*

$$M_n = \sum_{i=0}^{n} X_i \tag{2.5.6}$$

is a martingale.

Exercise 2.22. *Let* $M_n = \mathbb{E}[Y|X_0, \ldots, X_n]$ *for some random variables* $\{X_i\}_{i=0}^{\infty}$ *and* Y *with* $\mathbb{E}[|Y|] < \infty$ *and* $\{X_n\}_{n=0}^{\infty}$*. Show that* $\{M_n\}_{n=0}^{\infty}$ *is a martingale process with respect to* $\{X_n\}_{n=0}^{\infty}$*.* $\{M_n\}_{n=0}^{\infty}$ *is called a Doob martingale.*

In the following two sections, we state and prove two key results for martingales, the *martingale convergence theorem* and the *Azuma-Hoeffding inequality*. These results are a sign of the power of martingales. Martingale techniques play a central role in modern probability theory, partly due to these results.

2.5.1 Martingale convergence theorem

We start with the martingale convergence theorem:

Theorem 2.21 (Martingale convergence theorem)**.** *Let $\{M_n\}_{n=0}^{\infty}$ be a martingale process with respect to $\{X_n\}_{n=0}^{\infty}$ satisfying*

$$\mathbb{E}[|M_n|] \leq B \qquad \text{for all } n \geq 0. \tag{2.5.7}$$

Then, $M_n \xrightarrow{a.s.} M_\infty$, for some limiting random variable M_∞ which is finite with probability 1.

The martingale convergence theorem comes in various forms. There also is an L^2-version, for which it is assumed that $\mathbb{E}[M_n^2] \leq M$ uniformly for all $n \geq 1$. In this case, one also obtains the convergence $\lim_{n\to\infty} \mathbb{E}[M_n^2] = \mathbb{E}[M_\infty^2]$. Theorem 2.21 is an adaptation of the L^1-martingale convergence theorem, for which one only needs that $\{M_n\}_{n=0}^{\infty}$ is a *submartingale*, i.e., when we assume (2.5.7), but (2.5.4) is replaced with

$$\mathbb{E}[M_{n+1}|X_0, \ldots, X_n] \geq M_n \qquad \text{for all } n \geq 0. \tag{2.5.8}$$

See e.g., [111, Section 12.3].

Exercise 2.23. *Prove that when the martingale $\{M_n\}_{n=0}^{\infty}$ is non-negative, i.e., when $M_n \geq 0$ with probability 1 for all $n \geq 1$, then $M_n \xrightarrow{a.s.} M_\infty$ to some limiting random variable M_∞ which is finite with probability 1.*

Exercise 2.24. *Let $\{X_i\}_{i=0}^{\infty}$ be an independent sequence of random variables with $\mathbb{E}[X_i] = 1$ and for which $X_i \geq 0$ with probability 1. Show that the martingale*

$$M_n = \prod_{i=0}^{n} X_i \tag{2.5.9}$$

converges in probability to a random variable which is finite with probability 1. Hint: Prove that $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = 1$, and apply Exercise 2.23.

Exercise 2.25. *For $i = 1, \ldots, m$, let $\{M_n^{(i)}\}_{n=0}^{\infty}$ be a sequence of martingales with respect to $\{X_n\}_{n=0}^{\infty}$. Show that*

$$M_n = \max_{i=0}^{m} M_n^{(i)} \tag{2.5.10}$$

is a submartingale with respect to $\{X_n\}_{n=0}^{\infty}$.

Proof of Theorem 2.21. We shall prove Theorem 2.21 in the case where M_n is a *submartingale.*

We follow the proof of the martingale convergence theorem in [111, Section 12.3]. The key step in this classical probabilistic proof is 'Snell's up-crossings inequality'. Suppose that $\{m_n : n \geq 0\}$ is a real sequence, and $[a, b]$ is a real interval. An *up-crossing* of $[a, b]$ is defined to be a crossing by m of $[a, b]$ in the upwards direction. More precisely, let $T_1 = \min\{n : m_n \leq a\}$, the first time m hits the interval $(-\infty, a]$, and $T_2 = \min\{n > T_1 : m_n \geq b\}$, the first subsequent time when m hits $[b, \infty)$; we call the interval $[T_1, T_2]$ an up-crossing of $[a, b]$. In addition, for $k > 1$, define the stopping times T_n by

$$T_{2k-1} = \min\{n > T_{2k-2} : m_n \leq a\}, \qquad T_{2k} = \min\{n > T_{2k-1} : m_n \geq b\}, \tag{2.5.11}$$

so that the number of up-crossings of $[a, b]$ is equal to the number of intervals $[T_{2k-1}, T_{2k}]$ for $k \geq 1$. Let $U_n(a, b; m)$ be the number of up-crossings of $[a, b]$ by the subsequence $m_0, m_1, \ldots, m_n$, and let $U(a, b; m) = \lim_{n\to\infty} U_n(a, b; m)$ be the total number of up-crossings of m.

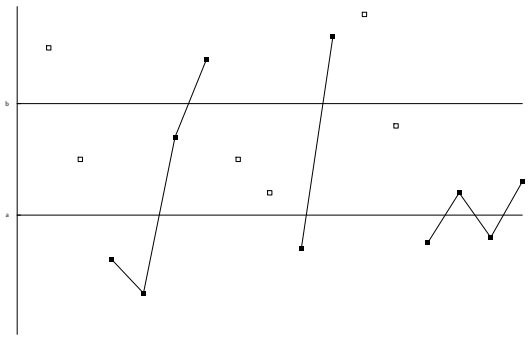

Figure 2.1: Up-crossings

Let $\{M_n\}_{n=0}^{\infty}$ be a submartingale, and let $U_n(a,b;M)$ be the number of up-crossings of $[a,b]$ by M up to time n. Then the up-crossing inequality gives a bound on the expected number of up-crossings of an interval $[a,b]$:

Proposition 2.22 (Up-crossing inequality)**.** *If $a<b$ then*

$$\mathbb{E}[U_n(a,b;M)] \leq \frac{\mathbb{E}[(M_n-a)^+]}{b-a},$$

where $(M_n-a)^+ = \max\{0, M_n-a\}$.

Proof. Setting $Z_n = (M_n-a)^+$, we have that Z_n is a non-negative submartingale because $\mathbb{E}[|M_n|] \leq \mathbb{E}[|M_n|] + |a| < \infty$. Furthermore, for every random variable X and $a \in \mathbb{R}$,

$$\mathbb{E}[(X-a)_+] \geq \mathbb{E}[X-a]_+, \tag{2.5.12}$$

so that

$$Z_n \leq \Big(\mathbb{E}[M_{n+1}|X_0,\ldots,X_n]-a\Big)_+ \leq \mathbb{E}[(M_{n+1}-a)_+|X_0,\ldots,X_n] = \mathbb{E}[Z_{n+1}|X_0,\ldots,X_n], \tag{2.5.13}$$

where we first used the submartingale property $\mathbb{E}[M_{n+1}|X_0,\ldots,X_n] \geq M_n$, followed by (2.5.12). Up-crossings of $[a,b]$ by M correspond to up-crossings of $[0,b-a]$ by Z, so that $U_n(a,b;M) = U_n(0,b-a;Z)$.

Let $[T_{2k-1}, T_{2k}]$, for $k \geq 1$, be the up-crossings of Z of $[0, b-a]$, and define the indicator functions

$$I_i = \begin{cases} 1 & \text{if } i \in (T_{2k-1}, T_{2k}] \text{ for some } k, \\ 0 & \text{otherwise} \end{cases} \tag{2.5.14}$$

Note that the event $\{I_i = 1\}$ depends on $M_0, M_1, \ldots, M_{i-1}$ only. Since $M_0, M_1, \ldots, M_{i-1}$ are measurable with respect to the the σ-algebra generated by $(X_0, \ldots, X_{i-1})$, also I_i is measurable with respect to the σ-algebra generated by $(X_0, \ldots, X_{i-1})$. Now

$$(b-a)U_n(0, b-a; Z) \leq \sum_{i=1}^{n} (Z_i - Z_{i-1}) I_i \tag{2.5.15}$$

since each up-crossing of $[0, b-a]$ by Z contributes an amount of at least $b-a$ to the summation. The expectation of the summands on the right-hand side of (2.5.15) is equal to

$$\begin{aligned} \mathbb{E}[(Z_i - Z_{i-1})I_i] =& \mathbb{E}\Big[\mathbb{E}\big[(Z_i - Z_{i-1})I_i | X_0, \ldots, X_{i-1}\big]\Big] = \mathbb{E}[I_i(\mathbb{E}[Z_i | X_0, \ldots, X_{i-1}] - Z_{i-1})] \\ \leq& \mathbb{E}[\mathbb{E}[Z_i | X_0, \ldots, X_{i-1}] - Z_{i-1}] = \mathbb{E}[Z_i] - \mathbb{E}[Z_{i-1}], \end{aligned}$$

where we use that I_i is measurable with respect to the σ-algebra generated by $(X_0, \ldots, X_{i-1})$ for the second equality, and we use that Z is a submartingale and $0 \leq I_i \leq 1$ to obtain the inequality. Summing over i and take expectations on both sides of (2.5.15), we obtain

$$(b-a)\mathbb{E}[U_n(0, b-a; Z)] \leq \mathbb{E}[Z_n] - \mathbb{E}[Z_0] \leq \mathbb{E}[Z_n], \tag{2.5.16}$$

which completes the proof of Proposition 2.22. □

Now we have the tools to give the proof of Theorem 2.21:

Proof of Theorem 2.21. Suppose $\{M_n\}_{n=0}^{\infty}$ is a submartingale and $\mathbb{E}[|M_n|] \leq B$ for all n. Let Λ be defined as follows

$$\Lambda = \{\omega : M_n(\omega) \text{ does not converge to a limit in } [-\infty, \infty]\}.$$

The claim that M_n converges is proved if we show that $\mathbb{P}(\Lambda) = 0$. The set Λ has an equivalent definition

$$\begin{aligned} \Lambda &= \{\omega : \liminf M_n(\omega) < \limsup M_n(\omega)\} \\ &= \bigcup_{a,b \in \mathbb{Q}: a<b} \{\omega : \liminf M_n(\omega) < a < b < \limsup M_n(\omega)\} \\ &= \bigcup_{a,b \in \mathbb{Q}: a<b} \Lambda_{a,b}. \end{aligned}$$

However,

$$\Lambda_{a,b} \subseteq \{\omega : U(a, b; M) = \infty\},$$

so that, by Proposition 2.22, $\mathbb{P}(\Lambda_{a,b}) = 0$. Since Λ is a countable union of sets $\Lambda_{a,b}$, it follows that $\mathbb{P}(\Lambda) = 0$. This concludes the first part of the proof that M_n converges almost surely to a limit M_∞.

To show that the limit is bounded, we use Fatou's lemma (see Theorem A.12 in the appendix) to conclude

$$\mathbb{E}[|M_\infty|] = \mathbb{E}[\liminf_{n\to\infty} |M_n|] \leq \liminf_{n\to\infty} \mathbb{E}[|M_n|] \leq \sup_{n \geq 0} \mathbb{E}[|M_n|] < \infty,$$

so that, by Markov's inequality (recall Theorem 2.14),

$$\mathbb{P}(M_\infty < \infty) = 1.$$

This completes the proof of Theorem 2.21. □

2.5.2 Azuma-Hoeffding inequality

We continue with the Azuma-Hoeffding inequality, which provides exponential bounds for the tails of a special class of martingales:

Theorem 2.23 (Azuma-Hoeffding inequality)**.** *Let $\{M_n\}_{n=0}^{\infty}$ be a martingale process with the property that, with probability 1, there exists $K_n \geq 0$ such that*

$$|M_n - M_{n-1}| \leq K_n \qquad \text{for all } n \geq 0, \tag{2.5.17}$$

where, by convention, we define $M_{-1} = \mu = \mathbb{E}[M_n]$ (recall also Exercise 2.19). Then, for every $a \geq 0$,

$$\mathbb{P}(|M_n - \mu| \geq a) \leq 2\exp\Big\{ - \frac{a^2}{2\sum_{i=0}^{n} K_i^2}\Big\}. \tag{2.5.18}$$

Theorem 2.23 is very powerful, as it provides tails on the distribution of M_n. In many cases, the bounds are close to optimal. The particular strength of Theorem 2.23 is that the bound is valid *for all* $n \geq 1$.

Proof. For $\psi > 0$, the function $g(d) = e^{\psi d}$ is convex, so that, for all d with $|d| \leq 1$,

$$e^{\psi d} \leq \frac{1}{2}(1-d)e^{-\psi} + \frac{1}{2}(1+d)e^{\psi}. \tag{2.5.19}$$

Applying this with $d = D$ to a random variable D having mean 0 and satisfying $\mathbb{P}(|D| \leq 1) = 1$, we obtain

$$\mathbb{E}[e^{\psi D}] \leq \mathbb{E}[\frac{1}{2}(1-D)e^{-\psi} + \frac{1}{2}(1+D)e^{\psi}] = \frac{1}{2}(e^{-\psi} + e^{\psi}). \tag{2.5.20}$$

We can use that $(2n)! \geq 2^n n!$ for all $n \geq 0$ to obtain that

$$\frac{1}{2}\ (e^{-\psi} + e^{\psi}) = \sum_{n\geq 0} \frac{\psi^{2n}}{(2n)!} \leq \sum_{n\geq 0} \frac{\psi^{2n}}{2^n n!} = e^{\psi^2/2}. \tag{2.5.21}$$

By Markov's inequality in Theorem 2.14, for any $\theta > 0$,

$$\mathbb{P}(M_n - \mu \geq x) = \mathbb{P}(e^{\theta(M_n-\mu)} \geq e^{\theta x}) \leq e^{-\theta x}\mathbb{E}[e^{\theta(M_n-\mu)}]. \tag{2.5.22}$$

Writing $D_n = M_n - M_{n-1}$, we obtain

$$\mathbb{E}[e^{\theta(M_n-\mu)}] = \mathbb{E}[e^{\theta(M_{n-1}-\mu)}e^{\theta D_n}].$$

Conditioning on $X_0, \ldots, X_{n-1}$ yields

$$\mathbb{E}[e^{\theta(M_n-\mu)} \mid X_0, \ldots, X_{n-1}] = e^{\theta(M_{n-1}-\mu)}\mathbb{E}[e^{\theta D_n} \mid X_0, \ldots, X_{n-1}] \leq e^{\theta(M_{n-1}-\mu)}\exp(\frac{1}{2}\,\theta^2 K_n^2), \tag{2.5.23}$$

where (2.5.20) and (2.5.21) are applied to the random variable D_n/K_n which satisfies

$$\mathbb{E}[D_n|X_0, \ldots, X_{n-1}] = \mathbb{E}[M_n|X_0, \ldots, X_{n-1}] - \mathbb{E}[M_{n-1}|X_0, \ldots, X_{n-1}] = M_{n-1} - M_{n-1} = 0. \tag{2.5.24}$$

Taking expectations on both sides of (2.5.23) and iterate to find

$$\mathbb{E}[e^{\theta(M_n-\mu)}] \leq \mathbb{E}[e^{\theta(M_{n-1}-\mu)}]\exp(\frac{1}{2}\,\theta^2 K_n^2) \leq \exp\left(\frac{1}{2}\,\theta^2 \sum_{i=0}^{n} K_i^2\right). \tag{2.5.25}$$

Therefore, by (2.5.22), for all $\theta > 0$,

$$\mathbb{P}(M_n - \mu \geq x) \leq \exp\left(-\theta x + \frac{1}{2}\,\theta^2 \sum_{i=0}^{n} K_i^2\right). \tag{2.5.26}$$

The exponential is minimized, with respect to θ, by setting $\theta = x/\sum_{i=0}^{n} K_i^2$. Hence,

$$\mathbb{P}(M_n - \mu \geq x) \leq \exp\left(-\frac{x^2}{\sum_{i=0}^{n} K_i^2}\right). \tag{2.5.27}$$

Using that also $-M_n$ is a martingale, we obtain by symmetry that

$$\mathbb{P}(M_n - \mu \leq -x) \leq \exp\left(-\frac{x^2}{\sum_{i=0}^{n} K_i^2}\right). \tag{2.5.28}$$

Adding the two bounds completes the proof. □

Exercise 2.26. *Show that Theorem 2.23 implies that for $X \sim \mathrm{Bin}(n,p)$ with $p \leq 1/2$*

$$\mathbb{P}(|X - np| \geq a) \leq 2\exp\Big\{-\frac{a^2}{2n(1-p)^2}\Big\}. \tag{2.5.29}$$

Exercise 2.27. *Let $\{X_i\}_{i=0}^{\infty}$ be an independent identically distributed sequence of random variables with $\mathbb{E}[X_i] = 0$ and $|X_i| \leq 1$, and define the martingale $\{M_n\}_{n=0}^{\infty}$ by*

$$M_n = \sum_{i=0}^{n} X_i. \tag{2.5.30}$$

Show that

$$\mathbb{P}(|M_n| \geq a) \leq 2\exp\Big(-\frac{a^2}{2n}\Big). \tag{2.5.31}$$

Take $a = x\sqrt{n}$, and prove by using the central limit theorem that $\mathbb{P}(|M_n| \geq a)$ converges. Compare the limit to the bound in (2.5.31).

2.6 Order statistics and extreme value theory

In this section, we study the largest values of a sequence of i.i.d. random variables. For more background on order statistics, we refer the reader to [93]. We will be particularly interested in the case where the random variables in question have heavy tails. We let $\{X_i\}_{i=1}^{n}$ be an i.i.d. sequence, and introduce the *order statistics* of $\{X_i\}_{i=1}^{n}$ by

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}, \tag{2.6.1}$$

so that $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(2)}$ is the second smallest of $\{X_i\}_{i=1}^{n}$, etc. In the notation in (2.6.1), we ignore the fact that the distribution of $X_{(i)}$ depends on n. Sometimes the notation $X_{(1:n)} \leq X_{(2:n)} \leq \cdots \leq X_{(n:n)}$ is used instead to make the dependence on n explicit. In this section, we shall mainly investigate $X_{(n)}$, i.e., the maximum of $X_1, \ldots, X_n$. We note that the results immediately translate to $X_{(1)}$, by changing to $-X_i$.

We denote the distribution function of the random variables $\{X_i\}_{i=1}^{n}$ by

$$F_X(x) = \mathbb{P}(X_1 \leq x). \tag{2.6.2}$$

Before stating the results, we introduce a number of special distributions. We say that the random variable Y has a *Fréchet distribution* if there exists an $\alpha > 0$ such that

$$\mathbb{P}(Y \leq y) = \begin{cases} 0, & y \leq 0, \\ \exp\{-y^{-\alpha}\} & y > 0. \end{cases} \tag{2.6.3}$$

We say that the random variable Y has a *Weibull distribution* if there exists an $\alpha > 0$ such that

$$\mathbb{P}(Y \leq y) = \begin{cases} \exp\{-(-y)^{\alpha}\}, & y \leq 0, \\ 1 & y > 0. \end{cases} \tag{2.6.4}$$

We say that the random variable Y has a *Gumbel distribution* if

$$\mathbb{P}(Y \leq y) = \exp\{-\exp\{-y\}\}, \qquad y \in \mathbb{R}. \tag{2.6.5}$$

One of the fundamental results in extreme value theory is the following characterization of possible limit distributions of $X_{(n)}$:

Theorem 2.24 (Fisher-Tippett theorem, limit laws for maxima)**.** *Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of i.i.d. random variables. If there exists norming constants $c_n > 0$ and $d_n \in \mathbb{R}$ and some non-degenerate distribution function H such that*

$$\frac{X_{(n)} - c_n}{d_n} \xrightarrow{d} Y, \tag{2.6.6}$$

where Y has distribution function H, then H is the distribution function of a Fréchet, Weibull or Gumbel distribution.

A fundamental role in extreme value statistics is played by approximate solutions u_n of $[1 - F_X(u_n)] = 1/n$. More precisely, we define u_n by

$$u_n = \inf\{u : 1 - F_X(u) \geq 1/n\}. \tag{2.6.7}$$

We shall often deal with random variables which have a power-law distribution. For such random variables, the following theorem identifies the Fréchet distribution as the only possible extreme value limit:

Theorem 2.25 (Maxima of heavy-tailed random variables)**.** *Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of i.i.d. unbounded random variables satisfying*

$$1 - F_X(x) = x^{1-\tau} L_X(x), \tag{2.6.8}$$

where $x \mapsto L_X(x)$ is a slowly varying function, and where $\tau > 1$. Then

$$\frac{X_{(n)}}{u_n} \xrightarrow{d} Y, \tag{2.6.9}$$

where Y has a Fréchet distribution with parameter $\alpha = \tau - 1$ and u_n is defined in (2.6.7)*.*

Exercise 2.28. *Show that when* (2.6.8) *holds, then u_n is regularly varying with exponent $\frac{1}{\tau - 1}$.*

For completeness, we also state two theorems identifying when the Weibull distribution or Gumbel distribution occur as the limiting distribution in extreme value theory:

Theorem 2.26 (Maxima of bounded random variables)**.** *Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of i.i.d. random variables satisfying that $F_X(x_X) = 1$ for some $x_X \in \mathbb{R}$ and*

$$1 - F_X(x_X - x^{-1}) = x^{-\alpha} L_X(x), \tag{2.6.10}$$

where $x \mapsto L_X(x)$ is a slowly varying function, and where $\alpha > 1$. Then

$$\frac{X_{(n)} - x_X}{d_n} \xrightarrow{d} Y, \tag{2.6.11}$$

where Y has a Weibull distribution with parameter α, and $d_n = x_X - u_n$ where u_n is defined in (2.6.7)*.*

Theorem 2.27 (Maxima of random variables with thin tails)**.** *Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of i.i.d. bounded random variables satisfying that $F(x_F) = 1$ for some $x_F \in [0, \infty]$, and*

$$\lim_{x \uparrow x_F} \frac{1 - F(x + ta(x))}{1 - F(x)} = e^{-t}, \qquad t \in \mathbb{R}, \tag{2.6.12}$$

where $x \mapsto a(x)$ is given by

$$a(x) = \int_x^{x_F} \frac{1 - F(t)}{1 - F(x)} dt. \tag{2.6.13}$$

Then

$$\frac{X_{(n)} - u_n}{d_n} \xrightarrow{d} Y, \tag{2.6.14}$$

where Y has a Gumbel distribution, and $d_n = a(u_n)$ where u_n is defined in (2.6.7)*.*

We next assume that the random variables $\{X_i\}_{i=1}^n$ have *infinite mean*. It is well known that the order statistics of the random variables, as well as their sum, are governed by u_n in the case that $\tau \in (1, 2)$. The following theorem shows this in detail. In the theorem below, $E_1, E_2, \ldots$ is an i.i.d. sequence of exponential random variables with unit mean and $\Gamma_j = E_1 + E_2 + \ldots + E_j$, so that Γ_j has a Gamma distribution with parameters j and 1.

It is well known that when the distribution function F of $\{X_i\}_{i=1}^n$ satisfies (2.6.8), then $\sum_{i=1}^n X_i$ has size approximately $n^{1/(\tau-1)}$, just as holds for the maximum, and the rescaled sum $n^{-1/(\tau-1)} \sum_{i=1}^n X_i$ converges to a stable distribution. The next result generalizes this statement to convergence of the sum together with the first order statistics:

Theorem 2.28 (Convergence in distribution of order statistics and sum)**.** *$\{X_n\}_{n=0}^{\infty}$ be a sequence of i.i.d. random variables satisfying* (2.6.8) *for some $\tau \in (1, 2)$. Then, for any $k \in \mathbb{N}$,*

$$\left(\frac{L_n}{u_n}, \left\{ \frac{X_{(n+1-i)}}{u_n} \right\}_{i=1}^n \right) \xrightarrow{d} \left(\eta, \{\xi_i\}_{i=1}^{\infty} \right), \text{ as } n \to \infty, \tag{2.6.15}$$

where $(\eta, \{\xi_i\}_{i=1}^{\infty})$ is a random vector which can be represented by

$$\eta = \sum_{j=1}^{\infty} \Gamma_j^{-1/(\tau-1)}, \qquad \xi_i = \Gamma_i^{-1/(\tau-1)}, \tag{2.6.16}$$

and where u_n is slowly varying with exponent $1/(\tau - 1)$ (recall Exercise 2.28). Moreover,

$$\xi_k k^{1/(\tau-1)} \xrightarrow{\mathbb{P}} 1 \text{ as } k \to \infty. \tag{2.6.17}$$

Proof. Because $\tau - 1 \in (0,1)$, the proof is a direct consequence of [157, Theorem 1'], and the continuous mapping theorem [37, Theorem 5.1], which together yield that on $\mathbb{R} \times \mathbb{R}^\infty$, equipped with the product topology, we have

$$(S_n^\#, Z^{(n)}) \xrightarrow{d} (S^\#, Z), \tag{2.6.18}$$

where $S_n^\# = u_n^{-1} L_n$, $Z^{(n)} = u_n^{-1}(D_{(n:n)}, \ldots, D_{(1:n)}, 0, 0, \ldots)$, and $Z_j = \Gamma_j^{-1/(\tau-1)}$, $j \geq 1$. Finally, (2.6.17) follows because by the weak law of large numbers,

$$\frac{\Gamma_k}{k} \xrightarrow{\mathbb{P}} 1, \tag{2.6.19}$$

and $\xi_k = \Gamma_k^{-1/(\tau-1)}$. □

Interestingly, much can be said about the random probability distribution $P_i = \xi_i/\eta$, which is called the *Poisson-Dirichlet distribution* (see e.g., [190]). For example, [190, Eqn. (10)] proves that for any $f\colon [0,1] \to \mathbb{R}$, and with $\alpha = \tau - 1 \in (0,1)$,

$$\mathbb{E}\big[\sum_{i=1}^{\infty} f(P_i)\big] = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \int_0^1 f(u) u^{-\alpha-1}(1-u)^{\alpha-1} du. \tag{2.6.20}$$

2.7 Notes and discussion

Notes on Section 2.1. For a through discussion on convergence issues of integer random variables including Theorems 2.4–2.6 and much more, see [44, Section 1.4].

Notes on Section 2.4. Theorem 2.16 has a long history. See e.g., [83, Theorem 2.2.3] for a more precise version of Cramér's Theorem, which states that (2.4.8)–(2.4.9) are sharp, in the sense that $-\frac{1}{n}\log \mathbb{P}(\frac{1}{n}\sum_{i=1}^n X_i \leq a)$ converges to $I(a)$. See [185, Theorem 1.1] for a version of Cramér's Theorem that includes also the Chernoff bound.

Notes on Section 2.5. This discussion is adapted after [111]. For interesting examples of martingale argument, as well as adaptations of the Azuma-Hoeffding inequality in Theorem 2.23, see [71].

Notes on Section 2.6. Theorem 2.24 is [93, Theorem 3.2.3]. Theorem 2.25 is [93, Theorem 3.3.7]. Theorem 2.26 is [93, Theorem 3.3.12]. Theorem 2.27 is [93, Theorem 3.3.27]. For a thorough discussion of extreme value results, as well as many examples, we refer to the standard work on the topic [93].

Chapter 3
Branching processes

Branching processes will be used in an essential way throughout these notes to describe the connected components of various random graphs. To prepare for this, we describe branching processes in quite some detail here. Special attention will be given to branching processes with a Poisson offspring distribution, as well as to branching processes with a binomial offspring distribution and their relation (see Sections 3.6 and 3.7 below). We start by describing the survival versus extinction transition in Section 3.1, and provide a useful random walk perspective on branching processes in Section 3.3. For more information about branching processes, we refer to the books [16, 115, 123].

3.1 Survival versus extinction

A branching process is the simplest possible model for a population evolving in time. Suppose each organism independently gives birth to a number of children with the same distribution. We denote the offspring distribution by $\{p_i\}_{i=0}^{\infty}$, where

$$p_i = \mathbb{P}(\text{individual has } i \text{ children}). \tag{3.1.1}$$

We denote by Z_n the number of individuals in the n^{th} generation, where, by convention, we let $Z_0 = 1$. Then Z_n satisfies the recursion relation

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i}, \tag{3.1.2}$$

where $\{X_{n,i}\}_{n,i\geq 1}$ is a doubly infinite array of i.i.d. random variables. We will often write X for the offspring distribution, so that $\{X_{n,i}\}_{n,i\geq 1}$ is a doubly infinite array of i.i.d. random variables with $X_{n,i} \sim X$ for all n, i.

One of the major results of branching processes is that when $\mathbb{E}[X] \leq 1$, the population dies out with probability one (unless $X_{1,1} = 1$ with probability one), while if $\mathbb{E}[X] > 1$, there is a non-zero probability that the population will not become extinct. In order to state the result, we denote the extinction probability by

$$\eta = \mathbb{P}(\exists n : Z_n = 0). \tag{3.1.3}$$

Theorem 3.1 (Survival v.s. extinction for branching processes)**.** *For a branching process with i.i.d. offspring X, $\eta = 1$ when $\mathbb{E}[X] < 1$, while $\eta < 1$ when $\mathbb{E}[X] > 1$. When $\mathbb{E}[X] = 1$, and $\mathbb{P}(X = 1) < 1$, then $\eta = 1$. Moreover, with G_X the probability generating function of the offspring distribution X, i.e.,*

$$G_X(s) = \mathbb{E}[s^X], \tag{3.1.4}$$

the extinction probability η is the smallest solution in $[0, 1]$ of

$$\eta = G_X(\eta). \tag{3.1.5}$$

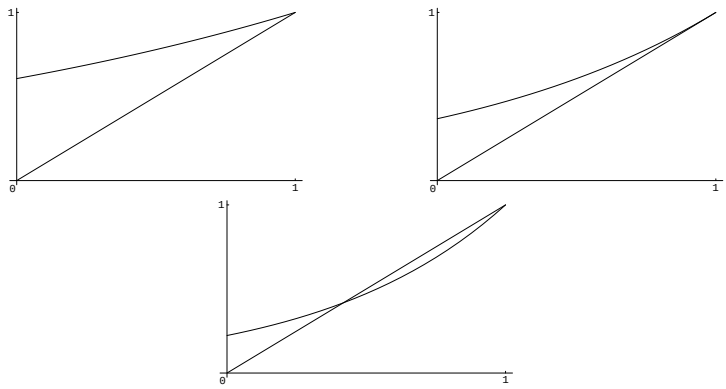

Figure 3.1: The solution of $s = G_X(s)$ when $\mathbb{E}[X] < 1, \mathbb{E}[X] = 1, \mathbb{E}[X] > 1$ respectively. Note that $\mathbb{E}[X] = G'_X(1)$, and $G'_X(1) > 1$ precisely when there is a solution $\eta < 1$ to $\eta = G_X(\eta)$.

Proof. We write

$$\eta_n = \mathbb{P}(Z_n = 0). \tag{3.1.6}$$

Because $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$, we have that $\eta_n \uparrow \eta$. Let

$$G_n(s) = \mathbb{E}[s^{Z_n}] \tag{3.1.7}$$

denote the generating function of the n^{th} generation. Then, since for an integer-valued random variable X, $\mathbb{P}(X = 0) = G_X(0)$, we have that

$$\eta_n = G_n(0). \tag{3.1.8}$$

By conditioning on the first generation, we obtain that

$$G_n(s) = \mathbb{E}[s^{Z_n}] = \sum_{i=0}^{\infty} p_i \mathbb{E}[s^{Z_n} | Z_1 = i] = \sum_{i=0}^{\infty} p_i G_{n-1}(s)^i. \tag{3.1.9}$$

Therefore, writing $G_X = G_1$ for the generating function of $X_{1,1}$, we have that

$$G_n(s) = G_X(G_{n-1}(s)). \tag{3.1.10}$$

When we substitute $s = 0$, we obtain that η_n satisfies the recurrence relation

$$\eta_n = G_X(\eta_{n-1}). \tag{3.1.11}$$

See Figure 3.2 for the evolution of $n \mapsto \eta_n$.

When $n \to \infty$, we have that $\eta_n \uparrow \eta$, so that, by continuity of $s \mapsto G_X(s)$, we have

$$\eta = G_X(\eta). \tag{3.1.12}$$

When $\mathbb{P}(X = 1) = 1$, then $Z_n = 1$ a.s., and there is nothing to prove. When, further, $\mathbb{P}(X \leq 1) = 1$, but $p = \mathbb{P}(X = 0) > 0$, then $\mathbb{P}(Z_n = 0) = 1 - (1 - p)^n \to 1$, so again there is nothing to prove. Therefore, for the remainder of this proof, we shall assume that $\mathbb{P}(X \leq 1) < 1$.

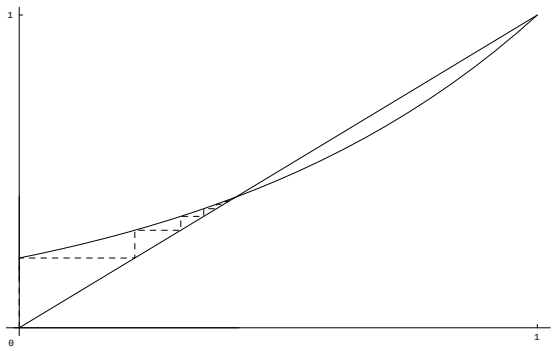

Figure 3.2: The iteration for $n \mapsto \eta_n$ in (3.1.11).

Suppose that $\psi \in [0,1]$ satisfies $\psi = G_X(\psi)$. We claim that $\eta \leq \psi$. We use induction to prove that $\eta_n \leq \psi$ for all n. Indeed, $\eta_0 = 0 \leq \psi$, which initializes the induction hypothesis. To advance the induction, we use (3.1.11), the induction hypothesis, as well as the fact that $s \mapsto G_X(s)$ is increasing on $[0,1]$, to see that

$$\eta_n = G_X(\eta_{n-1}) \leq G_X(\psi) = \psi, \tag{3.1.13}$$

where the final conclusion comes from the fact that ψ is a solution of $\psi = G_X(\psi)$. Therefore, $\eta_n \leq \psi$, which advances the induction. Since $\eta_n \uparrow \eta$, we conclude that $\eta \leq \psi$ for all solutions ψ of $\psi = G_X(\psi)$. Therefore, η is the smallest such solution.

We note that $s \mapsto G_X(s)$ is increasing and convex for $s \geq 0$, since

$$G_X''(s) = \mathbb{E}[X(X-1)s^{X-2}] \geq 0. \tag{3.1.14}$$

When $\mathbb{P}(X \leq 1) < 1$, then $\mathbb{E}[X(X-1)s^{X-2}] > 0$, so that $s \mapsto G_X(s)$ is strictly increasing and strictly convex for $s > 0$. Therefore, there can be at most two solutions of $s = G_X(s)$ in $[0,1]$. Note that $s = 1$ is always a solution of $s = G_X(s)$, since G is a probability generating function. Since $G_X(0) > 0$, there is precisely one solution when $G_X'(1) < 1$, while there are two solutions when $G_X'(1) > 1$. The former implies that $\eta = 1$ when $G_X'(1) > 1$, while the latter implies that $\eta < 1$ when $G_X'(1) < 1$. When $G_X'(1) = 1$, again there is precisely one solution, except when $G_X(s) = s$, which is equivalent to $\mathbb{P}(X = 1) = 1$. Since $G_X'(1) = \mathbb{E}[X]$, this proves the claim. □

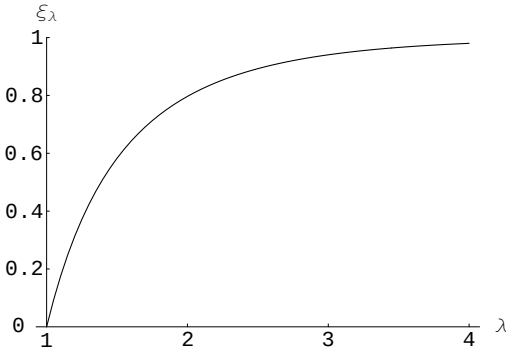

Figure 3.3: The survival probability $\zeta = \zeta_\lambda$ for a Poisson branching process with mean offspring equal to λ. The survival probability equals $\zeta = 1 - \eta$, where η is the extinction probability.

In many cases, we shall be interested in the *survival probability*, denoted by $\zeta = 1-\eta$, which is the probability that the branching process survives forever, i.e., $\zeta = \mathbb{P}(Z_n > 0 \ \forall n \geq 0)$. See Figure 3.3 for the survival probability of a Poisson branching process with parameter λ, as a function of λ.

Exercise 3.1. *Show that $\eta = 0$ precisely when $p_0 = 0$.*

Exercise 3.2. *When the offspring distribution is given by*

$$p_x = (1-p)\mathbb{1}_{\{x=0\}} + p\mathbb{1}_{\{x=2\}}, \tag{3.1.15}$$

we speak of binary branching. Prove that $\eta = 1$ when $p \leq 1/2$ and, for $p > 1/2$,

$$\eta = \frac{1-p}{p}. \tag{3.1.16}$$

Exercise 3.3 ([16], Pages 6-7.)**.** *Let the probability distribution $\{p_k\}_{k=0}^{\infty}$ be given by*

$$\begin{cases} p_k = b(1-p)^{k-1} & \text{for } k = 1, 2, \ldots; \\ p_0 = 1 - b/p & \text{for } k = 0, \end{cases} \tag{3.1.17}$$

so that, for $b = p$, the offspring distribution has a geometric distribution with success probability p. Show that the extinction probability η is given by $\eta = 1$ if $\mu = \mathbb{E}[X] = b/p^2 \leq 1$, while, with the abbreviation $q = 1 - p$,

$$\eta = \frac{1-\mu p}{q}. \tag{3.1.18}$$

Exercise 3.4 (Exercise 3.3 cont.)**.** *Let the probability distribution $\{p_k\}_{k=0}^{\infty}$ be given by* (3.1.17)*. Show that $G_n(s)$, the generating function of Z_n is given by*

$$G_n(s) = \begin{cases} 1 - \mu^n \frac{1-\eta}{\mu^n - \eta} + \frac{\mu^n \left(\frac{1-\eta}{\mu^n - \eta}\right)^2 s}{1 - \left(\frac{\mu^n - 1}{\mu^n - \eta}\right)} & \text{when } b \neq p^2; \\ \frac{nq - (nq-p)s}{p + nq - nps} & \text{when } b = p^2. \end{cases} \tag{3.1.19}$$

Exercise 3.5 (Exercise 3.4 cont.)**.** *Conclude from Exercise 3.4 that, for* $\{p_k\}_{k=0}^{\infty}$ *in* (3.1.17)*,*

$$\mathbb{P}(Z_n > 0,\ \exists m > n \text{ such that } Z_m = 0) = \begin{cases} \mu^n \frac{1-\eta}{\mu^n - \eta} & \text{when } b < p^2; \\ \frac{p}{p+nq} & \text{when } b = p^2; \\ \frac{(1-\eta)\eta}{\mu^n - \eta} & \text{when } b > p^2. \end{cases} \tag{3.1.20}$$

We continue by studying the total progeny T of the branching process, which is defined as

$$T = \sum_{n=0}^{\infty} Z_n. \tag{3.1.21}$$

We denote by $G_T(s)$ the probability generating function of T, i.e.,

$$G_T(s) = \mathbb{E}[s^T]. \tag{3.1.22}$$

The main result is the following:

Theorem 3.2 (Total progeny probability generating function)**.** *For a branching process with i.i.d. offspring* X *having probability generating function* $G_X(s) = \mathbb{E}[s^X]$*, the probability generating function of the total progeny* T *satisfies the relation*

$$G_T(s) = sG_X(G_T(s)). \tag{3.1.23}$$

Proof. We again condition on the size of the first generation, and use that when $Z_1 = i$, for $j = 1, \ldots, i$, the total progeny T_j of the j^{th} child of the initial individual satisfies that $\{T_j\}_{j=1}^{i}$ is an i.i.d. sequence of random variables with law equal to the one of T. Therefore, using also that

$$T = 1 + \sum_{j=1}^{i} T_j, \tag{3.1.24}$$

where, by convention, the empty sum, arising when $i = 0$, is equal to zero, we obtain

$$G_T(s) = \sum_{i=0}^{\infty} p_i \mathbb{E}[s^T | Z_1 = i] = s \sum_{i=0}^{\infty} p_i \mathbb{E}[s^{T_1 + \cdots + T_i}] = s \sum_{i=0}^{\infty} p_i G_T(s)^i = sG_X(G_T(s)). \tag{3.1.25}$$

This completes the proof. □

Exercise 3.6 (Exercise 3.2 cont.)**.** *In the case of binary branching, i.e., when* p *is given by* (3.1.15)*, show that*

$$G_T(s) = \frac{1 - \sqrt{1 - 4s^2pq}}{2sp}. \tag{3.1.26}$$

Exercise 3.7 (Exercise 3.5 cont.)**.** *Show, using Theorem 3.2, that, for* $\{p_k\}_{k=0}^{\infty}$ *in* (3.1.17)*,*

$$G_T(s) = \frac{\sqrt{(p + s(b-pq))^2 - 4pqs(p-b)} - (p + sbq)}{2pq} \tag{3.1.27}$$

3.2 Family moments

In this section, we compute the mean generation size of a branching process, and use this to compute the mean family size or the mean total progeny. The main result is the following theorem:

Theorem 3.3 (Moments of generation sizes)**.** *For all $n \geq 0$, and with $\mu = \mathbb{E}[Z_1] = \mathbb{E}[X]$ the expected offspring of a given individual,*

$$\mathbb{E}[Z_n] = \mu^n. \tag{3.2.1}$$

Proof. Recall that

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i}, \tag{3.2.2}$$

where $\{X_{n,i}\}_{n,i\geq 1}$ is a doubly infinite array of i.i.d. random variables. In particular, $\{X_{n,i}\}_{i\geq 1}$ is independent of Z_{n-1}.

Exercise 3.8. *Complete the proof of Theorem 3.3 by conditioning on Z_{n-1} and showing that*

$$\mathbb{E}\big[\sum_{i=1}^{Z_{n-1}} X_{n,i} | Z_{n-1} = m\big] = m\mu, \tag{3.2.3}$$

so that

$$\mathbb{E}[Z_n] = \mu\mathbb{E}[Z_{n-1}]. \tag{3.2.4}$$

Exercise 3.9. *Prove that $\{\mu^{-n} Z_n\}_{n\geq 1}$ is a martingale.*

Exercise 3.10. *When the branching process is critical, note that $Z_n \xrightarrow{\mathbb{P}} 0$. On the other hand, conclude that $\mathbb{E}[Z_n] = 1$ for all $n \geq 1$.*

Theorem 3.4. *Fix $n \geq 0$. Let $\mu = \mathbb{E}[Z_1] = \mathbb{E}[X]$ be the expected offspring of a given individual, and assume that $\mu < 1$. Then,*

$$\mathbb{P}(Z_n > 0) \leq \mu^n. \tag{3.2.5}$$

Exercise 3.11. *Prove Theorem 3.4 by using Theorem 3.3, together with the Markov inequality* (2.4.1).

Theorem 3.4 implies that in the subcritical regime, i.e., when the expected offspring $\mu < 1$, the probability that the population survives up to time n is exponentially small in n.

Theorem 3.5 (Expected total progeny)**.** *For a branching process with i.i.d. offspring X having mean offspring $\mu < 1$,*

$$\mathbb{E}[T] = \frac{1}{1-\mu}. \tag{3.2.6}$$

Exercise 3.12. *Prove* (3.2.6).

3.3 Random-walk perspective to branching processes

In branching processes, it is common to study the number of descendants of each generation. For random graph purposes, it is often convenient to use a different construction of a branching process by sequentially investigating the number of children of each member of the population. This picture leads to a *random walk formulation of branching processes*. For more background on random walks, we refer the reader to [211] or [111, Section 5.3].

We now give the random walk representation of a branching process. Let $X_1, X_2, \ldots$ be independent random variables with the same distribution as $X_{1,1}$ in (3.1.2). Define $S_0, S_1, \ldots$ by the recursion

$$\begin{aligned} &S_0 = 1, \\ &S_i = S_{i-1} + X_i - 1 = X_1 + \ldots + X_i - (i-1). \end{aligned} \tag{3.3.1}$$

Let T be the smallest t for which $S_t = 0$, i.e., (recall (1.5.10))

$$T = \min\{t : S_t = 0\} = \min\{t : X_1 + \ldots + X_t = t - 1\}. \tag{3.3.2}$$

If such a t does not exist, then we define $T = +\infty$.

The above description is *equivalent* to the normal definition of a branching process, but records the branching process tree in a different manner. For example, in the random walk picture, it is slightly more difficult to extract the distribution of the generation sizes. To see that the two pictures agree, we shall show that the distribution of the random variable T is equal to the *total progeny* of the branching process as defined in (3.1.21), and it is equal to the total number of individuals in the family tree of the initial individual.

To see this, we note that we can explore the branching process family tree as follows. We let X_1 denote the children of the original individual, and set S_1 as in (3.3.1). Then, there are $S_1 = X_1 - 1$ unexplored individuals, i.e., individuals of whom we have not yet explored how many children they have. We claim that after exploring i individuals, and on the event that there are at least i individuals in the family tree, the random variable S_i denotes the number of individuals of whom the children have not yet been explored:

Lemma 3.6 (The interpretation of $\{S_i\}_{i=0}^{\infty}$)**.** *The random process $\{S_i\}_{i=0}^{\infty}$ in (3.3.1) has the same distribution as the random process $\{S_i'\}_{i=0}^{\infty}$, where S_i' denotes the number of unexplored individuals in the exploration of a branching process population after exploring i individuals successively.*

Proof. We shall prove this by induction on i. Clearly, it is correct when $i = 0$. We next advance the induction hypothesis. For this, suppose this is true for S_{i-1}. We are done when $S_{i-1} = 0$, since then all individuals have been explored, and the total number of explored individuals is clearly equal to the size of the family tree, which is T by definition. Thus, assume that $S_{i-1} > 0$. Then we pick an arbitrary unexplored individual and denote the number of its children by X_i. By the independence property of the offspring of different individuals in a branching process, we have that the distribution of X_i is equal to the distribution of Z_1, say. Also, after exploring the children of the i^{th} individual, we have added X_i individuals that still need to be explored, and have explored a single individual, so that now the total number of unexplored individuals is equal to $S_{i-1} + X_i - 1$, which, by (3.3.1) is equal to S_i. This completes the proof using induction. □

Lemma 3.6 gives a nice interpretation of the random process $\{S_i\}_{i=0}^{\infty}$ in (3.3.1). Finally, since the branching process total progeny is explored *precisely* at the moment that all of its individuals have been explored, it follows that T in (3.3.2) has the same distribution as the total progeny of the branching process.

Exercise 3.13. *Compute $\mathbb{P}(T = k)$ for T in (3.3.2) and $\mathbb{P}(T = k)$ for T in (3.1.21) explicitly, for $k = 1, 2$ and 3.*

The branching process belonging to the recursion in (3.3.1) is the following. The population starts with one active individual. At time i, we select one of the active individuals in the population, and give it X_i children. The children (if any) are set to active, and the individual becomes inactive.

This process is continued as long as there are active individuals in the population. Then, the process S_i describes the number of active individuals after the first i individuals have been explored. The process stops when $S_t = 0$, but the recursion can be defined for all t since this leaves the value of T unaffected. Note that, for a branching process, (3.3.1) only makes sense as long as $i \leq T$, since only then $S_i \geq 0$ for all $i \leq T$. However, (3.3.1) in itself can be defined for *all* $i \geq 0$, also when $S_i < 0$. This fact will be useful in the sequel.

Exercise 3.14 (Exercise 3.2 cont.)**.** *In the case of binary branching, i.e., when the offspring distribution is given by* (3.1.15)*, show that*

$$\mathbb{P}(T = k) = \frac{1}{p}\mathbb{P}\big(S_0 = S_{k+1} = 0, S_i > 0 \ \forall 1 \leq i \leq k\big), \tag{3.3.3}$$

where $\{S_i\}_{i=1}^{\infty}$ *is a simple random walk, i.e.,*

$$S_i = Y_1 + \cdots + Y_i, \tag{3.3.4}$$

where $\{Y_i\}_{i=1}^{\infty}$ *are i.i.d. random variables with distribution*

$$\mathbb{P}(Y_1 = 1) = 1 - \mathbb{P}(Y_1 = -1) = p. \tag{3.3.5}$$

This gives a one-to-one relation between random walks excursions and the total progeny of a binary branching process.

Denote by $H = (X_1, \ldots, X_T)$ the history of the process up to time T. We include the case where $T = \infty$, in which case the vector H has infinite length. A sequence $(x_1, \ldots, x_t)$ is a possible history if and only if the sequence x_i satisfies (3.3.1), i.e., when $s_i > 0$ for all $i < t$, while $s_t = 0$, where $s_i = x_1 + \cdots + x_i - (i-1)$. Then, for any $t < \infty$,

$$\mathbb{P}(H = (x_1, \ldots, x_t)) = \prod_{i=1}^{t} p_{x_i}. \tag{3.3.6}$$

Note that (3.3.6) determines the law of the branching process when conditioned on extinction.

We will use the random walk perspective in order to describe the distribution of a branching process *conditioned on extinction*. Call the distributions p and p' a *conjugate pair* if

$$p'_x = \eta^{x-1} p_x, \tag{3.3.7}$$

where η is the extinction probability belonging to the offspring distribution $\{p_x\}_{x=0}^{\infty}$, so that $\eta = G_X(\eta)$.

Exercise 3.15. *Prove that* $p' = \{p'_x\}_{x=0}^{\infty}$ *defined in* (3.3.7) *is a probability distribution.*

The relation between a supercritical branching process conditioned on extinction and its conjugate branching process is as follows:

Theorem 3.7 (Duality principle for branching processes)**.** *Let* p *and* p' *be conjugate offspring distributions. The branching process with distribution* p*, conditional on extinction, has the same distribution as the branching process with offspring distribution* p'*.*

The duality principle takes a particularly appealing form for Poisson branching processes, see Theorem 3.15 below.

Proof. It suffices to show that for every finite history $H = (x_1, \ldots, x_t)$, the probability (3.3.6) is the same for the branching process with offspring distribution p, when conditioned on extinction, and the branching process with offspring distribution p'. Fix a $t < \infty$. First observe that

$$\begin{aligned}\mathbb{P}(H = (x_1, \ldots, x_t) | \text{extinction}) &= \frac{\mathbb{P}(\{H = (x_1, \ldots, x_t)\} \cap \text{extinction})}{\mathbb{P}(\text{extinction})} \\ &= \eta^{-1} \mathbb{P}(H = (x_1, \ldots, x_t)), \end{aligned} \tag{3.3.8}$$

since a finite history implies that the population becomes extinct. Then, we use (3.3.6), together with the fact that

$$\prod_{i=1}^{t} p_{x_i} = \prod_{i=1}^{t} p'_{x_i} \eta^{-(x_i - 1)} = \eta^{t - \sum_{i=1}^{t} x_i} \prod_{i=1}^{t} p'_{x_i} = \eta \prod_{i=1}^{t} p'_{x_i}, \tag{3.3.9}$$

since $x_1 + \ldots + x_t = t - 1$. Substitution into (3.3.8) yields that

$$\mathbb{P}(H = (x_1, \ldots, x_t) | \text{extinction}) \quad = \quad \mathbb{P}'(H = (x_1, \ldots, x_t)), \tag{3.3.10}$$

where $\mathbb{P}'$ is the distribution of the branching process with offspring distribution p'. □

Exercise 3.16. *Let $G_d(s) = \mathbb{E}'[s^{X_1}]$ be the probability generating function of the offspring of the dual branching process. Show that*

$$G_d(s) = \frac{1}{\eta} G_X(\eta s). \tag{3.3.11}$$

Exercise 3.17. *Let X' have probability mass function $p' = \{p'_x\}_{x=0}^{\infty}$ defined in (3.3.7). Show that when $\eta < 1$, then*

$$\mathbb{E}[X'] < 1. \tag{3.3.12}$$

Thus, the branching process with offspring distribution p' is subcritical.

Another convenient feature of the random walk perspective for branching processes is that it allows one to study what the probability is of extinction when the family tree has at least a given size. The main result in this respect is given below:

Theorem 3.8 (Extinction probability with large total progeny)**.** *For a branching process with i.i.d. offspring X having mean $\mu = \mathbb{E}[X] > 1$,*

$$\mathbb{P}(k \leq T < \infty) \leq \frac{\mathrm{e}^{-Ik}}{1 - \mathrm{e}^{-I}}, \tag{3.3.13}$$

where the exponential rate I is given by

$$I = \sup_{t \leq 0} \left(t - \log \mathbb{E}[\mathrm{e}^{tX}] \right) > 0. \tag{3.3.14}$$

Theorem 3.8 can be reformulated by saying that when the total progeny is large, then the branching process will survive with high probability.

Note that when $\mu = \mathbb{E}[X] > 1$ and when $\mathbb{E}[\mathrm{e}^{tX}] < \infty$ for all $t \in \mathbb{R}$, then we can also write

$$I = \sup_t \big(t - \log \mathbb{E}[\mathrm{e}^{tX}]\big), \tag{3.3.15}$$

(see also (2.4.12)). However, in Theorem 3.8, it is *not* assumed that $\mathbb{E}[\mathrm{e}^{tX}] < \infty$ for all $t \in \mathbb{R}$! Since $X \geq 0$, we clearly *do* have that $\mathbb{E}[\mathrm{e}^{tX}] < \infty$ for all $t \leq 0$. Therefore, since also the derivative of $t \mapsto t - \log \mathbb{E}[\mathrm{e}^{tX}]$ in $t = 0$ is equal to $1 - \mathbb{E}[X] < 0$, the supremum is attained at a $t < 0$, and, therefore, we obtain that $I > 0$ under no assumptions on the existence of the moment generating function of the offspring distribution. We now give the full proof:

Proof. We use the fact that $T = s$ implies that $S_s = 0$, which in turn implies that $X_1 + \ldots + X_s = s - 1 \leq s$. Therefore,

$$\mathbb{P}(k \leq T < \infty) \leq \sum_{s=k}^{\infty} \mathbb{P}(S_s = 0) \leq \sum_{s=k}^{\infty} \mathbb{P}(X_1 + \ldots + X_s \leq s). \tag{3.3.16}$$

For the latter probability, we use (2.4.9) and (2.4.11) in Theorem 2.16 with $a = 1 < \mathbb{E}[X]$. Then, we arrive at

$$\mathbb{P}(k \leq T < \infty) \leq \sum_{s=k}^{\infty} \mathrm{e}^{-sI} = \frac{\mathrm{e}^{-Ik}}{1 - \mathrm{e}^{-I}}. \tag{3.3.17}$$

□

3.4 Supercritical branching processes

In this section, we prove a convergence result for the population in the n^{th} generation. Clearly, in the (sub)critical case, the limit of Z_n is equal to 0, and there is nothing to prove. In the supercritical case, when the expected offspring is equal to $\mu > 1$, it is also known that (see e.g., [16, Theorem 2, p. 8]) $\lim_{n\to\infty} \mathbb{P}(Z_n = k) = 0$ unless $k = 0$, and $\mathbb{P}(\lim_{n\to\infty} Z_n = 0) = 1 - \mathbb{P}(\lim_{n\to\infty} Z_n = \infty) = \eta$, where η is the extinction probability of the branching process. In particular, the branching process population cannot stabilize. It remains to investigate what happens when $\eta < 1$, in which case $\lim_{n\to\infty} Z_n = \infty$ occurs with positive probability. We prove the following convergence result:

Theorem 3.9 (Convergence for supercritical branching processes)**.** *For a branching process with i.i.d. offspring X having mean $\mu = \mathbb{E}[X] > 1$, $\mu^{-n} Z_n \xrightarrow{a.s.} W_\infty$ for some random variable W_∞ which is finite with probability 1.*

Proof. We use the martingale convergence theorem (Theorem 2.21), and, in particular, its consequence formulated in Exercise 2.23. Denote $M_n = \mu^{-n} Z_n$, and recall that by Exercise 3.9, $\{M_n\}_{n=1}^{\infty}$ is a martingale. By Theorem 3.3, we have that $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = 1$, so that Theorem 2.21 gives the result. □

Unfortunately, not much is known about the limiting distribution W_∞. Its probability generating function $G_W(s) = \mathbb{E}[s^{W_\infty}]$ satisfies the implicit relation, for $s \in [0, 1]$,

$$G_W(s) = G_X\big(G_W(s^{1/\mu})\big). \tag{3.4.1}$$

Exercise 3.18. *Prove* (3.4.1).

We next investigate when $\mathbb{P}(W_\infty > 0) = 1 - \eta = \zeta$:

Theorem 3.10 (Kesten-Stigum Theorem)**.** *For a branching process with i.i.d. offspring X having mean $\mu = \mathbb{E}[X] > 1$, $\mathbb{P}(W_\infty = 0) = \eta$ precisely when $\mathbb{E}[X \log X] < \infty$. When $\mathbb{E}[X \log X] < \infty$, also $\mathbb{E}[W_\infty] = 1$, while, when $\mathbb{E}[X \log X] = \infty$, $\mathbb{P}(W_\infty = 0) = 1$.*

Theorem 3.10 implies that $\mathbb{P}(W_\infty > 0) = 1 - \eta$, where η is the extinction probability of the branching process, so that conditionally on survival, the probability that $W_\infty > 0$ is equal to one. Theorem 3.10 was first proved by Kesten and Stigum in [142, 143, 144]. It is remarkable that the precise condition when $W_\infty = 0$ a.s. can be so easily expressed in terms of a moment condition on the offspring distribution. A proof of Theorem 3.10 is given in [16, Pages 24-26], while in [165] a conceptual proof is given. See [88, Proof of Theorem 2.16] for a simple proof of the statement under the stronger condition that $\mathbb{E}[X^2] < \infty$, using the L^2-martingale convergence theorem (see also below Theorem 2.21).

Theorem 3.10 leaves us with the question what happens when $\mathbb{E}[X \log X] = \infty$. In this case, Seneta [199] has shown that there always exists a proper renormalization, i.e., there exists a sequence $\{c_n\}_{n=1}^\infty$ with $\lim_{n\to\infty} c_n^{1/n} = \mu$ such that Z_n/c_n converges to a non-degenerate limit. However, $c_n = o(\mu^n)$, so that $\mathbb{P}(W_\infty = 0) = 1$.

Exercise 3.19. *Prove that $\mathbb{P}(W_\infty > 0) = 1 - \eta$ implies that $\mathbb{P}(W_\infty > 0 |$ survival$) = 1$.*

Exercise 3.20. *Prove, using Fatou's lemma (Theorem A.12), that $\mathbb{E}[W_\infty] \le 1$ always holds.*

We continue by studying the number of particles with an infinite line of descent, i.e., the particles of whom the family tree survives forever. Interestingly, these particles form a branching process again, as we describe now. In order to state the result, we start with some definitions. We let $Z_n^{(1)}$ denote those particles from the n^{th} generation of $\{Z_k\}_{k=0}^\infty$ that survive forever. Then, the main result is as follows:

Theorem 3.11 (Individuals with an infinite line of descent)**.** *Conditionally on survival, the process $\{Z_n^{(\infty)}\}_{n=0}^\infty$ is again a branching process with offspring distribution $p^{(\infty)} = \{p_k^{(\infty)}\}_{k=0}^\infty$ given by $p_0^{(\infty)} = 0$ and, for $k \ge 1$,*

$$p_k^{(\infty)} = \frac{1}{\zeta} \sum_{j=k}^\infty \binom{j}{k} \eta^{j-k} (1-\eta)^j p_j. \tag{3.4.2}$$

Moreover, since

$$\mu^{(\infty)} = \mathbb{E}[Z_1^{(\infty)}] = \mu = \mathbb{E}[Z_1], \tag{3.4.3}$$

this branching process is supercritical with the same expected offspring as $\{Z_n\}_{n=0}^\infty$ itself.

Comparing Theorem 3.11 to Theorem 3.7, we see that in the supercritical regime, the branching process conditioned on extinction is a branching process with the dual (subcritical) offspring distribution, while, conditional on survival, the individuals with an infinite line of descent for a (supercritical) branching process.

Exercise 3.21. *Prove that $p^{(\infty)}$ is a probability distribution.*

Proof of Theorem 3.11. We let $\mathcal{A}_\infty$ be the event that $Z_n \to \infty$. We shall prove, by induction on $n \ge 0$, that the distribution of $\{Z_k^{(\infty)}\}_{k=0}^n$ conditionally on $\mathcal{A}_\infty$ is equal to that of a $\{\hat{Z}_k\}_{k=0}^n$, where $\{\hat{Z}_k\}_{k=0}^\infty$ is a branching process with offspring distribution $p^{(\infty)}$ given in (3.4.2). We start by initializing the induction hypothesis. For this, we note that, on $\mathcal{A}_\infty$, we have that $Z_0^{(\infty)} = 1$, whereas, by convention, $\hat{Z}_0 = 1$. This initializes the induction hypothesis.

To advance the induction hypothesis, we argue as follows. Suppose that the distribution of $\{Z_k^{(\infty)}\}_{k=0}^n$, conditionally on $\mathcal{A}_\infty$, is equal to that of $\{\hat{Z}_k\}_{k=0}^n$. Then, we shall show that also the distribution of $\{Z_k^{(\infty)}\}_{k=0}^{n+1}$, conditionally on $\mathcal{A}_\infty$, is equal to that of $\{\hat{Z}_k\}_{k=0}^{n+1}$. By the induction hypothesis, this immediately follows if the *conditional* distributions of $Z_{n+1}^{(\infty)}$ given $\{Z_k^{(\infty)}\}_{k=0}^n$ is equal to the conditional distribution of $\hat{Z}_{n+1}$ given $\{\hat{Z}_k\}_{k=0}^n$.

The law of $\hat{Z}_{n+1}$ given $\{\hat{Z}_k\}_{k=0}^n$ is that of an independent sum of $\hat{Z}_n$ i.i.d. random variables with law $p^{(\infty)}$. Now, the law of $Z_{n+1}^{(\infty)}$ given $\{Z_k^{(\infty)}\}_{k=0}^n$ is equal to the law of $Z_{n+1}^{(\infty)}$ given $Z_n^{(\infty)}$, and each individual with infinite line of descent in the n^{th} generation gives rise to a random and i.i.d. number of individuals with infinite line of descent in the $(n+1)^{\text{st}}$ generation with the same law as $Z_1^{(\infty)}$ conditionally on $\mathcal{A}_\infty$. As a result, to complete the proof of (3.4.2), we must show that

$$\mathbb{P}\big(Z_1^{(\infty)} = k \,\big|\mathcal{A}_\infty\big) = p_k^{(\infty)}. \tag{3.4.4}$$

For $k = 0$, this is trivial, since, conditionally on $\mathcal{A}_\infty$, we have that $Z_1^{(\infty)} \geq 1$, so that both sides are equal to 0 for $k = 0$. For $k \geq 1$, on the other hand, the proof follows by conditioning on Z_1. We have that, for $k \geq 1$, $Z_1^{(\infty)} = k$ implies that $Z_1 \geq k$ and that $\mathcal{A}_\infty$ occurs, so that

$$\begin{aligned}\mathbb{P}\big(Z_1^{(\infty)} = k \,\big|\mathcal{A}_\infty\big) &= \zeta^{-1}\mathbb{P}\big(Z_1^{(\infty)} = k\big) = \zeta^{-1}\sum_{j\geq k}\mathbb{P}\big(Z_1^{(\infty)} = k \,\big|Z_1 = j\big)\mathbb{P}(Z_1 = j)\\ &= \zeta^{-1}\sum_{j\geq k}\binom{j}{k}\eta^{j-k}(1-\eta)^j p_j,\end{aligned} \tag{3.4.5}$$

since each of the j particles has infinite line of descent with probability $\zeta = 1 - \eta$, so that $\mathbb{P}\big(Z_1^{(\infty)} = k \,\big|Z_1 = j\big) = \mathbb{P}(\mathrm{Bin}(j, 1-\eta) = k)$.

We complete the proof of Theorem 3.11 by proving (3.4.3). We start by proving (3.4.2) when $\mu < \infty$. For this, we write, using that for $k = 0$, we may substitute the right-hand side of (3.4.2) instead of $p_0^{(\infty)} = 0$, to obtain

$$\begin{aligned}\mu^{(\infty)} &= \sum_{k=0}^\infty k p_k^{(\infty)} = \sum_{k=0}^\infty k\frac{1}{\zeta}\sum_{j=k}^\infty\binom{j}{k}\eta^{j-k}(1-\eta)^j p_j\\ &= \frac{1}{\zeta}\sum_{j=0}^\infty p_j\sum_{k=0}^j k\binom{j}{k}\eta^{j-k}(1-\eta)^j = \frac{1}{\zeta}\sum_{j=0}^\infty p_j(\zeta j) = \sum_{j=0}^\infty j p_j = \mu.\end{aligned} \tag{3.4.6}$$

This proves (3.4.2) when $\mu < \infty$. When $\mu = \infty$, on the other hand, we only need to show that $\mu^{(\infty)} = \infty$ as well. This can easily be seen by an appropriate truncation argument, and is left to the reader. □

Exercise 3.22. *Prove* (3.4.2) *when* $\mu = \infty$.

With Theorems 3.11 and 3.9 at hand, we see an interesting picture emerging. Indeed, by Theorem 3.9, we have that $Z_n\mu^{-n} \xrightarrow{a.s.} W_\infty$, where, if the $X \log X$-condition in Theorem 3.10 is satisfied, $\mathbb{P}(W_\infty > 0) = \zeta$, the branching process survival probability. On the other hand, by Theorem 3.11 and conditionally on $\mathcal{A}_\infty$, $\{Z_n^{(\infty)}\}_{n=0}^\infty$ is also a branching process with expected offspring μ, which survives with probability 1. As a result, $Z_n^{(\infty)}\mu^{-n} \xrightarrow{a.s.} W_\infty^{(\infty)}$, where, conditionally on $\mathcal{A}_\infty$, $\mathbb{P}(W_\infty^{(\infty)} > 0) = 1$, while, yet, $Z_n^{(\infty)} \leq Z_n$ for all

$n \geq 0$, by definition. This raises the question what the relative size is of $Z_n^{(\infty)}$ and Z_n, conditionally on $\mathcal{A}_\infty$. This question is answered in the following theorem:

Theorem 3.12 (Proportion of particles with infinite line of descent)**.** *Conditionally on survival,*

$$\frac{Z_n^{(\infty)}}{Z_n} \xrightarrow{a.s.} \zeta. \tag{3.4.7}$$

Theorem 3.12 will prove to be quite useful, since it allows us sometimes to transfer results on branching processes which survive with probability 1, such as $\{Z_n^{(\infty)}\}_{n=0}^{\infty}$ conditionally on survival, to branching processes which have a non-zero extinction probability, such as $\{Z_n\}_{n=0}^{\infty}$.

Proof of Theorem 3.12. We first give the proof in the case where the mean offspring μ is finite. Applying Theorem 3.11 together with Theorem 3.9 and the fact that, conditionally on survival, $\mathbb{E}[Z_1^{(\infty)}] = \mu$ (see (3.4.3)), we obtain that there exists $W^{(\infty)}$ such that $Z_n^{(\infty)}\mu^{-n} \to W^{(\infty)}$. Moreover, by Theorem 3.10 and the fact that the survival probability of the branching process in $\{Z_n^{(\infty)}\}_{n=0}^{\infty}$ equals 0 (recall Exercise 3.1), we have that $\mathbb{P}(W^{(\infty)} > 0) = 1$. Further, again by Theorem 3.9 now applied to $\{Z_n\}_{n=0}^{\infty}$, conditionally on survival, Z_n/μ^n converges in distribution to the conditional distribution of W_∞ conditionally on $W_\infty > 0$. Thus, we obtain that $Z_n^{(\infty)}/Z_n$ converges a.s. to a finite and positive limit R.

In order to see that this limit in fact equals ζ, we use that the distribution of $Z_n^{(\infty)}$ given that $Z_n = k$ is binomial with parameters k probability of success ζ. As a result, since as $n \to \infty$ and conditionally on survival $Z_n \to \infty$, we have that $Z_n^{(\infty)}/Z_n$ converges in probability to ζ. This implies that $R = \zeta$ a.s.

Add proof when $\mu = \infty$!

□

3.5 Hitting-time theorem and the total progeny

In this section, we derive a general result for the law of the total progeny for branching processes, by making use of the Hitting-time theorem for random walks. The main result is the following:

Theorem 3.13 (Law of total progeny)**.** *For a branching process with i.i.d. offspring distribution $Z_1 = X$,*

$$\mathbb{P}(T = n) = \frac{1}{n}\mathbb{P}(X_1 + \cdots + X_n = n - 1), \tag{3.5.1}$$

where $\{X_i\}_{i=1}^n$ are i.i.d. copies of X.

Exercise 3.23 (Total progeny for binomial branching processes)**.** *Compute the probability mass function of a branching process with a binomial offspring distribution using Theorem 3.13.*

Exercise 3.24 (Total progeny for geometric branching processes)**.** *Compute the probability mass function of a branching process with a geometric offspring distribution using Theorem 3.13. Hint: note that when $\{X_i\}_{i=1}^n$ are i.i.d. geometric, then $X_1 + \cdots + X_n$ has a negative binomial distribution.*

We shall prove Theorem 3.13 below. In fact, we shall prove a more general version of Theorem 3.13, which states that

$$\mathbb{P}(T_1 + \cdots + T_k = n) = \frac{k}{n}\mathbb{P}(X_1 + \cdots + X_n = n - k), \tag{3.5.2}$$

where $T_1, \ldots, T_k$ are k independent random variables with the same distribution as T. Alternatively, we can think of $T_1 + \cdots + T_k$ as being the total progeny of a branching process starting with k individuals, i.e., when $Z_0 = k$.

The proof is based on the random walk representation of a branching process, together with the random-walk hitting time theorem. In its statement, we write $\mathbb{P}_k$ for the law of a random walk starting in k, we let $\{Y_i\}_{i=1}^{\infty}$ be the i.i.d. steps of the random walk, and we let $S_n = k + Y_1 + \cdots + Y_n$ be the position of the walk, starting in k, after n steps. We finally let

$$T_0 = \inf\{n \colon S_n = 0\} \tag{3.5.3}$$

denote the first hitting time of the origin of the walk. Then, the hitting-time theorem is the following result:

Theorem 3.14 (Hitting-time theorem)**.** *For a random walk with i.i.d. steps $\{Y_i\}_{i=1}^{\infty}$ satisfying that*

$$\mathbb{P}(Y_i \geq -1) = 1, \tag{3.5.4}$$

the distribution of T_0 is given by

$$\mathbb{P}_k(T_0 = n) = \frac{k}{n}\mathbb{P}_k(S_n = 0). \tag{3.5.5}$$

Theorem 3.14 is a remarkable result, since it states that, conditionally on the event $\{S_n = 0\}$, and regardless of the precise distribution of the steps of the walk $\{Y_i\}_{i=1}^{\infty}$ satisfying (3.5.4), the probability of the walk to be at 0 *for the first time* at time n is equal to $\frac{k}{n}$. Equation (3.5.2) follows from Theorem 3.14 since the law of $T_1 + \cdots + T_k$ is that of a the hitting time of a random walk starting in k with step distribution $Y_i = X_i - 1$, where $\{X_i\}_{i=1}^{\infty}$ are the offsprings of the vertices. Since $X_i \geq 0$, we have that $Y_i \geq -1$, which completes the proof of (3.5.2) and hence of Theorem 3.13. The details are left as an exercise:

Exercise 3.25 (The total progeny from the hitting-time theorem)**.** *Prove that Theorem 3.14 implies* (3.5.2)*.*

Exercise 3.26 (The condition in the hitting-time theorem)**.** *Is Theorem 3.14 still true when the restriction that $\mathbb{P}(Y_i \geq -1) = 1$ is dropped?*

Proof of Theorem 3.14. We prove (3.5.5) for all $k \geq 0$ by induction on $n \geq 1$. When $n = 1$, then both sides are equal to 0 when $k > 1$ and $k = 0$, and are equal to $\mathbb{P}(Y_1 = -1)$ when $k = 1$. This initializes the induction.

To advance the induction, we take $n \geq 2$, and note that both sides are equal to 0 when $k = 0$. Thus, we may assume that $k \geq 1$. We condition on the first step to obtain

$$\mathbb{P}_k(T_0 = n) = \sum_{s=-1}^{\infty} \mathbb{P}_k(T_0 = n \big| Y_1 = s)\mathbb{P}(Y_1 = s). \tag{3.5.6}$$

By the random-walk Markov property,

$$\mathbb{P}_k(T_0 = n \big| Y_1 = s) = \mathbb{P}_{k+s}(T_0 = n - 1) = \frac{k+s}{n-1}\mathbb{P}_{k+s}(S_{n-1} = 0), \tag{3.5.7}$$

where in the last equality we used the induction hypothesis, which is allowed since $k \geq 1$ and $s \geq -1$, so that $k + s \geq 0$. This leads to

$$\mathbb{P}_k(T_0 = n) = \sum_{s=-1}^{\infty} \frac{k+s}{n-1} \mathbb{P}_{k+s}(S_{n-1} = 0)\mathbb{P}(Y_1 = s). \tag{3.5.8}$$

We undo the law of total probability, using that $\mathbb{P}_{k+s}(S_{n-1} = 0) = \mathbb{P}_k(S_n = 0 \big| Y_1 = s)$, to arrive at

$$\mathbb{P}_k(T_0 = n) = \sum_{s=-1}^{\infty} (k+s)\mathbb{P}_k(S_n = 0 \big| Y_1 = s)\mathbb{P}(Y_1 = s) = \mathbb{P}_k(S_n = 0)\Big(k + \mathbb{E}_k[Y_1 | S_n = 0]\Big), \tag{3.5.9}$$

where $\mathbb{E}_k[Y_1|S_n = 0]$ is the conditional expectation of Y_1 given that $S_n = 0$ occurs. We next note that the conditional expectation of $\mathbb{E}_k[Y_i|S_n = 0]$ is independent of i, so that

$$\mathbb{E}_k[Y_1|S_n = 0] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_k[Y_i|S_n = 0] = \frac{1}{n}\mathbb{E}_k\Big[\sum_{i=1}^{n} Y_i \big| S_n = 0\Big] = -\frac{k}{n}, \tag{3.5.10}$$

since $\sum_{i=1}^{n} Y_i = S_n - k = -k$ when $S_n = 0$. Therefore, we arrive at

$$\mathbb{P}_k(T_0 = n) = \frac{1}{n-1}\Big[k - \frac{k}{n}\Big]\mathbb{P}_k(S_n = 0) = \frac{k}{n}\mathbb{P}_k(S_n = 0). \tag{3.5.11}$$

This advances the induction, and completes the proof of Theorem 3.14. □

Exercise 3.27 (Extension of Theorem 3.14)**.** *Extend the hitting-time theorem, Theorem 3.14, to the case where $\{Y_i\}_{i=1}^n$ is an exchangeable sequence rather than an i.i.d. sequence, where a sequence $\{Y_i\}_{i=1}^n$ is called exchangeable when its distribution is the same as the distribution of any permutation of the sequence. Hint: if $\{Y_i\}_{i=1}^n$ is exchangeable, then so is $\{Y_i\}_{i=1}^n$ conditioned on $\sum_{i=1}^n Y_i = -k$.*

3.6 Properties of Poisson branching processes

In this section, we specialize the discussion of branching processes to branching processes with Poisson offspring distributions. We will denote the distribution of a Poisson branching process by $\mathbb{P}^*_\lambda$. We also write T^* for the total progeny of the Poisson branching process, and X^* for a Poisson random variable.

For a Poisson random variable X^* with mean λ, we have that the probability generating function of the offspring distribution is equal to

$$G^*_\lambda(s) = \mathbb{E}^*_\lambda[s^{X^*}] = \sum_{i=0}^{\infty} s^i \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!} = \mathrm{e}^{\lambda(s-1)}. \tag{3.6.1}$$

Therefore, the relation for the extinction probability η in (3.1.5) becomes

$$\eta_\lambda = \mathrm{e}^{\lambda(\eta_\lambda - 1)}, \tag{3.6.2}$$

where we add the subscript λ to make the dependence on λ explicit.

For $\lambda \leq 1$, the equation (3.6.2) has the unique solution $\eta_\lambda = 1$, which corresponds to certain extinction. For $\lambda > 1$ there are two solutions, of which the smallest satisfies $\eta_\lambda \in (0, 1)$. As $\mathbb{P}^*_\lambda(T^* < \infty) < 1$, we know

$$\mathbb{P}^*_\lambda(T^* < \infty) = \eta_\lambda. \tag{3.6.3}$$

We recall that $H = (X_1^*, \ldots, X_T^*)$ is the history of the branching process, where again we have added superscripts $*$ to indicate that we mean a Poisson branching process. Then, conditionally on extinction, a Poisson branching process has law p' given by

$$p'_i = \eta_\lambda^{i-1} p_i = \mathrm{e}^{-\lambda\eta_\lambda} \frac{(\lambda\eta_\lambda)^i}{i!}, \tag{3.6.4}$$

where we have used (3.6.2). Note that this offspring distribution is again Poisson with mean

$$\mu_\lambda = \lambda\eta_\lambda, \tag{3.6.5}$$

and, again by (3.6.2),

$$\mu_\lambda \mathrm{e}^{-\mu_\lambda} = \lambda\eta_\lambda \mathrm{e}^{-\lambda\eta_\lambda} = \lambda \mathrm{e}^{-\lambda}. \tag{3.6.6}$$

Therefore, we call $\mu < 1 < \lambda$ a *conjugate pair* if

$$\mu \mathrm{e}^{-\mu} = \lambda \mathrm{e}^{-\lambda}. \tag{3.6.7}$$

Since $x \mapsto x\mathrm{e}^{-x}$ is first increasing and then decreasing, with a maximum of e^{-1} at $x = 1$, the equation $\mu\mathrm{e}^{-\mu} = \lambda\mathrm{e}^{-\lambda}$ has precisely two solutions, a solution $\mu < 1$ and a solution $\lambda > 1$. Therefore, for Poisson offspring distributions, the duality principle in Theorem 3.7 can be reformulated as follows:

Theorem 3.15 (Poisson duality principle)**.** *Let $\mu < 1 < \lambda$ be conjugates. The Poisson branching process with mean λ, conditional on extinction, has the same distribution as a Poisson branching process with mean μ.*

We further describe the law of the total progeny of a Poisson branching process:

Theorem 3.16 (Total progeny for Poisson BP)**.** *For a branching process with i.i.d. offspring X, where X has a Poisson distribution with mean λ,*

$$\mathbb{P}_\lambda^*(T^* = n) = \frac{(\lambda n)^{n-1}}{n!} \mathrm{e}^{-\lambda n}, \qquad (n \geq 1). \tag{3.6.8}$$

Exercise 3.28 (The total progeny of a Poisson branching process)**.** *Prove Theorem 3.16 using Theorem 3.13.*

Exercise 3.29 (Large, but finite, Poisson total progeny)**.** *Use Theorem 3.16 to show that, for any λ, and for k sufficiently large,*

$$\mathbb{P}_\lambda^*(k \leq T^* < \infty) \leq \mathrm{e}^{-I_\lambda k}, \tag{3.6.9}$$

where $I_\lambda = \lambda - 1 - \log\lambda$.

We use Theorem 3.16 to prove Cayley's Theorem. on the number of labeled trees [61]. In its statement, we define a *labeled tree* on $[n]$ to be a tree of size n where all vertices have a label in $[n]$ and each label occurs precisely once. We now make this definition precise. An *edge* of a labeled tree is a pair $\{v_1, v_2\}$, where v_1 and v_2 are the labels of two connected vertices in the tree. The *edge set* of a tree of size n is the collection of its $n-1$ edges. Two labeled trees are equal if and only if they consist of the same edge sets. A labeled tree of n vertices is equivalent to a spanning tree of the complete graph K_n on the vertices $[n]$. Cayley's Theorem reads as follows:

Theorem 3.17 (Cayley's Theorem)**.** *The number of labeled trees of size n is equal to n^{n-2}. Equivalently, the number of spanning trees of the complete graph of size n equals n^{n-2}.*

Proof. In is convenient to label the vertices of a tree in terms of *words*. These words arise inductively as follows. The root is the word $\varnothing$. The children of the root are the words $1, 2, \ldots, d_\varnothing$, where, for a word w, we let d_w denote the number of children of w. The children of 1 are $11, 12, \ldots, 1d_1$, etc. A tree is then uniquely represented by its set of words. For example, the word 1123 represents the third child of the second child of the first child of the first child of root.

Two trees are the same if and only if they are represented by the same set of words. We obtain a branching process when the variables $(d_w)_w$ are equal to a collection i.i.d. random variables. For a word w, we let $|w|$ be its length, where $|0| = 0$. The length of a word w is the number of steps the word is away from the root, and equals its generation.

Let $\mathcal{T}$ denote the family tree of a branching process with Poisson offspring distribution with parameter 1. We compute the probability of obtaining a given tree t by

$$\mathbb{P}(\mathcal{T} = t) = \prod_{w \in t} \mathbb{P}(\xi = d_w), \tag{3.6.10}$$

where ξ is a Poisson random variable with parameter 1, and d_w is the number of children of the word w in the tree t. For a Poisson branching process, $\mathbb{P}(\xi = c_w) = \mathrm{e}^{-1}/d_w!$, so that

$$\mathbb{P}(\mathcal{T} = t) = \prod_{w \in t} \frac{\mathrm{e}^{-1}}{d_w!} = \frac{\mathrm{e}^{-n}}{\prod_{w \in t} d_w!}, \tag{3.6.11}$$

where n denotes the number of vertices in t. We note that the above probability is the same for each tree with the same number of vertices of degree k for each k.

Conditionally on having total progeny $T^* = n$, we introduce a labeling as follows. We give the root label 1, and give all other vertices a label from the set $\{2, \ldots, n\}$, giving a labeled tree on n vertices. Given $\mathcal{T}$, there are precisely

$$\prod_{w \in \mathcal{T}} d_w! \tag{3.6.12}$$

possible ways to put down the labels that give rise to the *same* labeled tree, since permuting the children of any vertex does not change the labeled tree. Also, the probability that w receives label i_w with $i_\varnothing = 1$ is precisely equal to $1/(n-1)!$, where $n = |\mathcal{T}|$. For a labeled tree ℓ, let t_ℓ be any tree, i.e., a collection of words, from which ℓ can be obtained by labeling the vertices. Then, the probability of obtaining a given labeled tree ℓ of arbitrary size equals

$$\mathbb{P}(\mathcal{L} = \ell) = \mathbb{P}(\mathcal{T} = t_\ell) \frac{\prod_{w \in t_\ell} d_w!}{(|\ell| - 1)!} = \frac{\mathrm{e}^{-|\ell|}}{\prod_{w \in t_\ell} d_w!} \frac{\prod_{w \in t_\ell} d_w!}{(|\ell| - 1)!} = \frac{\mathrm{e}^{-|\ell|}}{(|\ell| - 1)!}. \tag{3.6.13}$$

Therefore, conditionally on $T^* = n$, the probability of a given labeled tree $\mathcal{L}$ of size n equals

$$\mathbb{P}(\mathcal{L} \mid |\mathcal{L}| = n) = \frac{\mathbb{P}(\mathcal{L} = \ell)}{\mathbb{P}(|\mathcal{L}| = n)}. \tag{3.6.14}$$

By Theorem 3.16,

$$\mathbb{P}(|\mathcal{L}| = n) = \mathbb{P}(T^* = n) = \frac{\mathrm{e}^{-n} n^{n-2}}{(n-1)!}. \tag{3.6.15}$$

As a result, for each labeled tree $\mathcal{L}$ of size $|\mathcal{L}| = n$

$$\mathbb{P}(\mathcal{L} = \ell \mid |\mathcal{L}| = n) = \frac{\mathbb{P}(\mathcal{L} = \ell)}{\mathbb{P}(|\mathcal{L}| = n)} = \frac{\mathrm{e}^{-n}}{(n-1)!}\frac{\mathrm{e}^{-n}n^{n-2}}{(n-1)!} = \frac{1}{n^{n-2}}. \tag{3.6.16}$$

The obtained probability is *uniform* over all labeled trees. Therefore, the number of labeled trees equals

$$\mathbb{P}(\mathcal{L} = \ell \mid |\mathcal{L}| = n)^{-1} = n^{n-2}. \tag{3.6.17}$$

The above not only proves Cayley's Theorem, but also gives an explicit construction of a uniform labeled tree from a Poisson branching process. □

Exercise 3.30. *Verify* (3.6.8) *for* $n = 1, 2$ *and* $n = 3$.

Theorem 3.18 (Asymptotics for total progeny for Poisson BP)**.** *For a branching process with i.i.d. offspring* X*, where* X *has a Poisson distribution with mean* λ*, as* $n \to \infty$*,*

$$\mathbb{P}^*_\lambda(T^* = n) = \frac{1}{\lambda\sqrt{2\pi n}}\mathrm{e}^{-I_\lambda n}(1 + o(1)), \tag{3.6.18}$$

where

$$I_\lambda = \lambda - 1 - \log\lambda. \tag{3.6.19}$$

In particular, when $\lambda = 1$*,*

$$\mathbb{P}^*_1(T^* = n) = (2\pi)^{-1/2}n^{-3/2}[1 + O(n^{-1})]. \tag{3.6.20}$$

d proof Theorem 3.18.

Proof.

□Equation (3.6.20) is an example of a *power-law* relationship that often holds at criticality. The above $n^{-3/2}$ behavior is associated more generally with the distribution of the total progeny whose offspring distribution has finite variance (see e.g., [10, Proposition 24]).

In Chapter 4, we will investigate the behavior of the Erdős-Rényi random graph by making use of couplings to branching processes. There, we also need the fact that, for $\lambda > 1$, the survival probability is sufficiently smooth (see Section 4.4):

Corollary 3.19 (Differentiability of the survival probability)**.** *Let* η_λ *denote the extinction probability of a branching process with a mean* λ *Poisson offspring distribution. Then, for all* $\lambda > 1$*,*

$$|\frac{d}{d\lambda}\zeta_\lambda| = \frac{\eta_\lambda(\lambda - \mu_\lambda)}{\lambda(1 - \mu_\lambda)} < \infty, \tag{3.6.21}$$

where μ_λ *is the dual of* λ*. When* $\lambda \downarrow 1$*,*

$$\zeta_\lambda = 2(\lambda - 1)(1 + o(1)). \tag{3.6.22}$$

Proof. The function η_λ, which we denote in this proof by $\eta(\lambda)$, is decreasing and satisfies

$$\eta(\lambda) = \mathbb{P}^*_\lambda(T^* < \infty) = \sum_{n=1}^{\infty}\mathrm{e}^{-\lambda n}\frac{(\lambda n)^{n-1}}{n!}, \tag{3.6.23}$$

and thus

$$0 \le -\frac{d}{d\lambda}\eta(\lambda) = \sum_{n=1}^{\infty}\mathrm{e}^{-n\lambda}\left[\frac{(\lambda n)^{n-1}}{(n-1)!}\right] - \sum_{n=2}^{\infty}\mathrm{e}^{-n\lambda}\left[\frac{(\lambda n)^{n-2}}{(n-2)!}\right]. \tag{3.6.24}$$

On the other hand,

$$\mathbb{E}^*_\lambda[T^*|T^*<\infty] = \frac{1}{\mathbb{P}^*_\lambda(T^*<\infty)}\sum_{n=1}^{\infty} n\cdot \mathrm{e}^{-\lambda n}\frac{(\lambda n)^{n-1}}{n!} = \frac{1}{\eta(\lambda)}\sum_{n=1}^{\infty}\mathrm{e}^{-\lambda n}\frac{(\lambda n)^{n-1}}{(n-1)!}, \quad (3.6.25)$$

so that

$$-\frac{d}{d\lambda}\eta(\lambda) = \eta(\lambda)\mathbb{E}^*_\lambda[T^*|T^*<\infty] - \frac{\eta(\lambda)}{\lambda}\mathbb{E}^*_\lambda[T^*|T^*<\infty] + \frac{\eta(\lambda)}{\lambda}, \quad (3.6.26)$$

where we have made use of the fact that

$$\begin{aligned}\sum_{n=2}^{\infty}\mathrm{e}^{-\lambda n}\frac{(\lambda n)^{n-2}}{(n-2)!} &= \sum_{n=1}^{\infty}\mathrm{e}^{-\lambda n}(n-1)\frac{(\lambda n)^{n-2}}{(n-1)!} = \sum_{n=1}^{\infty}\mathrm{e}^{-\lambda n}\frac{1}{\lambda}\frac{(\lambda n)^{n-1}}{(n-1)!} - \sum_{n=1}^{\infty}\mathrm{e}^{-\lambda n}\frac{(\lambda n)^{n-2}}{(n-1)!}\\ &= \frac{\eta(\lambda)}{\lambda}\mathbb{E}^*_\lambda[T^*|T^*<\infty] - \sum_{n=1}^{\infty}\mathrm{e}^{-\lambda n}\frac{1}{\lambda}\frac{(\lambda n)^{n-1}}{n!}\\ &= \frac{\eta(\lambda)}{\lambda}\mathbb{E}^*_\lambda[T^*|T^*<\infty] - \frac{1}{\lambda}\mathbb{P}^*_\lambda(T^*<\infty). \quad (3.6.27)\end{aligned}$$

By the duality principle and Theorem 3.5,

$$\mathbb{E}[T^*|T^*<\infty] = \frac{1}{1-\mu_\lambda}$$

where $\mu_\lambda = \lambda\eta(\lambda)$, by (3.6.5). Hence,

$$\begin{aligned}0 \le -\frac{d}{d\lambda}\eta(\lambda) &= \frac{\eta(\lambda)}{1-\mu_\lambda}\left(1-\frac{1}{\lambda}\right) + \frac{\eta(\lambda)}{\lambda}\\ &= \frac{\eta(\lambda)(\lambda-\mu(\lambda))}{\lambda(1-\mu_\lambda)}. \quad (3.6.28)\end{aligned}$$

□ Add proof of (3.6.22)!

3.7 Binomial and Poisson branching processes

When dealing with random graphs where the probability of keeping an edge is λ/n, the total number of vertices incident to a given vertex has a binomial distribution with parameters n and success probability λ/n. By Theorem 2.9, this distribution is close to a Poisson distribution with parameter λ. This suggests that also the corresponding branching processes, the one with a binomial offspring distribution with parameters n and λ/n, and the one with Poisson offspring distribution with mean λ, are close. In the following theorem, we make this statement more precise. In its statement, we write $\mathbb{P}_{n,p}$ for the law of a Binomial branching process with parameters n and success probability p.

Theorem 3.20 (Poisson and binomial branching processes)**.** *For a branching process with binomial offspring distribution with parameters n and p, and the branching process with Poisson offspring distribution with parameter $\lambda = np$, for each $k \ge 1$,*

$$\mathbb{P}_{n,p}(T \ge k) = \mathbb{P}^*_\lambda(T^* \ge k) + e_k(n), \quad (3.7.1)$$

where T and T^* are the total progenies of the binomial and Poisson branching processes, respectively, and where

$$|e_k(n)| \le \frac{2\lambda^2}{n}\sum_{s=1}^{k-1}\mathbb{P}^*_\lambda(T^* \ge s). \quad (3.7.2)$$

In particular, $|e_k(n)| \le \frac{2k\lambda^2}{n}$.

Proof. We use a coupling proof. The branching processes are described by their offspring distributions, which are binomial and Poisson random variables respectively. We use the coupling in Theorem 2.9 for each of the random variables X_i and X_i^* determining the branching processes, where $X_i \sim \mathrm{Bin}(n, \lambda/n)$, $X_i^* \sim \mathrm{Poi}(\lambda)$, and where

$$\mathbb{P}(X_i \neq X_i^*) \leq \frac{\lambda^2}{n}. \tag{3.7.3}$$

We use $\mathbb{P}$ to denote the joint probability distributions of the binomial and Poisson branching processes, where the offspring is coupled in the above way.

We start by noting that

$$\mathbb{P}_{n,p}(T \geq k) = \mathbb{P}(T \geq k, T^* \geq k) + \mathbb{P}(T \geq k, T^* < k), \tag{3.7.4}$$

and

$$\mathbb{P}^*_\lambda(T^* \geq k) = \mathbb{P}(T \geq k, T^* \geq k) + \mathbb{P}(T^* \geq k, T < k). \tag{3.7.5}$$

Subtracting the two probabilities yields

$$|\mathbb{P}_{n,p}(T \geq k) - \mathbb{P}^*_\lambda(T^* \geq k)| \leq \mathbb{P}(T \geq k, T^* < k) + \mathbb{P}(T^* \geq k, T < k). \tag{3.7.6}$$

We then use Theorem 2.9, as well as the fact that the event $\{T \geq k\}$ is determined by the values of $X_1, \ldots, X_{k-1}$ only. Indeed, by (3.3.1), by investigating $X_1, \ldots, X_{k-1}$, we can verify whether there exists a $t < k$ such that $X_1 + \cdots + X_t = t - 1$, implying that $T < k$. When there is no such t, then $T \geq k$. Similarly, by investigating $X_1^*, \ldots, X_{k-1}^*$, we can verify whether there exists a $t < k$ such that $X_1^* + \cdots + X_t^* = t - 1$, implying that $T < k$.

When $T \geq k$ and $T^* < k$, or when $T^* \geq k$ and $T < k$, there must be a value of $s < k$ for which $X_s \neq X_s^*$. Therefore, we can bound, by splitting depending on the first value $s < k$ where $X_s \neq X_s^*$,

$$\mathbb{P}(T \geq k, T^* < k) \leq \sum_{s=1}^{k-1} \mathbb{P}(X_i = X_i^* \ \forall i \leq s-1, X_s \neq X_s^*, T \geq k), \tag{3.7.7}$$

where $\{X_i^*\}_{i=1}^\infty$ are i.i.d. Poisson random variables with mean λ and $\{X_i\}_{i=1}^\infty$ are i.i.d. binomial random variables with parameters n and p. Now we note that when $X_i = X_i^*$ for all $i \leq s-1$ and $T \geq k$, this implies in particular that $X_1^* + \ldots + X_i^* \geq i$ for all $i \leq s-1$, which in turn implies that $T^* \geq s$. Moreover, the event $\{T^* \geq s\}$ depends only on $X_1^*, \ldots, X_{s-1}^*$, and, therefore, is independent of the event that $X_s \neq X_s^*$. Thus, we arrive at the fact that

$$\begin{aligned}\mathbb{P}(T \geq k, T^* < k) &\leq \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s, X_s \neq X_s^*) \\ &= \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s)\mathbb{P}(X_s \neq X_s^*).\end{aligned} \tag{3.7.8}$$

By Theorem 2.9,

$$\mathbb{P}(X_s \neq X_s^*) \leq \frac{\lambda^2}{n}, \tag{3.7.9}$$

so that

$$\mathbb{P}(T \geq k, T^* < k) \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s). \tag{3.7.10}$$

An identical argument yields that

$$\begin{aligned}\mathbb{P}(T^* \geq k, T < k) &\leq \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s)\mathbb{P}(X_s \neq X_s^*) \\ &\leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s). \end{aligned} \tag{3.7.11}$$

We conclude from (3.7.6) that

$$|\mathbb{P}_{n,p}(T \geq k) - \mathbb{P}_\lambda^*(T^* \geq k)| \leq \frac{2\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}_\lambda^*(T^* \geq s). \tag{3.7.12}$$

This completes the proof of Theorem 3.20. □

3.8 Notes and discussion

Notes on Section 3.6. The proof of Theorem 3.17 is taken from [159]. Theorem 3.16, together with (3.6.2), can also be proved making use of Lambert's W function. Indeed, we use that the generating function of the total progeny in (3.1.23), for Poisson branching process, reduces to

$$G_T(s) = s\mathrm{e}^{\lambda(G_T(s)-1)}. \tag{3.8.1}$$

Equation (3.8.1) actually defines a function analytic in $\mathbb{C}\backslash[1,\infty)$, and we are taking the principal branch. Equation (3.8.1) can be written in terms of the Lambert W function, which is defined by $W(x)\mathrm{e}^{W(x)} = x$, as $G_T(s) = -W(-s\lambda\mathrm{e}^{-\lambda})/\lambda$. The branches of W are described in [77], where also the fact that

$$W(x) = -\sum_{n=1}^{\infty} \frac{n^{n-1}}{n!}(-x)^n. \tag{3.8.2}$$

is derived. Theorem 3.17 follows immediately from this equation upon substituting $x = \lambda\mathrm{e}^{-\lambda}$ and using that the coefficient of s^n in $G_T(s)$ equals $\mathbb{P}(T = n)$. Also, since $\eta_\lambda = \lim_{s\uparrow 1} G_T(s) = -W(-\lambda\mathrm{e}^{-\lambda})/\lambda$. This also allows for a more direct proof of Corollary 3.19, since

$$\frac{d}{d\lambda}\eta_\lambda = -\frac{d}{d\lambda}\Big[\frac{W(-\lambda\mathrm{e}^{-\lambda})}{\lambda}\Big], \tag{3.8.3}$$

and where, since $W(x)\mathrm{e}^{W(x)} = x$,

$$W'(x) = \frac{1}{x}\frac{W(x)}{1+W(x)}. \tag{3.8.4}$$

We omit the details of this proof, taking a more combinatorial approach instead.

Notes on Section 3.5. The current proof is taken from [120], where also an extension is proved by conditioning on the numbers of steps of various sizes. The first proof of the special case of Theorem 3.14 for $k = 1$ can be found in [186]. The extension to $k \geq 2$ is in [141], or in [91] using a result in [90]. Most of these proofs make unnecessary use of generating functions, in particular, the Lagrange inversion formula, which the simple proof given here does not employ. See also [111, Page 165-167] for a more recent version of the generating function proof. In [222], various proofs of the hitting-time theorem are given,

including a combinatorial proof making use of a relation in [89]. A proof for random walks making only steps of size ± 1 using the reflection principle can for example be found in [111, Page 79].

The hitting-time theorem is closely related to the ballot theorem, which has a long history dating back to Bertrand in 1887 (see [150] for an excellent overview of the history and literature). The version of the ballot theorem in [150] states that, for a random walk $\{S_n\}_{n=0}^{\infty}$ starting at 0, with exchangeable, nonnegative steps, the probability that $S_m < m$ for all $m = 1, \ldots, n$, conditionally on $S_n = k$, equals k/n. This proof borrows upon queueing theory methodology, and is related to, yet slightly different from, our proof.

The ballot theorem for random walks with independent steps is the following result:

Theorem 3.21 (Ballot theorem)**.** *Consider a random walk with i.i.d. steps $\{X_i\}_{i=1}^{\infty}$ taking non-negative integer values. Then, with $S_m = X_1 + \cdots + X_m$ the position of the walk after m steps,*

$$\mathbb{P}_0(S_m < m \text{ for all } 1 \leq m \leq n | S_n = n - k) = \frac{k}{n}. \tag{3.8.5}$$

Exercise 3.31. *Prove the ballot theorem using the hitting-time theorem. Hint: Let $S'_m = k + (S_n - n) - (S_{n-m} - n + m)$, and note that $S_m < m$ for all $1 \leq m \leq n$ precisely when $S'_m > 0$ for all $0 \leq m < n$, and $\{S'_m\}_{m=0}^{n}$ is a random walk taking steps $Y_m = S'_m - S'_{m-1} = X_{n-m} - 1$.*

Chapter 4
Phase transition for the Erdős-Rényi random graph

In this chapter, we study the connected components of the Erdős-Rényi random graph. In the introduction in Section 4.1, we will argue that these connected components can be described in a similar way as for branching processes. As we have seen in Chapter 3, branching processes have a phase transition: when the mean offspring is below 1, the branching process dies out almost surely, while when the expected offspring exceeds 1, then it will survive with positive probability. The Erdős-Rényi random graph has a related phase transition. Indeed, when the expected degree is smaller than 1, the components are small, the largest one being of order $\log n$. On the other hand, when the expected degree exceeds 1, the there is a giant connected component which contains a positive proportion of all vertices. This phase transition can already be observed for relatively small graphs. For example, Figure 4.1 shows two realizations of Erdős-Rényi random graphs with 100 elements and expected degree close to 1/2, respectively, 3/2. The left picture is in the subcritical regime, and the connected components are tiny, while the right picture is in the supercritical regime, and the largest connected component is already substantial. The aim of this chapter is to quantify these facts.

The link between the Erdős-Rényi random graph and branching processes is described in more detail in Section 4.2, where we prove upper and lower bounds for the tails of the cluster size (or connected component size) distribution. The *connected component containing* v is denoted by $\mathcal{C}(v)$, and consists of all vertices that can be reached from v using occupied edges. We sometimes also call $\mathcal{C}(v)$ the *cluster of* v. The connection between branching processes and clusters is used extensively in the later sections, Section 4.3–4.5. In Section 4.3, we study the subcritical regime of the Erdős-Rényi random graph. In Sections 4.4 and 4.5 we study the supercritical regime of the Erdős-Rényi random graph, by proving a law of large numbers for the largest connected component in Section 4.4 and a central limit theorem in Section 4.5.

In Chapter 5, we shall investigate several more properties of the Erdős-Rényi random graph. In particular, in Section 5.1, we study the bounds on the component sizes of the critical Erdős-Rényi random graph, in Section 5.1.3 we describe the weak limits of the connected components ordered in size at criticality, in Section 5.2 we study the connectivity threshold of the Erdős-Rényi random graph, while in Section 5.3 we prove that the Erdős-Rényi random graph is sparse and identify its asymptotic degree sequence.

4.1 Introduction

In this section, we introduce some notation for the Erdős-Rényi random graph, and prove some elementary properties. We recall from Section 1.5 that the Erdős-Rényi random graph has vertex set $[n] = \{1, \ldots, n\}$, and, denoting the edge between vertices $s, t \in [n]$ by st, st is occupied or present with probability p, and absent or vacant otherwise, independently of all the other edges. The parameter p is called the *edge probability*. The above random graph is denoted by $\mathrm{ER}_n(p)$.

Exercise 4.1 (Number of edges in $\mathrm{ER}_n(p)$)**.** *What is the distribution of the number of edges in the Erdős-Rényi random graph* $\mathrm{ER}_n(p)$*?*

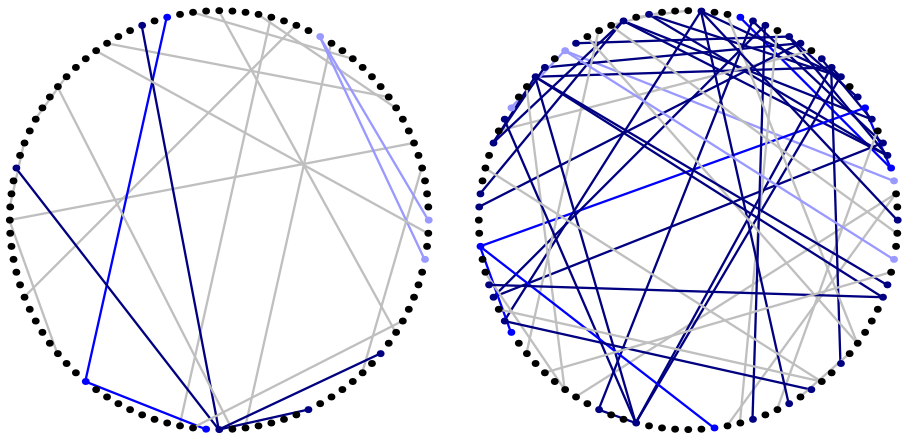

Figure 4.1: Two realizations of Erdős-Rényi random graphs with 100 elements and edge probabilities 1/200, respectively, 3/200. The three largest connected components are ordered by the darkness of their edge colors, the remaining connected components have edges with the lightest shade.

Exercise 4.2 (CLT for number of edges in $\mathrm{ER}_n(p)$)**.** *Prove that the number of edges in* $\mathrm{ER}_n(p)$ *satisfies a central limit theorem and compute its asymptotic mean and variance.*

We now introduce some notation. For two vertices $s, t \in [n]$, we write $s \longleftrightarrow t$ when there exists a path of occupied edges connecting s and t. By convention, we always assume that $v \longleftrightarrow v$. For $v \in [n]$, we denote the *connected component containing* v or *cluster of* v by

$$\mathcal{C}(v) = \big\{x \in [n] \colon v \longleftrightarrow x\big\}. \tag{4.1.1}$$

We denote the size of $\mathcal{C}(v)$ by $|\mathcal{C}(v)|$. The *largest connected component* is equal to any cluster $\mathcal{C}(v)$ for which $|\mathcal{C}(v)|$ is maximal, so that

$$|\mathcal{C}_{\max}| = \max\{|\mathcal{C}(v)| \colon v = 1, \dots, n\}. \tag{4.1.2}$$

Note that the above definition does identify $|\mathcal{C}_{\max}|$ uniquely, but it may not identify $\mathcal{C}_{\max}$ uniquely. We can make this definition unique, by requiring that $\mathcal{C}_{\max}$ is the cluster of maximal size containing the vertex with the smallest label. As we will see, the typical size of $\mathcal{C}_{\max}$ will depend sensitively on the value λ.

We first define a procedure to find the connected component $\mathcal{C}(v)$ containing a given vertex v in a given graph G. This procedure is closely related to the random walk perspective for branching processes described in Section 3.3, and works as follows. In the course of the exploration, vertices can have three different statuses: vertices are active, neutral or inactive. The status of vertices is changed in the course of the exploration of the connected component of v, as follows. At time $t = 0$, only v is active and all other vertices are neutral, and we set $S_0 = 1$. At each time t, we choose an active vertex w in an arbitrary way (for example, by taking the smallest active vertex) and explore all the edges ww', where w' runs over all the neutral vertices. If there is an edge in G connecting the active vertex w and some neutral vertex w', then we set w' active, otherwise it remains neutral. After searching the entire set of neutral vertices, we set w inactive and we let S_t equal the new number of active vertices at time t. When there are no more active vertices,

i.e., when $S_t = 0$ for the first time, the process terminates and $\mathcal{C}(v)$ is the set of all inactive vertices, i.e., $|\mathcal{C}(v)| = t$. Note that at any stage of the process, the size of $\mathcal{C}(v)$ is bounded from below by the sum of the number of active and inactive vertices.

Let w_t be the t^{th} active vertex of which all edges to neutral vertices are explored. Let X_t denote the number of neutral vertices w' with $w_t w' \in G$. Let S_t be the total number of active vertices at time t. Similarly as for the branching process in (3.3.1), we can represent this procedure with the recursive relation

$$S_0 = 1, \qquad S_t = S_{t-1} + X_t - 1. \tag{4.1.3}$$

The variable X_t is the number of vertices that become active due to the exploration of the t^{th} vertex, and after its exploration, the t^{th} explored vertex becomes inactive. Thus, if S_{t-1} denotes the number of active vertices after the exploration of $(t-1)$ vertices, then $S_t = S_{t-1} + X_t - 1$ denotes the number of active vertices after the exploration of t vertices. This explains (4.1.3).

The above description is true for *any* graph G. We now specialize to the random graph $\mathrm{ER}_n(p)$, where each edge can be independently occupied or vacant. As a result, the distribution of X_t depends on the number of active vertices at time $t-1$, i.e., on S_{t-1}, and not in any other way on which vertices are active, inactive or neutral. More precisely, each neutral w' in the random graph has probability p to become active. The edges ww' are examined precisely once, so that the conditional probability for $ww' \in \mathrm{ER}_n(p)$ is always equal to p. After $t-1$ explorations of active vertices, we have $t-1$ inactive vertices and S_{t-1} active vertices. This leaves $n-(t-1)-S_{t-1}$ neutral vertices. Therefore, conditionally on S_{t-1},

$$X_t \sim \mathrm{Bin}\big(n-(t-1)-S_{t-1}, p\big). \tag{4.1.4}$$

We note that the recursion in (4.1.3) is identical to the recursive relation (3.3.1). The only difference is the distribution of the process $\{X_i\}_{i=1}^n$, as described in (4.1.4). For branching processes, $\{X_i\}_{i=1}^n$ is an i.i.d. sequence, but for the exploration of connected components, we see that this is not quite true. However, by (4.1.4), it is 'almost' true as long as the number of active vertices is not too large. We see in (4.1.4) that the parameter of the binomial distribution decreases. This is due to the fact that after more explorations, fewer neutral vertices remain, and is sometimes called the *depletion of points* effect.

Let T be the least t for which $S_t = 0$, i.e.,

$$T = \inf\{t : S_t = 0\}, \tag{4.1.5}$$

then $|\mathcal{C}(v)| = T$, see also (1.5.10) for a similar result in the branching process setting. This describes the exploration of a single connected component. While of course the recursion in (4.1.3) and (4.1.4) only makes sense when $S_{t-1} \geq 1$, that is, when $t \leq T$, there is no harm in continuing it formally for $t > T$. This will be prove to be extremely useful later on.

Exercise 4.3 (Verification of cluster size description)**.** *Verify that $T = |\mathcal{C}(v)|$ by computing the probabilities of the events that $\{|\mathcal{C}(v)| = 1\}$, $\{|\mathcal{C}(v)| = 2\}$ and $\{|\mathcal{C}(v)| = 3\}$ directly, and by using* (4.1.4)*,* (4.1.3) *and* (4.1.5)*.*

We end this section by introducing some notation. For the Erdős-Rényi random graph, the status of all edges $\{st : 1 \leq s < t \leq n\}$ are i.i.d. random variables taking the value 1 with probability p and the value 0 with probability $1-p$, 1 denoting that the edge is occupied and 0 that it is vacant. We will sometimes call the edge probability p, and sometimes λ/n. We will always use the convention that

$$p = \frac{\lambda}{n}. \tag{4.1.6}$$

We shall write $\mathbb{P}_\lambda$ for the distribution of $\mathrm{ER}_n(p) = \mathrm{ER}_n(\lambda/n)$.

Exercise 4.4 (CLT for number of edges in $\mathrm{ER}_n(\lambda/n)$). *Prove that the number of edges in* $\mathrm{ER}_n(\lambda/n)$ *satisfies a central limit theorem with asymptotic mean and variance equal to* $\lambda n/2$.

Exercise 4.5 (Mean number of triangles in $\mathrm{ER}_n(\lambda/n)$). *We say that the distinct vertices* (i,j,k) *form an occupied triangle when the edges* ij, jk *and* ki *are all occupied. Note that* (i,j,k) *is the same triangle as* (i,k,j) *and as any other permutation. Compute the expected number of occupied triangles in* $\mathrm{ER}_n(\lambda/n)$.

Exercise 4.6 (Mean number of squares in $\mathrm{ER}_n(\lambda/n)$). *We say that the distinct vertices* (i,j,k,l) *form an occupied square when the edges* ij, jk, kl *and* li *are all occupied. Note that the squares* (i,j,k,l) *and* (i,k,j,l) *are different. Compute the expected number of occupied squares in* $\mathrm{ER}_n(\lambda/n)$.

Exercise 4.7 (Poisson limits for number of triangles and squares in $\mathrm{ER}_n(\lambda/n)$). *Show that the number of occupied triangles in an Erdős-Rényi random graph with edge probability* $p=\lambda/n$ *has an asymptotic Poisson distribution. Do the same for the number of occupied squares. Hint: use the method of moments in Theorem 2.4.*

Exercise 4.8 (Clustering of $\mathrm{ER}_n(\lambda/n)$). *Define the clustering coefficient of a random graph* G *to be*

$$CC_G = \frac{\mathbb{E}[\Delta_G]}{\mathbb{E}[W_G]}, \tag{4.1.7}$$

where

$$\Delta_G = \sum_{i,j,k\in G} \mathbb{1}_{\{ij,ik,jk \text{ occupied}\}}, \qquad W_G = \sum_{i,j,k\in G} \mathbb{1}_{\{ij,ik \text{ occupied}\}}. \tag{4.1.8}$$

Thus, Δ_G *is six times the number of triangles in* G, *and* W_G *is two times the number of open wedges in* G, *and* CC_G *is the ratio of the number of expected closed triangles to the expected number of open wedges. Compute* CC_G *for* $\mathrm{ER}_n(\lambda/n)$.

Exercise 4.9 (Asymptotic clustering of $\mathrm{ER}_n(\lambda/n)$). *Show that* $W_G/n \xrightarrow{\mathbb{P}} \lambda^2$ *by using the second moment method. Use Exercise 4.7 to conclude that*

$$\frac{n\Delta_G}{W_G} \xrightarrow{d} \frac{3}{\lambda^2} Y, \tag{4.1.9}$$

where $Y \sim \mathrm{Poi}(\lambda^3/6)$.

4.1.1 Monotonicity of Erdős-Rényi random graphs in the edge probability

In this section, we investigate Erdős-Rényi random graphs with different values of p, and show that the Erdős-Rényi random graph is *monotonically increasing in* p, using a coupling argument. The material in this section makes it clear that components of the Erdős-Rényi random graph are growing with the edge probability p, as one would intuitively expect. This material shall also play a crucial role in determining the critical behavior of the Erdős-Rényi random graph in Section 5.1 below.

We use a coupling of all random graphs $\mathrm{ER}_n(p)$ for all $p \in [0,1]$. For this, we draw independent uniform random variables for each edge st, and, for fixed p, we declare an edge to be p-occupied if and only if $U_{st} \le p$. The above coupling shows that the number of occupied bonds increases when p increases. Therefore, the Erdős-Rényi random graph $\mathrm{ER}_n(p)$ is monotonically increasing in p. Because of the monotone nature of $\mathrm{ER}_n(p)$ one expects that certain events and random variables grow larger when p increases. This is formalized in the following definition:

Definition 4.1 (Increasing events and random variables)**.** *We say that an event is increasing when, if the event occurs for a given set of occupied edges, it remains to hold when we make some more edges occupied.*
We say that a random variable X is increasing when the events $\{X \geq x\}$ are increasing for each $x \in \mathbb{R}$.

An example of an increasing event is $\{s \longleftrightarrow t\}$. An example of a monotone random variable is $|\mathcal{C}(v)|$ and the maximal cluster $|\mathcal{C}_{\max}|$, where

$$|\mathcal{C}_{\max}| = \max_{v=1}^{n} |\mathcal{C}(v)|. \tag{4.1.10}$$

Exercise 4.10. *Show that $|\mathcal{C}_{\max}|$ is an increasing random variable.*

Exercise 4.11. *Is the event $\{v \in \mathcal{C}_{\max}\}$ an increasing event?*

4.1.2 Informal link to Poisson branching processes

We now describe the link to Poisson branching processes in an informal manner. The results in this section will not be used in the remainder of the chapter, even though the philosophy forms the core of the argument. Fix $\lambda > 0$. Let $S_0^*, S_1^*, \ldots, X_1^*, X_2^*, \ldots, H^*$ refer to the history of a branching process with Poisson offspring distribution with mean λ and $S_0, S_1, \ldots, X_1, X_2, \ldots, H$ refer to the history of the random graph, where $S_0, S_1, \ldots$ are defined in (4.1.3) above. The event $\{H^* = (x_1, \ldots, x_t)\}$ is the event that the total progeny T^* of the Poisson branching process is equal to t, and the values of $X_1^*, \ldots, X_t^*$ are given by $x_1, \ldots, x_t$. Recall that $\mathbb{P}_\lambda^*$ denotes the law of a Poisson branching process with mean offspring distribution λ. Naturally, by (3.3.2), we have that

$$t = \min\{i : s_i = 0\} = \min\{i : x_1 + \ldots + x_i = i - 1\}, \tag{4.1.11}$$

where

$$s_0 = 1, \qquad s_i = s_{i-1} + x_i - 1. \tag{4.1.12}$$

For any possible history $(x_1, \ldots, x_t)$, we have that (recall (3.3.6))

$$\mathbb{P}_\lambda^*(H^* = (x_1, \ldots, x_t)) = \prod_{i=1}^{t} \mathbb{P}_\lambda^*(X_i^* = x_i), \tag{4.1.13}$$

where $\{X_i^*\}_{i=1}^{\infty}$ are i.i.d. Poisson random variables with mean λ, while

$$\mathbb{P}_\lambda(H = (x_1, \ldots, x_t)) = \prod_{i=1}^{t} \mathbb{P}_\lambda(X_i = x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}),$$

where, conditionally on $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}$, the random variable X_i is binomially distributed $\mathrm{Bin}(n - (i-1) - s_{i-1}, \lambda/n)$, recall (4.1.4) and (4.1.12).

As shown in Theorem 2.9, the Poisson distribution is the limiting distribution of binomials when n is large and $p = \lambda/n$. When $m(n) = n(1 + o(1))$ and λ, i are fixed, then we can extend this to

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathrm{Bin}\big(m(n), \lambda/n\big) = i\Big) = \mathrm{e}^{-\lambda} \frac{\lambda^i}{i!}. \tag{4.1.14}$$

Therefore, for every $t < \infty$,

$$\lim_{n\to\infty} \mathbb{P}_\lambda\big(H = (x_1, \ldots, x_t)\big) = \mathbb{P}_\lambda^*\big(H^* = (x_1, \ldots, x_t)\big). \tag{4.1.15}$$

Thus, the distribution of finite connected components in the random graph $\mathrm{ER}_n(\lambda/n)$ is closely related to a Poisson branching process with mean λ. This relation shall be explored further in the remainder of this chapter.

4.2 Comparisons to branching processes

In this section, we investigate the relation between connected components and binomial branching processes. We start by proving two stochastic domination results for connected components in the Erdős-Rényi random graph. In Theorem 4.2, we give a stochastic upper bound on $|\mathcal{C}(v)|$, and in Theorem 4.3 a lower bound on the cluster tails. These bounds will be used in the following sections to prove results concerning $|\mathcal{C}_{\max}|$.

4.2.1 Stochastic domination of connected components

We prove the following upper bound, which shows that each connected component is bounded from above by the total progeny of a branching process with binomial offspring distribution:

Theorem 4.2 (Stochastic domination of the cluster size)**.** *For each $k \geq 1$,*

$$\mathbb{P}_{np}(|\mathcal{C}(1)| \geq k) \leq \mathbb{P}_{n,p}(T^{\geq} \geq k), \qquad \textit{i.e.,} \qquad |\mathcal{C}(1)| \preceq T^{\geq}, \tag{4.2.1}$$

where $T^{\geq}$ is the total progeny of a binomial branching process with parameters n and p.

Proof. Let $N_i = n - i - S_i$ denote the number of neutral vertices after i explorations, so that, conditionally on N_{i-1}, $X_i \sim \mathrm{Bin}(N_{i-1}, p)$. Let $Y_i \sim \mathrm{Bin}(n - N_{i-1}, p)$, and write

$$X_i^{\geq} = X_i + Y_i. \tag{4.2.2}$$

Then, conditionally on $(X_j)_{j=1}^{i-1}$, $X_i^{\geq} \sim \mathrm{Bin}(n, p)$. Since this distribution is independent of $(X_j)_{j=1}^{i-1}$, the sequence $(X_j^{\geq})_{j\geq 1}$ is in fact i.i.d. Also, $X_i^{\geq} \leq X_i$ a.s. since $Y_i \geq 0$ a.s.

Denote

$$S_i^{\geq} = X_1^{\geq} + \ldots + X_i^{\geq} - (i-1). \tag{4.2.3}$$

Then,

$$\mathbb{P}_{np}(|\mathcal{C}(1)| \geq k) = \mathbb{P}(S_t > 0 \forall t \leq k-1) \leq \mathbb{P}(S_t^{\geq} > 0 \forall t \leq k-1) = \mathbb{P}(T^{\geq} \geq k), \tag{4.2.4}$$

where $T^{\geq} = \min\{t\colon S_t^{\geq} = 0\}$ is the total progeny of a branching process with binomial distribution with parameters n and success probability p. □

Exercise 4.12 (Upper bound for mean cluster size)**.** *Show that, for $\lambda < 1$, $\mathbb{E}_\lambda[|\mathcal{C}(v)|] \leq 1/(1-\lambda)$.*

4.2.2 Lower bound on the cluster tail

We prove the following lower bound, which shows that the probability that a connected component has size at least k is bounded from below by the probability that the total progeny of a branching process with binomial offspring distribution exceeds k, where now the parameters of the binomial distribution are $n-k$ and p:

Theorem 4.3 (Lower bound on cluster tail)**.** *For every $k \in [n]$,*

$$\mathbb{P}_{np}(|\mathcal{C}(1)| \geq k) \geq \mathbb{P}_{n-k,p}(T^{\leq} \geq k), \tag{4.2.5}$$

where $T^{\leq}$ is the total progeny of a branching process with binomial distribution with parameters $n-k$ and success probability $p = \lambda/n$.

Note that, since the parameter $n-k$ on the right-hand side of (4.2.5) depends explicitly on k, Theorem 4.3 does *not* imply a stochastic lower bound on $|\mathcal{C}(1)|$.

Proof. We again use a coupling approach. Recall that N_i denotes the number of neutral vertices after i explorations. Denote the stopping time $\mathcal{T}_k$ by

$$\mathcal{T}_k = \min\{t\colon N_t \leq n-k\}. \tag{4.2.6}$$

Then, $\mathcal{T}_k \leq k-1$, since $N_{k-1} \leq n-(k-1)-1 = n-k$. We let $X_i^{\leq}$ denote an i.i.d. sequence of $\mathrm{Bin}(n-k,p)$ random variables. For $i \leq \mathcal{T}_k+1$, and conditionally on N_{i-1}, let Y_i have a $\mathrm{Bin}(N_{i-1}-(n-k),p)$ distribution independently of all other random variables involved. Define

$$X_i = X_i^{\leq} + Y_i. \tag{4.2.7}$$

Then, clearly, $X_i \leq X_i^{\leq}$ a.s. for all $i \leq \mathcal{T}_k+1$.

Now, we can rewrite

$$\mathbb{P}_{np}(|\mathcal{C}(1)| \geq k) = \mathbb{P}_{np}(S_t > 0 \forall t \leq k-1) = \mathbb{P}_{np}(S_t > 0 \forall t \leq \mathcal{T}_k). \tag{4.2.8}$$

Indeed, if $S_t > 0$ for all $t \leq \mathcal{T}_k$, then also $S_t > 0$ for all $t \leq k-1$, since then (a) $S_{\mathcal{T}_k} = n - \mathcal{T}_k - N_{\mathcal{T}_k} \geq n - \mathcal{T}_k - (n-k) = k - \mathcal{T}_k$ and (b) $t \mapsto S_t$ goes down by at most 1, so that $S_t > 0$ for all $t \leq \mathcal{T}_k + (k-\mathcal{T}_k) - 1 = k-1$, as required.

Denote

$$S_i^{\leq} = X_1^{\leq} + \ldots + X_i^{\leq} - (i-1). \tag{4.2.9}$$

Then, $S_t^{\leq} \leq S_t$ for all $t \leq \mathcal{T}_k$. Using the above coupling and the fact that $\mathcal{T}_k \leq k-1$, we can therefore bound

$$\{S_t > 0 \forall t \leq \mathcal{T}_k - 1\} \supseteq \{S_t^{\leq} > 0 \forall t \leq \mathcal{T}_k\} \supseteq \{S_t^{\leq} > 0 \forall t \leq k-1\} = \{T^{\leq} \geq k\}, \tag{4.2.10}$$

where $T^{\leq} = \min\{t\colon S_t^{\leq} = 0\}$ is the total progeny of a branching process with binomial distribution with parameters $n-k$ and success probability p. □

The general strategy for the investigation of the largest connected component $|\mathcal{C}_{\max}|$ is as follows. We make use of the stochastic bounds in Theorems 4.2–4.3 in order to compare the cluster sizes to binomial branching processes. Then, using Theorem 3.20, we can make the comparison to a Poisson branching process with a parameter that is close to the parameter λ in $\mathrm{ER}_n(\lambda/n)$. Using the results on branching processes in Chapter 3 then allows us to complete the proofs.

By Theorems 4.2–4.3, the connected components of the Erdős-Rényi random graph are closely related to binomial branching processes with a binomial offspring with parameters n and $p = \lambda/n$. By Theorem 3.1, the behavior of branching processes is rather different when the expected offspring is larger than 1 or smaller than or equal to 1. In Theorems 4.2–4.3, when $k = o(n)$, the expected offspring is close to $np \approx \lambda$. Therefore, for the Erdős-Rényi random graph, we expect different behavior in the subcritical regime $\lambda < 1$, in the supercritical regime $\lambda > 1$ and in the critical regime $\lambda = 1$.

The proof of the behavior of the largest connected component $|\mathcal{C}_{\max}|$ is substantially different in the subcritical regime where $\lambda < 1$, which is treated in Section 4.3, compared to the supercritical regime $\lambda > 1$, which is treated in Section 4.4. In Section 4.5, we prove a central limit theorem for the giant supercritical component. The critical regime $\lambda = 1$ requires some new ideas, and is treated in Section 5.1.

4.3 The subcritical regime

In this section, we derive bounds for the size of the largest connected component for the Erdős-Rényi random graph in the subcritical regime, i.e., when $\lambda = np < 1$. Let I_λ denote the large deviation rate function for Poisson random variables with mean λ, given by

$$I_\lambda = \lambda - 1 - \log(\lambda). \tag{4.3.1}$$

Recall Exercise 2.17 to see an upper bound on Poisson random variables involving I_λ, as well as the fact that $I_\lambda > 0$ for all $\lambda \neq 1$.

The main results when $\lambda < 1$ are Theorem 4.4, which proves that $|\mathcal{C}_{\max}| \leq a \log n$ with high probability, for any $a > I_\lambda^{-1}$, and Theorem 4.5, where a matching lower bound on $|\mathcal{C}_{\max}|$ is provided by proving that $|\mathcal{C}_{\max}| \geq a \log n$ with high probability, for any $a < I_\lambda^{-1}$. These results are stated now:

Theorem 4.4 (Upper bound on largest subcritical component)**.** *Fix $\lambda < 1$. Then, for every $a > I_\lambda^{-1}$, there exists a $\delta = \delta(a, \lambda) > 0$ such that*

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| \geq a \log n) = O(n^{-\delta}). \tag{4.3.2}$$

Theorem 4.5 (Lower bound on largest subcritical component)**.** *Fix $\lambda < 1$. Then, for every $a < I_\lambda^{-1}$, there exists a $\delta = \delta(a, \lambda) > 0$ such that*

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| \leq a \log n) = O(n^{-\delta}). \tag{4.3.3}$$

Theorems 4.4 and 4.5 will be proved in Sections 4.3.2 and 4.3.3 below. Together, they prove that $|\mathcal{C}_{\max}|/\log n \xrightarrow{\mathbb{P}} I_\lambda^{-1}$:

Exercise 4.13 (Convergence in probability of largest subcritical cluster)**.** *Prove that Theorems 4.4 and 4.5 imply $|\mathcal{C}_{\max}|/\log n \xrightarrow{\mathbb{P}} I_\lambda^{-1}$.*

4.3.1 Largest subcritical cluster: strategy of proof of Theorems 4.4 and 4.5

We start by describing the strategy of proof. We denote by

$$Z_{\geq k} = \sum_{v=1}^{n} \mathbb{1}_{\{|\mathcal{C}(v)| \geq k\}} \tag{4.3.4}$$

the number of vertices that are contained in connected components of size at least k. We can identify $|\mathcal{C}_{\max}|$ as

$$|\mathcal{C}_{\max}| = \max\{k : Z_{\geq k} \geq k\}, \tag{4.3.5}$$

which allows us to prove bounds on $|\mathcal{C}_{\max}|$ by investigating $Z_{\geq k}$ for an appropriately chosen k. In particular, (4.3.5) implies that $\{|\mathcal{C}_{\max}| \geq k\} = \{Z_{\geq k} \geq k\}$:

Exercise 4.14 (Relation $|\mathcal{C}_{\max}|$ and $Z_{\geq k}$)**.** *Prove* (4.3.5) *and conclude that $\{|\mathcal{C}_{\max}| \geq k\} = \{Z_{\geq k} \geq k\}$.*

To prove Theorem 4.4, we use the first moment method or Markov inequality (Theorem 2.14). We compute that

$$\mathbb{E}_\lambda[Z_{\geq k}] = n\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k), \tag{4.3.6}$$

and we use Theorem 4.2 to bound $\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k_n)$ for $k_n = a \log n$ for any $a > I_\lambda^{-1}$. Therefore, with high probability, $Z_{\geq k_n} = 0$, so that, again with high probability, $|\mathcal{C}_{\max}| \leq k_n$. This proves Theorem 4.4. For the details we refer to the formal argument in Section 4.3.2.

To prove Theorem 4.5, we use the second moment method or Chebychev inequality (Theorem 2.15). In order to be able to apply this result, we first prove an upper bound on the variance of $Z_{\geq k}$, see Proposition 4.7 below. We further use Theorem 4.3 to prove a lower bound on $\mathbb{E}_\lambda[Z_{\geq k_n}]$, now for $k_n = a \log n$ for any $a < I_\lambda^{-1}$. Then, (2.4.5) in Theorem 2.15 proves that with high probability, $Z_{\geq k_n} > 0$, so that, again with high probability, $|\mathcal{C}_{\max}| \geq k_n$. We now present the details of the proofs.

4.3.2 Upper bound on the largest subcritical cluster: proof of Theorem 4.4

By Theorem 4.2,

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| > t) \leq \mathbb{P}_{n,p}(T > t), \tag{4.3.7}$$

where T is the total progeny of a branching process with a binomial offspring distribution with parameters n and $p = \lambda/n$. To study $\mathbb{P}_{n,p}(T > t)$, we let $\{\hat{X}_i\}_{i=1}^\infty$ be an i.i.d. sequence of binomial random variables with parameters n and success probability p, and let

$$\hat{S}_t = \hat{X}_1 + \ldots + \hat{X}_t - (t-1). \tag{4.3.8}$$

Then, by (3.3.2) and (3.3.1), we have that

$$\mathbb{P}_{n,p}(T > t) \leq \mathbb{P}_{n,p}(\hat{S}_t > 0) = \mathbb{P}_{n,p}(\hat{X}_1 + \ldots + \hat{X}_t \geq t) \leq \mathrm{e}^{-tI_\lambda}, \tag{4.3.9}$$

by Corollary 2.17 and using the fact that $\hat{X}_1 + \ldots + \hat{X}_t \sim \mathrm{Bin}(nt, \lambda/n)$. We conclude that

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| > t) \leq \mathrm{e}^{-tI_\lambda}. \tag{4.3.10}$$

Therefore, using Exercise 4.14, the Markov inequality (Theorem 2.14) and again with $k_n = a \log n$,

$$\begin{aligned}\mathbb{P}_\lambda(|\mathcal{C}_{\max}| > a \log n) &\leq \mathbb{P}_\lambda(Z_{\geq k_n} \geq 1) \leq \mathbb{E}_\lambda[Z_{\geq k_n}] \\ &= n\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq a \log n) \leq n^{1-aI_\lambda}\mathrm{e}^{I_\lambda} = O(n^{-\delta}),\end{aligned} \tag{4.3.11}$$

whenever $a > 1/I_\lambda$ and with $\delta = aI_\lambda - 1$. This proves that with high probability the largest connected component is bounded by $a \log n$ for every $a > I_\lambda^{-1}$. □

We now give a second proof of (4.3.10), which is based on a distributional equality of S_t, and which turns out to be useful in the analysis of the Erdős-Rényi random graph with $\lambda > 1$ as well. The result states that S_t is also binomially distributed, but with a different success probability. In the statement of Proposition 4.6 below, we make essential use of the formal continuation of the recursions in (4.1.3) and (4.1.4) for the breadth-first search, defined right below (4.1.4). Note that, in particular, S_t need not be non-negative.

Proposition 4.6 (The law of S_t)**.** *For all $t \in [n]$,*

$$S_t + (t-1) \sim \mathrm{Bin}\Big(n-1, 1-(1-p)^t\Big). \tag{4.3.12}$$

We shall only make use of Proposition 4.6 when $|\mathcal{C}(v)| \geq t$, in which case $S_t \geq 0$ *does* hold.

Proof. Let N_t represent the number of unexplored vertices, i.e.,

$$N_t = n - t - S_t. \tag{4.3.13}$$

Note that $X \sim \mathrm{Bin}(m,p)$ holds precisely when $Y = m - X \sim \mathrm{Bin}(m, 1-p)$. It is more convenient to show the equivalent statement that for all t

$$N_t \sim \mathrm{Bin}\Big(n-1, (1-p)^t\Big). \tag{4.3.14}$$

Heuristically, (4.3.14) can be understood by noting that each of the vertices $\{2, \ldots, n\}$ has, independently of all other vertices, probability $(1-p)^t$ to stay neutral in the first t

explorations. More formally, conditionally on S_t, we have that $X_t \sim \mathrm{Bin}\big(n-(t-1)-S_{t-1},p\big)$ by (4.1.4). Thus, noting that $N_0 = n-1$ and

$$\begin{aligned} N_t &= n-t-S_t = n-t-S_{t-1}-X_t+1 \\ &= n-(t-1)-S_{t-1}-\mathrm{Bin}(n-(t-1)-S_{t-1},p) \\ &= N_{t-1}-\mathrm{Bin}(N_{t-1},p) = \mathrm{Bin}(N_{t-1},1-p), \end{aligned} \tag{4.3.15}$$

the conclusion follows by recursion on t. □

Exercise 4.15 (A binomial number of binomial trials)**.** *Show that if $N \sim \mathrm{Bin}(n,p)$ and, conditionally on N, $M \sim \mathrm{Bin}(N,q)$, then $M \sim \mathrm{Bin}(n,pq)$. Use this to complete the proof that $N_t \sim \mathrm{Bin}(n-1,(1-p)^t)$.*

To complete the second proof of (4.3.10), we use Proposition 4.6 to see that

$$\mathbb{P}_\lambda(|\mathcal{C}(v)|>t) \le \mathbb{P}(S_t>0) \le \mathbb{P}_\lambda\big(\mathrm{Bin}(n-1,1-(1-p)^t)\ge t\big). \tag{4.3.16}$$

Using Bernoulli's inequality $1-(1-p)^t \le tp$, we therefore arrive at

$$\begin{aligned} \mathbb{P}_\lambda(|\mathcal{C}(v)|>t) &\le \mathbb{P}_\lambda\big(\mathrm{Bin}(n,\frac{t\lambda}{n})\ge t\big) \le \min_{s\ge 0} \mathrm{e}^{-st}\mathbb{E}_\lambda[\mathrm{e}^{s\mathrm{Bin}(n,\frac{t\lambda}{n})}] \\ &= \min_{s\ge 0} \mathrm{e}^{-st}\big[1+\frac{t\lambda}{n}(\mathrm{e}^s-1)\big]^n \le \min_{s\ge 0} \mathrm{e}^{-st}\mathrm{e}^{t\lambda(\mathrm{e}^s-1)}, \end{aligned} \tag{4.3.17}$$

where we have used the Markov inequality (Theorem 2.14) in the second inequality, and $1+x\le \mathrm{e}^x$ in the last. We arrive at the bound

$$\mathbb{P}_\lambda(|\mathcal{C}(v)|>t) \le \mathrm{e}^{-I_\lambda t}, \tag{4.3.18}$$

which reproves (4.3.10). □

4.3.3 Lower bound on the largest subcritical cluster: proof of Theorem 4.5

The proof of Theorem 4.5 makes use of a variance estimate on $Z_{\ge k}$. We use the notation

$$\chi_{\ge k}(\lambda) = \mathbb{E}_\lambda\Big[|\mathcal{C}(v)|\mathbb{1}_{\{|\mathcal{C}(v)|\ge k\}}\Big]. \tag{4.3.19}$$

Note that, by exchangeability of the vertices, $\chi_{\ge k}(\lambda)$ does not depend on v.

Proposition 4.7 (A variance estimate for $Z_{\ge k}$)**.** *For every n and $k\in[n]$,*

$$\mathrm{Var}_\lambda(Z_{\ge k}) \le n\chi_{\ge k}(\lambda). \tag{4.3.20}$$

Proof. We use that

$$\mathrm{Var}_\lambda(Z_{\ge k}) = \sum_{i,j=1}^n \Big[\mathbb{P}_\lambda(|\mathcal{C}(i)|\ge k,|\mathcal{C}(j)|\ge k) - \mathbb{P}_\lambda(|\mathcal{C}(i)|\ge k)\mathbb{P}_\lambda(|\mathcal{C}(j)|\ge k)\Big]. \tag{4.3.21}$$

We split the probability $\mathbb{P}_\lambda(|\mathcal{C}(i)|\ge k,|\mathcal{C}(j)|\ge k)$, depending on whether $i \longleftrightarrow j$ or not:

$$\mathbb{P}_\lambda(|\mathcal{C}(i)|\ge k,|\mathcal{C}(j)|\ge k) = \mathbb{P}_\lambda(|\mathcal{C}(i)|\ge k, i\longleftrightarrow j) + \mathbb{P}_\lambda(|\mathcal{C}(i)|\ge k,|\mathcal{C}(j)|\ge k, i\not\longleftrightarrow j). \tag{4.3.22}$$

Clearly,

$$\begin{aligned} &\mathbb{P}_\lambda(|\mathcal{C}(i)|=l,|\mathcal{C}(j)|\ge k, i\not\longleftrightarrow j) \\ &\quad = \mathbb{P}_\lambda(|\mathcal{C}(i)|=l, i\not\longleftrightarrow j)\mathbb{P}_\lambda\big(|\mathcal{C}(j)|\ge k \;\big|\; |\mathcal{C}(i)|=l, i\not\longleftrightarrow j\big). \end{aligned} \tag{4.3.23}$$

When $|\mathcal{C}(i)| = l$ and $i \not\longleftrightarrow j$, then all vertices in the components different from the one of i, which includes the components of j, form a random graph where the size n is replaced by $n-l$. Since the probability that $|\mathcal{C}(j)| \geq k$ in $\mathrm{ER}_n(p)$ is increasing in n, we have that

$$\mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k \big| |\mathcal{C}(i)| = l, i \not\longleftrightarrow j) \leq \mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k). \tag{4.3.24}$$

We conclude that

$$\mathbb{P}_\lambda(|\mathcal{C}(i)| = l, |\mathcal{C}(j)| \geq k, i \not\longleftrightarrow j) - \mathbb{P}_\lambda(|\mathcal{C}(i)| = l)\mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k) \leq 0, \tag{4.3.25}$$

which in turn implies that

$$\mathrm{Var}_\lambda(Z_{\geq k}) \leq \sum_{i,j=1}^{n} \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, i \longleftrightarrow j). \tag{4.3.26}$$

Therefore, we arrive at the fact that, by the exchangeability of the vertices,

$$\begin{aligned}
\mathrm{Var}_\lambda(Z_{\geq k}) &\leq \sum_{i,j=1}^{n} \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, i \longleftrightarrow j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}_\lambda\left[\mathbb{1}_{\{|\mathcal{C}(i)| \geq k\}}\mathbb{1}_{\{j \in \mathcal{C}(i)\}}\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_\lambda\Big[\mathbb{1}_{\{|\mathcal{C}(i)| \geq k\}} \sum_{j=1}^{n} \mathbb{1}_{\{j \in \mathcal{C}(i)\}}\Big].
\end{aligned} \tag{4.3.27}$$

Since $\sum_{j=1}^{n} \mathbb{1}_{\{j \in \mathcal{C}(i)\}} = |\mathcal{C}(i)|$, we arrive at

$$\mathrm{Var}_\lambda(Z_{\geq k}) \leq \sum_i \mathbb{E}_\lambda[|\mathcal{C}(i)|\mathbb{1}_{\{|\mathcal{C}(i)| \geq k\}}] = n\mathbb{E}_\lambda[|\mathcal{C}(1)|\mathbb{1}_{\{|\mathcal{C}(1)| \geq k\}}] = n\chi_{\geq k}(\lambda). \tag{4.3.28}$$

□

Proof of Theorem 4.5. To prove Theorem 4.5, it suffices to prove that $\mathbb{P}_\lambda(Z_{\geq k_n} = 0) = O(n^{-\delta})$, where $k_n = a \log n$ with $a < I_\lambda^{-1}$. For this, we use the Chebychev inequality (Theorem 2.15). In order to apply Theorem 2.15, we need to derive a lower bound on $\mathbb{E}_\lambda[Z_{\geq k}]$ and an upper bound on $\mathrm{Var}_\lambda(Z_{\geq k})$.

We start by giving a lower bound on $\mathbb{E}_\lambda[Z_{\geq k}]$. We use that

$$\mathbb{E}_\lambda[Z_{\geq k}] = nP_{\geq k}(\lambda), \qquad \text{where} \qquad P_{\geq k}(\lambda) = \mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k). \tag{4.3.29}$$

We take $k = k_n = a \log n$. We use Theorem 4.3 to see that, with T a binomial branching process with parameters $n - k_n$ and $p = \lambda/n$,

$$P_{\geq k}(\lambda) \geq \mathbb{P}_{n-k_n,p}(T \geq a \log n). \tag{4.3.30}$$

By Theorem 3.20, with T^* the total progeny of a Poisson branching process with mean $\lambda_n = \lambda \frac{n-k_n}{n}$,

$$\mathbb{P}_{n-k_n,p}(T \geq a \log n) = \mathbb{P}^*_{\lambda_n}(T^* \geq a \log n) + O\big(\frac{a\lambda^2 \log n}{n}\big). \tag{4.3.31}$$

Also, by Theorem 3.16, we have that

$$\mathbb{P}^*_{\lambda_n}(T^* \geq a \log n) = \sum_{k=a \log n}^{\infty} \mathbb{P}^*_{\lambda_n}(T^* = k) = \sum_{k=a \log n}^{\infty} \frac{(\lambda_n k)^{k-1}}{k!}\mathrm{e}^{-\lambda_n k}. \tag{4.3.32}$$

By Stirling's formula,

$$k! = \Big(\frac{k}{e}\Big)^k \sqrt{2\pi k}\big(1+o(1)\big), \tag{4.3.33}$$

so that, recalling (4.3.1), and using that $I_{\lambda_n} = I_\lambda + o(1)$,

$$\mathbb{P}(T^* \geq a\log n) = \lambda^{-1} \sum_{k=a\log n}^{\infty} \frac{1}{\sqrt{2\pi k^3}} \mathrm{e}^{-I_{\lambda_n} k}(1+o(1)) = \mathrm{e}^{-I_\lambda a \log n(1+o(1))}. \tag{4.3.34}$$

As a result, it follows that, with $k_n = a\log n$ and any $0 < \alpha < 1 - I_\lambda a$,

$$\mathbb{E}_\lambda[Z_{\geq k_n}] = nP_{\geq k_n}(\lambda) \geq n^{(1-I_\lambda a)(1+o(1))} \geq n^\alpha. \tag{4.3.35}$$

We next bound the variance of $Z_{\geq k_n}$ using Proposition 4.7. By (4.3.10),

$$\begin{aligned}\chi_{\geq k_n}(\lambda) &= \sum_{t=k_n}^{n} P_{\geq t}(\lambda) \leq \sum_{t=k_n}^{n} \mathrm{e}^{-I_\lambda(t-1)} \\ &\leq \frac{\mathrm{e}^{-(k_n-1)I_\lambda}}{1-\mathrm{e}^{-I_\lambda}} = O(n^{-aI_\lambda}).\end{aligned} \tag{4.3.36}$$

We conclude that, by Proposition 4.7,

$$\mathrm{Var}_\lambda(Z_{\geq k_n}) \leq n\chi_{\geq k_n}(\lambda) \leq O(n^{1-aI_\lambda}), \tag{4.3.37}$$

while

$$\mathbb{E}_\lambda[Z_{\geq k_n}] \geq n^\alpha. \tag{4.3.38}$$

Therefore, by the Chebychev inequality (Theorem 2.14),

$$\mathbb{P}_\lambda(Z_{\geq k_n} = 0) \leq \frac{\mathrm{Var}_\lambda(Z_{\geq k_n})}{\mathbb{E}_\lambda[Z_{\geq k_n}]^2} \leq O(n^{1-aI-2\alpha}) = O(n^{-\delta}), \tag{4.3.39}$$

when we pick $\delta = 2\alpha - (1 - I_\lambda a)$, and $0 < \alpha < 1 - I_\lambda a$ such that $\delta = 2\alpha - (1 - I_\lambda a) > 0$. Finally, we use that

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| < k_n) = \mathbb{P}_\lambda(Z_{\geq k_n} = 0), \tag{4.3.40}$$

to complete the proof of Theorem 4.5. □

4.4 The supercritical regime

In this section, we fix $\lambda > 1$. The main result proved in this section is the following theorem. In its statement, we write $\zeta_\lambda = 1 - \eta_\lambda$ for the survival probability of a Poisson branching process with mean offspring λ.

Theorem 4.8 (Law of large numbers for giant component)**.** *Fix $\lambda > 1$. Then, for every $\nu \in (\frac{1}{2}, 1)$, there exists a $\delta = \delta(\nu, \lambda) > 0$ such that*

$$\mathbb{P}_\lambda\Big(\Big||\mathcal{C}_{\max}| - \zeta_\lambda n\Big| \geq n^\nu\Big) = O(n^{-\delta}). \tag{4.4.1}$$

Theorem 4.8 can be interpreted as follows. A vertex has a large connected component with probability ζ_λ. Therefore, there are of the order $\zeta_\lambda n$ vertices with large connected components. Theorem 4.8 implies that all these vertices in large components are in fact in the *same* connected component, which is called the *giant component*. We first give an overview of the proof of Theorem 4.8.

4.4.1 Strategy of proof of law of large numbers for the giant component

In this section, we give an overview of the proof of Theorem 4.8. We again crucially rely on an analysis of the number of vertices in connected components of size at least k,

$$Z_{\geq k} = \sum_{v=1}^{n} \mathbb{1}_{\{|\mathcal{C}(v)|\geq k\}}. \tag{4.4.2}$$

The proof contains 4 main steps. In the **first step**, we compute

$$\mathbb{E}[Z_{\geq k_n}] = n\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n). \tag{4.4.3}$$

We evaluate $\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n)$ using the bound in Theorem 4.3. Indeed, Proposition 4.9 below states that for $k_n = K\log n$ and K sufficiently large

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) = \zeta_\lambda(1+o(1)). \tag{4.4.4}$$

In the **second step**, we use a variance estimate on $Z_{\geq k}$ in Proposition 4.10, which implies that, **whp**and for all $\nu \in (\frac{1}{2}, 1)$,

$$|Z_{\geq k_n} - \mathbb{E}_\lambda[Z_{\geq k_n}]| \leq n^\nu. \tag{4.4.5}$$

In the **third step,** we show that, for $k = k_n = K\log n$, for some $K > 0$ sufficiently large, there is with high probability no connected component with size in between k_n and αn for any $\alpha < \zeta_\lambda$. This is done by a first moment argument: the expected number of vertices in such connected components is equal to $\mathbb{E}_\lambda[Z_{\geq k_n} - Z_{\geq \alpha n}]$, and we use the bound in Proposition 4.9 described above, as well as Proposition 4.12, which states that, for any $\alpha < \zeta_\lambda$, there exists $J > 0$ such that

$$\mathbb{P}_\lambda\big(k_n \leq |\mathcal{C}(v)| < \alpha n\big) \leq \mathrm{e}^{-k_n J}. \tag{4.4.6}$$

In the **fourth step**, we prove that for $2\alpha > \zeta_\lambda$, and on the event that there are no clusters with size in between k_n and αn, and on the event in (4.4.5),

$$Z_{\geq k_n} = |\mathcal{C}_{\max}|. \tag{4.4.7}$$

The proof of Theorem 4.8 follows by combining (4.4.3), (4.4.5) and (4.4.7). We now give the details of the proof of Theorem 4.8.

Step 1: The expected number of vertices in large components. In the first step, we show that the probability that $|\mathcal{C}(v)| \geq k$ is, for $k_n \geq a\log n$, close to the survival probability of a Poisson branching process with mean λ. Proposition 4.9 implies (4.4.4).

Proposition 4.9 (Cluster tail is branching process survival probability)**.** *Fix $\lambda > 1$ and let $n \to \infty$. Then, for $k_n \geq a\log n$ where $a > I_\lambda^{-1}$ and I_λ is defined in* (4.3.1)*,*

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) = \zeta_\lambda + O(k_n/n). \tag{4.4.8}$$

Proof. For the upper bound on $\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k)$, we first use Theorem 4.2, followed by Theorem 3.20, to deduce

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) \leq \mathbb{P}_{n,\lambda/n}(T \geq k_n) \leq \mathbb{P}^*_\lambda(T^* \geq k_n) + O(k_n/n), \tag{4.4.9}$$

where T and T^*, respectively, are the total progeny of a binomial branching process with parameters n and λ/n and a Poisson mean λ branching process, respectively. To complete the upper bound, we use Theorem 3.8 to see that

$$\begin{aligned}\mathbb{P}^*_\lambda(T^* \geq k_n) &= \mathbb{P}^*_\lambda(T^* = \infty) + \mathbb{P}^*_\lambda(k_n \leq T^* < \infty)\\ &= \zeta_\lambda + O(\mathrm{e}^{-k_n I_\lambda}) = \zeta_\lambda + O(k_n/n),\end{aligned} \tag{4.4.10}$$

as required.

For the lower bound, we use Theorem 4.3 again followed by Theorem 3.20, so that, with $\lambda_n = \lambda(1 - k_n/n)$,

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) \geq \mathbb{P}_{n-k_n,\lambda/n}(T \geq k_n) \geq \mathbb{P}^*_{\lambda_n}(T^* \geq k_n) + O(k_n/n), \tag{4.4.11}$$

where now T and T^*, respectively, are the the total progeny of a binomial branching process with parameters $n - k_n$ and λ/n and a Poisson mean λ_n branching process, respectively.

By Exercise 3.29 for $k_n \geq a \log n$ with $a > I_\lambda^{-1}$,

$$\mathbb{P}^*_{\lambda_n}(T^* \geq k_n) = \zeta_{\lambda_n} + O(\mathrm{e}^{-k_n I_{\lambda_n}}) = \zeta_{\lambda_n} + O(k_n/n). \tag{4.4.12}$$

Now, furthermore, by the mean-value theorem,

$$\eta_{\lambda_n} = \eta_\lambda + (\lambda_n - \lambda)\frac{d}{d\lambda}\eta_\lambda\big|_{\lambda=\lambda_n^*} = \eta_\lambda + O(k_n/n), \tag{4.4.13}$$

for some $\lambda_n^* \in (\lambda_n, \lambda)$, where we use Corollary 3.19 for $\lambda > 1$ and $\lambda_n - \lambda = k_n/n$. Therefore, also $\zeta_{\lambda_n} = \zeta_\lambda + O(k_n/n)$. Putting these estimates together proves the lower bound. Together, the upper and lower bound complete the proof of Proposition 4.9. □

Step 2: Concentration of the number of vertices in large clusters. The proof of Theorem 4.8 makes use of a variance estimate on $Z_{\geq k}$. In its statement, we use the notation

$$\chi_{<k}(\lambda) = \mathbb{E}_\lambda[|\mathcal{C}(v)|\mathbb{1}_{\{|\mathcal{C}(v)|<k\}}]. \tag{4.4.14}$$

Proposition 4.10 (A second variance estimate on $Z_{\geq k}$)**.** *For every n and $k \in [n]$,*

$$\mathrm{Var}_\lambda(Z_{\geq k}) \leq (\lambda k + 1)n\chi_{<k}(\lambda). \tag{4.4.15}$$

Note that the variance estimate in Proposition 4.10 is, in the supercritical regime, much better than the variance estimate in Proposition 4.7. Indeed, the bound in Proposition 4.7 reads

$$\mathrm{Var}_\lambda(Z_{\geq k}) \leq n\chi_{\geq k}(\lambda). \tag{4.4.16}$$

However, when $\lambda > 1$, according to Theorem 4.8 (which is currently not yet proved), $|\mathcal{C}(1)| = \Theta(n)$ with positive probability. Therefore,

$$n\chi_{\geq k}(\lambda) = \Theta(n^2), \tag{4.4.17}$$

which is a trivial bound. The bound in Proposition 4.10 is at most $\Theta(k^2 n)$, which is much smaller when k is not too large.

Proof. Define

$$Z_{<k} = \sum_{v=1}^n \mathbb{1}_{\{|\mathcal{C}(v)|<k\}}. \tag{4.4.18}$$

Then, since $Z_{<k} = n - Z_{\geq k}$, we have

$$\mathrm{Var}_\lambda(Z_{\geq k}) = \mathrm{Var}_\lambda(Z_{<k}). \tag{4.4.19}$$

Therefore, it suffices to prove that $\mathrm{Var}(Z_{<k}) \leq (\lambda k + 1)n\chi_{<k}(\lambda)$. For this, we compute

$$\mathrm{Var}_\lambda(Z_{<k}) = \sum_{i,j=1}^n \big[\mathbb{P}_\lambda(|\mathcal{C}(i)| < k, |\mathcal{C}(j)| < k) - \mathbb{P}_\lambda(|\mathcal{C}(i)| < k)\mathbb{P}_\lambda(|\mathcal{C}(j)| < k)\big]. \tag{4.4.20}$$

We again split, depending on whether $i \longleftrightarrow j$ or not:

$$\begin{aligned}\mathrm{Var}_\lambda(Z_{<k}) = \sum_{i,j=1}^{n} \big[\mathbb{P}_\lambda(|\mathcal{C}(i)|<k, |\mathcal{C}(j)|<k, i \longleftrightarrow\!\!\!\!\!\!/\;\; j) - \mathbb{P}_\lambda(|\mathcal{C}(i)|<k)\mathbb{P}_\lambda(|\mathcal{C}(j)|<k)\big] \\ + \sum_{i,j=1}^{n} \mathbb{P}_\lambda(|\mathcal{C}(i)|<k, |\mathcal{C}(j)|<k, i \longleftrightarrow j). \qquad (4.4.21)\end{aligned}$$

We compute explicitly, using that $|\mathcal{C}(i)| = |\mathcal{C}(j)|$ when $i \longleftrightarrow j$,

$$\begin{aligned}\sum_{i,j=1}^{n} \mathbb{P}_\lambda(|\mathcal{C}(i)|<k, |\mathcal{C}(j)|<k, i \longleftrightarrow j) &= \sum_{i,j=1}^{n} \mathbb{E}_\lambda\big[\mathbb{1}_{\{|\mathcal{C}(i)|<k\}}\mathbb{1}_{\{i\longleftrightarrow j\}}\big] \\ &= \sum_{i=1}^{n} \mathbb{E}_\lambda\big[\mathbb{1}_{\{|\mathcal{C}(i)|<k\}}\sum_{j=1}^{n}\mathbb{1}_{\{i\longleftrightarrow j\}}\big] \\ &= \sum_{i=1}^{n} \mathbb{E}_\lambda[|\mathcal{C}(i)|\mathbb{1}_{\{|\mathcal{C}(i)|<k\}}] = n\chi_{<k}(\lambda). \qquad (4.4.22)\end{aligned}$$

To compute the first sum on the right hand-side of (4.4.21) we write that, for $l < k$,

$$\begin{aligned}&\mathbb{P}_\lambda(|\mathcal{C}(i)|=l, |\mathcal{C}(j)|<k, i \longleftrightarrow\!\!\!\!\!\!/\;\; j) \\ &\quad = \mathbb{P}_\lambda(|\mathcal{C}(i)|=l)\mathbb{P}_\lambda\big(i \longleftrightarrow\!\!\!\!\!\!/\;\; j\big||\mathcal{C}(i)|=l\big)\mathbb{P}_\lambda\big(|\mathcal{C}(j)|<k\big||\mathcal{C}(i)|=l, i \longleftrightarrow\!\!\!\!\!\!/\;\; j\big). \qquad (4.4.23)\end{aligned}$$

See Exercise 4.16 below for an explicit formula for $\mathbb{P}_\lambda\big(i \longleftrightarrow\!\!\!\!\!\!/\;\; j\big||\mathcal{C}(i)|=l\big)$. We bound $\mathbb{P}_\lambda\big(i \longleftrightarrow\!\!\!\!\!\!/\;\; j\big||\mathcal{C}(i)|=l\big) \le 1$, to obtain

$$\mathbb{P}_\lambda(|\mathcal{C}(i)|=l, |\mathcal{C}(j)|<k, i \longleftrightarrow\!\!\!\!\!\!/\;\; j) \le \mathbb{P}_\lambda(|\mathcal{C}(i)|=l)\mathbb{P}_\lambda\big(|\mathcal{C}(j)|<k\big||\mathcal{C}(i)|=l, i \longleftrightarrow\!\!\!\!\!\!/\;\; j\big). \qquad (4.4.24)$$

Now we use that, when $|\mathcal{C}(i)| = l$ and when $i \longleftrightarrow\!\!\!\!\!\!/\;\; j$, the law of $|\mathcal{C}(j)|$ is identical to the law of $|\mathcal{C}(1)|$ in a random graph with $n-l$ vertices and edge probability $p = \lambda/n$, i.e.,

$$\mathbb{P}_{n,\lambda}(|\mathcal{C}(j)|<k\big||\mathcal{C}(i)|=l, i \longleftrightarrow\!\!\!\!\!\!/\;\; j) = \mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)|<k), \qquad (4.4.25)$$

where we write $\mathbb{P}_{m,\lambda}$ for the distribution of $\mathrm{ER}(m, \lambda/n)$. Therefore,

$$\begin{aligned}&\mathbb{P}_\lambda(|\mathcal{C}(j)|<k\big||\mathcal{C}(i)|=l, i \longleftrightarrow\!\!\!\!\!\!/\;\; j) \qquad (4.4.26)\\ &\quad = \mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)|<k) = \mathbb{P}_{n,\lambda}(|\mathcal{C}(1)|<k) + \mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)|<k) - \mathbb{P}_{n,\lambda}(|\mathcal{C}(1)|<k).\end{aligned}$$

We can couple $\mathrm{ER}(n-l,p)$ and $\mathrm{ER}_n(p)$ by adding the vertices $\{n-l+1,\dots,n\}$, and by letting st, for $s \in \{n-l+1,\dots,n\}$ and $t \in [n]$ be independently occupied with probability p. In this coupling, we note that $\mathbb{P}_{n-l,\lambda}(|\mathcal{C}(1)|<k) - \mathbb{P}_{n,\lambda}(|\mathcal{C}(1)|<k)$ is equal to the probability of the event that $|\mathcal{C}(1)|<k$ in $\mathrm{ER}(n-l,p)$, but $|\mathcal{C}(1)| \ge k$ in $\mathrm{ER}_n(p)$. If $|\mathcal{C}(1)|<k$ in $\mathrm{ER}(n-l,p)$, but $|\mathcal{C}(1)| \ge k$ in $\mathrm{ER}_n(p)$, it follows that at least one of the vertices $\{n-l+1,\dots,n\}$ must be connected to one of the at most k vertices in the connected component of vertex 1 in $\mathrm{ER}(n-l,p)$. This has probability at most lkp, so that, by Boole's inequality,

$$\mathbb{P}_\lambda(|\mathcal{C}(j)|<k, i \longleftrightarrow\!\!\!\!\!\!/\;\; j\big||\mathcal{C}(i)|=l) - \mathbb{P}_\lambda(|\mathcal{C}(j)|<k) \le lk\lambda/n. \qquad (4.4.27)$$

Therefore,

$$\sum_{i,j=1}^{n}\big[\mathbb{P}_\lambda(|\mathcal{C}(i)|<k,|\mathcal{C}(j)|<k,i\longleftrightarrow j)-\mathbb{P}_\lambda(|\mathcal{C}(i)|<k)\mathbb{P}_\lambda(|\mathcal{C}(j)|<k)\big]$$
$$\leq\sum_{l=1}^{k-1}\sum_{i,j}\frac{\lambda kl}{n}\mathbb{P}_\lambda(|\mathcal{C}(i)|=l)=\frac{\lambda k}{n}\sum_{i,j}\mathbb{E}_\lambda[|\mathcal{C}(i)|\mathbb{1}_{\{|\mathcal{C}(i)|<k\}}]=nk\lambda\chi_{<k}(\lambda),\quad(4.4.28)$$

which, together with (4.4.21)–(4.4.22), completes the proof. □

Exercise 4.16 (Connectivity with given expected cluster size)**.** *Show that*

$$\mathbb{P}_\lambda\big(1\longleftrightarrow 2\big||\mathcal{C}(1)|=l\big)=1-\frac{l-1}{n-1}.\quad(4.4.29)$$

Corollary 4.11 (Concentration of the number of vertices in large components)**.** *Fix $k_n=K\log n$. Then, for K sufficiently large,*

$$\mathbb{P}_\lambda(|Z_{\geq k_n}-n\zeta_\lambda|>n^\nu)=O(n^{-\delta}).\quad(4.4.30)$$

Proof. We use Proposition 4.9 to note that

$$\mathbb{E}_\lambda[Z_{\geq k_n}]=n\mathbb{P}_\lambda(|\mathcal{C}(v)|\geq k_n)=n\zeta_\lambda+O(k_n),\quad(4.4.31)$$

and therefore, for n sufficiently large and since $k_n=o(n^\nu)$,

$$\{|Z_{\geq k_n}-\mathbb{E}_\lambda[Z_{\geq k_n}]|\leq n^\nu/2\}\subseteq\{|Z_{\geq k_n}-n\zeta_\lambda|\leq n^\nu\}.\quad(4.4.32)$$

By the Chebychev inequality (Theorem 2.15), and using Proposition 4.10 as well as $\chi_{<k_n}(\lambda)\leq k_n$, we then obtain that

$$\mathbb{P}_\lambda(|Z_{\geq k_n}-n\zeta_\lambda|\leq n^\nu)\geq\mathbb{P}_\lambda(|Z_{\geq k_n}-\mathbb{E}_\lambda[Z_{\geq k_n}]|\leq n^\nu/2)\geq 1-4n^{-2\nu}\mathrm{Var}(Z_{\geq k_n})$$
$$\geq 1-4n^{1-2\nu}\big(\lambda k_n^2+k_n\big)\geq 1-n^{-\delta},\quad(4.4.33)$$

for any $\delta<2\nu-1$ and n sufficiently large, since $k_n=K\log n$. □

Step 3: No middle ground. We next show that the probability that $k_n\leq|\mathcal{C}(v)|\leq\alpha n$ is exponentially small in k_n:

Proposition 4.12 (Exponential bound for supercritical clusters smaller than $\zeta_\lambda n$)**.** *Fix $\lambda>1$ and let $k_n\to\infty$. Then, for any $\alpha<\zeta_\lambda$, and with*

$$J(\alpha,\lambda)=I_{g(\alpha;\lambda)}>0,\qquad\text{with}\qquad g(\alpha;\lambda)=(1-\mathrm{e}^{-\lambda\alpha})/\alpha,\quad(4.4.34)$$

such that

$$\mathbb{P}_\lambda(k_n\leq|\mathcal{C}(v)|\leq\alpha n)\leq\mathrm{e}^{-k_nJ(\alpha,\lambda)}/[1-\mathrm{e}^{-J(\alpha,\lambda)}].\quad(4.4.35)$$

Proof. We start by bounding

$$\mathbb{P}_\lambda(k_n\leq|\mathcal{C}(v)|\leq\alpha n)=\sum_{t=k_n}^{\alpha n}\mathbb{P}_\lambda(|\mathcal{C}(v)|=t)\leq\sum_{t=k_n}^{\alpha n}\mathbb{P}_\lambda(S_t=0),\quad(4.4.36)$$

where we recall (4.1.3). By Proposition 4.6, we have that $S_t\sim\mathrm{Bin}(n-1,1-(1-p)^t)+1-t$. Therefore, with $p=\lambda/n$,

$$\mathbb{P}_\lambda(S_t=0)=\mathbb{P}_\lambda\Big(\mathrm{Bin}\big(n-1,1-(1-p)^t\big)=t-1\Big).\quad(4.4.37)$$

To explain the exponential decay, we note that, for $p = \lambda/n$ and $t = \alpha n$,

$$1 - (1-p)^t = 1 - \left(1 - \frac{\lambda}{n}\right)^{\alpha n} = (1 - \mathrm{e}^{-\lambda\alpha})(1 + o(1)) \qquad \text{for large } n. \tag{4.4.38}$$

The unique solution to the equation $1 - \mathrm{e}^{-\lambda\alpha} = \alpha$ is $\alpha = \zeta_\lambda$:

Exercise 4.17 (Uniqueness solution of Poisson survival probability equation)**.** *Prove that the unique solution to the equation $1 - \mathrm{e}^{-\lambda\alpha} = \alpha$ is $\alpha = \zeta_\lambda$, where ζ_λ is the survival probability of a Poisson branching process with parameter λ.*

If $\alpha < \zeta_\lambda$, then $\alpha < 1 - \mathrm{e}^{-\lambda\alpha}$, and thus the probability in (4.4.37) drops exponentially. We now fill in the details. First, by (4.4.37) and using that $1-p \le \mathrm{e}^{-p}$, so that $1-(1-p)^t \ge 1 - \mathrm{e}^{-pt}$,

$$\begin{aligned}\mathbb{P}_\lambda(S_t = 0) &= \mathbb{P}_\lambda\big(\mathrm{Bin}(n-1, 1-(1-p)^t) = t-1\big) \le \mathbb{P}_\lambda\big(\mathrm{Bin}(n-1, 1-(1-p)^t) \le t-1\big) \\ &\le \mathbb{P}_\lambda\big(\mathrm{Bin}(n, 1-(1-p)^t) \le t\big) \le \mathbb{P}_\lambda\big(\mathrm{Bin}(n, 1-\mathrm{e}^{-pt}) \le t\big).\end{aligned} \tag{4.4.39}$$

We bound, for each $s \ge 0$, and using that $1 - x \le \mathrm{e}^{-x}$,

$$\begin{aligned}\mathbb{P}_\lambda(S_t = 0) &\le \mathrm{e}^{st}\mathbb{E}\Big[\mathrm{e}^{s\mathrm{Bin}\left(n, 1-\mathrm{e}^{-pt}\right)}\Big] = \mathrm{e}^{st}\Big((1-\mathrm{e}^{-pt})\mathrm{e}^{-s} + \mathrm{e}^{-pt}\Big)^n \\ &= \mathrm{e}^{st}\Big(1 + (1-\mathrm{e}^{-pt})(\mathrm{e}^{-s} - 1)\Big)^n \le \mathrm{e}^{st + n(1-\mathrm{e}^{-pt})(1-\mathrm{e}^{-s})} \\ &= \mathrm{e}^{st + n(1-\mathrm{e}^{-\lambda t/n})(1-\mathrm{e}^{-s})}.\end{aligned} \tag{4.4.40}$$

The minimizer of $s \mapsto st + n(1-\mathrm{e}^{-pt})(1-\mathrm{e}^{-s})$ over s is equal to

$$s^* = \log\Big(n(1-\mathrm{e}^{-\lambda t/n})/t\Big), \tag{4.4.41}$$

Write $t = \beta n$ and $g(\beta;\lambda) = (1-\mathrm{e}^{-\lambda\beta})/\beta$. Note that $\lim_{\beta\downarrow 0} g(\beta;\lambda) = \lambda > 1$, and, by Exercise 4.17, $g(\zeta_\lambda;\lambda) = 1$. Further, $\beta \mapsto g(\beta;\lambda)$ is decreasing, since

$$\frac{\partial}{\partial\beta} g(\beta;\lambda) = \frac{g(\beta;\lambda)}{\beta}\big[(\beta\lambda)\mathrm{e}^{-\beta\lambda} - 1\big] < 0. \tag{4.4.42}$$

As a result, $s^* \ge 0$ precisely when $\beta < \zeta_\lambda$.

Substitution of $s^* = \log\Big(n(1-\mathrm{e}^{-\lambda t/n})/t\Big)$ yields

$$\mathbb{P}_\lambda(S_t = 0) \le \mathrm{e}^{-t\left(\log g(t/n;\lambda) - 1 - g(t/n;\lambda)\right)} = \mathrm{e}^{-tI_{g(t/n;\lambda)}}. \tag{4.4.43}$$

Since $\lambda \mapsto I_\lambda$ is increasing and $t/n \le \alpha < \zeta_\lambda$, it follows that

$$\mathbb{P}_\lambda(S_t = 0) \le \mathrm{e}^{-tI_{g(\alpha;\lambda)}} = \mathrm{e}^{-tJ(\alpha;\lambda)}. \tag{4.4.44}$$

We conclude that

$$\mathbb{P}_\lambda(k_n \le |\mathcal{C}(v)| \le \alpha n) \le \sum_{t=k_n}^{\alpha n} \mathbb{P}_\lambda(S_t = 0) \le \sum_{t=k_n}^{\alpha n} \mathrm{e}^{-tJ(\alpha;\lambda)} \le \mathrm{e}^{-k_n J(\alpha;\lambda)}/[1 - \mathrm{e}^{-J(\alpha;\lambda)}]. \tag{4.4.45}$$

This completes the proof of Proposition 4.12. □

We finally state a consequence of Proposition 4.12 that shows that there is, with high probability, no cluster with intermediate size, i.e., size in between $k_n = K\log n$ and αn. Corollary 4.13 implies (4.4.6):

Corollary 4.13 (No intermediate clusters)**.** *Fix $k_n = K \log n$ and $\alpha < \zeta_\lambda$. Then, for K sufficiently large, and with probability at least $1 - n^{-\delta}$, there is no connected component with size in between k_n and αn.*

Proof. We use that the expected number of clusters with sizes in between k_n and αn, for any $\alpha < \zeta_\lambda$, is equal to

$$\mathbb{E}_\lambda[Z_{\geq k_n} - Z_{\geq \alpha n+1}] = n\mathbb{P}_\lambda(k_n \leq |\mathcal{C}(v)| \leq \alpha n) \leq Cn\mathrm{e}^{-k_n J}, \tag{4.4.46}$$

where we have used Proposition 4.12 for the last estimate. When $k_n = K \log n$, and K is sufficiently large, the right-hand side is $O(n^{-\delta})$. By the Markov inequality (Theorem 2.14),

$$\begin{aligned}\mathbb{P}_\lambda(\exists v : k_n \leq |\mathcal{C}(v)| \leq \alpha n) &= \mathbb{P}_\lambda(Z_{\geq k_n} - Z_{\geq \alpha n+1} \geq 1) \\ &\leq \mathbb{E}_\lambda[Z_{\geq k_n} - Z_{\geq \alpha n+1}] = O(n^{-\delta}).\end{aligned} \tag{4.4.47}$$

This completes the proof of Corollary 4.13. □

Exercise 4.18 (Connectivity and expected cluster size)**.** *Prove that the expected cluster size of a given vertex*

$$\chi(\lambda) = \mathbb{E}_\lambda[|\mathcal{C}(1)|], \tag{4.4.48}$$

satisfies

$$\chi(\lambda) = 1 + (n-1)\mathbb{P}_\lambda(1 \longleftrightarrow 2). \tag{4.4.49}$$

Exercise 4.19 (Connectivity function)**.** *Prove that* (4.4.1) *and Corollary 4.13 imply that, for $\lambda > 1$,*

$$\mathbb{P}_\lambda(1 \longleftrightarrow 2) = \zeta_\lambda^2[1 + o(1)]. \tag{4.4.50}$$

Exercise 4.20 (Supercritical expected cluster size)**.** *Prove that* (4.4.1) *implies that the expected cluster size satisfies, for $\lambda > 1$,*

$$\chi(\lambda) = \zeta_\lambda^2 n(1 + o(1)). \tag{4.4.51}$$

Step 4: proof of law of large numbers of giant component in Theorem 4.8. We fix $\nu \in (\frac{1}{2}, 1)$, $\alpha \in (\zeta_\lambda/2, \zeta_\lambda)$ and take $k_n = K \log n$ with K sufficiently large. Let $\mathcal{E}_n$ be the event that

(1) $|Z_{\geq k_n} - n\zeta_\lambda| \leq n^\nu$;

(2) there does not exist a $v \in [n]$ such that $k_n \leq |\mathcal{C}(v)| \leq \alpha n$.

In the proof of Theorem 4.8 we use the following lemma:

Lemma 4.14 ($|\mathcal{C}_{\max}|$ equals $Z_{\geq k_n}$ with high probability)**.** *The event $\mathcal{E}_n$ occurs with high probability, i.e., $\mathbb{P}_\lambda(\mathcal{E}_n^c) = O(n^{-\delta})$, and $|\mathcal{C}_{\max}| = Z_{\geq k_n}$ on the event $\mathcal{E}_n$.*

Proof. We start by proving that $\mathcal{E}_n$ occurs with high probability. For this, we note that $\mathcal{E}_n^c$ equals the union of complements of the events in (1) and (2) above, and we shall bound these complements one by one. By Corollary 4.11, $\mathbb{P}_\lambda(|Z_{\geq k_n} - n\zeta_\lambda| > n^\nu) = O(n^{-\delta})$. By Corollary 4.13, $\mathbb{P}_\lambda(\exists v \in [n] \text{ such that } k_n \leq |\mathcal{C}(v)| \leq \alpha n) \leq n^{-\delta}$. Together, these estimates imply that $\mathbb{P}_\lambda(\mathcal{E}_n^c) = O(n^{-\delta})$.

To prove that $|\mathcal{C}_{\max}| = Z_{\geq k_n}$, we start by noting that $\{|Z_{\geq k_n} - \zeta_\lambda n| \leq n^\nu\} \subseteq \{Z_{\geq k_n} \geq 1\}$. Thus, $|\mathcal{C}_{\max}| \leq Z_{\geq k_n}$ when the event $\mathcal{E}_n$ holds. In turn, $|\mathcal{C}_{\max}| < Z_{\geq k_n}$ implies that there are two connected components with size at least k_n. Furthermore, since $\mathcal{E}_n$ occurs, there are no connected components with sizes in between k_n and αn. Therefore, there must

be two connected components with size at least αn, which in turn implies that $Z_{\geq k_n} \geq 2\alpha n$. When $2\alpha > \zeta_\lambda$ and n is sufficiently large, this is in contradiction with $Z_{\geq k_n} \leq \zeta_\lambda n + n^\nu$. We conclude that $|\mathcal{C}_{\max}| = Z_{\geq k_n}$. □

Proof of Theorem 4.8. By the fact that $|\mathcal{C}_{\max}| = Z_{\geq k_n}$, we have

$$\mathbb{P}_\lambda\big(\big||\mathcal{C}_{\max}| - \zeta_\lambda n\big| \leq n^\nu\big) \geq \mathbb{P}_\lambda\big(\{\big||\mathcal{C}_{\max}| - \zeta_\lambda n\big| \leq n^\nu\} \cap \mathcal{E}_n\big) = \mathbb{P}_\lambda(\mathcal{E}_n) \geq 1 - O(n^{-\delta}), \tag{4.4.52}$$

since, by Lemma 4.14 and on the event $\mathcal{E}_n$, $|\mathcal{C}_{\max}| = Z_{\geq k_n}$ and $|Z_{\geq k_n} - n\zeta_\lambda| \leq n^\nu$. This completes the proof of the law of large number of the giant component in Theorem 4.8. □

4.4.2 The discrete duality principle

Using the results we can construct a duality principle for Erdős-Rényi random graphs similar to the duality principle for branching processes:

Theorem 4.15 (Discrete duality principle)**.** *Let $\mu_\lambda < 1 < \lambda$ be conjugates as in* (3.6.7)*. Conditionally, the graph $\mathrm{ER}_n(\lambda/n)$ with the giant component removed is close in law to the random graph $\mathrm{ER}(m, \frac{\mu_\lambda}{m})$, where the variable $m = \lceil n\eta_\lambda \rceil$ is the asymptotic number of vertices outside the giant component.*

We will see that the proof follows from Theorem 4.8, since this implies that the giant component has size $n - m = \zeta_\lambda n(1 + o(1))$. In the statement of Theorem 4.15 we make use of the informal notion 'close in law'. This notion can be made precise as follows. Let $\mathrm{ER}_n(\lambda/n)'$ be $\mathrm{ER}_n(\lambda/n)$ with the giant component removed. We write $\mathbb{P}'_\lambda$ for the law of $\mathrm{ER}_n(\lambda/n)'$, and we recall that $\mathbb{P}_{m,\mu}$ denotes the law of $\mathrm{ER}(m, \mu)$. Let $\mathcal{E}$ be an event which is determined by the edges variables. Then, if $\lim_{m\to\infty} \mathbb{P}_{m,\mu_\lambda}(\mathcal{E})$ exists, then

$$\lim_{n\to\infty} \mathbb{P}'_{n,\lambda}(\mathcal{E}) = \lim_{m\to\infty} \mathbb{P}_{m,\mu_\lambda}(\mathcal{E}). \tag{4.4.53}$$

We shall sketch a proof of Theorem 4.15. First of all, all the edges in the complement of the giant component in $\mathrm{ER}_n(p)$ are independent. Furthermore, the conditional probability that an edge st is occupied in $\mathrm{ER}_n(p)$ with the giant component removed is, conditionally on $|\mathcal{C}_{\max}| = n - m$, equal to

$$\frac{\lambda}{n} = \frac{\lambda}{m}\frac{m}{n}. \tag{4.4.54}$$

Now, $m \approx \eta_\lambda n$, so that the conditional probability that an edge st is occupied in $\mathrm{ER}_n(p)$ with the giant component removed, conditionally on $|\mathcal{C}_{\max}| \approx \zeta_\lambda n$, is equal to

$$\frac{\lambda}{n} \approx \frac{\lambda\eta_\lambda}{m} = \frac{\mu_\lambda}{m}, \tag{4.4.55}$$

where we have used (3.6.2) and (3.6.5), which implies that $\lambda\eta_\lambda = \mu_\lambda$. Therefore, the conditional probability that an edge st is occupied in $\mathrm{ER}_n(p)$ with the giant component removed, conditionally on $|\mathcal{C}_{\max}| \approx \zeta_\lambda n$, is equal to $\frac{\mu_\lambda}{m}$.

Exercise 4.21 (Second largest supercritical cluster)**.** *Use the duality principle to show that the second largest component of a supercritical Erdős-Rényi random graph $\mathcal{C}_{(2)}$ satisfies*

$$\frac{|\mathcal{C}_{(2)}|}{\log n} \xrightarrow{\mathbb{P}} I^{-1}_{\mu_\lambda}. \tag{4.4.56}$$

4.5 The CLT for the giant component

In this section, we prove a central limit theorem for the giant component in the supercritical regime, extending the law of large numbers for the giant component in Theorem 4.8. The main result is as follows:

Theorem 4.16 (Central limit theorem for giant component)**.** *Fix $\lambda > 1$. Then,*

$$\frac{|\mathcal{C}_{\max}| - \zeta_\lambda n}{\sqrt{n}} \xrightarrow{d} Z, \tag{4.5.1}$$

where Z is a normal random variable with mean 0 and variance $\sigma_\lambda^2 = \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda+\lambda\zeta_\lambda)^2}$.

We shall make use of the exploration of connected components to prove Theorem 4.16. In the proof, we shall make essential use of Theorem 4.8.

In order to present the proof, we start with some introductions. Fix $k = k_n$, which will be chosen later on. We shall explore the union of the connected components of the vertices $[k] = \{1, \ldots, k\}$. When $k \to \infty$ and using Theorem 4.8, this union contains the largest connected component $\mathcal{C}_{\max}$, and it cannot be larger than $|\mathcal{C}_{\max}| + kb_n$, where $b_n \leq K \log n$ is an upper bound on the second largest component. Taking $k = o(n^\nu)$ with $\nu < \frac{1}{2}$, this union of components is equal to $|\mathcal{C}_{\max}| + o(\sqrt{n})$. As a result, a central limit theorem for the union of components implies one for $|\mathcal{C}_{\max}|$. We now describe the size of the union of the components of $\{1, \ldots, k\}$.

Let $S_0 = k$ and, for $t \geq 1$, let

$$S_t = S_{t-1} + X_t - 1, \tag{4.5.2}$$

where

$$X_t \sim \mathrm{Bin}\Big(n - S_{t-1} - (t-1), p\Big). \tag{4.5.3}$$

Equations (4.5.2) and (4.5.3) are similar to the ones in (4.1.3) and (4.1.4). We next derive the distribution of S_t in a similar way as in Proposition 4.6:

Proposition 4.17 (The law of S_t revisited)**.** *For all $t \in [n]$,*

$$S_t + (t-1) \sim \mathrm{Bin}(n-k, 1-(1-p)^{t-1}). \tag{4.5.4}$$

Moreover, for all $l, t \in [n]$ satisfying $l \geq t$, and conditionally on S_t,

$$S_l + (l-t) - S_t \sim \mathrm{Bin}\big(n - (t-1) - S_t, 1-(1-p)^{l-t}\big). \tag{4.5.5}$$

For $k = 1$, the equality in distribution (4.5.4) in Proposition 4.17 reduces to Proposition 4.6.

Proof. For $t \geq 1$, let N_t represent the number of unexplored vertices, i.e.,

$$N_t = n - (t-1) - S_t. \tag{4.5.6}$$

It is more convenient to show the equivalent statement that for all $t \geq 1$

$$N_t \sim \mathrm{Bin}\big(n-k, (1-p)^{t-1}\big). \tag{4.5.7}$$

To see this, we note that each of the vertices $\{k+1, \ldots, n\}$ has, independently of all other vertices, probability $(1-p)^{t-1}$ to stay neutral in the first t explorations. More formally,

conditionally on S_{t-1}, and by (4.5.3), we have that $X_t \sim \mathrm{Bin}\big(n - S_{t-1} - (t-1), p\big) = \mathrm{Bin}\big(N_{t-1}, p\big)$ by (4.5.3). Thus, noting that $N_1 \sim \mathrm{Bin}(n-k, (1-p)^k)$ and

$$\begin{aligned} N_t &= n - (t+k-1) - S_t = n - (t+k-1) - S_{t-1} - X_t + 1 \\ &= n - (t+k-2) - S_{t-1} - \mathrm{Bin}(N_{t-1}, p) \\ &= N_{t-1} - \mathrm{Bin}(N_{t-1}, p) = \mathrm{Bin}(N_{t-1}, 1-p), \end{aligned} \tag{4.5.8}$$

the conclusion follows by recursion on $t \geq 1$ and Exercise 4.15. We note that (4.5.8) also implies that for any $l \geq t$,

$$N_l \sim \mathrm{Bin}(N_t, (1-p)^{l-t}). \tag{4.5.9}$$

Substituting $N_t = n - (t-1) - S_t$, this implies that

$$\begin{aligned} n - (l-1) - S_l &\sim \mathrm{Bin}\big(n - (t-1) - S_t, (1-p)^{l-t}\big) \\ &= n - (t-1) - S_t - \mathrm{Bin}\big(n - (t-1) - S_t, 1 - (1-p)^{l-t}\big), \end{aligned} \tag{4.5.10}$$

which, in turn, is equivalent to the statement that, for all $l \geq t$ and, conditionally on S_t,

$$S_l + (l-t) - S_t \sim \mathrm{Bin}\big(n - (t-1) - S_t, 1 - (1-p)^{l-t}\big). \tag{4.5.11}$$

□

We now state a corollary of Proposition 4.17 which states that $S_{\lfloor nt \rfloor}$ satisfies a central limit theorem. In its statement, we make use of the asymptotic mean

$$\mu_t = 1 - t - \mathrm{e}^{-\lambda t} \tag{4.5.12}$$

and asymptotic variance

$$v_t = \mathrm{e}^{-\lambda t}(1 - \mathrm{e}^{-\lambda t}). \tag{4.5.13}$$

The central limit theorem for $S_{\lfloor nt \rfloor}$ reads as follows:

Corollary 4.18 (CLT for $S_{\lfloor nt \rfloor}$)**.** *Fix $k = k_n = o(\sqrt{n})$. Then, for every $t > 0$, the random variable $\frac{S_{\lfloor nt \rfloor} - n\mu_t}{\sqrt{nv_t}}$ converges in distribution to a standard normal random variable.*

Proof. The statement follows immediately from the central limit theorem for the binomial distribution when we can show that

$$\mathbb{E}[S_{\lfloor nt \rfloor}] = n\mu_t + o(\sqrt{n}), \qquad \mathrm{Var}(S_{\lfloor nt \rfloor}) = nv_t + o(n). \tag{4.5.14}$$

Indeed, by the central limit theorem for the binomial distribution we have that

$$\frac{S_{\lfloor nt \rfloor} - \mathbb{E}[S_{\lfloor nt \rfloor}]}{\sqrt{\mathrm{Var}(S_{\lfloor nt \rfloor})}} \xrightarrow{d} Z, \tag{4.5.15}$$

where Z is a standard normal random variable.

Exercise 4.22 (CLT for binomials with general parameters)**.** *Prove that if $X_n = \mathrm{Bin}(a_n, p_n)$, where $\mathrm{Var}(X) = a_n p_n (1-p_n) \to \infty$, then*

$$\frac{X_n - a_n p_n}{\sqrt{a_n p_n (1-p_n)}} \xrightarrow{d} Z, \tag{4.5.16}$$

where Z is a standard normal random variable. Use this to conclude that (4.5.14) *implies* (4.5.15).

Write

$$\frac{S_{\lfloor nt\rfloor}-n\mu_t}{\sqrt{nv_t}}=\sqrt{\frac{\mathrm{Var}(S_{\lfloor nt\rfloor})}{nv_t}}\,\frac{S_{\lfloor nt\rfloor}-\mathbb{E}[S_{\lfloor nt\rfloor}]}{\sqrt{\mathrm{Var}(S_{\lfloor nt\rfloor})}}+\frac{\mathbb{E}[S_{\lfloor nt\rfloor}]-n\mu_t}{\sqrt{\mathrm{Var}(S_{\lfloor nt\rfloor})}}. \tag{4.5.17}$$

By (4.5.14), we have that the last term converges to zero, and the factor $\sqrt{\frac{\mathrm{Var}(S_{\lfloor nt\rfloor})}{nv_t}}$ converges to one. Therefore, (4.5.14) implies the central limit theorem.

To see the asymptotics of the mean in (4.5.14), we note that

$$\mathbb{E}[S_{\lfloor nt\rfloor}]=(n-k)\Big(1-(1-\frac{\lambda}{n})^{\lfloor nt\rfloor+k-1}\Big)-\big(\lfloor nt\rfloor-1\big)=n\mu_t+o(\sqrt{n}), \tag{4.5.18}$$

as long as $k=o(\sqrt{n})$. For the asymptotics of the variance in (4.5.14), we note that

$$\mathrm{Var}(S_{\lfloor nt\rfloor})=(n-k)(1-\frac{\lambda}{n})^{\lfloor nt\rfloor+k-1}\big(1-(1-\frac{\lambda}{n})^{\lfloor nt\rfloor+k-1}\big)=nv_t+o(n), \tag{4.5.19}$$

as long as $k=o(n)$. □

Proof of Theorem 4.16. Let $|\mathcal{C}([k])|$ be the size of the union of the components of the vertices $1,\dots,k$. Then,

$$|\mathcal{C}([k])|\sim\min\{m\colon S_m=0\}. \tag{4.5.20}$$

Let $k=k_n\to\infty$. We will prove below that $|\mathcal{C}([k])|$ satisfies a CLT with asymptotic mean $n\zeta_\lambda$ and asymptotic variance $n\sigma_\lambda^2$. By Corollary 4.13, for $2\alpha>\zeta_\lambda$ and Theorem 4.8, **whp**, the second largest cluster has size at most $K\log n$. Hence, **whp**,

$$|\mathcal{C}_{\max}|\le|\mathcal{C}([k])|\le|\mathcal{C}_{\max}|+(k-1)K\log n. \tag{4.5.21}$$

We conclude that a central limit theorem for $|\mathcal{C}_{\max}|$ follows from that for $|\mathcal{C}([k])|$ for any $k=k_n\to\infty$.

The central limit theorem for $|\mathcal{C}([k])|$ is proved by upper and upper bounds on the probabilities

$$\mathbb{P}_\lambda\Big(\frac{|\mathcal{C}([k])|-\zeta_\lambda n}{\sqrt{n}}>x\Big).$$

For the upper bound, we use that (4.5.20) implies that, for every ℓ,

$$\mathbb{P}_\lambda(|\mathcal{C}([k])|>\ell)=\mathbb{P}_\lambda(\forall m\le\ell\colon S_m>0). \tag{4.5.22}$$

Applying (4.5.22) to $\ell=m_x=\lfloor n\zeta_\lambda+x\sqrt{n}\rfloor$, we obtain

$$\mathbb{P}_\lambda\Big(\frac{|\mathcal{C}([k])|-\zeta_\lambda n}{\sqrt{n}}>x\Big)=\mathbb{P}_\lambda(\forall m\le m_x\colon S_m>0)\le\mathbb{P}_\lambda(S_{m_x}>0). \tag{4.5.23}$$

Now we use (4.5.12), (4.5.14) and $\mu_{\zeta_\lambda}=0$, and writing μ_t' for the derivative of $t\mapsto\mu_t$, to see that

$$\mathbb{E}[S_{m_x}]=n\mu_{\zeta_\lambda}+\sqrt{n}x\mu_{\zeta_\lambda}'+o(\sqrt{n})=\sqrt{n}x(\lambda\mathrm{e}^{-\lambda\zeta_\lambda}-1)+o(\sqrt{n})=\sqrt{n}x(\lambda\mathrm{e}^{-\lambda\zeta_\lambda}-1)+o(\sqrt{n}), \tag{4.5.24}$$

where we note that $\lambda\mathrm{e}^{-\lambda\zeta_\lambda}-1<0$ for $\lambda>1$.

Exercise 4.23 (Asymptotic mean and variance at $t=\zeta_\lambda$)**.** *Prove that* $\mu_{\zeta_\lambda}=0$ *and* $\mu_{\zeta_\lambda}'=\lambda\mathrm{e}^{-\lambda\zeta_\lambda}-1<0$ *for* $\lambda>1$.

The variance of S_{m_x} is, by (4.5.13) and (4.5.14),

$$\mathrm{Var}(S_{m_x}) = nv_{\zeta_\lambda} + o(n), \tag{4.5.25}$$

where $v_{\zeta_\lambda} > 0$. Thus,

$$\mathbb{P}_\lambda(S_{m_x} > 0) = \mathbb{P}_\lambda\Big(\frac{S_{m_x} - \mathbb{E}[S_{m_x}]}{\sqrt{\mathrm{Var}(S_{m_x})}} > \frac{x(1-\lambda \mathrm{e}^{-\lambda\zeta_\lambda})}{\sqrt{v_{\zeta_\lambda}}}\Big) + o(1). \tag{4.5.26}$$

By Corollary 4.18, the right-hand side converges to

$$\mathbb{P}\Big(Z > \frac{x(1-\lambda \mathrm{e}^{-\lambda\zeta_\lambda})}{\sqrt{v_{\zeta_\lambda}}}\Big) = \mathbb{P}(Z' > x), \tag{4.5.27}$$

where Z' has a normal distribution with mean 0 and variance $v_{\zeta_\lambda}(1-\lambda \mathrm{e}^{-\lambda\zeta_\lambda})^{-2}$. We finally note that, by (3.6.2) and $\zeta_\lambda = 1-\eta_\lambda$, we have that $1-\zeta_\lambda = \mathrm{e}^{-\lambda\zeta_\lambda}$, so that

$$v_{\zeta_\lambda} = \mathrm{e}^{-\lambda\zeta_\lambda}(1-\mathrm{e}^{-\lambda\zeta_\lambda}) = \zeta_\lambda(1-\zeta_\lambda). \tag{4.5.28}$$

By (4.5.28), the variance of the normal distribution appearing in the lower bound can be rewritten as

$$\frac{v_{\zeta_\lambda}}{(1-\lambda \mathrm{e}^{-\lambda\zeta_\lambda})^2} = \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda+\lambda\zeta_\lambda)^2}. \tag{4.5.29}$$

By (4.5.23), this completes the upper bound.

For the lower bound, we again use the fact that

$$\mathbb{P}_\lambda\Big(|\mathcal{C}([k])| - \zeta_\lambda n > x\Big) = \mathbb{P}_\lambda(\forall m \le m_x \colon S_m > 0), \tag{4.5.30}$$

where we recall that $m_x = \lfloor n\zeta_\lambda + x\sqrt{n}\rfloor$. Then, for any $\varepsilon > 0$, we bound from below

$$\begin{aligned}\mathbb{P}_\lambda(\forall m \le m_x \colon S_m > 0) &\ge \mathbb{P}_\lambda(\forall m < m_x \colon S_m > 0, S_{m_x} > \varepsilon\sqrt{n}) \\ &= \mathbb{P}_\lambda(S_{m_x} > \varepsilon\sqrt{n}) - \mathbb{P}_\lambda(S_{m_x} > \varepsilon\sqrt{n}, \exists m < m_x \colon S_m = 0).\end{aligned} \tag{4.5.31}$$

The first term can be handled in a similar way as for the upper bound. Indeed, repeating the steps in the upper bound, we obtain that, for every $\varepsilon > 0$,

$$\mathbb{P}_\lambda(S_{m_x} > \varepsilon\sqrt{n}) = \mathbb{P}\Big(Z > \frac{x(1-\lambda \mathrm{e}^{-\lambda\zeta_\lambda}) + \varepsilon}{\sqrt{v_{\zeta_\lambda}}}\Big) + o(1). \tag{4.5.32}$$

The quantity in (4.5.32) converges to $\mathbb{P}(Z' > x)$, where Z' has a normal distribution with mean 0 and variance σ_λ^2, as $\varepsilon \downarrow 0$.

We conclude that it suffices to prove that

$$\mathbb{P}_\lambda(S_{m_x} > \varepsilon\sqrt{n}, \exists m < m_x \colon S_m = 0) = o(1). \tag{4.5.33}$$

To bound the probability in (4.5.33), we first use Boole's inequality to get

$$\mathbb{P}_\lambda(S_{m_x} > \varepsilon\sqrt{n}, \exists m < m_x \colon S_m = 0) \le \sum_{m=1}^{m_x-1} \mathbb{P}_\lambda(S_m = 0, S_{m_x} > \varepsilon\sqrt{n}). \tag{4.5.34}$$

For $m \le \alpha n$ with $\alpha < \zeta_\lambda$, we can show that, when $k = K\log n$ and K sufficiently large, and uniformly in $m \le \alpha n$,

$$\mathbb{P}_\lambda(S_m = 0) \le \mathrm{e}^{-mJ(m/n,\lambda)}. \tag{4.5.35}$$

Add proof of (4.5.35) myself!

Exercise 4.24 (The probability that $S_m = 0$)**.** *Prove that* (4.5.35) *holds uniformly in* $m \le \alpha n$ *with* $\alpha < \zeta_\lambda$*, by using* (4.5.4) *in Proposition 4.17 and adapting the proof in* (4.4.44)*.*

We continue by proving a similar bound for $m > \alpha n$, where $\alpha < \zeta_\lambda$ can be chosen arbitrarily close to ζ_λ. Here we shall make use of the fact that, for m close to $\zeta_\lambda n$, $\mathbb{E}_\lambda[X_m] < 1$, so that $m \mapsto S_m$, for $m \ge \alpha n$ is close to a random walk with negative drift. As a result, the probability that $S_m = 0$, yet $S_{m_x} > \varepsilon\sqrt{n}$ is exponentially small.

We now present the details of this argument. We bound

$$\begin{aligned}\mathbb{P}_\lambda\big(S_m = 0, S_{m_x} > \varepsilon\sqrt{n}\big) &= \mathbb{P}_\lambda\big(S_{m_x} > \varepsilon\sqrt{n} \mid S_m = 0\big)\mathbb{P}_\lambda(S_m = 0) \qquad (4.5.36)\\ &= \mathbb{P}_\lambda\Big(\mathrm{Bin}\big(n-(m-1), 1-(1-p)^{m_x-m}\big) > (m_x - m) + \varepsilon\sqrt{n}\Big)\mathbb{P}_\lambda(S_m = 0),\end{aligned}$$

since, by (4.5.5) in Proposition 4.17 and conditionally on $S_m = 0$,

$$S_l + (l-m) \sim \mathrm{Bin}\Big(n-(m-1), 1-(1-p)^{l-m}\Big).$$

We pick $\kappa = \zeta_\lambda - \varepsilon$, for some $\varepsilon > 0$ which is very small. Then, using that $1-(1-a)^b \le ab$ for every a, b with $0 < a < 1, b \ge 1$, we arrive at

$$1-(1-p)^{m_x-m} = 1 - \big(1-\frac{\lambda}{n}\big)^{m_x-m} \le \frac{\lambda(m_x-m)}{n}. \qquad (4.5.37)$$

As a result, with $X = \mathrm{Bin}\Big(n-(m-1), 1-(1-p)^{m_x-m}\Big)$, and using that $n-(m-1) \le n-m \le n(1-\zeta_\lambda+\varepsilon)$ and $p = \lambda/n$,

$$\mathbb{E}_\lambda[X] = [n-(m-1)][1-(1-p)^{m_x-m}] \le (m_x-m)\lambda(1-\zeta_\lambda+\varepsilon). \qquad (4.5.38)$$

Since $\lambda > 1$, we can use that $\lambda(1-\zeta_\lambda) = \lambda \mathrm{e}^{-\lambda\zeta_\lambda} < 1$ by Exercise 4.23, so that, taking $\varepsilon > 0$ so small that $\lambda(1-\zeta_\lambda+\varepsilon) < 1-\varepsilon$, we have

$$\mathbb{E}[X] \le (1-\varepsilon)(m_x - m). \qquad (4.5.39)$$

Therefore,

$$\mathbb{P}_\lambda\big(S_m = 0, S_{m_x} > \varepsilon\sqrt{n}\big) \le \mathbb{P}_\lambda\Big(X - \mathbb{E}[X] > \varepsilon\big((m_x-m)+\sqrt{n}\big)\Big). \qquad (4.5.40)$$

By Theorem 2.18, with $t = \varepsilon\big((m_x-m)+\sqrt{n}\big)$ and using (4.5.39), we obtain

$$\begin{aligned}\mathbb{P}_\lambda\big(S_m = 0, S_{m_x} > \varepsilon\sqrt{n}\big) &\le \exp\left(-\frac{t^2}{2\big((1-\varepsilon)(m_x-m)+t/3\big)}\right)\\ &\le \exp\left(-\frac{t^2}{2\big((m_x-m)+2\varepsilon\sqrt{n}/3\big)}\right). \qquad (4.5.41)\end{aligned}$$

Thus, for $m_x - m \ge \varepsilon\sqrt{n}$, since $t \ge \varepsilon(m_x-m)$, we have

$$\mathbb{P}_\lambda\big(S_m = 0, S_{m_x} > \varepsilon\sqrt{n}\big) \le \exp\big(-3\varepsilon^2(m_x-m)/8\big) = o(n^{-1}), \qquad (4.5.42)$$

while, for $m_x - m \le \varepsilon\sqrt{n}$, since $t \ge \varepsilon\sqrt{n}$, we have

$$\mathbb{P}_\lambda\big(S_m = 0, S_{m_x} > \varepsilon\sqrt{n}\big) \le \exp\big(-3\varepsilon\sqrt{n}/8\big) = \exp\big(-\varepsilon\sqrt{n}/2\big) = o(n^{-1}). \qquad (4.5.43)$$

apt this proof further, by
olying Chernoff bounds on
omials...

The bounds (4.5.35), (4.5.42) and (4.5.43) complete the proof of Theorem 4.16. □

4.6 Notes and discussion

Notes on Section 4.1. There are several possible definitions of the Erdős-Rényi random graph. Many of the classical results are proved for $\mathrm{ER}(n, M)$, which is the random graph on the vertices $[n]$ obtained by adding M edges uniformly at random. Since the number of edges in the Erdős-Rényi random graph has a binomial distribution with parameters $n(n-1)/2$ and p, we should think of M corresponding roughly to $pn(n-1)/2$. Also, writing $\mathbb{P}_M$ for the distribution of $ER(n, M)$, we have that $\mathbb{P}_\lambda$ and $\mathbb{P}_M$ are related as

$$\mathbb{P}_\lambda(E) = \sum_{M=1}^{n(n-1)/2} \mathbb{P}_M(E)\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2, p) = M\big), \tag{4.6.1}$$

where E is any event. This allows one to deduce results for $\mathrm{ER}(n, M)$ from the ones for $\mathrm{ER}(n, p)$ and vice versa. The model $\mathrm{ER}(n, M)$ was first studied in [96], the model $\mathrm{ER}(n, p)$ was introduced in [108], and a model with possibly multiple edges between vertices in [17].

The random graph $\mathrm{ER}(n, M)$ has the advantage that we can think of the graph as evolving as a process, by adding the edges one at a time, which also allows us to investigate *dynamical properties*, such as when the first cycle appears. This is also possible for $\mathrm{ER}(n, p)$ using the coupling in Section 4.1.1, but is slightly less appealing.

We refer to the books [12, 44, 134] for more detailed references of the early literature on random graphs.

Notes on Section 4.2.

Notes on Section 4.3. The strategy in the proof of Theorems 4.4 and 4.5 is close in spirit to the proof in [12], with ingredients taken from [51], which, in turn, was inspired by [55, 56]. In particular, the use of the random variable $Z_{\geq k}$ has appeared in these references. The random variable $Z_{\geq k}$ also plays a crucial role in the analysis of $|\mathcal{C}_{\max}|$ both when $\lambda > 1$ and when $\lambda = 1$.

Exercise 4.25 (Subcritical clusters for $\mathrm{ER}(n, M)$)**.** *Use* (4.6.1) *and Theorems 4.4–4.5 to show that* $|\mathcal{C}_{\max}|/\log n \xrightarrow{\mathbb{P}} I_\lambda^{-1}$ *for* $\mathrm{ER}(n, M)$ *when* $M = n\lambda/2$.

Notes on Section 4.4.

Exercise 4.26 (Supercritical clusters for $\mathrm{ER}(n, M)$)**.** *Use* (4.6.1) *and Theorem 4.8 to show that* $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\lambda$ *for* $\mathrm{ER}(n, M)$ *when* $M = n\lambda/2$.

Exercises 4.25 and 4.26 show that $\mathrm{ER}(n, M)$ has a phase transition when $M = n\lambda/2$ at $\lambda = 1$.

Notes on Section 4.5. The central limit theorem for the largest supercritical cluster was proved in [166],[191] and [23]. In [191], the result follows as a corollary of the main result, involving central limit theorems for various random graph quantities, such as the number tree components of various size. Martin-Löf [166] studies the giant component in the context of epidemics. His proof makes clever use of a connection to asymptotic stochastic differential equations, and is reproduced in [88]. Since we do not assume familiarity with stochastic differential equations, we have produced an independent proof which only relies on elementary techniques.

Chapter 5
Erdős-Rényi random graph revisited

In the previous chapter, we have proved that the largest connected component of the Erdős-Rényi random graph exhibits a phase transition. In this chapter, we investigate several more properties of the Erdős-Rényi random graph. We start by investigating the critical behavior of the size of largest connected component in the Erdős-Rényi random graph by studying $p = 1/n$ in Section 5.1. After this, in Section 5.2, we investigate the phase transition for the connectivity of $\mathrm{ER}_n(p)$, and for p inside the *critical window*, compute the asymptotic probability that the Erdős-Rényi random graph is connected. Finally, in Section 5.3, we study the degree sequence of an Erdős-Rényi random graph.

5.1 The critical behavior

In this section, we study the behavior of the largest connected component for the critical value $p = 1/n$. In this case, it turns out that there is interesting behavior, where the size of the largest connected component is large, yet much smaller than the size of the volume.

Theorem 5.1 (Largest critical cluster)**.** *Take $\lambda = 1 + \theta n^{-1/3}$, where $\theta \in \mathbb{R}$. There exists a constant $b = b(\theta) > 0$ such that, for all $\omega > 1$,*

$$\mathbb{P}_{1+\theta n^{-1/3}}\left(\omega^{-1} n^{2/3} \leq |\mathcal{C}_{\max}| \leq \omega n^{2/3}\right) \geq 1 - \frac{b}{\omega}. \tag{5.1.1}$$

Theorem 5.1 shows that the largest critical cluster obeys a non-trivial scaling result. While $|\mathcal{C}_{\max}|$ is logarithmically small in the subcritical regime $\lambda < 1$ by Theorem 4.4, and $|\mathcal{C}_{\max}| = \Theta(n)$ in the supercritical regime $\lambda > 1$ by Theorem 4.8, at the critical value $\lambda = 1$, we see that the largest cluster is $\Theta(n^{2/3})$. The result in Theorem 5.1 shows that the random variable $|\mathcal{C}_{\max}|n^{-2/3}$ is *tight*, in the sense that with high probability, we have $|\mathcal{C}_{\max}|n^{-2/3} \leq \omega$ for ω sufficiently large. Also, with high probability, $|\mathcal{C}_{\max}|n^{-2/3} \geq \omega^{-1}$, so that with substantial probability, $|\mathcal{C}_{\max}| = \Theta(n^{2/3})$.

5.1.1 Strategy of the proof

We start by studying the tail of the distribution of $|\mathcal{C}(v)|$ for the critical case $\lambda = 1$ in the following theorem. We generalize the setting to values of $\lambda = 1 + \theta n^{-1/3}$ that are close to the critical value $\lambda = 1$:

Proposition 5.2 (Critical cluster tails)**.** *Take $\lambda = 1 + \theta n^{-1/3}$, where $\theta \in \mathbb{R}$, and let $r > 0$. For $k \leq r n^{2/3}$, there exist constants $0 < c_1 < c_2 < \infty$ with $c_1 = c_1(r, \theta)$ such that $\min_{r \leq 1} c_1(r) > 0$, and c_2 independent of r and θ, such that, for n sufficiently large,*

$$\frac{c_1}{\sqrt{k}} \leq \mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \geq k) \leq c_2(\theta_+ n^{-1/3} + \frac{1}{\sqrt{k}}). \tag{5.1.2}$$

Proposition 5.2 implies that the tails of the critical cluster size distribution obey similar asymptotics as the tails of the total progeny of a critical branching process (see (3.6.20)). The tail in (5.1.2) is only valid for values of k that are not too large. Indeed, when $k > n$, then $\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k) = 0$. Therefore, there must be a cut-off above which the asymptotics

fails to hold. As it turns out, this cut-off is given by $rn^{2/3}$. The upper bound in (5.1.2) holds for a wider range of k, in fact, the proof yields that (5.1.2) is valid for *all* k.

We next study the critical expected cluster size:

Proposition 5.3 (Bound on critical expected cluster size)**.** *Take $\lambda = 1 + \theta n^{-1/3}$, where $\theta \in \mathbb{R}$. Then, for all $n \geq 1$,*

$$\mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}(1)|] \leq n^{1/3}/\theta_{-}. \tag{5.1.3}$$

Proposition 5.3 is intuitively consistent with Theorem 5.1. Indeed, in the critical regime, the expected cluster size receives a substantial amount from the largest cluster. Therefore, intuitively, for any $v \in [n]$,

$$\mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}(1)|] \sim \mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}(v)|\mathbb{1}_{\{v\in\mathcal{C}_{\max}\}}] = \mathbb{E}_1[|\mathcal{C}_{\max}|\mathbb{1}_{\{v\in\mathcal{C}_{\max}\}}], \tag{5.1.4}$$

where $\sim$ denotes an equality with an uncontrolled error.

When $|\mathcal{C}_{\max}| = \Theta(n^{2/3})$, then

$$\mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}_{\max}|\mathbb{1}_{\{v\in\mathcal{C}_{\max}\}}] \sim n^{2/3}\mathbb{P}_1(v \in \mathcal{C}_{\max}). \tag{5.1.5}$$

Furthermore, when $|\mathcal{C}_{\max}| = \Theta(n^{2/3})$, then

$$\mathbb{P}_{1+\theta n^{-1/3}}\big(v \in \mathcal{C}_{\max}\big) \sim \frac{n^{2/3}}{n} = n^{-1/3}. \tag{5.1.6}$$

Therefore, one is intuitively lead to the conclusion

$$\mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}(1)|] \sim n^{1/3}. \tag{5.1.7}$$

Exercise 5.1 (Critical expected cluster size)**.** *Prove that Proposition 5.2 also implies that $\chi(1) \geq cn^{1/3}$ for some $c > 0$. Therefore, for $\lambda = 1$, the bound in Proposition 5.3 is asymptotically sharp.*

Propositions 5.2 and 5.3 are proved in Section 5.1.2 below. We will first prove Theorem 5.1 subject to them.

Proof of Theorem 5.1 subject to Propositions 5.2 and 5.3. The statement in Theorem 5.1 is vacuous when $\omega \leq b$, so that, by taking b sufficiently large, we may assume that $\omega \geq 1$ is large. In turn, the statement in Theorem 5.1 is vacuous when $\omega^{-1}n^{2/3} \leq 1$ and $\omega n^{2/3} \leq n$, i.e., for $n \leq \omega^{2/3}$. Since ω is large, we may also assume that $n \geq N$, where N is large.

We start with the upper bound on $|\mathcal{C}_{\max}|$. We again make use of the fundamental equality $\{|\mathcal{C}_{\max}| \geq k\} = \{Z_{\geq k} \geq k\}$, where we recall that

$$Z_{\geq k} = \sum_{v\in[n]} \mathbb{1}_{\{|\mathcal{C}(v)|\geq k\}}. \tag{5.1.8}$$

By the Markov inequality (Theorem 2.14), we obtain

$$\mathbb{P}_{1+\theta n^{-1/3}}\big(|\mathcal{C}_{\max}| \geq \omega n^{2/3}\big) = \mathbb{P}_{1+\theta n^{-1/3}}\big(Z_{\geq \omega n^{2/3}} \geq \omega n^{2/3}\big) \leq \omega^{-1}n^{-2/3}\mathbb{E}_{1+\theta n^{-1/3}}[Z_{\geq \omega n^{2/3}}]. \tag{5.1.9}$$

By Proposition 5.2,

$$\mathbb{E}_{1+\theta n^{-1/3}}[Z_{\geq \omega n^{2/3}}] = n\mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \geq \omega n^{2/3}) \leq c_2 n^{2/3}(\theta_{+} + 1/\sqrt{\omega}), \tag{5.1.10}$$

so that

$$\mathbb{P}_{1+\theta n^{-1/3}}\big(|\mathcal{C}_{\max}| > \omega n^{2/3}\big) \le c_2 n^{2/3}(\theta_+ + 1/\sqrt{\omega})/(\omega n^{2/3}) \le \frac{c_2}{\omega}(\theta_+ + 1/\sqrt{\omega}). \quad (5.1.11)$$

For the lower bound on $|\mathcal{C}_{\max}|$, we first note that for $\omega < b$, there is nothing to prove. The constant $b > 0$ will be taken large, so that we shall assume that $\omega > \kappa^{-1}$, where $\kappa > 0$ is the constant appearing in Proposition 5.2.

We start by proving the result for $\theta \le -1$. We use the Chebychev inequality (Theorem 2.15), as well as $\{|\mathcal{C}_{\max}| < k\} = \{Z_{\ge k} = 0\}$, to obtain that

$$\mathbb{P}_{1+\theta n^{-1/3}}\big(|\mathcal{C}_{\max}| < \omega^{-1} n^{2/3}\big) = \mathbb{P}_{1+\theta n^{-1/3}}\big(Z_{\ge \omega^{-1} n^{2/3}} = 0\big) \le \frac{\mathrm{Var}_{1+\theta n^{-1/3}}(Z_{\ge \omega^{-1} n^{2/3}})}{\mathbb{E}_{1+\theta n^{-1/3}}[Z_{\ge \omega^{-1} n^{2/3}}]^2}. \quad (5.1.12)$$

By (5.1.2), we have that

$$\mathbb{E}_{1+\theta n^{-1/3}}[Z_{\ge \omega^{-1} n^{2/3}}] = n\mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \ge k) \ge c_1 \sqrt{\omega} n^{2/3}, \quad (5.1.13)$$

where we used that $\omega \ge \kappa^{-1}$, and $c_1 = \min_{r \le \kappa} c_1(r) > 0$. Also, by Proposition 4.7, with $k_n = \omega^{-1} n^{2/3}$,

$$\mathrm{Var}_{1+\theta n^{-1/3}}(Z_{\ge \omega^{-1} n^{2/3}}) \le n\mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}(1)| \mathbb{1}_{\{|\mathcal{C}(1)| \ge \omega^{-1} n^{2/3}\}}]. \quad (5.1.14)$$

By Proposition 5.3, we can further bound, using that $\theta \le -1$,

$$\mathrm{Var}_{1+\theta n^{-1/3}}(Z_{\ge \omega^{-1} n^{2/3}}) \le n\mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}(1)|] \le n^{4/3}. \quad (5.1.15)$$

Substituting (5.1.12)–(5.1.15), we obtain, for n sufficiently large,

$$\mathbb{P}_{1+\theta n^{-1/3}}\big(|\mathcal{C}_{\max}| < \omega^{-1} n^{2/3}\big) \le \frac{n^{4/3}}{c_1^2 \omega n^{4/3}} = \frac{1}{c_1^2 \omega}. \quad (5.1.16)$$

We conclude that

$$\begin{aligned} \mathbb{P}_{1+\theta n^{-1/3}}\Big(\omega^{-1} n^{2/3} \le |\mathcal{C}_{\max}| \le \omega n^{2/3}\Big) &= 1 - \mathbb{P}_{1+\theta n^{-1/3}}\big(|\mathcal{C}_{\max}| < \omega^{-1} n^{2/3}\big) \quad (5.1.17)\\ &\quad - \mathbb{P}_{1+\theta n^{-1/3}}\big(|\mathcal{C}_{\max}| > \omega n^{2/3}\big)\\ &\ge 1 - \frac{1}{c_1^2 \omega} - \frac{c_2}{\omega^{3/2}} \ge 1 - \frac{b}{\omega}, \quad (5.1.18) \end{aligned}$$

when $b = c_1^{-2} + c_2$.

Finally, we make use of *monotonicity* in λ. The random variable $|\mathcal{C}_{\max}|$ is *increasing*, and therefore,

$$\mathbb{P}_{1+\theta n^{-1/3}}\Big(|\mathcal{C}_{\max}| < \omega^{-1} n^{2/3}\Big) \le \mathbb{P}_{1-\bar{\theta} n^{-1/3}}\Big(|\mathcal{C}_{\max}| < \omega^{-1} n^{2/3}\Big), \quad (5.1.19)$$

where we define $\bar{\theta} = |\theta| \vee 1$. Thus, $-\bar{\theta} \le -1$, and the result follows from our previous calculation. This completes the proof of Theorem 5.1 subject to Propositions 5.2 and 5.3. □

5.1.2 Proofs of Propositions 5.2 and 5.3

Proof of Proposition 5.2. Theorem 4.2 gives

$$\mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \geq k) \leq \mathbb{P}_{n,p}(T \geq k), \tag{5.1.20}$$

where we recall that $\mathbb{P}_{n,p}$ is the law of a binomial branching process with parameters n and $p = \lambda/n = (1 + \theta n^{-1/3})/n$, and T its total progeny. By Theorem 3.20, for $\lambda = 1$,

$$\mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \geq k) \leq \mathbb{P}^*_\lambda(T^* \geq k) + e_k(n), \tag{5.1.21}$$

where, by (3.7.2),

$$|e_k(n)| \leq \frac{2}{n} \sum_{s=1}^{k-1} \mathbb{P}^*_\lambda(T^* \geq s), \tag{5.1.22}$$

and where we recall that $\mathbb{P}^*_\lambda$ is the law of a Poisson branching process with parameter λ, and T^* is its total progeny.

By Theorem 3.18, it follows that there exists a $C > 0$ such that for all $s \geq 1$,

$$\mathbb{P}^*_{1+\theta n^{-1/3}}(T^* \geq s) \leq \zeta_\lambda + \sum_{t=s}^{\infty} \mathbb{P}^*_{1+\theta n^{-1/3}}(T^* = t) \leq C\big(\theta_+ n^{-1/3} + 1/\sqrt{s}\big), \tag{5.1.23}$$

by (3.6.22) in Corollary 3.19. Therefore, for all $k \leq n$,

$$|e_k(n)| \leq \frac{4}{n} \sum_{s=1}^{k} C\big(\theta_+ n^{-1/3} + 1/\sqrt{s}\big) \leq 4C\big(\theta_+ k n^{-4/3} + \frac{\sqrt{k}}{n}\big) \leq 4C\big(\theta_+ n^{-1/3} + 1/\sqrt{k}\big). \tag{5.1.24}$$

We conclude that, for all $k \leq n$,

$$\mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \geq k) \leq 5C\big(\theta_+ n^{-1/3} + 1/\sqrt{k}\big). \tag{5.1.25}$$

We proceed with the lower bound in (5.1.2), for which we make use of Theorem 4.3 with $k \leq r n^{2/3}$. This gives that

$$\mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \geq k) \geq \mathbb{P}_{n-k,p}(T \geq k). \tag{5.1.26}$$

where T is the total progeny of a binomial branching process with parameters $n - k \leq n - r n^{2/3}$ and $p = \lambda/n = (1+\theta n^{-1/3})/n$. We again use Theorem 3.20 for $\lambda_n = 1+(\theta - r)n^{-1/3}$, as in (5.1.21) and (5.1.22). We apply the one-but-last bound in (5.1.24), so that

$$\mathbb{P}_{1+\theta n^{-1/3}}(|\mathcal{C}(1)| \geq k) \geq \mathbb{P}^*_{\lambda_n}(T^* \geq k) - \frac{4C\sqrt{k}}{n} \geq \mathbb{P}^*_{\lambda_n}(T^* \geq k) - \frac{4C\sqrt{r}}{n^{2/3}}. \tag{5.1.27}$$

We then use Theorem 3.16 to obtain, since $\lambda_n \leq 1$,

$$\begin{aligned}
\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k) &\geq \sum_{t=k}^{\infty} \mathbb{P}^*_{\lambda_n}(T^* = t) - \frac{4C\sqrt{r}}{n^{2/3}} \\
&= \sum_{t=k}^{\infty} \frac{(\lambda_n t)^{t-1}}{t!} \mathrm{e}^{-\lambda_n t} - \frac{4C\sqrt{r}}{n^{2/3}} \\
&= \frac{1}{\lambda_n} \sum_{t=k}^{\infty} \mathbb{P}^*_1(T^* = t) \mathrm{e}^{-I_{\lambda_n} t} - \frac{4C\sqrt{r}}{n^{2/3}},
\end{aligned} \tag{5.1.28}$$

where, for $\lambda_n = 1 + (\theta - r)n^{-1/3}$ and by (4.3.1),

$$I_{\lambda_n} = \lambda_n - 1 - \log \lambda_n = \frac{1}{2}(\lambda_n - 1)^2 + O(|\lambda_n - 1|^3). \tag{5.1.29}$$

Therefore,

$$\begin{aligned}\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k) &\geq \sum_{t=k}^{2k} \mathbb{P}_1^*(T^* = t)\mathrm{e}^{-\frac{1}{2}(\lambda_n-1)^2 t(1+o(1))} - \frac{4C\sqrt{r}}{n^{2/3}} \\ &\geq \sum_{t=k}^{2k} \frac{C}{\sqrt{t^3}}\mathrm{e}^{-\frac{1}{2}(\lambda_n-1)^2 t(1+o(1))} - \frac{4C\sqrt{r}}{n^{2/3}} \\ &\geq \frac{2^{-3/2}C}{\sqrt{k}}\mathrm{e}^{-k(\lambda_n-1)^2(1+o(1))} - \frac{4C\sqrt{r}}{n^{2/3}},\end{aligned} \tag{5.1.30}$$

since $\lambda_n - 1 = (\theta - r)n^{-1/3}$. Now use that, for $n \geq N$,

$$\sqrt{r}n^{-2/3} = \sqrt{rk}n^{-2/3}/\sqrt{k} \leq rn^{-1/3}/\sqrt{k} \leq rN^{-1/3}/\sqrt{k}, \tag{5.1.31}$$

so that

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k) \geq \frac{c_1(r)}{\sqrt{k}}, \tag{5.1.32}$$

with $c_1(r) = C(2^{-3/2}\mathrm{e}^{-r(\theta-r)^2} - 4\sqrt{r}/N^{-1/3}) > 0$ for $r \leq 1$, whenever N is sufficiently large. This completes the proof of Proposition 5.2. □

Proof of Proposition 5.3. Theorem 4.2 gives that $|\mathcal{C}(1)| \preceq T$, where T is the total progeny of a branching process with a $\mathrm{Bin}(n, \lambda/n)$ branching process, and where $\lambda = 1 + \theta n^{-1/3}$. As a result, for $\theta < 0$,

$$\mathbb{E}_{1+\theta n^{-1/3}}[|\mathcal{C}(1)|] \leq \mathbb{E}[T] = 1/(1-\lambda) = n^{1/3}/|\theta|. \tag{5.1.33}$$

This proves the claim. □

5.1.3 Connected components in the critical window revisited

In this section, we discuss the *critical window* of the Erdős-Rényi random graph. By Theorem 5.1, we know that, for $p = 1/n$, the largest connected component has size roughly equal to $n^{2/3}$. As it turns out, such behavior is also seen for related values of p. Namely, if we choose $p = (1 + tn^{-1/3})/n$, then we see similar behavior appearing for the largest connected component size. Therefore, the values of p for which $p = (1+tn^{-1/3})/n$ are called the *critical window*. We start by discussing the most detailed work on this problem, which is by Aldous [11], following previous work on the critical window in [42, 132, 162, 164].

The point in [11] is to prove simultaneous weak convergence of all connected components at once. We start by introducing some notation. Let $|\mathcal{C}_{(j)}(t)|$ denote the j^{th} largest cluster of $\mathrm{ER}_n(p)$ for $p = (1 + tn^{-1/3})/n$. Then one of the main results in [11] is the following theorem:

Theorem 5.4 (Weak convergence of largest clusters in critical window)**.** *For $p = (1 + tn^{-1/3})/n$, and any $t \in \mathbb{R}$, the vector $\mathbf{C}(t) \equiv (n^{-2/3}|\mathcal{C}_{(1)}(t)|, n^{-2/3}|\mathcal{C}_{(2)}(t)|, n^{-2/3}|\mathcal{C}_{(3)}(t)|, \ldots)$ converges in distribution to a random vector $\gamma \equiv (\gamma_i(t))_{i\geq 1}$.*

Thus, Theorem 5.4 is stronger than Theorem 5.1 in three ways: (1) Theorem 5.4 proves weak convergence, rather than tightness only; (2) Theorem 5.4 considers *all* connected components, ordered by size, rather than only the first one; (3) Theorem 5.4 investigates all values inside the critical window at once.

While [11] is the first paper where a result as in Theorem 5.4 is stated explicitly, similar results had been around before [11], which explains why Aldous calls Theorem 5.4 a 'Folk Theorem'. The beauty of [11] is that Aldous gives two explicit descriptions of the distribution of the limiting random variable $(C^t_{(1)}|, \mathcal{C}^t_{(2)}, \mathcal{C}^t_{(3)}|, \ldots)$, the first being in terms of lengths of excursions of Brownian motion, the second in terms of the so-called *multiplicative coalescent process*. We shall intuitively explain these constructions now.

We start by explaining the construction in terms of excursions of Brownian motion. Let $\{W(s)\}_{s\geq 0}$ be standard Brownian motion, and define

$$W^t(s) = W(s) + ts - s^2/2 \tag{5.1.34}$$

be Brownian motion with an (inhomogeneous) drift $t - s$ at time s. Let

$$B^t(s) = W^t(s) - \min_{0\leq s'\leq s} W^t(s') \tag{5.1.35}$$

correspond to the process $\{W^t(s)\}_{s\geq 0}$ reflected at 0. We now consider the *excursions* of this process, ordered in their length. Here an excursion γ of $\{B^t(s)\}_{s\geq 0}$ is a time interval $[l(\gamma), r(\gamma)]$ for which $B^t(l(\gamma)) = B^t(r(\gamma)) = 0$, but $B^t(s) > 0$ for all $s \in (l(\gamma), r(\gamma))$. Let the length $|\gamma|$ of the excursion γ be given by $r(\gamma) - l(\gamma)$. As it turns out (see [11, Section 1] for details), the excursions of $\{B^t(s)\}_{s\geq 0}$ can be ordered by decreasing length, so that $\{\gamma^t_j : j \geq 1\}$ are the excursions. Then, the limiting random vector $\mathbf{C}^n$ has the *same* distribution as the ordered excursions $\{\gamma^t_j : j \geq 1\}$. The idea behind this is as follows. We make use of the random walk representation of the various clusters, which connects the cluster exploration to random walks. However, as for example (4.5.3) shows, the step size distribution is decreasing as we explore more vertices, which means that we arrive at an inhomogeneous and ever decreasing drift, as in (5.1.34). Since, in general, random walks converge to Brownian motions, this way the connection between these precise processes can be made.

To explain the connection to the multiplicative coalescent, we shall interpret the t-variable in $p = (1 + tn^{-1/3})/n$ as *time*. We note that when we have two clusters of size $xn^{2/3}$ and $yn^{2/3}$ say, and we increase t to $t + dt$, then the probability that these two clusters merge is roughly equal to the number of possible connecting edges, which is $xn^{2/3} \times yn^{2/3} = xyn^{4/3}$ times the probability that an edge turns from vacant to occupied when p increases from $p = (1+tn^{-1/3})/n$ to $(1+(t+dt)n^{-1/3})/n$, which is $dtn^{-4/3}$. Thus, this probability is, for small dt close to

$$xydt. \tag{5.1.36}$$

Thus, distinct clusters meet at a rate proportional to the rescaled product of their sizes. The continuous process which does this precisely is called the multiplicative coalescent, and using the above ideas, Aldous is able to show that the limit of $\mathbf{C}^{t,n}$ equals such a multiplicative coalescent process.

5.2 Connectivity threshold

In this section, we investigate the connectivity threshold for the Erdős-Rényi random graph. As we can see in Theorem 4.8, for every $1 < \lambda < \infty$, the largest cluster for the Erdős-Rényi random graph when $p = \lambda/n$ is $\zeta_\lambda n(1 + o(1))$, where $\zeta_\lambda > 0$ is the survival

probability of a Poisson branching process with parameter λ. Since extinction is certain when the root has no offspring, we have

$$\zeta_\lambda \le 1 - \mathbb{P}^*(Z_1^* = 0) = 1 - \mathrm{e}^{-\lambda} < 1. \tag{5.2.1}$$

Therefore, the Erdős-Rényi random graph with edge probability $p = \lambda/n$ is with high probability *disconnected* for each fixed $\lambda < \infty$. Here, we use the terminology "with high probability" to denote an event of which the probability tends to 1. We now investigate the threshold for connectivity for an appropriate choice $\lambda = \lambda_n \to \infty$. Theorem 5.5 and its extension, Theorem 5.8, were first proved in [96].

Theorem 5.5 (Connectivity threshold)**.** *For $\lambda - \log n \to \infty$, the Erdős-Rényi random graph is with high probability connected, while for $\lambda - \log n \to -\infty$, the Erdős-Rényi random graph is with high probability disconnected.*

In the proof, we investigate the number of isolated vertices. Define

$$Y = \sum_{i=1}^{n} I_i, \qquad \text{where} \qquad I_i = \mathbb{1}_{\{|\mathcal{C}(i)|=1\}} \tag{5.2.2}$$

for the number of isolated vertices. Clearly, when $Y \ge 1$, then there exists at least one isolated vertex, so that the graph is disconnected. Remarkably, it turns out that when there is no isolated vertex, i.e., when $Y = 0$, then the random graph is also with high probability connected. See Proposition 5.7 below for the precise formulation of this result. By Proposition 5.7, we need to investigate the probability that $Y \ge 1$. In the case where $|\lambda - \log n| \to \infty$, we make use of the Markov and Chebychev inequality (Theorems 2.14 and 2.15) combined with a first and second moment argument using a variance estimate in Proposition 5.6. We will extend the result to the case that $\lambda = \log n + t$, in which case we need a more precise result in Theorem 5.8 below. The main ingredient to the proof of Theorem 5.8 is to show that, for $\lambda = \log n + t$, Y converges to a Poisson random variable with parameter e^{-t} when $n \to \infty$.

To prove that $Y \ge 1$ with high probability when $\lambda - \log n \to -\infty$, and $Y = 0$ with high probability when $\lambda - \log n \to \infty$ we use the Markov inequality (Theorem 2.14). We make use of an estimate on the mean and variance of Y:

Proposition 5.6 (Mean and variance of number of isolated vertices)**.** *For every $\lambda \le n/2$,*

$$\mathbb{E}_\lambda[Y] = n\mathrm{e}^{-\lambda}(1 + O(\mathrm{e}^{-\frac{\lambda^2}{n}})), \tag{5.2.3}$$

and, for every $\lambda \le n$,

$$\mathrm{Var}_\lambda(Y) \le \mathbb{E}_\lambda[Y] + \frac{\lambda}{n-\lambda}\mathbb{E}_\lambda[Y]^2. \tag{5.2.4}$$

Proof. Since $|\mathcal{C}(i)| = 1$ precisely when all edges emanating from i are vacant, we have, using $1 - x \le \mathrm{e}^{-x}$,

$$\mathbb{E}_\lambda[Y] = n\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1) = n(1 - \frac{\lambda}{n})^{n-1} \le n\mathrm{e}^{-\lambda}\mathrm{e}^{\frac{\lambda}{n}}. \tag{5.2.5}$$

Also, using that $1 - x \ge \mathrm{e}^{-x-x^2}$ for $0 \le x \le \frac{1}{2}$, we obtain

$$\begin{aligned}\mathbb{E}_\lambda[Y] &= n\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1) \ge n\mathrm{e}^{-(n-1)\frac{\lambda}{n}(1+\frac{\lambda}{n})} \\ &\ge n\mathrm{e}^{-\lambda(1+\frac{\lambda}{n})} = n\mathrm{e}^{-\lambda}\mathrm{e}^{-\frac{\lambda^2}{n}}.\end{aligned} \tag{5.2.6}$$

This proves (5.2.3).

To prove (5.2.4), we use the exchangeability of the vertices to compute

$$\mathbb{E}_\lambda[Y^2] = n\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1) + n(n-1)\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1, |\mathcal{C}(2)| = 1). \tag{5.2.7}$$

Therefore, we obtain

$$\begin{aligned}\mathrm{Var}_\lambda(Y) = n\big[\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1) - \mathbb{P}_\lambda(|\mathcal{C}(1)| = 1, |\mathcal{C}(2)| = 1)\big] \\ + n^2\big[\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1, |\mathcal{C}(2)| = 1) - \mathbb{P}_\lambda(|\mathcal{C}(1)| = 1)^2\big].\end{aligned} \tag{5.2.8}$$

The first term is bounded above by $\mathbb{E}_\lambda[Y]$. The second term can be computed by using (5.2.5), together with

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1, |\mathcal{C}(2)| = 1) = (1 - \frac{\lambda}{n})^{2n-3}. \tag{5.2.9}$$

Therefore, by (5.2.5) and (5.2.8), we obtain

$$\begin{aligned}\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1, |\mathcal{C}(2)| = 1) - \mathbb{P}_\lambda(|\mathcal{C}(1)| = 1)^2 &= \mathbb{P}_\lambda(|\mathcal{C}(1)| = 1)^2\big[(1 - \frac{\lambda}{n})^{-1} - 1\big] \\ &= \frac{\lambda}{n(1-\frac{\lambda}{n})}\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1)^2.\end{aligned} \tag{5.2.10}$$

We conclude that

$$\mathrm{Var}_\lambda(Y) \le \mathbb{E}_\lambda[Y] + \frac{\lambda}{n-\lambda}\mathbb{E}_\lambda[Y]^2. \tag{5.2.11}$$

□

Proposition 5.7 (Connectivity and isolated vertices)**.** *For all $0 \le \lambda \le n$,*

$$\mathbb{P}_\lambda\Big(\mathrm{ER}_n(\lambda/n) \text{ connected}\Big) \le \mathbb{P}_\lambda(Y = 0). \tag{5.2.12}$$

Moreover, if there exists an $a > 1/2$ such that $\lambda \ge a \log n$, then, for $n \to \infty$,

$$\mathbb{P}_\lambda\Big(\mathrm{ER}_n(\lambda/n) \text{ connected}\Big) = \mathbb{P}_\lambda(Y = 0) + o(1). \tag{5.2.13}$$

Proof. We use that

$$\mathbb{P}_\lambda\Big(\mathrm{ER}_n(\lambda/n) \text{ disconnected}\Big) = \mathbb{P}_\lambda(Y > 0) + \mathbb{P}_\lambda\Big(\mathrm{ER}_n(\lambda/n) \text{ disconnected}, Y = 0\Big). \tag{5.2.14}$$

This immediately proves (5.2.12).

To prove (5.2.13), we make use of a computation involving *trees*. For $k = 2, \dots, n$, we denote by X_k the number of occupied trees of size equal to k on the vertices $1, \dots, n$ that cannot be extended to a tree of larger size. Thus, each tree which is counted in X_k has size precisely equal to k, and when we denote it's vertices by $v_1, \dots, v_k$, then all the edges between v_i and $v \notin \{v_1, \dots, v_k\}$ are vacant. Moreover, there are precisely $k-1$ occupied edges between the v_i that are such that these occupied edges form a tree. Note that a connected component of size k can contain more than one tree of size k, since the connected component may contain cycles. Note furthermore that, when $\mathrm{ER}_n(\lambda/n)$ is disconnected, but $Y = 0$, there must be a $k \in \{2, \dots, n/2\}$ for which $X_k \ge 1$.

We conclude from Boole's inequality and the Markov inequality (Theorem 2.14) that

$$\mathbb{P}_\lambda\Big(\mathrm{ER}_n(\lambda/n) \text{ disconnected}, Y=0\Big) \leq \mathbb{P}_\lambda\Big(\cup_{k=2}^{n/2}\{X_k \geq 1\}\Big) \leq \sum_{k=2}^{n/2} \mathbb{P}_\lambda(X_k \geq 1) \leq \sum_{k=2}^{n/2} \mathbb{E}_\lambda[X_k]. \tag{5.2.15}$$

Therefore, we need to bound $\mathbb{E}_\lambda[X_k]$. For this, we note that there are $\binom{n}{k}$ ways of choosing k vertices, and, by Cayley's Theorem 3.17, there are k^{k-2} labeled trees containing k vertices. Therefore,

$$\mathbb{E}_\lambda[X_k] = \binom{n}{k} k^{k-2} q_k, \tag{5.2.16}$$

where q_k is the probability that any tree of size k is occupied and all the edges from the tree to other vertices are vacant, which is equal to

$$q_k = \Big(\frac{\lambda}{n}\Big)^{k-1}\Big(1-\frac{\lambda}{n}\Big)^{k(n-k)} \leq \Big(\frac{\lambda}{n}\Big)^{k-1} \mathrm{e}^{-\lambda k(n-k)/n}. \tag{5.2.17}$$

We conclude that

$$\mathbb{E}_\lambda[X_k] \leq n\lambda^{k-1}\frac{k^{k-2}}{k!}\mathrm{e}^{-\frac{\lambda}{n}k(n-k)}. \tag{5.2.18}$$

If we further use that $k! \geq k^k \mathrm{e}^{-k}$, and also use that $\lambda \geq 1$, then we arrive at

$$\mathbb{E}_\lambda[X_k] \leq n(e\lambda)^k \frac{1}{k^2}\mathrm{e}^{-\frac{\lambda}{n}k(n-k)}. \tag{5.2.19}$$

Since $\lambda \mapsto \mathrm{e}^{-\frac{\lambda}{n}k(n-k)}$ is decreasing in λ, it suffices to investigate $\lambda = a\log n$ for some $a > 1/2$. For $k \in \{2,3,4\}$, for $\lambda = a\log n$ for some $a > 1/2$,

$$\mathbb{E}_\lambda[X_k] \leq n(e\lambda)^4 \mathrm{e}^{-\lambda k}\mathrm{e}^{o(1)} = o(1). \tag{5.2.20}$$

For all $k \leq n/2$ with $k \geq 5$, we bound $k(n-k) \geq kn/2$, so that

$$\mathbb{E}_\lambda[X_k] \leq n(e\lambda\mathrm{e}^{-\lambda/2})^k. \tag{5.2.21}$$

As a result, for $\lambda = a\log n$ with $a > 1/2$, and all $k \geq 5$, and using that $\lambda \mapsto \lambda\mathrm{e}^{-\lambda/2}$ is decreasing for $\lambda \geq 2$,

$$\mathbb{E}_\lambda[X_k] \leq n^{1-k/4}. \tag{5.2.22}$$

We conclude that

$$\mathbb{P}_\lambda\Big(\mathrm{ER}_n(\lambda/n) \text{ disconnected}, Y=0\Big) \leq \sum_{k=2}^{n/2}\mathbb{E}_\lambda[X_k] \leq \sum_{k=2}^{n/2} n^{1-k/4} = o(1). \tag{5.2.23}$$

□

Proof of Theorem 5.5. The proof makes essential use of Proposition 5.7. We start by proving that for $\lambda - \log n \to -\infty$, the Erdős-Rényi random graph is with high probability disconnected. We use (5.2.3) to note that

$$\mathbb{E}_\lambda[Y] = n\mathrm{e}^{-\lambda}(1+o(1)) = \mathrm{e}^{-\lambda+\log n}(1+o(1)) \to \infty. \tag{5.2.24}$$

By the Chebychev inequality (Theorem 2.15), and the fact that $\lambda \leq \log n$,

$$\mathbb{P}_\lambda(Y=0) \leq \frac{\mathbb{E}_\lambda[Y] + \frac{\lambda}{n-\lambda}\mathbb{E}_\lambda[Y]^2}{\mathbb{E}_\lambda[Y]^2} = \mathbb{E}_\lambda[Y]^{-1} + \frac{\lambda}{n-\lambda} \to 0. \tag{5.2.25}$$

Proposition 5.7 completes the proof that for $\lambda - \log n \to -\infty$, the Erdős-Rényi random graph is with high probability disconnected.

When $\lambda - \log n \to \infty$ with $\lambda \leq 2\log n$, then, by the Markov inequality (Theorem 2.14) and (5.2.5),

$$\mathbb{P}_\lambda(Y=0) = 1 - \mathbb{P}_\lambda(Y \geq 1) \geq 1 - \mathbb{E}_\lambda[Y] \geq 1 - n\mathrm{e}^{-\lambda}O(1) \to 1. \tag{5.2.26}$$

Since the connectivity is an increasing property, this also prove the claim for $\lambda - \log n \to \infty$ with $\lambda \geq 2\log n$. Therefore, the claim again follows from Proposition 5.7. □

5.2.1 Critical window for connectivity*

In this section, we investigate the critical window for connectivity, by considering connectivity of $\mathrm{ER}_n(\lambda/n)$ when $\lambda = \log n + t$ for *fixed* $t \in \mathbb{R}$. The main result in this section is as follows:

Theorem 5.8 (Critical window for connectivity)**.** *For $\lambda = \log n + t \to \infty$, the Erdős-Rényi random graph is connected with probability* $\mathrm{e}^{-\mathrm{e}^{-t}}(1+o(1))$.

Proof. In the proof, we again rely on Proposition 5.7. We fix $\lambda = \log n + t$ for some $t \in \mathbb{R}$.

We prove a Poisson approximation for Y that reads that $Y \xrightarrow{d} Z$, where Z is a Poisson random variable with parameter

$$\lim_{n\to\infty} \mathbb{E}_\lambda[Y] = \mathrm{e}^{-t}, \tag{5.2.27}$$

where we recall (5.2.3). Therefore, the convergence in distribution of Y to a Poisson random variable with mean e^{-t} implies that

$$\mathbb{P}_\lambda(Y=0) = \mathrm{e}^{-\lim_{n\to\infty}\mathbb{E}_\lambda[Y]} + o(1) = \mathrm{e}^{-\mathrm{e}^{-t}} + o(1), \tag{5.2.28}$$

and the result follows by Proposition 5.7.

In order to show that $Y \xrightarrow{d} Z$, we use Theorem 2.4 and Theorem 2.5, so that it suffices to prove, recalling that $I_i = \mathbb{1}_{\{|\mathcal{C}(i)|=1\}}$, for all $r \geq 1$,

$$\lim_{n\to\infty} \mathbb{E}[(Y)_r] = \lim_{n\to\infty} \sum_{i_1,\ldots,i_r}^{*} \mathbb{P}_\lambda\big(I_{i_1} = \cdots = I_{i_r} = 1\big) = \mathrm{e}^{-tr}, \tag{5.2.29}$$

where the sum ranges over all $i_1, \ldots, i_r \in [n]$ which are *distinct*. By exchangeability of the vertices, $\mathbb{P}_\lambda\big(I_{i_1} = \cdots = I_{i_r} = 1\big)$ is independent of the precise choice of the indices $i_1, \ldots, i_r$, so that

$$\mathbb{P}_\lambda\big(I_{i_1} = \cdots = I_{i_r} = 1\big) = \mathbb{P}_\lambda\big(I_1 = \cdots = I_r = 1\big). \tag{5.2.30}$$

Using that there are $n(n-1)\cdots(n-r+1)$ distinct choices of $i_1, \ldots, i_r \in [n]$, we arrive at

$$\mathbb{E}[(Y)_r] = \frac{n!}{(n-r)!}\mathbb{P}_\lambda\big(I_1 = \cdots = I_r = 1\big). \tag{5.2.31}$$

The event $\{I_1 = \cdots = I_r = 1\}$ occurs precisely when all edges st with $s \in [r]$ and $t \in [n]$ are vacant. There are $r(r-1)/2 + r(n-r) = r(2n-r-1)/2$ of such edges, and since these edges are all independent, we arrive at

$$\begin{aligned}\mathbb{P}_\lambda\big(I_1 = \cdots = I_r = 1\big) &= (1-\frac{\lambda}{n})^{r(2n-r-1)/2} \\ &= (1-\frac{\lambda}{n})^{nr}(1-\frac{\lambda}{n})^{-r(r+1)/2} = n^{-r}\mathbb{E}_\lambda[Y]^r(1+o(1)), \qquad (5.2.32)\end{aligned}$$

using that $\mathbb{E}_\lambda[Y] = n(1-\lambda/n)^{n-1}$. Thus,

$$\lim_{n\to\infty}\mathbb{E}[(Y)_r] = \lim_{n\to\infty}\frac{n!}{(n-r)!}n^{-r}\mathbb{E}_\lambda[Y]^r = \mathrm{e}^{-tr}, \qquad (5.2.33)$$

where we use (5.2.27). This completes the proof of Theorem 5.8. □

Exercise 5.2 (Second moment of the number of isolated vertices)**.** *Prove directly that the second moment of Y converges to the second moment of Z, by using* (5.2.10)*.*

5.3 Degree sequence of the Erdős-Rényi random graph

As described in Chapter 1, the degree sequences of various real networks obey power laws. Therefore, in this section, we investigate the degree sequence of the Erdős-Rényi random graph for fixed $\lambda > 0$. In order to be able to state the result, we first introduce some notation. We write

$$p_k = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \qquad k \geq 0, \qquad (5.3.1)$$

for the Poisson distribution with parameter λ. Let D_i denote the degree of vertex i and write

$$P_k^{(n)} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{D_i=k\}} \qquad (5.3.2)$$

for the empirical degree distribution of the degrees. The main result is as follows:

Theorem 5.9 (Degree sequence of the Erdős-Rényi random graph)**.** *Fix $\lambda > 0$. Then, for every ε_n such that $\sqrt{n}\varepsilon_n \to \infty$,*

$$\mathbb{P}_\lambda\big(\max_k |p_k^{(n)} - p_k| \geq \varepsilon_n\big) \to 0. \qquad (5.3.3)$$

Proof. We note that

$$\mathbb{E}_\lambda[P_k^{(n)}] = \mathbb{P}_\lambda(D_1 = k) = \binom{n-1}{k}\Big(\frac{\lambda}{n}\Big)^k\Big(1-\frac{\lambda}{n}\Big)^{n-k-1}. \qquad (5.3.4)$$

Furthermore,

$$\sum_{k=0}^{\infty}\Big|p_k - \binom{n-1}{k}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k-1}\Big| = \sum_{k=0}^{\infty}\big|\mathbb{P}(X^* = k) - \mathbb{P}(X_n = k)\big|, \qquad (5.3.5)$$

where X^* is a Poisson random variable with mean λ, and X_n is a binomial random variable with parameters $n-1$ and $p = \lambda/n$. We will use a coupling argument to bound this difference. Indeed, we let X denote a binomial random variable with parameters n and

$p = \lambda/n$. Since we can couple X and X_n such that the probability that they are different is precisely equal to $p = \lambda/n$, we obtain

$$\sum_{k=0}^{\infty} \big|\mathbb{P}(X_n = k) - \mathbb{P}(X = k)\big| \leq \frac{\lambda}{n}. \tag{5.3.6}$$

Therefore, for all $k \geq 0$,

$$\sum_{k=0}^{\infty} \big|\mathbb{P}(X^* = k) - \mathbb{P}(X_n = k)\big| \leq \frac{\lambda}{n} + \mathbb{P}(X^* \neq X) \leq \frac{\lambda + \lambda^2}{n}, \tag{5.3.7}$$

where we have also used Theorem 2.9. Since $\frac{\lambda+\lambda^2}{n} \leq \frac{\varepsilon_n}{2}$, we have just shown that $\sum_{k=0}^{\infty} |p_k - \mathbb{E}_\lambda[P_k^{(n)}]| \leq \varepsilon_n/2$ for n sufficiently large. Thus, it suffices to prove that

$$\mathbb{P}_\lambda\Big(\sum_k |P_k^{(n)} - \mathbb{E}_\lambda[P_k^{(n)}]| \geq \frac{\varepsilon_n}{2}\Big) = o(1). \tag{5.3.8}$$

For this, we use Boole's inequality to bound

$$\mathbb{P}_\lambda\Big(\max_k |P_k^{(n)} - \mathbb{E}_\lambda[P_k^{(n)}]| \geq \frac{\varepsilon_n}{2}\Big) \leq \sum_{k=1}^{\infty} \mathbb{P}_\lambda\big(|P_k^{(n)} - \mathbb{E}_\lambda[P_k^{(n)}]| \geq \frac{\varepsilon_n}{2}\big). \tag{5.3.9}$$

By the Chebychev inequality (Theorem 2.15),

$$\mathbb{P}_\lambda\big(|P_k^{(n)} - \mathbb{E}_\lambda[P_k^{(n)}]| \geq \frac{\varepsilon_n}{2}\big) \leq 4\varepsilon_n^{-2}\mathrm{Var}_\lambda(P_k^{(n)}). \tag{5.3.10}$$

We then note that

$$\begin{aligned}\mathrm{Var}_\lambda(P_k^{(n)}) = \frac{1}{n}\Big[\mathbb{P}_\lambda(D_1 = k) - \mathbb{P}_\lambda(D_1 = k)^2\Big] \\ + \frac{n-1}{n}\Big[\mathbb{P}_\lambda(D_1 = D_2 = k) - \mathbb{P}_\lambda(D_1 = k)^2\Big].\end{aligned} \tag{5.3.11}$$

We now use a coupling argument. We let X_1, X_2 be two independent $\mathrm{Bin}(n-2, \lambda/n)$ random variables, and I_1, I_2 two independent Bernoulli random variables with success probability λ/n. Then, the law of (D_1, D_2) is the same as the one of $(X_1 + I_1, X_2 + I_1)$ while $(X_1 + I_1, X_2 + I_2)$ are two independent copies of the D_1. Then,

$$\mathbb{P}_\lambda(D_1 = D_2 = k) = \mathbb{P}_\lambda\Big((X_1 + I_1, X_2 + I_1) = (k, k)\Big), \tag{5.3.12}$$

$$\mathbb{P}_\lambda(D_1 = k)^2 = \mathbb{P}_\lambda\Big((X_1 + I_1, X_2 + I_2) = (k, k)\Big), \tag{5.3.13}$$

t quite correct?

so that

$$\mathbb{P}_\lambda(D_1 = D_2 = k) - \mathbb{P}_\lambda(D_1 = k)^2 \leq \mathbb{P}_\lambda\Big((X_1 + I_1, X_2 + I_1) = (k, k), (X_1 + I_1, X_2 + I_2) \neq (k, k)\Big). \tag{5.3.14}$$

When $(X_1 + I_1, X_2 + I_1) = (k, k)$, but $(X_1 + I_1, X_2 + I_2) \neq (k, k)$, we must have that $I_1 \neq I_2$. If $I_1 = 1$, then $I_2 = 0$ and $X_2 = k$, while, if $I_1 = 0$, then $I_2 = 1$ and $X_1 = k$. Therefore, since X_1 and X_2 have the same distribution,

$$\mathbb{P}_\lambda(D_1 = D_2 = k) - \mathbb{P}_\lambda(D_1 = k)^2 \leq \frac{2\lambda}{n}\mathbb{P}_\lambda(X_1 = k). \tag{5.3.15}$$

We conclude from (5.3.11) that

$$\mathrm{Var}_\lambda(P_k^{(n)}) \leq \frac{(2\lambda+1)}{n}\mathbb{P}_\lambda(X_1=k), \tag{5.3.16}$$

so that, by (5.3.9)–(5.3.10),

$$\begin{aligned}\mathbb{P}_\lambda\big(\max_k |P_k^{(n)} - \mathbb{E}_\lambda[P_k^{(n)}]| \geq \varepsilon_n/2\big) &\leq \frac{4(2\lambda+1)}{\varepsilon_n^2 n}\sum_{k=0}^{\infty}\mathbb{P}_\lambda(X_1=k)\\ &= \frac{4(2\lambda+1)}{\varepsilon_n^2 n} \to 0.\end{aligned} \tag{5.3.17}$$

This completes the proof of Theorem 5.9. □

In Chapter 6 below, we give an alternative proof of Theorem 5.9, allowing for weaker bounds on ε_n. In that proof, we use that the Erdős-Rényi random graph is a special case of the generalized random graph with equal weights. See Theorem 6.11 below.

5.4 Notes and discussion

Notes on Section 5.1. We list some more recent results. In [135], a point process description is given of the sizes and number of components of size $\varepsilon n^{2/3}$. In [193], an explicit, yet involved, description is given for the distribution of the limit of $|\mathcal{C}_{\max}|n^{-2/3}$. The proof makes use of generating functions, and the relation between the largest connected component and the number of labeled graphs with a given complexity l. Here, the complexity of a graph is its number of edges minus its number of vertices. Relations between Erdős-Rényi random graphs and the problem of counting the number of labeled graphs has received considerable attention, see e.g. [43, 121, 163, 210, 225, 226] and the references therein. Consequences of the result by Pittel [193] are for example that the probability that $|\mathcal{C}_{\max}|n^{-2/3}$ exceeds a for large a decays as $\mathrm{e}^{-a^3/8}$ (in fact, the asymptotics are much stronger than this!), and for very small $a>0$, the probability that $|\mathcal{C}_{\max}|n^{-2/3}$ is smaller than a decays as $\mathrm{e}^{-ca^{-3/2}}$ for some explicit constant $c>0$. The bound on the upper tails of $|\mathcal{C}_{\max}|n^{-2/3}$ is also proved in [174], and is valid for all n and a, with the help of relatively simple martingale arguments. In [174], the bound (5.1.11) is also explicitly proved.

The relation between the Erdős-Rényi random graph and coalescing processes can also be found in [30, Section 5.2] and the references therein. In fact, $\mathrm{ER}_n(p)$ for the entire regime of $p\in[0,1]$ can be understood using coalescent processes, for which the multiplicative coalescent is most closely related to random graphs.

Notes on Section 5.2. Connectivity of the Erdős-Rényi random graph was investigated in the early papers on the subject. In [96], versions of Theorems 5.5–5.8 were proved for $\mathrm{ER}(n,M)$. Bollobás gives two separate proofs in [44, Pages 164-165].

Notes on Section 5.3. The degrees of Erdős-Rényi random graphs have attracted considerable attention. In particular, when ordering the degrees by size as $d_1\geq d_2\geq\cdots\geq d_n$, various properties have been shown, such as the fact that there is, with high probability, a *unique* vertex with degree d_1 [100]. See [41] or [44] for more details. The result on the degree sequence proved here is a weak consequence of the result in [127, Theorem 4.1], where even asymptotic normality was shown for the number of vertices with degree k, for all k simultaneously.

Intermezzo: Back to real networks I...

Theorem 5.9 shows that the degree sequence of the Erdős-Rényi random graph is close to a Poisson distribution with parameter λ. A Poisson distribution has thin tails, for example, its moment generating function is always finite. As a result, the Erdős-Rényi random graph cannot be used to model real networks where power law degree sequences are observed. Therefore, several related models have been proposed. In this intermezzo, we shall discuss three of them.

The first model is the so-called *generalized random graph* (GRG), and was first introduced in [58]. In this model, each vertex $i \in \{1, \ldots, n\}$ receives a *weight* W_i. Given the weights, edges are present independently, but the occupation probabilities for different edges are not identical, but moderated by the weights of the vertices. Naturally, this can be done in several different ways. The most general version is presented in [47], which we explain in detail in Chapter 9. In the generalized random graph, the edge probability of the edge between vertex i and j (conditionally on the weights $\{W_i\}_{i=1}^n$) is equal to

$$p_{ij} = \frac{W_i W_j}{L_n + W_i W_j}, \tag{I.1}$$

where the random variables $\{W_i\}_{i=1}^n$ are the *weights of the vertices,* and L_n is the total weight of all vertices given by

$$L_n = \sum_{i=1}^n W_i. \tag{I.2}$$

We shall assume that the weights $\{W_i\}_{i=1}^n$ are independent and identically distributed.

The second model is the *configuration model*, in which the degrees of the vertices are fixed. Indeed, we write D_i for the degree of vertex i, and let, similarly to (I.2), $L_n = \sum_{i=1}^n D_i$ denote the total degree. We assume that L_n is even. We will make a graph where vertex i has degree D_i. For this, we think of each vertex having D_i stubs attached to it. Two stubs can be connected to each other to form an edge. The configuration model is the model where all stubs are connected in a uniform fashion, i.e., where the stubs are uniformly matched.

The third model is the so-called *preferential attachment model*, in which the *growth of the random graph* is modeled by adding edges to the already existing graph in such a way that vertices with large degree are more likely to be connected to the newly added edges. See Chapter 8 for details.

All these models have in common that the degree sequence converges to some limiting distribution which can have various shapes, particularly including power laws. For the generalized random graph and the configuration model, this is proved in Chapter 6 and Chapter 7 respectively. For the preferential attachment models, we will defer this proof to Chapter 8. In Chapters 6–8, we shall focus on properties of the degree sequence of the random graphs involved. We shall study further properties, namely, the connected components and distances in these models, in Chapters 9–11, respectively.

In Chapters 6–8 we shall be interested in the properties of the *degree sequence of a graph*. A natural question is which sequences of numbers can occur as the degree sequence of a simple graph. A sequence $\{d_1, d_2, \ldots, d_n\}$ with $d_1 \leq d_2 \leq \cdots \leq d_n$ is called *graphic* if it is the degree sequence of a simple graph. Thus, the question is which degree sequences are graphic? Erdős and Gallai [95] proved that a degree sequence $\{d_1, d_2, \ldots, d_n\}$ is graphic if and only if $\sum_{i=1}^n d_i$ is even and

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min(k, d_i), \tag{I.3}$$

for each integer $k \leq n-1$. The fact that the total degree of a graph needs to be even is fairly obvious:

Exercise 5.3 (Handshake lemma)**.** *Show that for every graph, and d_j the degree of vertex j we have that $\sum_{j=1}^{n} d_j$ is even.*

The necessity of (I.3) is relatively easy to see. The left side of (I.3) is the degree of the first k vertices. The first term on the right-hand side of (I.3) is the twice the maximal number of edges between the vertices in $\{1, \ldots, k\}$. The second term is a bound on the total degree of the vertices $\{1, \ldots, k\}$ coming from edges that connect to vertices in $\{k+1, \ldots, n\}$. The sufficiency is harder to see, see [64] for a simple proof of this fact, and [201] for seven different proofs. Arratia and Liggett [15] investigate the asymptotic probability that an i.i.d. sequence of n integer random variables is graphical, the result being in many cases equal to 0 or 1/2, at least when $\mathbb{P}(D \text{ even}) \neq 1$. The limit is equal to 0 when $\lim_{n\to\infty} n\mathbb{P}(D_i \geq n) = \infty$ and 1/2 when $\lim_{n\to\infty} n\mathbb{P}(D_i \geq n) = 0$. Interestingly, when $\lim_{n\to\infty} n\mathbb{P}(D_i \geq n) = c$ for some constant $c > 0$, then the set of limit points of the probability that $\{D_1, \ldots, D_n\}$ is graphical is a subset of $(0, 1/2)$. The proof is by verifying that (I.3) holds.

Chapter 6
Inhomogeneous random graphs

In this chapter, we discuss *inhomogeneous random graphs*, in which the equal edge probabilities of the Erdős-Rényi random graph are replaced by edge occupation statuses that are independent, and are moderated by certain vertex weights. These weights can be taken to be *deterministic* or *random*, and both options have been considered in the literature. An important example, on which we shall focus in this chapter, is the so-called *generalized random graph*. We shall see that this model gives rise to random graphs having a power-law degree sequence when the weights have a power law distribution. As such, this is one of the simplest adaption of the Erdős-Rényi random graph having a power-law degree sequence.

This chapter is organized as follows. In Section 6.1, we introduce the model. In Section 6.2, we investigate the degree of a fixed vertex in the generalized random graph, and in Section 6.3, we investigate the degree sequence of the generalized random graph. In Section 6.4, we study the generalized random graph with i.i.d. vertex weights. In Section 6.5 we show that the generalized random graph, conditioned on its degrees, is a uniform random graph with these degrees. In Section 6.6, we study when two inhomogeneous random graphs are asymptotically equivalent, meaning that they have the same asymptotic probabilities. Finally, in Section 6.7, we introduce several more models of inhomogeneous random graphs similar to the generalized random graph that have been studied in the literature, such as the so-called Chung-Lu or random graph with prescribed expected degrees and the Norros-Reittu or Poisson graph process model. We close this chapter with notes and discussion in Section 6.8.

6.1 Introduction of the model

In the generalized random graph, each vertex has a weight associated to it. Edges are present independently given these weights, but the occupation probabilities for edges are *not* identical, but are rather moderated by the vertex weights. We start with a simple example.

Example 6.1 (Population of two types)**.** Suppose that we have a complex network in which two distinct types of vertices are present. The first type has on average m_1 neighbors, the second type m_2, where $m_1 \neq m_2$. How can we construct a random graph in which such heterogeneity can be incorporated?

Example 6.2 (Power-law or related degrees)**.** In Chapter 1, many examples of real-world networks were given where the degrees are quite variable, including hubs having quite high degrees. How can we construct a random graph with power-law degrees?

Example 6.1 deals with a population having only two types. In many situations, there can be many different types. An important example is the case where the degrees obey a power law as in Example 6.2, in which case any finite number of types is insufficient. We model this inhomogeneity by adding *vertex weights*. Vertices with higher weights are more likely to have many neighbors than vertices with small weights. Vertices with extremely high weights could act as the hubs observed in many real-world networks.

In the generalized random graph model, the edge probability of the edge between vertices i and j is equal to

$$p_{ij} = p_{ij}^{(\mathrm{GRG})} = \frac{w_i w_j}{\ell_n + w_i w_j}, \tag{6.1.1}$$

where $\boldsymbol{w} = (w_i)_{i\in[n]}$ are the *weights of the vertices,* and ℓ_n is the total weight of all vertices given by

$$\ell_n = \sum_{i\in[n]} w_i. \tag{6.1.2}$$

We denote the resulting graph by $\mathrm{GRG}_n(\boldsymbol{w})$. A special case of the generalized random graph is when we take $w_i \equiv \frac{n\lambda}{n-\lambda}$, in which case $p_{ij} = \lambda/n$ for all $i, j \in [n]$, so that we retrieve the Erdős-Rényi random graph $\mathrm{ER}_n(\lambda/n)$.

Exercise 6.1 (The Erdős-Rényi random graph)**.** *Prove that $p_{ij} = \lambda/n$ when $w_i = n\lambda/(n-\lambda)$ for all $i \in [n]$.*

Without loss of generality, we assume that $w_i > 0$. Note that when, for a particular $i \in [n]$, $w_i = 0$, then vertex i will be isolated with probability 1, and, therefore, we can omit i from the graph. The vertex weights moderate the inhomogeneity in the random graph, vertices with high weights have higher edge occupation probabilities than vertices with low weights. Therefore, by choosing the weights in an appropriate way, this suggests that we can create graphs with flexible degree sequences. We investigate the degree structure in more detail in this chapter.

Example 6.3 (Population of two types (Cont.))**.** *In Example 6.1, we let the vertices of type 1 have weight m_1, and the vertices of type 2 have weight m_2. Let n_1 and n_2 denote the number of vertices of weight 1 and 2, respectively. Then, we compute that*

$$\ell_n = n_1 m_1 + n_2 m_2. \tag{6.1.3}$$

Further, the probability that a vertex of type 1 is connected to another vertex of type 1 is equal to $m_1^2/(\ell_n + m_1^2)$, while the probability that it is connected to a vertex of type 2 equals $m_1 m_2/(\ell_n + m_1 m_2)$. Therefore, the expected degree of a vertex of type 1 is equal to

$$n_1 \frac{m_1^2}{\ell_n + m_1^2} + n_2 \frac{m_1 m_2}{\ell_n + m_1 m_2} = m_1 [\frac{n_1 m_1}{\ell_n + m_1^2} + \frac{n_2 m_2}{\ell_n + m_1 m_2}] \approx m_1, \tag{6.1.4}$$

whenever $m_1^2 + m_2^2 = o(\ell_n)$. Thus, our graph is such that the first type has on average approximately m_1 neighbors, the second type m_2.

Naturally, the topology of the generalized random graph sensitively depends upon the choice of the vertex weights $\boldsymbol{w} = (w_i)_{i\in[n]}$. These vertex weights can be rather general. In order to describe the empirical proporties of the weights, we define their *empirical distribution function* to be

$$F_n(x) = \frac{1}{n} \sum_{i\in[n]} \mathbb{1}_{\{w_i \le x\}}, \qquad x \ge 0. \tag{6.1.5}$$

We can interpret F_n as the distribution of the weight of a uniformly chosen vertex in $[n]$:

Exercise 6.2 (The weight of a uniformly chosen vertex)**.** *Let V be a uniformly chosen vertex in $[n]$. Show that the weight w_V of V has distribution function F_n.*

We denote the weight of a uniformly chosen vertex in $[n]$ by $W_n = w_V$, so that, by Exercise 6.2, W_n has distribution function F_n. We often assume that the vertex weights satisfy the following *regularity conditions:*

Condition 6.4 (Regularity conditions for vertex weights)**.**
(a) Weak convergence of vertex weight.
There exists a distribution function F such that

$$W_n \xrightarrow{d} W, \tag{6.1.6}$$

where W_n and W have distribution functions F_n and F, respectively.
Equivalently, for any x for which $x \mapsto F(x)$ is continuous,

$$\lim_{n\to\infty} F_n(x) = F(x). \tag{6.1.7}$$

(b) Convergence of average vertex weight.

$$\lim_{n\to\infty} \mathbb{E}[W_n] = \mathbb{E}[W], \tag{6.1.8}$$

where W_n and W have distribution functions F_n and F, respectively. Further, we assume that $\mathbb{E}[W] > 0$.
(c) Convergence of second moment vertex weight.

$$\lim_{n\to\infty} \mathbb{E}[W_n^2] = \mathbb{E}[W^2]. \tag{6.1.9}$$

Condition 6.4(a) guarantees that the weight of a 'typical' vertex is close to a random variable W. Condition 6.4(b) implies that the average degree in $\mathrm{GRG}_n(\boldsymbol{w})$ converges, while Condition 6.4(c) ensures also the convergence of the second moment of the degree. In most of our results, we assume Condition 6.4(a)-(b), in some we also need Condition 6.4(c).

Exercise 6.3 (Bound on the maximal weight assuming Condition 6.4(b)-(c))**.** *Prove that Condition 6.4(b) implies that $\max_{i\in[n]} w_i = o(n)$, while Condition 6.4(c) implies that $\max_{i\in[n]} w_i = o(\sqrt{n})$.*

We now discuss two key examples of choices of vertex weights.

Key example of generalized random graph with deterministic weights. Let F be a distribution function for which $F(0) = 0$ and fix

$$w_i = [1-F]^{-1}(i/n), \tag{6.1.10}$$

where $[1-F]^{-1}$ is the generalized inverse function of $1-F$ defined, for $u \in (0,1)$, by

$$[1-F]^{-1}(u) = \inf\{s\colon [1-F](s) \le u\}. \tag{6.1.11}$$

By convention, we set $[1-F]^{-1}(1) = 0$. Here the definition of $[1-F]^{-1}$ is chosen such that

$$[1-F]^{-1}(1-u) = F^{-1}(u) = \inf\{x : F(x) \ge u\}. \tag{6.1.12}$$

We shall often make use of (6.1.12), in particular since it implies that $[1-F]^{-1}(U)$ has distribution function F when U is uniform on $(0,1)$. For this choice,

$$\begin{aligned} F_n(x) &= \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{w_i\le x\}} = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{[1-F]^{-1}(i/n)\le x\}} = \frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}_{\{[1-F]^{-1}(1-\frac{j}{n})\le x\}} \\ &= \frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}_{\{F^{-1}(\frac{j}{n})\le x\}} = \frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}_{\{\frac{j}{n}\le F(x)\}} = \frac{1}{n}\big(\lfloor nF(x)\rfloor + 1\big) \wedge 1, \end{aligned} \tag{6.1.13}$$

where we write $j = n - i$ in the third equality and use (6.1.12) in the fourth equality.

Exercise 6.4 (Condition 6.4(a)). *Prove that Condition 6.4(a) holds for* $(w_i)_{i\in[n]}$ *as in* (6.1.10).

Note that by (6.1.13), we obtain $F_n(x) \geq F(x)$ for every $x \geq 0$, which shows that W_n is stochastically dominated by W. In particular, this implies that for *increasing* functions $x \mapsto h(x)$,

$$\frac{1}{n}\sum_{j=1}^{n} h(w_j) \leq \mathbb{E}[h(W)]. \tag{6.1.14}$$

We now study some properties of the weights in (6.1.10):

Exercise 6.5 (Moments of $\boldsymbol{w}$ and F [102]). *Prove that* $u \mapsto [1-F]^{-1}(u)$ *is non-increasing, and conclude that, for every non-decreasing function* $x \mapsto h(x)$ *and for* w_i *as in* (6.1.10),

$$\frac{1}{n}\sum_{i\in[n]} h(w_i) \leq \mathbb{E}[h(W)], \tag{6.1.15}$$

where W *is a random variable with distribution function* F.

Exercise 6.6 (Moments of $\boldsymbol{w}$ and F [102] (Cont.)). *Set* $\alpha > 0$, *assume that* $\mathbb{E}[W^\alpha] < \infty$ *where* W *is a random variable with distribution function* F. *Use Lebesgue's dominated convergence theorem (Theorem A.9) to prove that for* w_i *as in* (6.1.10),

$$\frac{1}{n}\sum_{i\in[n]} w_i^\alpha \to \mathbb{E}[W^\alpha]. \tag{6.1.16}$$

Conclude that Condition 6.4(a) holds when $\mathbb{E}[W] < \infty$, *and Condition 6.4(b) when* $\mathbb{E}[W^2] < \infty$.

An example of the generalized random graph arises when we take, for some $a \geq 0$ and $\tau > 1$,

$$F(x) = \begin{cases} 0 & \text{for } x \leq a, \\ 1-(a/x)^{\tau-1} & \text{for } x > a, \end{cases} \tag{6.1.17}$$

for which

$$[1-F]^{-1}(u) = au^{-1/(\tau-1)}, \tag{6.1.18}$$

so that

$$w_i = a\big(i/n\big)^{-1/(\tau-1)}. \tag{6.1.19}$$

Exercise 6.7 (Bounds on $\boldsymbol{w}$). *Fix* $(w_i)_{i\in[n]}$ *as in* (6.1.10). *Prove that when*

$$1-F(x) \leq cx^{-(\tau-1)}, \tag{6.1.20}$$

then there exists a $c' > 0$ *such that* $w_j \leq w_1 \leq c'n^{\frac{1}{\tau-1}}$ *for all* $j \in [n]$, *and all large enough* n.

The generalized random graph with i.i.d. weights. GRG can be studied both with deterministic weights as well as with independent and identically distributed (i.i.d.) weights. The GRG with deterministic weights is denoted by $\mathrm{GRG}_n(\boldsymbol{w})$, the GRG with i.i.d. weights by $\mathrm{GRG}_n(\boldsymbol{W})$. Since we often deal with ratios of the form $W_iW_j/(\sum_{k\in[n]} W_k)$, we shall assume that $\mathbb{P}(W=0)=0$ to avoid situations where all weights are zero.

Both models have their own merits (see Section 6.8 for more details). The great advantage of independent and identically distributed weights is that the vertices in the resulting graph are, in distribution, the same. More precisely, the vertices are completely *exchangeable*, like in the Erdős-Rényi random graph $\mathrm{ER}_n(p)$. Unfortunately, when we take the weights to be i.i.d., then in the resulting graph the edges are no longer independent (despite the fact that they are *conditionally* independent given the weights):

Exercise 6.8 (Dependence edges in $\mathrm{GRG}_n(\boldsymbol{W})$)**.** *Let $(W_i)_{i\in[n]}$ be an i.i.d. sequence of weights for which $\mathbb{E}[W^2]<\infty$. Assume further that there exists $\varepsilon>0$ such that $\mathbb{P}(W\le\varepsilon)=0$. Prove that*

$$n\mathbb{P}(12 \text{ present}) = n\mathbb{P}(23 \text{ present}) \to \mathbb{E}[W], \tag{6.1.21}$$

while

$$n^2\mathbb{P}(12 \text{ and } 23 \text{ present}) \to \mathbb{E}[W^2]. \tag{6.1.22}$$

Conclude that the status of different edges that share a vertex are dependent *whenever* $\mathrm{Var}(W)>0$.

When the weights are *random*, we need to specify the kind of convergence in Condition 6.4, and we shall assume that the limits hold *in probability*. We now investigate the conditions under which Condition 6.4(a)-(c) hold. The empirical distribution function F_n of the weights is given by

$$F_n(x) = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{W_i\le x\}}. \tag{6.1.23}$$

When the weights are independently and identically distributed with distribution function F, then it is well-known that this empirical distribution function is close to F (this is the Glivenko-Cantelli Theorem). Therefore, Condition 6.4(a) holds.

We close this introductory section by investigating the total number of edges $E(\mathrm{GRG}_n(\boldsymbol{w}))$ in the GRG:

Theorem 6.5 (Total number of edges in $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *Assume that Condition 6.4(a)-(b) hold. Then,*

$$\frac{1}{n}E(\mathrm{GRG}_n(\boldsymbol{w})) \xrightarrow{\mathbb{P}} \tfrac{1}{2}\mathbb{E}[W]. \tag{6.1.24}$$

Proof. We apply a second moment method. For this, we note

$$\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))] = \tfrac{1}{2}\sum_{i,j\in[n]:\, i\neq j} p_{ij} = \sum_{i,j\in[n]:\, i\neq j} \frac{w_iw_j}{\ell_n + w_iw_j}. \tag{6.1.25}$$

We start by bounding this from above by

$$\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))] \le \tfrac{1}{2}\sum_{i,j\in[n]} \frac{w_iw_j}{\ell_n} = \ell_n, \tag{6.1.26}$$

which proves an upper bound on $\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))]$. Further, for any sequence $a_n\to\infty$ and since $x\mapsto x/(\ell_n+x)$ is increasing,

$$\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))] \ge \tfrac{1}{2}\sum_{i,j\in[n]:\, i\neq j} \frac{(w_i\wedge a_n)(w_j\wedge a_n)}{\ell_n + (w_i\wedge a_n)(w_j\wedge a_n)}. \tag{6.1.27}$$

Therefore, with $\ell_n(a_n) = \sum_{i\in[n]}(w_i \wedge a_n)$,

$$\begin{aligned}\ell_n(a_n)^2/\ell_n - 2\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))] &\le \sum_{i\in[n]} \frac{(w_i\wedge a_n)^2}{\ell_n + (w_i\wedge a_n)^2} \\ &\quad + \sum_{i,j\in[n]} (w_i\wedge a_n)(w_j\wedge a_n)\big[\frac{1}{\ell_n} - \frac{1}{\ell_n + (w_i\wedge a_n)(w_j\wedge a_n)}\big] \\ &\le \sum_{i\in[n]} \frac{(w_i\wedge a_n)^2}{\ell_n + (w_i\wedge a_n)^2} + \sum_{i,j\in[n]} \frac{(w_i\wedge a_n)^2(w_j\wedge a_n)^2}{\ell_n(\ell_n + (w_i\wedge a_n)(w_j\wedge a_n))} \\ &\le \sum_{i\in[n]} \frac{(w_i\wedge a_n)^2}{\ell_n}\Big(1 + \sum_{i\in[n]} \frac{(w_i\wedge a_n)^2}{\ell_n}\Big).\end{aligned} \tag{6.1.28}$$

Since

$$\sum_{i\in[n]} \frac{(w_i\wedge a_n)^2}{\ell_n} \le a_n, \tag{6.1.29}$$

the right-hand side of (6.1.28) is $o(n)$ when we choose $a_n = o(\sqrt{n})$, so that it suffices to prove that $\ell_n(a_n)^2/(n\ell_n) \to \mathbb{E}[W]$ and $\ell_n/n \to \mathbb{E}[W]$.

By Condition 6.4(b), $\ell_n/n = \mathbb{E}[W_n] \to \mathbb{E}[W]$, and also $\ell_n(a_n)/n = \mathbb{E}[(W_n\wedge a_n)] \to \mathbb{E}[W]$ by Condition 6.4(a)-(b). We conclude that

$$\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))]/n \to \tfrac{1}{2}\mathbb{E}[W]. \tag{6.1.30}$$

We continue by bounding $\mathrm{Var}(E(\mathrm{GRG}_n(\boldsymbol{w})))$, for which we use that the edge statuses are independent to obtain

$$\begin{aligned}\mathrm{Var}(E(\mathrm{GRG}_n(\boldsymbol{w}))) &= \tfrac{1}{2}\sum_{i,j\in[n]:\ i\neq j} \mathrm{Var}(I_{ij}) = \tfrac{1}{2}\sum_{i,j\in[n]:\ i\neq j} p_{ij}(1-p_{ij}) \\ &\le \tfrac{1}{2}\sum_{i,j\in[n]:\ i\neq j} p_{ij} = \mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))].\end{aligned} \tag{6.1.31}$$

As a result, $\mathrm{Var}(E(\mathrm{GRG}_n(\boldsymbol{w}))) \le \mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))]$, which is $o(\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))]^2)$, by (6.1.30) and the fact that $\mathbb{E}[W] > 0$ by Condition 6.4(a). We conclude that

$$E(\mathrm{GRG}_n(\boldsymbol{w}))/\mathbb{E}[E(\mathrm{GRG}_n(\boldsymbol{w}))] \xrightarrow{\mathbb{P}} 1. \tag{6.1.32}$$

□

6.2 Degrees in the generalized random graph

In this section, we study the degrees of vertices in $\mathrm{GRG}_n(\boldsymbol{w})$. In order to state the main results, we start with some definitions. Given weights $\boldsymbol{w} = (w_i)_{i\in[n]}$, we let the probability that the edge ij is occupied be equal to p_{ij} in (6.1.1), and where we recall that $\ell_n = \sum_{i\in[n]} w_i$. We write $D_k = D_k^{(n)}$ for the degree of vertex k in $\mathrm{GRG}_n(\boldsymbol{w})$. Thus, D_k is given by

$$D_k = \sum_{j=1}^{n} X_{kj}, \tag{6.2.1}$$

where X_{kj} is the indicator that the edge kj is occupied. By convention, we set $X_{ij} = X_{ji}$. The main result concerning the degrees is as follows:

Theorem 6.6 (Degree of GRG with deterministic weights). *Assume that Condition 6.4(a)-(b) hold. Then,*

(a) there exists a coupling $(\hat{D}_k, \hat{Z}_k)$ of the degree D_k of vertex k and a Poisson random variable Z_k with parameter w_k, such that it satisfies

$$\mathbb{P}(\hat{D}_k \neq \hat{Z}_k) \leq \frac{w_k^2}{\ell_n}\Big(1 + 2\frac{\mathbb{E}[W_n^2]}{\mathbb{E}[W_n]}\Big). \tag{6.2.2}$$

In particular, D_k can be coupled to a Poisson random variable with parameter w_k.

(b) When p_{ij} given by (6.1.1) are all such that $\lim_{n\to\infty} p_{ij} = 0$, the degrees $D_1, \ldots, D_m$ of vertices $1, \ldots, m$ are asymptotically independent.

Before proving Theorem 6.6, we state a consequence for the degree sequence when the weights are given by (6.1.10). To be able to state this consequence, we need the following definition:

Definition 6.7 (Mixed Poisson distribution). *A random variable X has a mixed Poisson distribution with mixing distribution F when, for every $k \in \mathbb{N}$,*

$$\mathbb{P}(X = k) = \mathbb{E}[\mathrm{e}^{-W}\frac{W^k}{k!}], \tag{6.2.3}$$

where W is a random variable with distribution function F.

The next exercises investigate some properties of mixed Poisson random variables:

Not every random variable can be obtained as a mixed Poisson distribution (recall Definition 6.7). In the following exercises, aspects of mixed Poisson distributions are further investigated.

Exercise 6.9 (Not every random variable is mixed Poisson). *Give an example of a random variable that cannot be represented as a mixed Poisson distribution.*

Exercise 6.10 (Characteristic function of mixed Poisson distribution). *Let X have a mixed Poisson distribution with mixing distribution F and moment generating function M_W, i.e., for $t \in \mathbb{C}$,*

$$M_W(t) = \mathbb{E}[\mathrm{e}^{tW}], \tag{6.2.4}$$

where W has distribution function F. Show that the characteristic function of X is given by

$$\phi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = M_W(\mathrm{e}^{it} - 1). \tag{6.2.5}$$

Exercise 6.11 (Mean and variance mixed Poisson distribution). *Let X have a mixed Poisson distribution with mixing distribution F. Express the mean and variance of X into the moments of W, where W has distribution function F.*

Exercise 6.12 (Tail behavior mixed Poisson). *Suppose that there exist constants $0 < c_1 < c_2 < \infty$ such that*

$$c_1 x^{1-\tau} \leq 1 - F(x) \leq c_2 x^{1-\tau}. \tag{6.2.6}$$

Show that there exist $0 < c_1' < c_2' < \infty$ such that the distribution function G of a mixed Poisson distribution with mixing distribution F satisfies

$$c_1' x^{1-\tau} \leq 1 - G(x) \leq c_2' x^{1-\tau}. \tag{6.2.7}$$

By Theorem 6.6, the degree of vertex i is close to Poisson with parameter w_i. Thus, when we choose a vertex *uniformly at random*, and we denote the outcome by V, then the degree of that vertex is close to a Poisson distribution with *random* parameter $w_V = W_n$. Since $W_n \xrightarrow{d} W$ by Condition 6.4, this suggests the following result:

Corollary 6.8 (Degree of uniformly chosen vertex in GRG)**.** *Assume that Condition 6.4(a)-(b) hold. Then,*

(a) the degree of a uniformly chosen vertex converges in distribution to a mixed Poisson random variable with mixing distribution F;

(b) the degrees of m uniformly chosen vertices in $[n]$ are asymptotically independent.

We now prove Theorem 6.6 and Corollary 6.8:

Proof of Theorem 6.6. We make essential use of Theorem 2.9, in particular, the coupling of a sum of Bernoulli random variables with a Poisson random variable in (2.2.19). Throughout this proof, we shall omit the dependence on n of the weights, and abbreviate $w_i = w_i^{(n)}$. We recall that

$$D_k = \sum_{i\in[n]} X_{kj}, \tag{6.2.8}$$

where X_{kj} are independent Bernoulli random variables with success probabilities $p_{kj} = \frac{w_k w_j}{\ell_n + w_k w_j}$. By Theorem 2.9, there exists a Poisson random variable $\hat{Y}_k$ with parameter

$$\lambda_k = \sum_{j\neq k} \frac{w_k w_j}{\ell_n + w_k w_j}, \tag{6.2.9}$$

and a random variable $\hat{D}_k$ where $\hat{D}_k$ has the same distribution as D_k, such that

$$\mathbb{P}(\hat{D}_k \neq \hat{Y}_k) \leq \sum_{j\neq k} p_{kj}^2 = \sum_{j\neq k} \frac{w_k^2 w_j^2}{(\ell_n + w_k w_j)^2} \leq w_k^2 \sum_{j=1}^n \frac{w_j^2}{\ell_n^2}. \tag{6.2.10}$$

Thus, in order to prove the claim, it suffices to prove that we can, in turn, couple $\hat{Y}_k$ to a Poisson random variable $\hat{Z}_k$ with parameter w_k, such that

$$\mathbb{P}(\hat{Y}_k \neq \hat{Z}_k) \leq w_k^2 \sum_{j=1}^n \frac{w_j^2}{\ell_n^2} + \frac{w_k^2}{\ell_n^2}. \tag{6.2.11}$$

For this, we note that

$$\lambda_k \leq \sum_{\neq k} \frac{w_k w_j}{\ell_n} \leq \frac{w_k}{\ell_n} \sum_{j=1}^k w_j = w_k. \tag{6.2.12}$$

Let $\varepsilon_k = w_k - \lambda_k \geq 0$. Then, we let $\hat{V}_k \sim \mathrm{Poi}(\varepsilon_k)$ be independent of $\hat{Y}_k$, and write $\hat{Z}_k = \hat{Y}_k + \hat{V}_k$, so that

$$\mathbb{P}(\hat{Y}_k \neq \hat{Z}_k) = \mathbb{P}(\hat{V}_k \neq 0) = \mathbb{P}(\hat{V}_k \geq 1) \leq \mathbb{E}[\hat{V}_k] = \varepsilon_k. \tag{6.2.13}$$

To bound ε_k, we note that

$$\begin{aligned}\varepsilon_k &= w_k - \sum_{j\neq k} \frac{w_k w_j}{\ell_n + w_k w_j} = \sum_{j=1}^n w_k w_j \Big(\frac{1}{\ell_n} - \frac{1}{\ell_n + w_k w_j}\Big) + \frac{w_k^2}{\ell_n + w_k^2} \\ &= \sum_{j=1}^n \frac{w_j^2 w_k^2}{\ell_n(\ell_n + w_k w_j)} + \frac{w_k^2}{\ell_n + w_k^2} \leq \frac{w_k^2}{\ell_n} + \sum_{j=1}^n \frac{w_j^2 w_k^2}{\ell_n^2} = w_k^2\Big(\frac{1}{\ell_n} + \sum_{j=1}^n \frac{w_j^2}{\ell_n^2}\Big).\end{aligned} \tag{6.2.14}$$

We conclude that

$$\mathbb{P}(\hat{D}_k \neq \hat{Z}_k) \leq \mathbb{P}(\hat{D}_k \neq \hat{Y}_k) + \mathbb{P}(\hat{Y}_k \neq \hat{Z}_k) \leq 2w_k^2 \sum_{j=1}^{n} \frac{w_j^2}{\ell_n^2} + \frac{w_k^2}{\ell_n}, \tag{6.2.15}$$

as required. This proves Theorem 6.6(a).

To prove Theorem 6.6(b), it suffices to prove that we can couple $(D_i)_{i\in[m]}$ to an *independent* vector $(\hat{D}_i)_{i\in[m]}$ such that

$$\mathbb{P}\Big((D_i)_{i\in[m]} \neq (\hat{D}_i)_{i\in[m]}\Big) = o(1). \tag{6.2.16}$$

To this end, we recall that X_{ij} denotes the indicator that the edge ij is occupied. The random variables $(X_{ij})_{1\leq i<j\leq n}$ are independent Bernoulli random variables with parameters $(p_{ij})_{1\leq i<j\leq n}$ given in (6.1.1). We let $(X'_{ij})_{1\leq i<j\leq n}$ denote an *independent* copy of $(X_{ij})_{1\leq i<j\leq n}$, and let, for $i = 1, \ldots, n$,

$$\hat{D}_i = \sum_{j<i} X'_{ij} + \sum_{j=i+1}^{n} X_{ij}. \tag{6.2.17}$$

Then, we observe the following: (1) Since $(X'_{ij})_{1\leq i<j\leq n}$ is an independent copy of $(X_{ij})_{1\leq i<j\leq n}$, the distribution of $\hat{D}_i$ is equal to the one of D_i, for every $i = 1, \ldots, n$. (2) Set $i < j$. While D_i and D_j are *dependent* since they both contain $X_{ij} = X_{ji}$, $\hat{D}_i$ contains X_{ij}, while D'_j contains $X'_{ji} = X'_{ij}$, which is an independent copy of X_{ij}. We conclude that $(\hat{D}_i)_{i\in[m]}$ are sums of *independent* Bernoulli random variables, and, therefore, are *independent*. (3) Finally, $(D_i)_{i\in[m]} \neq (\hat{D}_i)_{i\in[m]}$ precisely when there exists at least one edge ij with $i, j \in [m]$ such that $X_{ij} \neq X'_{ij}$. Since X_{ij} and X'_{ij} are Bernoulli random variables, $X_{ij} \neq X'_{ij}$ implies that either $X_{ij} = 0, X'_{ij} = 1$ or $X_{ij} = 1, X'_{ij} = 0$. Thus, by Boole's inequality, we obtain that

$$\mathbb{P}\Big((D_i)_{i\in[m]} \neq (\hat{D}_i)_{i\in[m]}\Big) \leq 2 \sum_{i,j=1}^{m} \mathbb{P}(X_{ij} = 1) = 2 \sum_{i,j=1}^{m} p_{ij}. \tag{6.2.18}$$

By assumption, $\lim_{n\to\infty} p_{ij} = 0$, so that (6.2.16) holds for every $m \geq 2$ fixed. This proves Theorem 6.6(b). □

Exercise 6.13 (Independence of a growing number of degrees for bounded weights)**.** *Assume that the conditions in Corollary 6.8 hold, and further suppose that there exists a $\varepsilon > 0$ such that $\varepsilon \leq w_i \leq \varepsilon^{-1}$ for every i, so that the weights are uniformly bounded from above and below. Then, prove that we can couple $(D_i)_{i\in[m]}$ to an independent vector $(\hat{D}_i)_{i\in[m]}$ such that* (6.2.16) *holds whenever $m = o(\sqrt{n})$. As a result, even the degrees of a* growing *number of vertices can be coupled to independent degrees.*

Proof of Corollary 6.8. By (6.2.2) together with the fact that $\max_{i\in[n]} w_i = o(n)$ by Exercise 6.3 we have that the degree of vertex k is close to a Poisson random variable with parameter w_k. Thus, the degree of a uniformly chosen vertex in $[n]$ is close in distribution to a Poisson random variable with parameter w_V, where V is a uniform random variable in $[n]$. This is a mixed Poisson distribution with mixing distribution equal to w_V.

Since a mixed Poisson random variable converges to a limiting mixed Poisson random variable whenever the mixing distribution converges in distribution, it suffices to show that the weight $W_n = w_V$ of a uniform vertex has a limiting distribution given by F. This follows from Condition 6.4(a), whose validity follows by (6.1.13).

The proof of part (b) is a minor adaptation of the proof of Theorem 6.6(b). We shall only discuss the asymptotic independence. Let $(V_i)_{i\in[m]}$ be independent uniform random

variables. Then, the dependence between the degrees of the vertices $(V_i)_{i\in[m]}$ arises only through the edges between the vertices $(V_i)_{i\in[m]}$. Now, the expected number of occupied edges between the vertices $(V_i)_{i\in[m]}$, conditionally on $(V_i)_{i\in[m]}$, is bounded by

$$\sum_{i,j=1}^{m} \frac{w_{V_i}w_{V_j}}{\ell_n + w_{V_i}w_{V_j}} \le \sum_{i,j=1}^{m} \frac{w_{V_i}w_{V_j}}{\ell_n} = \frac{1}{\ell_n}\Big(\sum_{i=1}^{m} w_{V_i}\Big)^2. \tag{6.2.19}$$

The random variables $(w_{V_i})_{i\in[m]}$ are i.i.d., so that the expected number of occupied edges between m uniformly chosen vertices is equal to

$$\frac{1}{\ell_n}\mathbb{E}\Big[\Big(\sum_{i=1}^{m} w_{V_i}\Big)^2\Big] = \frac{m}{\ell_n}\mathrm{Var}(w_{V_1}) + \frac{m(m-1)}{\ell_n}\mathbb{E}[w_{V_1}]^2. \tag{6.2.20}$$

We can bound

$$\mathrm{Var}(w_{V_1}) \le \mathbb{E}[w_{V_1}^2] \le (\max_{i\in[n]} w_i)\mathbb{E}[w_{V_1}] = o(n), \tag{6.2.21}$$

by Exercise 6.3. Therefore, the expected number of edges between the vertices $(V_i)_{i\in[m]}$ is $o(1)$, so that with high probability there are none. We conclude that we can couple the degrees of m uniform vertices to m *independent* mixed Poisson random variables with mixing distribution $w_V^{(n)}$. Since these random variables converge in distribution to independent mixed Poisson random variables with mixing distribution F, this completes the argument. □

6.3 Degree sequence of generalized random graph

Theorem 6.6 investigates the degree of a single vertex in the generalized random graph. In this section, we extend the result to the convergence of the empirical degree sequence. For $k \ge 0$, we let

$$P_k^{(n)} = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{D_i=k\}} \tag{6.3.1}$$

denote the degree sequence of $\mathrm{GRG}_n(\boldsymbol{w})$. Due to Theorem 6.6, one would expect that this degree sequence is close to a mixed Poisson distribution. We denote the probability mass function of such a mixed Poisson distribution by p_k, i.e., for $k \ge 0$,

$$p_k = \mathbb{E}\big[\mathrm{e}^{-W}\frac{W^k}{k!}\big]. \tag{6.3.2}$$

Theorem 6.9 shows that indeed the degree sequence $(P_k^{(n)})_{k\ge0}$ is close to the mixed Poisson distribution with probability mass function $(p_k)_{k\ge0}$ in (6.3.2):

Theorem 6.9 (Degree sequence of $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *Assume that Condition 6.4(a)-(b) hold. Then, for every $\varepsilon > 0$,*

$$\mathbb{P}\Big(\sum_{k=0}^{\infty} |P_k^{(n)} - p_k| \ge \varepsilon\Big) \to 0, \tag{6.3.3}$$

where $(p_k)_{k=0}^{\infty}$ is given by (6.3.2)*.*

Proof of Theorem 6.9. By Exercise 2.14 and the fact that $(p_k)_{k=0}^{\infty}$ is a probability mass function, we have that $\sum_{k=0}^{\infty}|P_k^{(n)} - p_k| = 2d_{\mathrm{TV}}(P^{(n)}, p) \to 0$ if and only if $\max_{k=0}^{\infty}|P_k^{(n)} -$

$p_k| \to 0$. Thus, we need to show that, for every $\varepsilon > 0$, $\mathbb{P}\big(\max_{k=0}^{\infty} |P_k^{(n)} - p_k| \geq \varepsilon\big)$ converges to 0. We use that

$$\mathbb{P}\big(\max_{k=0}^{\infty} |P_k^{(n)} - p_k| \geq \varepsilon\big) \leq \sum_{k=0}^{\infty} \mathbb{P}\big(|P_k^{(n)} - p_k| \geq \varepsilon\big). \tag{6.3.4}$$

Note that

$$\mathbb{E}[P_k^{(n)}] = \mathbb{P}(D_V = k), \tag{6.3.5}$$

and, by Corollary 6.8(a), we have that

$$\lim_{n\to\infty} \mathbb{P}(D_V = k) = p_k. \tag{6.3.6}$$

Also, it is not hard to see that the convergence is *uniform in* k, that is, for every $\varepsilon > 0$, and for n sufficiently large, we have

$$\max_k |\mathbb{E}[P_k^{(n)}] - p_k| \leq \frac{\varepsilon}{2}. \tag{6.3.7}$$

Exercise 6.14 (Uniform convergence of mean degree sequence)**.** *Prove* (6.3.7).

By (6.3.4) and (6.3.7), it follows that, for n sufficiently large,

$$\mathbb{P}\Big(\max_k |P_k^{(n)} - p_k| \geq \varepsilon\Big) \leq \sum_{k=0}^{\infty} \mathbb{P}\Big(|P_k^{(n)} - \mathbb{E}[P_k^{(n)}]| \geq \varepsilon/2\Big). \tag{6.3.8}$$

Note that, by Chebychev inequality (Theorem 2.15),

$$\mathbb{P}\big(|P_k^{(n)} - \mathbb{E}[P_k^{(n)}]| \geq \varepsilon/2\big) \leq \frac{4}{\varepsilon^2}\mathrm{Var}(P_k^{(n)}), \tag{6.3.9}$$

so that

$$\mathbb{P}\Big(\max_k |P_k^{(n)} - p_k| \geq \varepsilon\Big) \leq \frac{4}{\varepsilon^2}\sum_{k=0}^{\infty} \mathrm{Var}(P_k^{(n)}). \tag{6.3.10}$$

We use the definition in (6.3.1) to see that

$$\begin{aligned}\mathbb{E}[(P_k^{(n)})^2] &= \frac{1}{n^2}\sum_{i,j\in[n]} \mathbb{P}(D_i = D_j = k) \qquad (6.3.11)\\ &= \frac{1}{n^2}\sum_{i\in[n]} \mathbb{P}(D_i = k) + \frac{1}{n^2}\sum_{i,j\in[n]:\, i\neq j} \mathbb{P}(D_i = D_j = k)\mathbb{P}(D_i = D_j = k).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{Var}(P_k^{(n)}) \leq \frac{1}{n^2}\sum_{i\in[n]} [\mathbb{P}(D_i = k) - \mathbb{P}(D_i = k)^2] \qquad (6.3.12)\\ + \frac{1}{n^2}\sum_{i,j\in[n]:\, i\neq j} [\mathbb{P}(D_i = D_j = k) - \mathbb{P}(D_i = k)\mathbb{P}(D_j = k)].\end{aligned}$$

We let

$$X_i = \sum_{k\in[n]:\, k\neq i,j} I_{ik}, \qquad X_j = \sum_{k\in[n]:\, k\neq i,j} I_{jk}, \tag{6.3.13}$$

where $(I_{ij})_{i,j\in[n]}$ are independent $\mathrm{Be}(p_{ij})$ random variables. Then, the law of (D_i, D_j) is the same as the one of (X_i+I_{ij}, X_j+I_{ij}) while $(X_{ij}+I_{ij}, X_j+I'_{ij})$, where I'_{ij} is independent of $(I_{ij})_{i,j\in[n]}$ has the same distribution as I_{ij}, are two independent random variables with the same marginals as D_i and D_j. Then,

$$\mathbb{P}(D_i = D_j = k) = \mathbb{P}\Big((X_i + I_{ij}, X_j + I_{ij}) = (k,k)\Big), \tag{6.3.14}$$

$$\mathbb{P}(D_i = k)\mathbb{P}(D_j = k) = \mathbb{P}\Big((X_i + I_{ij}, X_j + I'_{ij}) = (k,k)\Big), \tag{6.3.15}$$

so that

$$\begin{aligned}&\mathbb{P}(D_i = D_j = k) - \mathbb{P}(D_i = k)\mathbb{P}(D_j = k) \\ &\quad\le \mathbb{P}\Big((X_i + I_{ij}, X_j + I_{ij}) = (k,k), (X_i + I_{ij}, X_j + I'_{ij}) \neq (k,k)\Big).\end{aligned} \tag{6.3.16}$$

When $(X_I + I_{IJ}, X_j + I_{ij}) = (k,k)$, but $(X_i + I_{ij}, X_j + I'_{ij}) \neq (k,k)$, we must have that $I_{ij} \neq I'_{ij}$. If $I_{ij} = 1$, then $I'_{ij} = 0$ and $X_j = k$, while, if $I_{ij} = 0$, then $I'_{ij} = 1$ and $X_i = k$. Therefore,

$$\mathbb{P}(D_i = D_j = k) - \mathbb{P}(D_i = k)\mathbb{P}(D_j = k) \le 2p_{ij}[\mathbb{P}(D_i = k) + \mathbb{P}(D_j = k)]. \tag{6.3.17}$$

We conclude from (6.3.12) that

$$\sum_{k\ge 0} \mathrm{Var}(P_k^{(n)}) \le \frac{1}{n} + \frac{2}{n^2} \sum_{i,j\in[n]} p_{ij} \to 0, \tag{6.3.18}$$

since $\sum_{i,j\in[n]} p_{ij} = O(n)$ (recall Exercise 6.3). □ □

6.4 Generalized random graph with i.i.d. weights

We next state a consequence of Theorem 6.6, where we treat the special case where $(w_i)_{i\in[n]}$ are independent and identically distributed. To avoid confusion with W_n, which is the weight of a vertex chosen uniformly at random from $[n]$, we continue to write the weights as $(w_i)_{i\in[n]}$, bearing in mind that now these weights are *random*. Note that there now is *double randomness*. Indeed, there is randomness due to the fact that the weights $(w_i)_{i\in[n]}$ are random themselves, and then there is the randomness in the occupation status of the edges conditionally on the weights $(w_i)_{i\in[n]}$. We denote the resulting graph by $\mathrm{GRG}_n(\boldsymbol{w})$. By Exercise 6.8, the edge statuses are *not* independent.

We now investigate the degrees and degree sequence of $\mathrm{GRG}_n(\boldsymbol{w})$:

Corollary 6.10 (Degrees of $\mathrm{GRG}_n(\boldsymbol{W})$)**.** *When $(w_i)_{i\in[n]}$ are i.i.d. random variables with distribution function F with a finite mean, then*

(a) the degree D_k of vertex k converges in distribution to a mixed Poisson random variable with mixing distribution F;

(b) the degrees $D_1, \dots, D_m$ of vertices $1, \dots, m$ are asymptotically independent.

To see that Corollary 6.10 follows from Theorem 6.6, we note that when $(w_i)_{i\in[n]}$ are i.i.d. copies of a random variable W with distribution function F, we have that $\frac{1}{n^2}\sum_{i\in[n]} w_i^2 \to 0$, since $\frac{1}{n}\sum_{i\in[n]} w_i^2 = o_{\mathbb{P}}(n)$ follows when W has a finite mean:

Exercise 6.15 (Bound on sum of squares of i.i.d. random variables)**.** *Show that when $(w_i)_{i\in[n]}$ are i.i.d. random variables with distribution function F with a finite mean, then*

$$\frac{1}{n}\max_{i\in[n]} w_i \xrightarrow{\mathbb{P}} 0. \tag{6.4.1}$$

Conclude that

$$\frac{1}{n^2}\sum_{i\in[n]} w_i^2 \xrightarrow{\mathbb{P}} 0. \tag{6.4.2}$$

Hint: Use that

$$\begin{aligned}\mathbb{P}(\max_{i\in[n]} w_i \geq \varepsilon n) &\leq \sum_{i\in[n]} \mathbb{P}(w_i \geq \varepsilon n)\\ &= n\mathbb{P}(W \geq \varepsilon n).\end{aligned} \tag{6.4.3}$$

Then use a variant of the Markov inequality (Theorem 2.14) to show that $\mathbb{P}(W \geq \varepsilon n) = o(\frac{1}{n})$.

Theorem 6.6 is an extension of [58, Theorem 3.1], in which Corollary 6.10 was proved under the extra assumption that w_i have a finite $(1+\varepsilon)-$moment.

Theorem 6.11 (Degree sequence of $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *When $(w_i)_{i\in[n]}$ are i.i.d. random variables with distribution function F having finite mean, then, for every $\varepsilon > 0$,*

$$\mathbb{P}\Big(\sum_{k=0}^{\infty} |P_k^{(n)} - p_k| \geq \varepsilon\Big) \to 0, \tag{6.4.4}$$

where $(p_k)_{k=0}^{\infty}$ is the probability mass function of a mixed Poisson distribution with mixing distribution F.

We leave the proof of Theorem 6.11, which is quite similar to the proof of Theorem 6.9, to the reader:

Exercise 6.16 (Proof of Theorem 6.11)**.** *Complete the proof of Theorem 6.11, now using Corollary 6.8, as well as the equality*

$$\begin{aligned}\mathbb{E}[(P_k^{(n)})^2] &= \frac{1}{n^2}\sum_{1\leq i,j\leq n} \mathbb{P}(D_i = D_j = k)\\ &= \frac{1}{n}\mathbb{P}(D_1 = k) + \frac{2}{n^2}\sum_{1\leq i<j\leq n} \mathbb{P}(D_i = D_j = k).\end{aligned} \tag{6.4.5}$$

We next turn our attention to the case where the weights $(W_i)_{i\in[n]}$ are i.i.d. with *infinite mean.* We denote the distribution of W_i by F.

Exercise 6.17 (Condition for infinite mean)**.** *Show that the mean of W is infinite precisely when the distribution function F of W satisfies*

$$\int_0^{\infty} [1 - F(x)]dx = \infty. \tag{6.4.6}$$

Our next goal is to obtain a random graph which has a power-law degree sequence with a power-law exponent $\tau \in (1,2)$. We shall see that this is a non-trivial issue.

Theorem 6.12 (Degrees of $\mathrm{GRG}_n(\boldsymbol{w})$ with i.i.d. conditioned weights)**.** *Let $(w_i)_{i\in[n]}$ be i.i.d. random variables with distribution function F, and let $(w_i^{(n)})_{i\in[n]}$ be i.i.d. copies of the random variable W conditioned on $W \le a_n$. Then, for every $a_n \to \infty$ such that $a_n = o(n)$,*

(a) the degree $D_k^{(n)}$ of vertex k in the GRG with weights $(w_i^{(n)})_{i\in[n]}$, converges in distribution to a mixed Poisson random variable with mixing distribution F;

(b) the degrees $(D_i^{(n)})_{i\in[m]}$ of vertices $1,\dots,m$ are asymptotically independent.

Proof. Theorem 6.12 follows by a simple adaptation of the proof of Theorem 6.6 and will be left as an exercise:

Exercise 6.18 (Proof of Theorem 6.12)**.** *Prove Theorem 6.12.*

We finally show that the conditioning in Theorem 6.12 is necessary by proving that if we do not condition the weights to be at most a_n, then the degree distribution changes:

Theorem 6.13 (Degrees of $\mathrm{GRG}_n(\boldsymbol{w})$ with i.i.d. infinite mean weights)**.** *When $(w_i)_{i\in[n]}$ are i.i.d. random variables with distribution function F satisfying that for some $\tau \in (1,2)$,*

$$\lim_{x\to\infty} x^{\tau-1}[1-F(x)] = c. \tag{6.4.7}$$

Let the edge probabilities $(p_{ij})_{1\le i<j\le n}$ conditionally on the weights $(w_i)_{i\in[n]}$ be given by

$$p_{ij} = \frac{w_i w_j}{n^{1/(\tau-1)} + w_i w_j}. \tag{6.4.8}$$

Then,

(a) the degree D_k of vertex k converges in distribution to a mixed Poisson random variable with parameter $\gamma W^{\tau-1}$, where

$$\gamma = c\int_0^\infty (1+x)^{-2} x^{-(\tau-1)} dx. \tag{6.4.9}$$

(b) the degrees $(D_i)_{i\in[m]}$ of vertices $1,\dots,m$ are asymptotically independent.

The proof of Theorem 6.13 is deferred to Section 6.5 below. We note that a mixed Poisson distribution with mixing distribution γW^α does not obey a power law with exponent τ:

Exercise 6.19 (Tail of degree law for $\tau \in (1,2)$)**.** *Let the distribution function F satisfy* (6.4.7)*, and let Y be a mixed Poisson random variable with parameter $W^{\tau-1}$, where W has distribution function F. Show that Y is such that there exists a constant $c>0$ such that*

$$\mathbb{P}(Y \ge y) = cy^{-1}(1+o(1)). \tag{6.4.10}$$

As a result of Exercise 6.19, we see that if we do not condition on the weights to be at most a_n, and if the distribution function F of the weights satisfies (6.4.7), then the degree distribution always obeys a power law with exponent $\tau = 2$.

We note that the choice of the edge probabilities in (6.4.8) is different from the choice in (6.1.1). Indeed, the term $\ell_n = \sum_{i\in[n]} w_i$ in the denominator in (**??**) is replaced by $n^{1/(\tau-1)}$ in (6.4.8). Since, when (6.4.7) is satisfied,

$$n^{-1/(\tau-1)} \sum_{i\in[n]} w_i \xrightarrow{d} S, \tag{6.4.11}$$

where S is a stable random variable with parameter $\tau - 1 \in (0,1)$, we expect that the behavior for the choice (**??**) is similar (recall Theorem 2.28).

6.5 Generalized random graph conditioned on its degrees

In this section, we investigate the distribution of $\mathrm{GRG}_n(\boldsymbol{w})$ in more detail. The main result in this section is that the generalized random graph conditioned on its degree sequence is a uniform random graph with that degree sequence (see Theorem 6.14 below).

We start by introducing some notation. We let $X = (X_{ij})_{1\le i<j\le n}$, where X_{ij} are independent random variables with

$$\mathbb{P}(X_{ij} = 1) = 1 - \mathbb{P}(X_{ij} = 0) = p_{ij}, \tag{6.5.1}$$

where p_{ij} is given in (6.1.1). Then, with $q_{ij} = 1 - p_{ij}$, we have that, for $x = (x_{ij})_{1\le i<j\le n}$,

$$\mathbb{P}(X = x) = \prod_{1\le i<j\le n} p_{ij}^{x_{ij}} q_{ij}^{1-x_{ij}}. \tag{6.5.2}$$

We define the *odd-ratios* $(r_{ij})_{1\le i<j\le n}$ by

$$r_{ij} = \frac{p_{ij}}{q_{ij}}. \tag{6.5.3}$$

Then

$$p_{ij} = \frac{r_{ij}}{1+r_{ij}}, \qquad q_{ij} = \frac{1}{1+r_{ij}}, \tag{6.5.4}$$

so that

$$\mathbb{P}(X = x) = \prod_{1\le i<j\le n} \frac{1}{1+r_{ij}} \prod_{1\le i<j\le n} r_{ij}^{x_{ij}}. \tag{6.5.5}$$

We now specialize to the setting of the generalized random graph, and choose

$$r_{ij} = u_i u_j, \tag{6.5.6}$$

for some weights $\{u_i\}_{i\in[n]}$. Later, we shall choose

$$u_i = \frac{w_i}{\sqrt{\ell_n}}, \tag{6.5.7}$$

in which case we return to (6.1.1) since

$$p_{ij} = \frac{r_{ij}}{1+r_{ij}} = \frac{u_i u_j}{1+u_i u_j} = \frac{w_i w_j}{\ell_n + w_i w_j}. \tag{6.5.8}$$

Then, with

$$G(u) = \prod_{1\le i<j\le n} (1+u_i u_j), \tag{6.5.9}$$

we obtain

$$\mathbb{P}(X = x) = G(u)^{-1} \prod_{1\le i<j\le n} (u_i u_j)^{x_{ij}} = G(u)^{-1} \prod_{i\in[n]} u_i^{d_i(x)}, \tag{6.5.10}$$

where $\{d_i(x)\}_{i\in[n]}$ is given by

$$d_i(x) = \sum_{j=1}^{n} x_{ij}, \tag{6.5.11}$$

i.e., $d_i(x)$ is the degree of vertex i in the generalized random graph configuration $x = (x_{ij})_{1\le i<j\le n}$. By convention, we assume that $x_{ii} = 0$, and we recall that $x_{ij} = x_{ji}$.

Exercise 6.20 (Equality for probability mass function GRG)**.** *Prove the last equality in* (6.5.10).

From (6.5.10), and using that $\sum_x \mathbb{P}(X = x) = 1$, it follows that

$$\prod_{1\le i<j\le n}(1+u_iu_j) = G(u) = \sum_x \prod_{i\in[n]} u_i^{d_i(x)}. \tag{6.5.12}$$

Furthermore, it also follows from (6.5.10) that the distribution of X conditionally on $\{d_i(X) = d_i \forall 1 \le i \le n\}$ is *uniform*. That is, all graphs with the same degree sequence have the same probability. This wonderful result is formulated in the following theorem:

Theorem 6.14 (GRG conditioned on degrees has uniform law)**.** *The GRG with edge probabilities* $(p_{ij})_{1\le i<j\le n}$ *given by*

$$p_{ij} = \frac{u_iu_j}{1+u_iu_j}, \tag{6.5.13}$$

conditioned on $\{d_i(X) = d_i \forall i = 1,\dots,n\}$*, is uniform over all graphs with degree sequence* $\{d_i\}_{i\in[n]}$.

Proof. For x satisfying $d_i(x) = d_i$ for all $i = 1,\dots,n$, we can write out

$$\begin{aligned}\mathbb{P}(X = x | d_i(X) = d_i \forall i = 1,\dots,n) &= \frac{\mathbb{P}(X=x)}{\mathbb{P}(d_i(X) = d_i \forall i = 1,\dots,n)}\\ &= \frac{\mathbb{P}(X=x)}{\sum_{y:d_i(y)=d_i\forall i}\mathbb{P}(X=y)}.\end{aligned} \tag{6.5.14}$$

By (6.5.10), we have that (6.5.14) simplifies to

$$\begin{aligned}\mathbb{P}(X = x | d_i(X) = d_i \forall i = 1,\dots,n) &= \frac{\prod_{i\in[n]} u_i^{d_i(x)}}{\sum_{y:d_i(y)=d_i\forall i}\prod_{i\in[n]} u_i^{d_i(y)}}\\ &= \frac{\prod_{i\in[n]} u_i^{d_i}}{\sum_{y:d_i(y)=d_i\forall i}\prod_{i\in[n]} u_i^{d_i}}\\ &= \frac{1}{\#\{y : d_i(y) = d_i \forall i = 1,\dots,n\}},\end{aligned} \tag{6.5.15}$$

that is, the distribution is uniform over all graphs with the prescribed degree sequence. □

We next compute the generating function of all degrees, that is, for $t_1,\dots,t_n \in \mathbb{R}$, we compute, with $D_i = d_i(X)$,

$$\mathbb{E}\Big[\prod_{i\in[n]} t_i^{D_i}\Big] = \sum_x \mathbb{P}(X=x)\prod_{i\in[n]} t_i^{d_i(x)}. \tag{6.5.16}$$

By (6.5.10) and (6.5.12),

$$\mathbb{E}\Big[\prod_{i\in[n]} t_i^{D_i}\Big] = G(u)^{-1}\sum_x \prod_{i\in[n]} (u_it_i)^{d_i(x)} = \frac{G(tu)}{G(u)}, \tag{6.5.17}$$

where $(tu)_i = t_iu_i$. By (6.5.9), we obtain

$$\mathbb{E}\Big[\prod_{i\in[n]} t_i^{D_i}\Big] = \prod_{1\le i<j\le n}\frac{1+u_it_iu_jt_j}{1+u_iu_j}. \tag{6.5.18}$$

Therefore, we have proved the following nice property:

Proposition 6.15 (Generating function of degrees of $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *For the edge probabilities given by* (6.1.1) *and* (6.5.7)*,*

$$\mathbb{E}\Big[\prod_{i\in[n]} t_i^{D_i}\Big] = \prod_{1\le i<j\le n} \frac{\ell_n + w_i t_i w_j t_j}{\ell_n + w_i w_j}. \tag{6.5.19}$$

Exercise 6.21 (Alternative proof Theorem 6.6)**.** *Use Proposition 6.15 to give an alternative proof of Theorem 6.6.*

Exercise 6.22 (Degree of vertex 1 in $\mathrm{ER}_n(\lambda/n)$)**.** *Show that for the Erdős-Rényi random graph with $p = \lambda/n$, the degree of vertex 1 is close to a Poisson random variable with mean λ by using* (B.119)*. Hint: Use that the Erdős-Rényi random graph is obtained by taking $W_i \equiv \frac{\lambda}{1-\frac{\lambda}{n}}$.*

Exercise 6.23 (Asymptotic independence of vertex degrees in $\mathrm{ER}_n(\lambda/n)$)**.** *Show that for the Erdős-Rényi random graph with $p = \lambda/n$, the degrees of vertices $1, \dots, m$ are asymptotically independent.*

We finally make use of Proposition 6.15 to prove Theorem 6.13:

Proof of Theorem 6.13. We study the generating function of the degree D_k. We note that

$$\mathbb{E}[t^{D_k}] = \mathbb{E}\Big[\prod_{i\neq k} \frac{1 + tW_iW_kn^{-\frac{1}{\tau-1}}}{1 + W_iW_kn^{-\frac{1}{\tau-1}}}\Big]. \tag{6.5.20}$$

Denote $\phi_w : \mathbb{R} \mapsto \mathbb{R}$ by

$$\phi_w(x) = \frac{1+twx}{1+wx}. \tag{6.5.21}$$

Then, by the independence of the weights $(w_i)_{i\in[n]}$, we have that

$$\mathbb{E}[t^{D_k}|W_k = w] = \mathbb{E}\Big[\prod_{i\neq k} \phi_w\big(W_in^{-\frac{1}{\tau-1}}\big)\Big] = \psi_n(w)^{n-1}, \tag{6.5.22}$$

where

$$\psi_n(w) = \mathbb{E}\Big[\phi_w\big(W_in^{-\frac{1}{\tau-1}}\big)\Big]. \tag{6.5.23}$$

We claim that

$$\psi_n(w) = 1 + \frac{1}{n}(t-1)\gamma w^{\tau-1} + o(n^{-1}). \tag{6.5.24}$$

This completes the proof since it implies that

$$\mathbb{E}[t^{D_k}|W_k = w] = \psi_n(w)^{n-1} = \mathrm{e}^{(t-1)\gamma w^{\tau-1}}(1+o(1)), \tag{6.5.25}$$

which in turn implies that

$$\lim_{n\to\infty} \mathbb{E}[t^{D_k}] = \mathbb{E}[\mathrm{e}^{(t-1)\gamma W_k^{\tau-1}}]. \tag{6.5.26}$$

Since $\mathbb{E}[\mathrm{e}^{(t-1)\gamma W_k^{\tau-1}}]$ is the probability generating function of a mixed Poisson random variable with mixing distribution $\gamma W_k^{\tau-1}$ (see Exercise 6.24), (6.5.24) indeed completes the proof.

Exercise 6.24 (Identification of limiting vertex degree)**.** *Prove that* $\mathbb{E}[\mathrm{e}^{(t-1)\gamma W^{\tau-1}}]$ *is the probability generating function of a mixed Poisson random variable with mixing distribution* $\gamma W^{\tau-1}$

We complete the proof of Theorem 6.13 by showing that (6.5.24) holds. For this, we first note

$$\begin{aligned}\psi_n(w) &= \mathbb{E}\Big[\phi_w\big(W_1 n^{-\frac{1}{\tau-1}}\big)\Big]\\ &= 1+\mathbb{E}\Big[\phi_w\big(W_1 n^{-\frac{1}{\tau-1}}\big)-1\Big].\end{aligned} \tag{6.5.27}$$

Exercise 6.25 (A partial integration formula)**.** *Prove that for every function* $h\colon [0,\infty)\to\mathbb{R}$*, with* $h(0)=0$ *and every random variable* $X\geq 0$ *with distribution function* F*, we have the partial integration formula*

$$\mathbb{E}[h(X)] = \int_0^\infty h'(x)[1-F(x)]dx. \tag{6.5.28}$$

Applying (6.5.28) to $h(x)=\phi_w\big(xn^{-\frac{1}{\tau-1}}\big)-1$ and $X=W_1$ yields

$$\begin{aligned}\psi_n(w) &= 1+n^{-\frac{1}{\tau-1}}\int_0^\infty \phi_w'\big(xn^{-\frac{1}{\tau-1}}\big)[1-F(x)]dx\\ &= 1+\int_0^\infty \phi_w'(x)[1-F(xn^{\frac{1}{\tau-1}})]dx.\end{aligned} \tag{6.5.29}$$

Thus,

$$n(\psi_n(w)-1) = \int_0^\infty \frac{\phi_w'(x)}{x^{\tau-1}}(n^{\frac{1}{\tau-1}}x)^{\tau-1}[1-F(xn^{\frac{1}{\tau-1}})]dx. \tag{6.5.30}$$

By assumption, $x^{\tau-1}[1-F(x)]$ is a bounded function that converges to c. As a result, by the Dominated convergence theorem (Theorem A.9),

$$\lim_{n\to\infty}\int_0^\infty \frac{\phi_w'(x)}{x^{\tau-1}}(n^{\frac{1}{\tau-1}}x)^{\tau-1}[1-F(xn^{\frac{1}{\tau-1}})]dx = c\int_0^\infty \frac{\phi_w'(x)}{x^{\tau-1}}dx. \tag{6.5.31}$$

Exercise 6.26 (Conditions for dominated convergence)**.** *Verify the conditions for dominated convergence for the integral on the left-hand side of* (6.5.31)*.*

We complete the proof of (6.5.24) by noting that

$$\phi_w'(x) = \frac{tw}{1+wx}-\frac{w(1+twx)}{(1+wx)^2} = \frac{w(t-1)}{(1+wx)^2}, \tag{6.5.32}$$

so that

$$c\int_0^\infty \frac{\phi_w'(x)}{x^{\tau-1}}dx = c\int_0^\infty \frac{w(t-1)}{(1+wx)^2x^{\tau-1}}dx = \gamma(t-1)w^{\tau-1}. \tag{6.5.33}$$

□

6.6 Asymptotic equivalence of inhomogeneous random graphs

There are numerous papers that introduce models along the lines of the generalized random graph, in that they have (conditionally) independent edge statuses. The most general model has appeared in [47]. In this paper, the properties of such random graphs (such as diameter, phase transition and average distances) have been studied using comparisons to multitype branching processes. We shall return to [47] in Chapter 9. We start by investigating when two inhomogeneous random graph sequences are asymptotically equivalent, following the results of Janson in [131].

In this section, we shall investigate when two random graphs are asymptotically equivalent. We shall start by introducing this notion for general random variables. Before we can do so, we say that $(\mathcal{X}, \mathcal{F})$ is a measurable space when $\mathcal{X}$ is the state space, i.e., the space of all possible outcomes, and $\mathcal{F}$ the set of all possible events. We shall be particularly interested in *discrete* measurable spaces, in which case $\mathcal{X}$ is a discrete set and $\mathcal{F}$ can be taken to be the set of all subsets of $\mathcal{X}$. However, all notions that will be introduced in this section, can be more generally defined.

Definition 6.16 (Asymptotic equivalence of sequences of random variables)**.** *Let $(\mathcal{X}_n, \mathcal{F}_n)$ be a sequence of measurable spaces. Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two probability measures on $(\mathcal{X}_n, \mathcal{F}_n)$. Then, we say that the sequences $(\mathbb{P}_n)_{n=1}^{\infty}$ and $(\mathbb{Q}_n)_{n=1}^{\infty}$ are asymptotically equivalent if, for every sequence $\mathcal{E}_n \in \mathcal{F}_n$ of events, we have*

$$\lim_{n\to\infty} \mathbb{P}_n(\mathcal{E}_n) - \mathbb{Q}_n(\mathcal{E}_n) = 0. \tag{6.6.1}$$

Thus, $(\mathbb{P}_n)_{n=1}^{\infty}$ and $(\mathbb{Q}_n)_{n=1}^{\infty}$ are asymptotically equivalent when they have asymptotically equal probabilities. In practice, this means that there is asymptotically no difference between $(\mathbb{P}_n)_{n=1}^{\infty}$ and $(\mathbb{Q}_n)_{n=1}^{\infty}$.

The main result that we shall prove in this section is the following theorem that gives a sharp criterium on when two inhomogeneous random graph sequences are asymptotically equivalent. In its statement, we write $\boldsymbol{p} = (p_{ij})_{1\le i<j\le n}$ for the edge probabilities in the graph, and $\mathrm{IRG}_n(\boldsymbol{p})$ for the inhomogeneous random graph for which the edges are independent and the probability that the edge ij is present equals p_{ij}.

Theorem 6.17 (Asymptotic equivalence of inhomogeneous random graphs)**.** *Let $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ be two inhomogeneous random graphs with edge probabilities $\boldsymbol{p} = (p_{ij})_{1\le i<j\le n}$ and $\boldsymbol{q} = (q_{ij})_{1\le i<j\le n}$ respectively. Assume that there exists $\varepsilon > 0$ such that $\max_{1\le i<j\le n} p_{ij} \le 1-\varepsilon$. Then $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ are asymptotically equivalent when*

$$\lim_{n\to\infty} \sum_{1\le i<j\le n} \frac{(p_{ij}-q_{ij})^2}{p_{ij}} = 0. \tag{6.6.2}$$

When the edge probabilities $\boldsymbol{p} = (p_{ij})_{1\le i<j\le n}$ and $\boldsymbol{q} = (q_{ij})_{1\le i<j\le n}$ are themselves random variables, with $\max_{1\le i<j\le n} p_{ij} \le 1-\varepsilon$ a.s., then $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ are asymptotically equivalent when

$$\sum_{1\le i<j\le n} \frac{(p_{ij}-q_{ij})^2}{p_{ij}} \xrightarrow{\mathbb{P}} 0. \tag{6.6.3}$$

We note that, in particular, $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ are asymptotically equivalent when they can be coupled in such a way that $\mathbb{P}(\mathrm{IRG}_n(\boldsymbol{p}) \neq \mathrm{IRG}_n(\boldsymbol{q})) = o(1)$. Thus, Theorem 6.17 is a quite strong result. The remainder of this section shall be devoted to the proof of Theorem 6.17. We start by introducing the necessary ingredients.

There is a strong relation between asymptotic equivalence of random variables and coupling, in the sense that two sequences of random variables are asymptotically equivalent precisely when they can be coupled such that they agree with high probability. Recall the

results in Section 2.2 that we shall use and extend in this section. Let $p = (p_x)_{x\in\mathcal{X}}$ and $q = (q_x)_{x\in\mathcal{X}}$ be two discrete probability measures on the space $\mathcal{X}$, and recall that the total variation distance between p and q is given by

$$d_{\mathrm{TV}}(p,q) = \frac{1}{2}\sum_x |p_x - q_x|. \tag{6.6.4}$$

By (2.2.17)-(2.2.18), we see that two sequences of discrete probability measures $p^{(n)} = (p_x^{(n)})_{x\in\mathcal{X}}$ and $q^{(n)} = (q_x^{(n)})_{x\in\mathcal{X}}$ are asymptotically equivalent when

$$d_{\mathrm{TV}}(p^{(n)}, q^{(n)}) \to 0. \tag{6.6.5}$$

In fact, this turns out to be an equivalent definition:

Exercise 6.27 (Asymptotic equivalence and total variation distance)**.** *Use* (2.2.7) *and Definition 6.16 to prove that* $p^{(n)} = (p_x^{(n)})_{x\in\mathcal{X}}$ *and* $q^{(n)} = (q_x^{(n)})_{x\in\mathcal{X}}$ *are asymptotically equivalent if and only if* $d_{\mathrm{TV}}(p^{(n)}, q^{(n)}) \to 0$.

When p and q correspond to $\mathrm{Be}(p)$ and $\mathrm{Be}(q)$ distributions, then it is rather simple to show that

$$d_{\mathrm{TV}}(p,q) = |p-q|. \tag{6.6.6}$$

Now, for $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$, the edge occupation variables are all independent $\mathrm{Be}(p_{ij})$ and $\mathrm{Be}(q_{ij})$ random variables. Thus, we can couple each of the edges in such a way that the probability that a particular edge is distinct is equal to

$$d_{\mathrm{TV}}(p_{ij}, q_{ij}) = |p_{ij} - q_{ij}|, \tag{6.6.7}$$

so that we are led to the naive bound

$$d_{\mathrm{TV}}(\mathrm{IRG}_n(\boldsymbol{p}), \mathrm{IRG}_n(\boldsymbol{q})) \le \sum_{1\le i<j\le n} |p_{ij} - q_{ij}|, \tag{6.6.8}$$

which is far worse than (6.6.2). As we shall see later on, there are many examples for which $\sum_{1\le i<j\le n} \frac{(p_{ij}-q_{ij})^2}{p_{ij}} = o(1)$, but $\sum_{1\le i<j\le n} |p_{ij} - q_{ij}| \neq o(1)$. Thus, the coupling used in the proof of Theorem 6.17 is substantially stronger.

To explain this seeming contradiction, it is useful to investigate the setting of the Erdős-Rényi random graph $\mathrm{ER}_n(p)$. Fix p and q, assume that $q \le p$ and that $p \le 1-\varepsilon$. Then, by Theorem 6.17, $\mathrm{ER}_n(p)$ and $\mathrm{ER}_n(q)$ are asymptotically equivalent when

$$\sum_{1\le i<j\le n} \frac{(p_{ij}-q_{ij})^2}{p_{ij}} \le n^2(p-q)^2/p = O(n^3(p-q)^2), \tag{6.6.9}$$

when we assume that $p \ge \varepsilon/n$. Thus, it suffices that $p - q = o(n^{-3/2})$. On the other hand, the right-hand side of (6.6.8) is $o(1)$ when $p - q = o(n^{-2})$, which is rather stronger. This can be understood by noting that if we condition on the number of edges M, then the conditional distribution of $\mathrm{ER}_n(p)$ *conditionally on* $M = m$ does not depend on the precise value of p involved. As a result, we obtain that the asymptotic equivalence of $\mathrm{ER}_n(p)$ and $\mathrm{ER}_n(q)$ follows precisely when we have asymptotic equivalence of the number of edges in $\mathrm{ER}_n(p)$ and $\mathrm{ER}_n(q)$. For this, we note that $M \sim \mathrm{Bin}(n(n-1)/2, p)$ for $\mathrm{ER}_n(p)$, while the number of edges M' for $\mathrm{ER}_n(q)$ satisfies $M' \sim \mathrm{Bin}(n(n-1)/2, q)$. By Exercise 4.2 as well as Exercise 4.22, we have that binomial distributions with a variance that tends to infinity satisfy a central limit theorem. When M and M' both satisfy central limit theorems with equal asymptotic variances, it turns out that the asymptotic equivalence of M and M' follows when the asymptotic means are equal:

Exercise 6.28 (Asymptotic equivalence of binomials with increasing variances [131]). *Let M and M' be two binomial random variables with $M \sim \mathrm{Bin}(m,p)$ and $M' \sim \mathrm{Bin}(m,q)$ for some m. Show that M and M' are asymptotically equivalent when $m(p-q)/\sqrt{mp} = o(1)$.*

We apply Exercise 6.28 with $m = n(n-1)/2$ to obtain that $\mathrm{ER}_n(p)$ and $\mathrm{ER}_n(q)$ are asymptotically equivalent precisely when $n^2(p-q)^2/p = o(1)$, and, assuming that $p = \lambda/n$, this is equivalent to $p - q = o(n^{-3/2})$. This explains the result in Theorem 6.17 for the Erdős-Rényi random graph, and also shows that the result is optimal for the Erdős-Rényi random graph.

We now proceed by proving Theorem 6.17. In this section, rather than working with the total variation distance between two measures, it is more convenient to work with the so-called *Hellinger distance*, which is defined, for discrete measures $p = (p_x)_{x\in\mathcal{X}}$ and $q = (q_x)_{x\in\mathcal{X}}$ by

$$d_{\mathrm{H}}(p,q) = \sqrt{\frac{1}{2}\sum_x (\sqrt{p_x} - \sqrt{q_x})^2}. \tag{6.6.10}$$

It is readily seen that d_{H} and d_{TV} are quite intimately related:

Exercise 6.29 (Total variation and Hellinger distance). *Prove that, for discrete probability measures $p = (p_x)_{x\in\mathcal{X}}$ and $q = (q_x)_{x\in\mathcal{X}}$,*

$$d_{\mathrm{H}}(p,q)^2 \le d_{\mathrm{TV}}(p,q) \le 2^{1/2} d_{\mathrm{H}}(p,q). \tag{6.6.11}$$

Exercise 6.30 (Asymptotic equivalence and Hellinger distance). *Use Exercises 6.27 and 6.29 to prove that $p^{(n)} = (p_x^{(n)})_{x\in\mathcal{X}}$ and $q^{(n)} = (q_x^{(n)})_{x\in\mathcal{X}}$ are asymptotically equivalent if and only if $d_{\mathrm{H}}(p^{(n)}, q^{(n)}) \to 0$.*

We define

$$\rho(p,q) = 2d_{\mathrm{H}}(\mathrm{Be}(p), \mathrm{Be}(q))^2 = \big(\sqrt{p} - \sqrt{q}\big)^2 + \big(\sqrt{1-p} - \sqrt{1-q}\big)^2, \tag{6.6.12}$$

and note that

$$\rho(p,q) \le (p-q)^2\big(p^{-1} + (1-p)^{-1}\big). \tag{6.6.13}$$

Exercise 6.31 (Bound on Hellinger distance Bernoulli variables). *Prove that $\rho(p,q) \le (p-q)^2\big(p^{-1} + (1-p)^{-1}\big)$.*

In particular, Exercise 6.31 implies that when $p \le 1 - \varepsilon$, then

$$\rho(p,q) \le C(p-q)^2/p \tag{6.6.14}$$

for some $C = C(\varepsilon) > 0$. Now we are ready to complete the proof of Theorem 6.17:

Proof of Theorem 6.17. Let $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ with $\boldsymbol{p} = (p_{ij})_{1\le i<j\le n}$ and $\boldsymbol{q} = (q_{ij})_{1\le i<j\le n}$ be two inhomogeneous random graphs. The asymptotic equivalence of $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ is equivalent to the asymptotic equivalence of the edge variables, which are independent Bernoulli random variables with success probabilities $\boldsymbol{p} = (p_{ij})_{1\le i<j\le n}$ and $\boldsymbol{q} = (q_{ij})_{1\le i<j\le n}$. In turn, asymptotic equivalence of the edge variables is equivalent to the fact that $d_{\mathrm{H}}(\boldsymbol{p},\boldsymbol{q}) = o(1)$, which is what we shall prove now.

For two discrete probability measures $p = (p_x)_{x\in\mathcal{X}}$ and $q = (q_x)_{x\in\mathcal{X}}$, we denote

$$H(p,q) = 1 - \frac{1}{2}d_{\mathrm{H}}(p,q)^2 = \sum_{x\in\mathcal{X}} \sqrt{p_x}\sqrt{q_x}. \tag{6.6.15}$$

We shall assume that

$$\mathcal{X} = \mathcal{X}^{(1)} \times \cdots \times \mathcal{X}^{(m)} \tag{6.6.16}$$

is of product form, and, for $x = (x_1, \dots, x_m) \in \mathcal{X}$,

$$p_x = \prod_{i=1}^{m} p_{x_i}^{(i)}, \qquad q_x = \prod_{i=1}^{m} q_{x_i}^{(i)} \tag{6.6.17}$$

are product measures, so that p and q correspond to the probability mass functions of independent random variables. Then, due to the product structure of (6.6.15), we obtain

$$H(p,q) = \prod_{i=1}^{m} H(p^{(i)}, q^{(i)}). \tag{6.6.18}$$

For $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ with $\boldsymbol{p} = (p_{ij})_{1\le i<j\le n}$ and $\boldsymbol{q} = (q_{ij})_{1\le i<j\le n}$, the edges are independent, so that

$$H(\boldsymbol{p},\boldsymbol{q}) = \prod_{1\le i<j\le n} (1 - \frac{1}{2}\rho(p_{ij}, q_{ij})), \tag{6.6.19}$$

so that

$$d_{\mathrm{H}}(\boldsymbol{p},\boldsymbol{q}) = \sqrt{2 - 2H(\boldsymbol{p},\boldsymbol{q})}. \tag{6.6.20}$$

As a result, $d_{\mathrm{H}}(\boldsymbol{p},\boldsymbol{q}) = o(1)$ precisely when $H(\boldsymbol{p},\boldsymbol{q}) = 1 + o(1)$. By (6.6.19) and using that $(1-x)(1-y) \ge 1 - x - y$ and $1 - x \le \mathrm{e}^{-x}$, we obtain

$$1 - \frac{1}{2}\sum_{1\le i<j\le n} \rho(p_{ij}, q_{ij}) \le H(\boldsymbol{p},\boldsymbol{q}) \le \mathrm{e}^{-\frac{1}{2}\sum_{1\le i<j\le n}\rho(p_{ij},q_{ij})}, \tag{6.6.21}$$

so that $H(\boldsymbol{p},\boldsymbol{q}) = 1 - o(1)$ precisely when $\sum_{1\le i<j\le n}\rho(p_{ij},q_{ij}) = o(1)$. By (6.6.14), we further obtain that when $\max_{1\le i<j\le n} p_{ij} \le 1 - \varepsilon$ for some $\varepsilon > 0$, then

$$\sum_{1\le i<j\le n} \rho(p_{ij}, q_{ij}) \le C \sum_{1\le i<j\le n} \frac{(p_{ij} - q_{ij})^2}{p_{ij}} = o(1), \tag{6.6.22}$$

by (6.6.2). This completes the proof of the first part of Theorem 6.17. For the second part, we note that if (6.6.3) holds, then we can find a sequence ε_n such that

$$\mathbb{P}\Big(\sum_{1\le i<j\le n} \frac{(p_{ij} - q_{ij})^2}{p_{ij}} \le \varepsilon_n\Big) = 1 - o(1). \tag{6.6.23}$$

Then, the asymptotic equivalence of $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ is, in turn, equivalent to the asymptotic equivalence of $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ *conditionally on* $\sum_{1\le i<j\le n} \frac{(p_{ij}-q_{ij})^2}{p_{ij}} \le \varepsilon_n$. For the latter, we can use the first part of Theorem 6.17. □

In fact, tracing back the above proof, we see that under the assumptions of Theorem 6.17, we also obtain that $\rho(p,q) \ge c(p-q)^2/p$ for some $c = c(\varepsilon) \ge 0$. Thus, we can strengthen Theorem 6.17 to the fact that $\mathrm{IRG}_n(\boldsymbol{p})$ and $\mathrm{IRG}_n(\boldsymbol{q})$ are asymptotically equivalent if and only if (6.6.2) holds.

6.7 Related inhomogeneous random graph models

We now discuss two examples of inhomogeneous random graphs which have appeared in the literature, and are related to the generalized random graph. We start with the expected degree random graph.

6.7.1 Chung-Lu model or expected degree random graph

In this section, we prove a coupling result for the degrees of the Chung-Lu random graph, where the edge probabilities are given by

$$p_{ij}^{(\mathrm{CL})} = \frac{w_i w_j}{\ell_n} \wedge 1, \tag{6.7.1}$$

where again

$$\ell_n = \sum_{i\in[n]} w_i. \tag{6.7.2}$$

When $\max_{i\in[n]} w_i^2 \le \ell_n$, we may forget about the maximum with 1 in (6.7.1). We shall assume $\max_{i\in[n]} w_i^2 \le \ell_n$ throughout this section, and denote the resulting graph by $\mathrm{CL}_n(\boldsymbol{w})$.

Naturally, when $\frac{w_i}{\sqrt{\ell_n}}$ is quite small, there is hardly any difference between edge weights $p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j}$ and $p_{ij} = \frac{w_i w_j}{\ell_n}$. Therefore, one would expect that these models behave rather similarly. We shall make use of Theorem 6.17, and investigate the asymptotic equivalence of $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$:

Theorem 6.18 (Asymptotic equivalence of CL and GRG with deterministic weights)**.** *The random graphs $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ are asymptotically equivalent precisely when*

$$\sum_{i\in[n]} w_i^3 = o(n^{3/2}), \tag{6.7.3}$$

where W_n is the weight of a uniformly chosen vertex in $[n]$.

Proof. We make use of Theorem 6.17. For this, we compute, for fixed ij, and using the fact that $1 - 1/(1+x) \le x$,

$$p_{ij}^{(\mathrm{CL})} - p_{ij} = \frac{w_i w_j}{\ell_n} - \frac{w_i w_j}{\ell_n + w_i w_j} = \frac{w_i w_j}{\ell_n}\Big[1 - \frac{1}{1+\frac{w_i w_j}{\ell_n}}\Big] \le \frac{w_i^2 w_j^2}{\ell_n^2}. \tag{6.7.4}$$

Moreover, since $w_i = o(\sqrt{n})$ by Condition 6.4(a)-(c) and Exercise 6.3, for n sufficiently large

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j} \ge w_i w_j/(2\ell_n), \tag{6.7.5}$$

we arrive at

$$\sum_{1\le i<j\le n} \frac{(p_{ij} - p_{ij}^{(\mathrm{CL})})^2}{p_{ij}} \le 2\ell_n^{-3} \sum_{1\le i<j\le n} w_i^3 w_j^3 \le \ell_n^{-3}\Big(\sum_{i\in[n]} w_i^3\Big)^2 = o(1), \tag{6.7.6}$$

by (6.7.3). □

When Condition 6.4(a)-(c) hold, Exercise 6.3 implies that $\max_{i\in[n]} w_i = o(\sqrt{n})$, so that

$$\sum_{i\in[n]} w_i^3 = o(\sqrt{n}) \sum_{i\in[n]} w_i^2 = o(n^{3/2})\mathbb{E}[W_n^2] = o(n^{3/2}). \tag{6.7.7}$$

Thus, we have proved the following corollary:

Corollary 6.19 (Asymptotic equivalence of CL and GRG)**.** *Assume that Condition 6.4(a)-(c) hold. Then, the random graphs $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ are asymptotically equivalent.*

We can prove stronger results linking the degree sequences of $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ for deterministic weights given by (6.1.10) when $\mathbb{E}[W] < \infty$, by splitting between vertices with small and high weights, but we refrain from doing so.

6.7.2 Norros-Reittu model or the Poisson graph process

In [182], the authors introduce a random *multigraph* with a Poisson number of edges in between any two vertices i and j, with parameter equal to $w_i w_j/\ell_n$. The graph is defined as a *graph process*, where at each time t, a new vertex is born with an associated weight w_t. The number of edges between i and t is $\mathrm{Poi}(w_i w_t/\ell_t)$ distributed. Furthermore, at each time each of the older edges is erased with probability equal to w_t/ℓ_t. We claim that the number of edges between vertices i and j at time t is a Poisson random variable with mean $\frac{w_i w_j}{\ell_t}$, and that the number of edges between the various pairs of vertices are *independent*. To see this, we start by observing a useful property of Poisson random variables:

Exercise 6.32 (Poisson number of Bernoulli variables is Poisson)**.** *Let X be a Poisson random variable with mean λ, and let $(I_i)_{i=1}^{\infty}$ be an independent and identically distributed sequence of* $\mathrm{Be}(p)$ *random variables. Prove that*

$$Y = \sum_{i=1}^{X} I_i \tag{6.7.8}$$

has a Poisson distribution with mean λp.

We make use of Exercise 6.32 to prove that the number of edges between vertices i and j at time t is a Poisson random variable with mean $\frac{w_i w_j}{\ell_t}$, and that the number of edges between different pairs are independent. Indeed, making repeated use of Exercise 6.32 shows that the number of edges at time t between vertices i and j, for $i < j$, is Poisson with parameter

$$\frac{w_i w_j}{\ell_j} \prod_{s=j+1}^{t} (1 - \frac{w_s}{\ell_s}) = \frac{w_i w_j}{\ell_j} \prod_{s=j+1}^{t} (\frac{\ell_{s-1}}{\ell_s}) = \frac{w_i w_j}{\ell_t}, \tag{6.7.9}$$

as required. The independence of the number of edges between different pairs of vertices follows by the independence in the construction of the graph.

The Norros-Reittu graph process produces a multigraph. However, when the weights are sufficiently bounded, it can be seen that the resulting graph is with positive probability simple:

Exercise 6.33 (Simplicity of the Norros-Reittu random graph)**.** *Compute the probability that the Norros-Reittu random graph is simple at time n.*

Exercise 6.34 (The degree of a fixed vertex)**.** *Assume that Condition 6.4(a)-(b) hold. Prove that the degree of vertex k in the Norros-Reittu graph at time n has an asymptotic mixed Poisson distribution with mixing distribution F, the asymptotic distribution function of W_n.*

We now discuss the Norros-Reittu model at time n, ignoring the dynamic formulation given above. We shall denote this graph by $\mathrm{NR}_n(\boldsymbol{w})$. The Norros-Reittu is a *multigraph*, for which the probability that there is at least one edge between vertices i and j exists is, conditionally on the weights $(w_i)_{i\in[n]}$, given by

$$p_{ij}^{(\mathrm{NR})} = 1 - \mathrm{e}^{-\frac{w_i w_j}{\ell_n}}, \tag{6.7.10}$$

and the occupation status of different edges is independent.

We next return to the relation between the various random graph models discussed in this section. We shall fixe the weights to be equal to $(w_i)_{i\in[n]}$, and compare the generalized random graph, Chung-Lu model and Norros-Reittu model with these weights. The latter is denoted by $\mathrm{NR}_n(\boldsymbol{w})$.

We say that a random graph G_n is *stochastically dominated* by the random graph G'_n when, with $(X_{ij})_{1\le i<j\le n}$ and $(X'_{ij})_{1\le i<j\le n}$ denoting the occupation statuses of the edges in G_n and G'_n respectively, there exists a *coupling* $\Big((\hat{X}_{ij})_{1\le i<j\le n}, (\hat{X}'_{ij})_{1\le i<j\le n}\Big)$ of $(X_{ij})_{1\le i<j\le n}$ and $(X'_{ij})_{1\le i<j\le n}$ such that

$$\mathbb{P}\big(\hat{X}_{ij} \le \hat{X}'_{ij} \forall i,j \in [n]\big) = 1. \tag{6.7.11}$$

We write $G_n \preceq G'_n$ when the random graph G_n is stochastically dominated by the random graph G'_n.

Exercise 6.35 (Stochastic domination of increasing random variables). *Let $G_n \preceq G'_n$. Let the random variable $X(G)$ be an increasing random variable of the edge occupation random variables of the graph G. Let $X_n = X(G_n)$ and $X'_n = X(G'_n)$. Show that $X_n \preceq X'_n$.*

When the statuses of the edges are *independent*, then (6.7.11) is equivalent to the bound that, for all $i,j \in [n]$,

$$p_{ij} = \mathbb{P}(X_{ij} = 1) \le p'_{ij} = \mathbb{P}(X'_{ij} = 1). \tag{6.7.12}$$

We note that, by (6.7.12) and the fact that, for every $x \ge 0$,

$$\frac{x}{1+x} \le 1 - \mathrm{e}^{-x} \le \max\{x, 1\}, \tag{6.7.13}$$

we have that

$$\mathrm{GRG}_n(\boldsymbol{w}) \preceq \mathrm{NR}_n(\boldsymbol{w}) \preceq \mathrm{CL}_n(\boldsymbol{w}). \tag{6.7.14}$$

This provides a good way of comparing the various inhomogeneous random graph models discussed in this chapter.

Exercise 6.36 (Asymptotic equivalence of IRGs). *Assume that Condition 6.4(a)-(c) hold. Show that $\mathrm{NR}_n(\boldsymbol{w})$ is asymptotically equivalent to $\mathrm{GRG}_n(\boldsymbol{W})$.*

6.8 Notes and discussion

Notes on Section 6.1. In the generalized random graph studied in [58], the situation where the vertex weights are i.i.d. is investigated, and ℓ_n in the denominator of the edge probabilities in (6.1.1) is replaced by n, which leads to a minor change. Indeed, when the weights have finite mean, then $\ell_n = \mathbb{E}[W]n(1+o(1))$, by the law of large numbers. If we would replace ℓ_n by $\mathbb{E}[W]n$ in (6.1.1), then the edge occupation probabilities become

$$\frac{w_i w_j}{\mathbb{E}[W]n + w_i w_j}, \tag{6.8.1}$$

so that this change amounts to replacing w_i by $w_i/\sqrt{\mathbb{E}[W]}$. Therefore, at least on a heuristic level, there is hardly any difference between the definition of p_{ij} in (6.1.1), and the choice $p_{ij} = \frac{w_i w_j}{n+w_i w_j}$ in [58].

In the literature, both the cases with i.i.d. weights as well as the one with deterministic weights have been studied. In [66, 67, 68, 72, 161], the Chung-Lu model, as defined in Section 6.7, is studied with deterministic weights. In [102], general settings are studied, including the one with deterministic weights as in (6.1.10). In [58], on the other hand, the generalized random graph is studied where the weights are i.i.d., and in [102] for several cases including the one for i.i.d. degrees, in the case where the degrees have finite variance degrees, for the Chung-Lu model, the Norros-Reittu model, as well as the generalized random graph.

The advantage of deterministic weights is that there is no double randomness, which makes the model easier to analyse. The results are also more general, since often the results for random weights are a simple consequence of the ones for deterministic weights. On the other hand, the advantage of working with i.i.d. weights is that the vertices are exchangeable, and, in contrast to the deterministic weights case, not many assumptions need to be made. For deterministic weights, one often has to make detailed assumptions concerning the precise structure of the weights.

Notes on Section 6.2. The results in this section are novel, and are inspired by the ones in [58].

Notes on Section 6.3. The results in this section are novel, and are inspired by the ones in [58].

Notes on Section 6.4. Theorem 6.13 is [58, Proof of Theorem 3.2], whose proof we follow. Exercise 6.19 is novel.

Notes on Section 6.5. The proof in Section 6.5 follows the argument in [58, Section 3].

Notes on Section 6.6. Theorem 6.17 is [131, Corollary 2.12]. In [131], there are many more examples and results, also investigating the notion of *asymptotic contiguity* of random graphs, which is a slightly weaker notion than asymptotic equivalence, and holds when events that have vanishing probability under one measure also have vanishing probabilities under the other. There are deep relations between convergence in probability and in distribution and asymptotic equivalence and contiguity, see [131, Remark 1.4].

Notes on Section 6.7. The expected degree random graph, or Chung-Lu model, has been studied extensively by Chung and Lu in [66, 67, 68, 72, 161]. See in particular the recent book [70], in which many of these results are summarized.

Chapter 7
Configuration model

In this chapter, we investigate graphs with fixed degrees. Ideally, we would like to investigate uniform graphs having a prescribed degree sequence, i.e, a degree sequence which is given to us beforehand. An example of such a situation could arise from a real-world network, of which we know the degree sequence, and we would be interested in generating a random graph with *precisely the same* degrees. We start by discussion a few examples where such situation appears naturally.

Example 7.1 (Population of two types revisited)**.** Suppose that we have a complex network in which two distinct types of vertices are present. The first type has precisely m_1 neighbors, the second type precisely m_2. How can we construct a random graph satisfying these restrictions?

Example 7.2 (Regular graphs)**.** How many simple graphs are there in which every vertex has degree precisely r? How can we generate a random instance of such a graph?

Example 7.3 (Real-world network and its degrees)**.** Suppose that we have a complex network of size n in which vertex $i \in [n]$ has degree d_i. How can we decide whether this network resembles a uniform random graph with the same degree sequence, or whether it inherently has more structure? For this, we would need to be able to generate a random graph having precisely the right degree sequence.

As it turns out, it is not a trivial task to generate graphs having prescribed degrees, in particular, because they may not exist (recall (I.3) on page 114). We shall therefore introduce a model that produces a *multigraph* with the prescribed degrees, and which, when conditioned on simplicity, is uniform over all simple graphs with the prescribed degree sequence. This random multigraph is called the *configuration model*. We shall discuss the connections between the configuration model and a uniform simple random graph having the same degree sequence, and give an asymptotic formula for the number of simple graphs with a given degree sequence.

This chapter is organized as follows. In Section 7.1, we shall introduce the configuration model. In Sections 7.2, we shall investigate properties of the configuration model, given that the degrees satisfy some regularity conditions. We shall investigate two ways of turning the configuration model into a simple graph, namely, by erasing the self-loops and multiple edges, or by conditioning on obtaining a simple graph. For the latter, we compute the asymptotic probability of the configuration model to be simple. This also allows us to compute the asymptotic number of graphs with a given degree sequence in the case where the degrees are not too large. In Section 7.4, we shall discuss the tight relations that exist between the configuration model conditioned on being simple, and the generalized random graph conditioned on its degrees. This relation shall prove to be quite useful when deducing results for the generalized random graph from those for the configuration model. In Section 7.5, we treat the special case of i.i.d. degrees. We close this chapter in Section 7.6 with notes and discussion.

7.1 Introduction to the model

Fix an integer n. Consider a sequence $\boldsymbol{d} = (d_i)_{i\in[n]}$. The aim is to construct an undirected (multi)graph with n vertices, where vertex j has degree d_j. Without loss of generality, throughout this chapter, we shall assume that $d_j \geq 1$ for all $j \in [n]$, since when

$d_j = 0$, vertex j is isolated and can be removed from the graph. One possible random graph model is then to take the uniform measure over such undirected and simple graphs. Here, we call a graph simple when it has no self-loops and no multiple edges between any pair of vertices. However, the set of undirected simple graphs with n vertices where vertex j has degree d_j may be empty. For example, in order for such a graph to exist, we must assume that the total degree

$$\ell_n = \sum_{j\in[n]} d_j \tag{7.1.1}$$

is even. We wish to construct a simple graph such that $(d_i)_{i\in[n]}$ are the degrees of the n vertices. However, even when $\ell_n = \sum_{j\in[n]} d_j$ *is* even, this is not always possible, as explained in more detail in (I.3) on page 114.

Exercise 7.1 (Non-graphical degree sequence)**.** *Find a simple example of a $(d_i)_{i\in[n]}$ satisfying that $\ell_n = \sum_{j\in[n]} d_j$ is even, for which there is no simple graph where vertex i has degree d_i.*

Since it is not always possible to construct a simple graph with a given degree sequence, instead, we can construct a *multigraph*, that is, a graph *possibly* having self-loops and multiple edges between pairs of vertices. One way of obtaining a uniform multigraph with the given degree sequence is to pair the half-edges attached to the different vertices in a uniform way. Two half-edges together form an edge, thus creating the edges in the graph.

To construct the multigraph where vertex j has degree d_j for all $j \in [n]$, we have n separate vertices and incident to vertex j, we have d_j half-edges. Every half-edge needs to be connected to another half-edge to build the graph. The half-edges are numbered in an arbitrary order from 1 to ℓ_n. We start by randomly connecting the first half-edge with one of the $\ell_n - 1$ remaining half-edges. Once paired, two half-edges form a single edge of the multigraph. Hence, a half-edge can be seen as the left or the right half of an edge. We continue the procedure of randomly choosing and pairing the half-edges until all half-edges are connected, and call the resulting graph the *configuration model with degree sequence* $\boldsymbol{d}$, abbreviated as $\mathrm{CM}_n(\boldsymbol{d})$.

Unfortunately, vertices having self-loops, as well as multiple edges may occur. However, we shall see that self-loops and multiple edges are scarce when $n \to \infty$. Clearly, when the total degree $\ell_n = \sum_{j\in[n]} d_j$ is even, then the above procedure produces a multigraph with the right degree sequence. Here, in the degree sequence of the multigraph, a self-loop contributes two to the degree of the vertex incident to it, while each of the multiple edges contributes one to the degree of each of the two vertices incident to it.

To explain the term configuration model, we now present an equivalent way of defining the configuration model. For this, we construct a *second graph*, with vertices $1, \dots, \ell_n$. These vertices in the new graph will correspond to the edges of the random multigraph in the configuration model. We pair the vertices in a uniform way to produce a *uniform matching*. For this, we pair vertex 1 with a uniform other vertex. After this, we pair the first not yet paired vertex to a uniform vertex which is not yet paired. The procedure stops when all vertices are paired to another (unique) vertex. We denote the resulting graph by $\mathrm{Conf}_n(\boldsymbol{d})$. Thus, $\mathrm{Conf}_n(\boldsymbol{d})$ can be written as $\mathrm{Conf}_n(\boldsymbol{d}) = \{i\sigma(i)\colon i \in [\ell_n]\}$, where $\sigma(i)$ is the label of the vertex to which vertex $i \in [\ell_n]$ is paired. The pairing of the vertices $1, \dots, \ell_n$ is called a *configuration*, and each configuration has the same probability.

Exercise 7.2 (The number of configurations)**.** *Prove that there are $(2m-1)!! = (2m-1)(2m-3)\cdots 3\cdot 1$ different ways of pairing vertices $1, \dots, 2m$.*

To construct the graph of the configuration model from the above configuration, we *identify* vertices $1, \dots, d_1$ in $\mathrm{Conf}_n(\boldsymbol{d})$ to form vertex 1 in $\mathrm{CM}_n(\boldsymbol{d})$, and vertices $d_1 + 1, \dots, d_1 + d_2$ in $\mathrm{Conf}_n(\boldsymbol{d})$ to form vertex 2 in $\mathrm{CM}_n(\boldsymbol{d})$, etc. Therefore, precisely d_j vertices in $\mathrm{Conf}_n(\boldsymbol{d})$ are identified with vertex j in $\mathrm{CM}_n(\boldsymbol{d})$.

In the above identification, the number of edges in $\mathrm{CM}_n(\boldsymbol{d})$ between vertices $i, j \in [n]$ is the number of vertices in $\mathrm{Conf}_n(\boldsymbol{d})$ that are identified with $i \in \mathrm{CM}_n(\boldsymbol{d})$ and are paired to the vertex in $\mathrm{Conf}_n(\boldsymbol{d})$ that is identified with vertex $j \in \mathrm{CM}_n(\boldsymbol{d})$. As a consequence, the degree of vertex j in $\mathrm{CM}_n(\boldsymbol{d})$ is *precisely* equal to d_j. The resulting graph is a *multigraph*, since both self-loops and multiple edges between vertices are possible. We can identify the graph as $\mathrm{CM}_n(\boldsymbol{d}) = (X_{ij})_{i,j\in[n]}$, where X_{ij} is the number of edges between vertices $i, j \in [n]$ and X_{ii} is the number of self-loops of vertex $i \in [n]$, so that, for all $i \in [n]$,

$$d_i = X_{ii} + \sum_{j\in[n]} X_{ij}. \tag{7.1.2}$$

Here, the number of self-loops of vertex i, X_{ii}, appears twice, so that a self-loop contributes 2 to the degree. Since the uniform matching of the ℓ_n vertices in $\mathrm{Conf}_n(\boldsymbol{d})$ is sometimes referred to as the *configuration*, the resulting graph $\mathrm{CM}_n(\boldsymbol{d})$ is called *the configuration model*.

We note (see e.g. [132, Section 1]) that not all multigraph has the same probability, i.e., not every multigraph is equally likely and the measure obtained is not the uniform measure on all multigraphs with the prescribed degree sequence. Indeed, there is a weight $1/j!$ for every edge of multiplicity j, and a factor $1/2$ for every self-loop:

Proposition 7.4 (The law of $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Let $G = (x_{ij})_{i,j\in[n]}$ be a multigraph on the vertices $[n]$ which is such that*

$$d_i = x_{ii} + \sum_{j\in[n]} x_{ij}. \tag{7.1.3}$$

Then,

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) = G) = \frac{1}{(\ell_n - 1)!!} \frac{\prod_{i\in[n]} d_i!}{\prod_{i\in[n]} 2^{x_{ii}} \prod_{1\le i\le j\le n} x_{ij}!}. \tag{7.1.4}$$

Proposition 7.4 implies that if we condition on the graph as being simple, then the resulting graph is a uniform *simple* graph with the prescribed degree sequence. Here, we call a graph $G = (x_{ij})_{i,j\in[n]}$ simple whenever $x_{ij} \in \{0, 1\}$ for every $i, j \in [n]$ with $i \neq j$, and $x_{ii} = 0$ for every $i \in [n]$, i.e., there are no multiple edges and no self-loops.

Proof. By Exercise 7.2, the number of configurations is equal to $(\ell_n - 1)!!$. Each configuration has the same probability, so that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) = G) = \frac{1}{(\ell_n - 1)!!} N(G), \tag{7.1.5}$$

where $N(G)$ is the number of configurations that, after identifying the vertices, give the multigraph G. We note that if we permute the half-edges incident to a vertex, then the resulting multigraph remains unchanged, and there are precisely $\prod_{i\in[n]} d_i!$ ways to permute the half-edges incident to all vertices. Some of these permutations, however, give rise to the same configuration. The factor $x_{ij}!$ compensates for the multiple edges between vertices $i, j \in [n]$, and the factor $2^{x_{ii}}$ compensates for the fact that the paring kl and lk in $\mathrm{Conf}_n(\boldsymbol{d})$ give rise to the same configuration. □

Exercise 7.3 (Example of multigraph)**.** *Let $n = 2$, $d_1 = 2$ and $d_2 = 4$. Use the direct connection probabilities to show that the probability that $\mathrm{CM}_n(\boldsymbol{d})$ consists of 3 self-loops equals $1/5$. Hint: Note that when $d_1 = 2$ and $d_2 = 4$, the graph $\mathrm{CM}_n(\boldsymbol{d})$ consists only of self-loops* precisely *when the first half-edge of vertex 1 connects to the second half-edge of vertex 1.*

Exercise 7.4 (Example of multigraph (Cont.))**.** *Let $n = 2$, $d_1 = 2$ and $d_2 = 4$. Use Proposition 7.4 to show that the probability that $\mathrm{CM}_n(\boldsymbol{d})$ consists of 3 self-loops equals $1/5$.*

The flexibility in choosing the degree sequence $\boldsymbol{d}$ gives us a similar flexibility as in choosing the vertex weights $\boldsymbol{w}$ in Chapter 6. However, in this case, the choice of the vertex degrees gives a much more direct control over the topology of the graph. For example, for $\mathrm{CM}_n(\boldsymbol{d})$, it is possible to build graphs with fixed degrees, or where all degrees are at least a certain value. In many applications, such flexibility is rather convenient. For example, it allows us to generate a (multi)graph with *precisely* the same degrees as a real-world network, so that we can investigate whether the real-world network is similar to it or not.

The configuration model with fixed degrees has a long history, see e.g. [44, Section 2.4]. One specific example is to take the degrees all equal, in which case we speak of a *random regular graph*.

As in Chapter 6, we shall again impose *regularity conditions* on the degree sequence $\boldsymbol{d}$. In order to state these assumptions, we introduce some notation. We denote the degree of a uniformly chosen vertex V in $[n]$ by $D_n = d_V$. The random variable D_n has distribution function F_n given by

$$F_n(x) = \frac{1}{n} \sum_{j \in [n]} \mathbb{1}_{\{d_j \leq x\}}. \tag{7.1.6}$$

We assume that the vertex degrees satisfy the following *regularity conditions:*

Condition 7.5 (Regularity conditions for vertex degrees)**.**
(a) Weak convergence of vertex weight.
There exists a distribution function F such that

$$D_n \xrightarrow{d} D, \tag{7.1.7}$$

where D_n and D have distribution functions F_n and F, respectively. Equivalently, for any x,

$$\lim_{n \to \infty} F_n(x) = F(x). \tag{7.1.8}$$

(b) Convergence of average vertex degrees.

$$\lim_{n \to \infty} \mathbb{E}[D_n] = \mathbb{E}[D], \tag{7.1.9}$$

where D_n and D have distribution functions F_n and F, respectively. Further, we assume that $\mathbb{P}(D \geq 1) = 1$.
(c) Convergence of second moment vertex degrees.

$$\lim_{n \to \infty} \mathbb{E}[D_n^2] = \mathbb{E}[D^2]. \tag{7.1.10}$$

Similarly to Condition 6.4 in Chapter 6, we shall almost always assume that Condition 7.5(a)-(b) hold, and only sometimes assume Condition 7.5(c). We note that, since d_i only takes values in the integers, so does D_n, and therefore so must the limiting random variable D. As a result, the limiting distribution function F is constant between integers, and makes a jump $\mathbb{P}(D = x)$ at $x \in \mathbb{N}$. As a result, the distribution function F does have discontinuity points, and the weak convergence in (7.1.7) usually only implies (7.1.8) at continuity points. However, since F_n is constant in between integers, we do obtain the implication:

Exercise 7.5 (Weak convergence integer random variables)**.** *Let (D_n) be a sequence of integer random variables such that $D_n \xrightarrow{d} D$. Show that, for all $x \in \mathbb{R}$,*

$$\lim_{n \to \infty} F_n(x) = F(x), \tag{7.1.11}$$

and that also $\lim_{n \to \infty} \mathbb{P}(D_n = x) = \mathbb{P}(D = x)$ for every $x \in \mathbb{N}$.

Instead of defining $\mathrm{CM}_n(\boldsymbol{d})$ in terms of the degrees, we could have defined it in terms of the number of vertices with fixed degrees. Indeed, let

$$n_k = \sum_{i\in[n]} \mathbb{1}_{\{d_i=k\}} \tag{7.1.12}$$

denote the number of vertices with degree k. Then, clearly, apart from the vertex labels, the degree sequence $\boldsymbol{d}$ is uniquely determined by the sequence $(n_k)_{k\geq 0}$. Then, Condition 7.5(a) is equivalent to $\lim_{n\to\infty} n_k/n = \mathbb{P}(D=k)$, while Condition 7.5(b) is equivalent to $\lim_{n\to\infty}\sum_{k\geq 0} kn_k/n = \mathbb{E}[D]$.

We next describe two canonical ways of obtaining a degree sequence $\boldsymbol{d}$ such that Condition 7.5 holds.

The configuration model with fixed degrees moderated by F. Fix a distribution function F of an integer random variable D. We take the number of vertices with degree k to be equal to

$$n_k = \lceil nF(k)\rceil - \lceil nF(k-1)\rceil, \tag{7.1.13}$$

and take the corresponding degree sequence $\boldsymbol{d} = (d_i)_{i\in[n]}$ the unique ordered degree sequence compatible with $(n_k)_{k\geq 0}$. Clearly, for this sequence, Condition 7.5(a) is satisfied:

Exercise 7.6 (Regularity condition for configuration model moderated by F)**.** *Fix* $\mathrm{CM}_n(\boldsymbol{d})$ *be such that there are precisely* $n_k = \lceil nF(k)\rceil - \lceil nF(k-1)\rceil$ *vertices with degree* k*. Show that Condition 7.5(a) holds.*

The nice thing about our example is that

$$F_n(k) = \frac{1}{n}\lceil nF(k)\rceil. \tag{7.1.14}$$

In particular, $D_n \preceq D$, since $F_n(x) \geq F(x)$ for every x. As a result, Condition 7.5(b) holds whenever $\mathbb{E}[D] < \infty$, and Condition 7.5(c) whenever $\mathbb{E}[D^2] < \infty$:

Exercise 7.7 (Regularity condition for configuration model moderated by F (Cont.))**.** *Fix* $\mathrm{CM}_n(\boldsymbol{d})$ *be such that there are precisely* $n_k = \lceil nF(k)\rceil - \lceil nF(k-1)\rceil$ *vertices with degree* k*. Show that Condition 7.5(b) holds whenever* $\mathbb{E}[D] < \infty$*.*

The configuration model with i.i.d. degrees. The next canonical example arises by assuming that the degrees $\boldsymbol{D} = (D_i)_{i\in[n]}$ are an i.i.d. sequence of random variables. When we extend the construction of the configuration model to i.i.d. degrees $\boldsymbol{D}$, we should bear in mind that the total degree

$$L_n = \sum_{i\in[n]} D_i \tag{7.1.15}$$

is odd with probability close to $1/2$, as the following exercise shows:

Exercise 7.8 (Probability of i.i.d. sum to be odd)**.** *Assume that* $(D_i)_{i\geq 1}$ *is an i.i.d. sequence of random variables. Prove that* $L_n = \sum_{i\in[n]} D_i$ *is odd with probability close to* $1/2$*. For this, note that*

$$\mathbb{P}(L_n \text{ is odd}) = \frac{1}{2}\Big[1 - \mathbb{E}[(-1)^{L_n}]\Big]. \tag{7.1.16}$$

Then compute

$$\mathbb{E}[(-1)^{L_n}] = \phi_{D_1}(\pi)^n, \tag{7.1.17}$$

where

$$\phi_{D_1}(t) = \mathbb{E}[\mathrm{e}^{itD_1}] \tag{7.1.18}$$

is the characteristic function of the degree D_1. Prove that, when $\mathbb{P}(D \text{ even}) \neq 1$, $|\phi_{D_1}(\pi)| < 1$, so that $\mathbb{P}(L_n \text{ is odd})$ is exponentially close to $\frac{1}{2}$.

There are different possible solutions to overcome the problem of an odd total degree L_n, each producing a graph with similar characteristics. We make use of the following solution: If L_n is odd, then we add a half-edge to the n^{th} vertex, so that D_n is increased by 1, i.e., $d_i = D_i + \mathbb{1}_{\{L_n \text{ odd}, i=n\}}$. This single half-edge will make hardly any difference in what follows, and we will ignore this effect. Also, we warn the reader that now D_n has two distinct meanings. The first is the distribution of the degree of a random vertex $D_n = d_V$, the second the n^{th} element of the sequence $\boldsymbol{D} = (D_i)_{i\in[n]}$. In what follows, we shall always be clear about the meaning of D_n, which is always equal to $D_n = d_V$ unless explicitly stated otherwise.

It is not hard to see that Condition 7.5 follows from the Law of Large Numbers:

Exercise 7.9 (Regularity condition for configuration model with i.i.d. degrees)**.** *Fix $\mathrm{CM}_n(\boldsymbol{d})$ with degrees $\boldsymbol{d}$ given by $d_i = D_i + \mathbb{1}_{\{L_n \text{ odd}, i=n\}}$, where $(D_i)_{i\in[n]}$ is an i.i.d. sequence of integer random variables. Show that Condition 7.5(a) holds, whereas Condition 7.5(b) and (c) hold when $\mathbb{E}[D]$ and $\mathbb{E}[D^2]$, respectively, are finite. Here the convergence is replaced with convergence in probability.*

Organization of the remaining chapter. In this chapter, we study the configuration model both with fixed degrees, as well as with i.i.d. degrees. We focus on two main results. The first main result shows that when we erase all self-loops and combine the multiple edges into one, then we obtain a graph with asymptotically the same degree sequence. This model is also referred to as the *erased configuration model*, see also [58, Section 2.1].

In the second main result, we investigate the probability that the configuration model actually produces a *simple graph*. Remarkably, even though there *could* be many self-loops and multiple edges, in the case when the degrees are not too large, there is an asymptotically *positive* probability that the configuration model produces a simple graph. Therefore, we may obtain a uniform *simple* random graph by repeating the procedure until we obtain a simple graph. As a result, this model is sometimes called the *repeated* configuration model. The fact that the configuration model yields a simple graph with asymptotically positive probability has many interesting consequences that we shall explain in some detail. For example, it allows us to compute the asymptotics of the number of simple graphs with a given degree sequence.

7.2 Erased configuration model

We first define the erased configuration model. We fix the degrees $\boldsymbol{d}$. We start with the multigraph $\mathrm{CM}_n(\boldsymbol{d})$ and erase all self-loops, if any exist. After this, we merge all multiple edges into single edges. Therefore, the erased configuration model yields a simple random graph, where two vertices are connected by an edge if and only if there is (at least one) edge connecting them in the original multigraph definition of the configuration model.

We next introduce some notation. We denote the degrees in the erased configuration model by $\boldsymbol{D}^{(\mathrm{er})} = (D_i^{(\mathrm{er})})_{i\in[n]}$, so that

$$D_i^{(\mathrm{er})} = d_i - 2s_i - m_i, \tag{7.2.1}$$

where $(d_i)_{i\in[n]}$ are the degrees in the configuration model, $s_i = x_{ii}$ is the number of self-loops of vertex i in the configuration model, and

$$m_i = \sum_{j\neq i}(x_{ij}-1)\mathbb{1}_{\{x_{ij}\geq 2\}} \tag{7.2.2}$$

is the number of multiple edges removed from i.

Denote the empirical degree sequence $(p_k^{(n)})_{k\geq 1}$ in the configuration model by

$$p_k^{(n)} = \frac{1}{n}\sum_{i\in[n]}\mathbb{1}_{\{d_i=k\}}, \tag{7.2.3}$$

and denote the related degree sequence in the erased configuration model $(P_k^{(\mathrm{er})})_{k\geq 1}$ by

$$P_k^{(\mathrm{er})} = \frac{1}{n}\sum_{i\in[n]}\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}. \tag{7.2.4}$$

From the notation it is clear that $(p_k^{(n)})_{k\geq 1}$ is a *deterministic* sequence since $(d_i)_{i\in[n]}$ is deterministic, while $(P_k^{(\mathrm{er})})_{k\geq 1}$ is a *random* sequence, since the erased degrees $(D_i^{(\mathrm{er})})_{i\in[n]}$ is a random vector.

Exercise 7.10 (Mean degree sequence equals average degree)**.** *Prove that*

$$\sum_{k=1}^{\infty} k p_k^{(n)} = \frac{1}{n}\sum_{i\in[n]} d_i = \frac{\ell_n}{n}. \tag{7.2.5}$$

Now we are ready to state the main result concerning the degree sequence of the erased configuration model:

Theorem 7.6 (Degree sequence of erased configuration model with fixed degrees)**.** *For fixed degrees $\boldsymbol{d}$ satisfying Condition 7.5(a)-(b), the degree sequence of the erased configuration model $(P_k^{(\mathrm{er})})_{k\geq 1}$ converges to $(p_k)_{k\geq 1}$. More precisely, for every $\varepsilon > 0$,*

$$\mathbb{P}\Big(\sum_{k=1}^{\infty}|P_k^{(\mathrm{er})} - p_k| \geq \varepsilon\Big) \to 0. \tag{7.2.6}$$

Proof. By Condition 7.5(a) and the fact that pointwise convergence of a probability mass function is equivalent to convergence in total variation distance (recall Exercise 2.14), we obtain that

$$\lim_{n\to\infty}\sum_{k=1}^{\infty}|p_k^{(n)} - p_k| = 0. \tag{7.2.7}$$

Therefore, we can take n so large that

$$\sum_{k=1}^{\infty}|p_k^{(n)} - p_k| \leq \varepsilon/2. \tag{7.2.8}$$

We start by proving the result under the extra assumption that

$$\max_{i\in[n]} d_i = o(\sqrt{n}), \tag{7.2.9}$$

For this, we bound $\mathbb{P}(\sum_{k=1}^{\infty}|P_k^{(\mathrm{er})}-p_k^{(n)}|\geq\varepsilon/2)$. For this, we use (7.2.1), which implies that $D_i^{(\mathrm{er})}\neq d_i$ if and only if $2s_i+m_i\geq 1$. We use

$$\sum_{k=1}^{\infty}|P_k^{(\mathrm{er})}-p_k^{(n)}|\leq\frac{1}{n}\sum_{k=1}^{\infty}\sum_{i}|\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}-\mathbb{1}_{\{d_i=k\}}|, \tag{7.2.10}$$

and write out that

$$\begin{aligned}\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}-\mathbb{1}_{\{d_i=k\}}&=\mathbb{1}_{\{D_i^{(\mathrm{er})}=k,d_i>k\}}-\mathbb{1}_{\{D_i^{(\mathrm{er})}<k,d_i=k\}}\\&=\mathbb{1}_{\{s_i+m_i>0\}}\big(\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}-\mathbb{1}_{\{d_i=k\}}\big).\end{aligned} \tag{7.2.11}$$

Therefore,

$$|\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}-\mathbb{1}_{\{d_i=k\}}|\leq\mathbb{1}_{\{s_i+m_i>0\}}\big(\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}+\mathbb{1}_{\{d_i=k\}}\big), \tag{7.2.12}$$

so that

$$\begin{aligned}\sum_{k=1}^{\infty}|P_k^{(\mathrm{er})}-p_k^{(n)}|&\leq\frac{1}{n}\sum_{k=1}^{\infty}\sum_{i\in[n]}|\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}-\mathbb{1}_{\{d_i=k\}}|\\&\leq\frac{1}{n}\sum_{i\in[n]}\mathbb{1}_{\{s_i+m_i>0\}}\sum_{k=1}^{\infty}\big(\mathbb{1}_{\{D_i^{(\mathrm{er})}=k\}}+\mathbb{1}_{\{d_i=k\}}\big)\\&=\frac{2}{n}\sum_{i\in[n]}\mathbb{1}_{\{s_i+m_i>0\}}\leq\frac{2}{n}\sum_{i\in[n]}(s_i+m_i).\end{aligned} \tag{7.2.13}$$

We denote the number of self-loops by S_n and the number of multiple edges by M_n, that is

$$S_n=\sum_{i\in[n]}s_i,\qquad M_n=\frac{1}{2}\sum_{i\in[n]}m_i. \tag{7.2.14}$$

Then, by (7.2.13),

$$\mathbb{P}\Big(\sum_{k=1}^{\infty}|P_k^{(\mathrm{er})}-p_k^{(n)}|\geq\varepsilon/2\Big)\leq\mathbb{P}\big(2S_n+4M_n\geq\varepsilon n/2\big), \tag{7.2.15}$$

so that Theorem 7.6 follows if

$$\mathbb{P}(2S_n+4M_n\geq\varepsilon n/2)\to 0. \tag{7.2.16}$$

By the Markov inequality (Theorem 2.14), we obtain

$$\mathbb{P}(2S_n+4M_n\geq\varepsilon n/2)\leq\frac{4}{\varepsilon n}\big(\mathbb{E}[S_n]+2\mathbb{E}[M_n]\big). \tag{7.2.17}$$

Bounds on $\mathbb{E}[S_n]$ and $\mathbb{E}[M_n]$ are provided in the following proposition:

Proposition 7.7 (Bounds on the expected number of self-loops and multiple edge)**.** *The expected number of self-loops S_n in the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ satisfies*

$$\mathbb{E}[S_n]\leq\sum_{i\in[n]}\frac{d_i^2}{\ell_n}, \tag{7.2.18}$$

while the expected number of multiple edges M_n satisfies

$$\mathbb{E}[M_n]\leq 2\Big(\sum_{i\in[n]}\frac{d_i^2}{\ell_n}\Big)^2. \tag{7.2.19}$$

Proof. For a vertex i, and for $1 \leq s < t \leq d_i$, we define $I_{st,i}$ to be the indicator of the event that the half-edge s is paired to the half-edge t. Here we number the half-edges, or half-edges, of the vertices in an arbitrary way. Then

$$S_n = \sum_{i\in[n]} \sum_{1\leq s<t\leq d_i} I_{st,i}. \tag{7.2.20}$$

Therefore,

$$\mathbb{E}[S_n] = \sum_{i\in[n]} \sum_{1\leq s<t\leq d_i} \mathbb{E}[I_{st,i}] = \sum_{i\in[n]} \frac{1}{2} d_i(d_i-1)\mathbb{E}[I_{12,i}], \tag{7.2.21}$$

since the probability of producing a self-loop by pairing the half-edges s and t does not depend on s and t. Now, $\mathbb{E}[I_{12,i}]$ is equal to the probability that half-edges 1 and 2 are paired to each other, which is equal to $(\ell_n - 1)^{-1}$. Therefore,

$$\mathbb{E}[S_n] = \frac{1}{2} \sum_{i\in[n]} \frac{d_i(d_i-1)}{\ell_n - 1} \leq \sum_{i\in[n]} \frac{d_i^2}{\ell_n}. \tag{7.2.22}$$

Similarly, for vertices i and j, and for $1 \leq s_1 < s_2 \leq d_i$ and $1 \leq t_1 \neq t_2 \leq d_j$, we define $I_{s_1t_1,s_2t_2,ij}$ to be the indicator of the event that the half-edge s_1 is paired to the half-edge t_1 and half-edge s_2 is paired to the half-edge t_2. If $I_{s_1t_1,s_2t_2,ij} = 1$ for some s_1t_1 and s_2t_2, then there are multiple edges between vertices i and j. It follows that

$$M_n \leq \frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} I_{s_1t_1,s_2t_2,ij}, \tag{7.2.23}$$

so that

$$\begin{aligned}\mathbb{E}[M_n] &\leq \frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} \mathbb{E}[I_{s_1t_1,s_2t_2,ij}] \\ &= \frac{1}{4} \sum_{1\leq i\neq j\leq n} d_i(d_i-1)d_j(d_j-1)\mathbb{E}[I_{11,22,ij}].\end{aligned} \tag{7.2.24}$$

Now, since $I_{11,22,ij}$ is an indicator, $\mathbb{E}[I_{11,22,ij}]$ is the probability that $I_{11,22,ij} = 1$, which is equal to the probability that half-edge 1 of vertex i and half-edge 1 of vertex j, as well as half-edge 2 of vertex i and half-edge 2 of vertex j are paired, which is equal to

$$\mathbb{E}[I_{11,22,ij}] = \frac{1}{(\ell_n - 1)(\ell_n - 3)}. \tag{7.2.25}$$

Therefore,

$$\mathbb{E}[M_n] \leq \sum_{i,j=1}^{n} \frac{d_i(d_i-1)d_j(d_j-1)}{4(\ell_n-1)(\ell_n-3)} = \frac{\Big(\sum_{i\in[n]} d_i(d_i-1)\Big)^2}{4(\ell_n-1)(\ell_n-3)} \leq \frac{2\Big(\sum_{i\in[n]} d_i(d_i-1)\Big)^2}{\ell_n^2}, \tag{7.2.26}$$

where we use that $8(\ell_n - 1)(\ell_n - 3) \geq \ell_n^2$ since $\ell_n \geq 4$. Since $M_n = 0$ with probability one when $\ell_n \leq 3$, the claim follows. □

To complete the proof of Theorem 7.6 in the case that $\max_{i\in[n]} d_i = o(\sqrt{n})$ (recall (7.2.9)), we apply Proposition 7.7, we obtain

$$\mathbb{E}[S_n] \le \sum_{i\in[n]} \frac{d_i^2}{\ell_n} \le \max_{i\in[n]} d_i = o(\sqrt{n}). \tag{7.2.27}$$

The bound on $\mathbb{E}[M_n]$ is similar. By (7.2.17), this proves the claim.

To prove the result assuming only Condition 7.5(a)-(b), we start by noting that Condition 7.5(a)-(b) implies that $\max_{i\in[n]} d_i = o(n)$ (recall, e.g., Exercise 6.3). We note that $\sum_{k=1}^{\infty} |P_k^{(\mathrm{er})} - p_k^{(n)}| \ge \varepsilon$ implies that the degrees of at least εn vertices are changed by the erasure procedure. Take $a_n \to \infty$ arbitrarily slowly, such that there are at most $\varepsilon n/2$ vertices $i \in [n]$ of degree $d_i \ge a_n$. Then, $\sum_{k=1}^{\infty} |P_k^{(\mathrm{er})} - p_k^{(n)}| \ge \varepsilon$ implies that the number of vertices of degree at most a_n whose degrees are changed by the erasure procedure is at least $\varepsilon n/2$. Let

$$S_n(a_n) = \sum_{i\in[n]} s_i \mathbb{1}_{\{d_i \le a_n\}}, \qquad M_n(a_n) = \frac{1}{2}\sum_{i\in[n]} m_i \mathbb{1}_{\{d_i \le a_n\}} \tag{7.2.28}$$

denote the number of self-loops and multiple edge incident to vertices of degree at most a_n. Then, it is straightforward to adapt Proposition 7.7 to show that

$$\mathbb{E}[S_n(a_n)] \le \sum_{i\in[n]} \frac{d_i^2 \mathbb{1}_{\{d_i \le a_n\}}}{\ell_n}, \qquad \mathbb{E}[M_n(a_n)] \le 2\sum_{i\in[n]} \frac{d_i^2 \mathbb{1}_{\{d_i \le a_n\}}}{\ell_n} \sum_{j\in[n]} \frac{d_j^2}{\ell_n}. \tag{7.2.29}$$

Therefore, $\mathbb{E}[S_n(a_n)] \le a_n, \mathbb{E}[M_n(a_n)] \le a_n \max_{j\in[n]} d_j$. Take a_n so small that $a_n \max_{j\in[n]} d_j = o(n)$ (which is possible since $\max_{j\in[n]} d_j = o(n)$), then

$$\mathbb{P}(2S_n(a_n) + 4M_n(a_n) \ge \varepsilon n/2) \le \frac{4}{\varepsilon n}\big(\mathbb{E}[S_n(a_n)] + 2\mathbb{E}[M_n(a_n)]\big) = o(1), \tag{7.2.30}$$

as required. □

7.3 Repeated configuration model and probability simplicity

In this section, we investigate the probability that the configuration model yields a simple graph, i.e., the probability that the graph produced in the configuration model has no self-loops nor multiple edges. Then the asymptotics of the probability that the configuration model is simple is derived in the following theorem:

Theorem 7.8 (Probability of simplicity of $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies Condition 7.5(a)-(c). Then, the probability that $\mathrm{CM}_n(\boldsymbol{d})$ is a simple graph is asymptotically equal to $\mathrm{e}^{-\nu/2-\nu^2/4}$, where*

$$\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]. \tag{7.3.1}$$

Theorem 7.8 is a consequence of the following result:

Proposition 7.9 (Poisson limit of self-loops and multiple edges)**.** *Assume that $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies Condition 7.5(a)-(c). Then (S_n, M_n) converges in distribution to (S, M), where S and M are two* independent *Poisson random variables with means $\nu/2$ and $\nu^2/4$.*

Indeed, Theorem 7.8 is a simple consequence of Proposition 7.9, since $\mathrm{CM}_n(\boldsymbol{d})$ is simple precisely when $S_n = M_n = 0$. By the weak convergence result stated in Proposition 7.9 and the independence of S and M, the probability that $S_n = M_n = 0$ converges to $\mathrm{e}^{-\mu_S - \mu_M}$, where μ_S and μ_M are the means of the limiting Poisson random variables S and M. Using the identification of the means of S and M in Proposition 7.9, this completes the proof of Theorem 7.8. We are left to prove Proposition 7.9.

Proof of Proposition 7.9. Throughout the proof, we shall assume that S and M are two independent Poisson random variables with means $\nu/2$ and $\nu^2/4$.

We make use of Theorem 2.6 which imply that it suffices to prove that the factorial moments converge. Also, S_n is a sum of indicators, so that we can use Theorem 2.7 to identify its factorial moments. For M_n, this is not so clear. However, we define

$$\tilde{M}_n = \sum_{1\le i<j\le n} \sum_{1\le s_1<s_2\le d_i} \sum_{1\le t_1\neq t_2\le d_j} I_{s_1t_1,s_2t_2,ij}, \tag{7.3.2}$$

so that, by (7.2.23), $M_n \le \tilde{M}_n$. We shall first show that with high probability $M_n = M_n'$. Note that $M_n < \tilde{M}_n$ precisely when there exist vertices $i \neq j$ such that there are at least *three* edges between i and j. The probability that there are at least three edges between i and j is bounded above by

$$\frac{d_i(d_i-1)(d_i-2)d_j(d_j-1)(d_j-2)}{(\ell_n-1)(\ell_n-3)(\ell_n-5)}. \tag{7.3.3}$$

Thus, by Boole's inequality, the probability that there exist vertices $i \neq j$ such that there are at least *three* edges between i and j is bounded above by

$$\sum_{i,j=1}^{n} \frac{d_i(d_i-1)(d_i-2)d_j(d_j-1)(d_j-2)}{(\ell_n-1)(\ell_n-3)(\ell_n-5)} = o(1), \tag{7.3.4}$$

since $d_i = o(\sqrt{n})$ when Condition 7.5(a)-(c) holds (recall Exercise 6.3) as well as $\ell_n \ge n$. We conclude that the probability that there are $i, j \in [n]$ such that there are at least three edges between i and j is $o(1)$ as $n \to \infty$. As a result, (S_n, M_n) converges in distribution to (S, M) precisely when $(S_n, \tilde{M}_n)$ converges in distribution to (S, M).

To prove that $(S_n, \tilde{M}_n)$ converges in distribution to (S, M), we use Theorem 2.6 to see that we are left to prove that, for every $s, r \ge 0$,

$$\lim_{n\to\infty} \mathbb{E}[(S_n)_s(\tilde{M}_n)_r] = \Big(\frac{\nu}{2}\Big)^s \Big(\frac{\nu^2}{4}\Big)^r. \tag{7.3.5}$$

By Theorem 2.7,

$$\mathbb{E}[(S_n)_s(\tilde{M}_n)_r] = \sum_{\substack{m_1^{(1)},\ldots,m_s^{(1)}\in\mathcal{I}_1\\ m_1^{(2)},\ldots,m_r^{(2)}\in\mathcal{I}_2}}^{*} \mathbb{P}\big(I^{(1)}_{m_1^{(1)}} = \ldots = I^{(1)}_{m_s^{(1)}} = I^{(2)}_{m_1^{(2)}} = \ldots = I^{(2)}_{m_r^{(2)}} = 1\big), \tag{7.3.6}$$

where

$$\mathcal{I}_1 = \{(st, i) : i \in [n], 1 \le s < t \le d_i\}, \tag{7.3.7}$$

$$\mathcal{I}_2 = \{(s_1t_1, s_2t_2, i, j) : 1 \le i < j \le n, 1 \le s_1 < s_2 \le d_i, 1 \le t_1 \neq t_2 \le d_j\}, \tag{7.3.8}$$

and where, for $m^{(1)} = (st, i) \in \mathcal{I}_1$ and $m^{(2)} = (s_1t_1, s_2t_2, i, j) \in \mathcal{I}_2$,

$$I^{(1)}_{m^{(1)}} = I_{st,i}, \qquad I^{(2)}_{m^{(2)}} = I_{s_1t_1,s_2t_2,ij}. \tag{7.3.9}$$

Now, by the fact that all half-edges are uniformly paired, we have that

$$\mathbb{P}\big(I^{(1)}_{m^{(1)}_1} = \ldots = I^{(1)}_{m^{(1)}_s} = I^{(2)}_{m^{(2)}_1} = \ldots = I^{(2)}_{m^{(2)}_r} = 1\big) = \frac{1}{\prod_{i=0}^{s+2r}(\ell_n - 1 - 2i)}, \tag{7.3.10}$$

unless there is a conflict in the attachment rules, in which case

$$\mathbb{P}\big(I^{(1)}_{m^{(1)}_1} = \ldots = I^{(1)}_{m^{(1)}_s} = I^{(2)}_{m^{(2)}_1} = \ldots = I^{(2)}_{m^{(2)}_r} = 1\big) = 0. \tag{7.3.11}$$

Such a conflict arises precisely when a half-edge is required to be paired to *two* different other half-half-edges. Since the upper bound in (7.3.10) always holds, we arrive at

$$\begin{aligned}\mathbb{E}[(S_n)_s(\tilde{M}_n)_r] &\leq \sum^{*}_{m^{(1)}_1,\ldots,m^{(1)}_s \in \mathcal{I}_1} \sum^{*}_{m^{(2)}_1,\ldots,m^{(2)}_r \in \mathcal{I}_2} \frac{1}{(\ell_n-1)(\ell_n-3)\cdots(\ell_n-1-2s-4r)} \\ &= \frac{|\mathcal{I}_1|(|\mathcal{I}_1|-1)\cdots(|\mathcal{I}_1|-s+1)|\mathcal{I}_2|(|\mathcal{I}_2|-1)\cdots(|\mathcal{I}_2|-r+1)}{(\ell_n-1)(\ell_n-3)\cdots(\ell_n-1-2s-4r)}. \end{aligned} \tag{7.3.12}$$

Since $|\mathcal{I}_1|, |\mathcal{I}_2|, \ell_n$ all tend to infinity, and s, r remain fixed, we have that

$$\limsup_{n\to\infty} \mathbb{E}[(S_n)_s(\tilde{M}_n)_r] = \Big(\lim_{n\to\infty} \frac{|\mathcal{I}_1|}{\ell_n}\Big)^s \Big(\lim_{n\to\infty} \frac{|\mathcal{I}_2|}{\ell_n^2}\Big)^r. \tag{7.3.13}$$

Now,

$$\lim_{n\to\infty} \frac{|\mathcal{I}_1|}{\ell_n} = \lim_{n\to\infty} \frac{1}{\ell_n} \sum_{i\in[n]} \frac{d_i(d_i-1)}{2} = \nu/2, \tag{7.3.14}$$

by Condition 7.5(b)-(c). Further, again by Condition 7.5(b)-(c) and also using that $d_i = o(\sqrt{n})$ by Exercise 6.3, as well as $\ell_n \geq n$,

$$\begin{aligned}\lim_{n\to\infty} \frac{|\mathcal{I}_2|}{\ell_n^2} &= \lim_{n\to\infty} \frac{1}{\ell_n^2} \sum_{1\leq i<j\leq n} \frac{d_i(d_i-1)}{2} d_j(d_j-1) \\ &= \Big(\lim_{n\to\infty} \frac{1}{\ell_n} \sum_{i\in[n]} \frac{d_i(d_i-1)}{2}\Big)^2 - \lim_{n\to\infty} \sum_{i\in[n]} \frac{d_i^2(d_i-1)^2}{2\ell_n^2} = (\nu/2)^2. \end{aligned} \tag{7.3.15}$$

This provides the required upper bound.

To prove the matching lower bound, we note that, by (7.3.11),

$$\begin{aligned}&\sum^{*}_{m^{(1)}_1,\ldots,m^{(1)}_s \in \mathcal{I}_1} \sum^{*}_{m^{(2)}_1,\ldots,m^{(2)}_r \in \mathcal{I}_2} \frac{1}{\prod_{i=0}^{s+2r}(\ell_n-1-2i)} - \mathbb{E}[(S_n)_s(\tilde{M}_n)_r] \\ &\quad = \sum^{*}_{m^{(1)}_1,\ldots,m^{(1)}_s \in \mathcal{I}_1} \sum^{*}_{m^{(2)}_1,\ldots,m^{(2)}_r \in \mathcal{I}_2} \frac{I_{m^{(1)}_1,\ldots,m^{(1)}_s,m^{(2)}_1,\ldots,m^{(2)}_r}}{(\ell_n-1)(\ell_n-3)\cdots(\ell_n-1-2s-4r)}, \end{aligned} \tag{7.3.16}$$

where the indicator $I_{m^{(1)}_1,\ldots,m^{(1)}_s \in \mathcal{I}_1, m^{(2)}_1,\ldots,m^{(2)}_r}$ is equal to one precisely when there is a conflict in $m^{(1)}_1,\ldots,m^{(1)}_s, m^{(2)}_1,\ldots,m^{(2)}_r$. There is a conflict precisely when there exist a vertex i such that one of its half-edges s must be paired to two *different* half-edges. For this, there has to be a pair of indices in $m^{(1)}_1,\ldots,m^{(1)}_s$, $m^{(2)}_1,\ldots,m^{(2)}_r$ which create the conflict. There are three such possibilities: (a) the conflict is created by $m^{(1)}_a, m^{(1)}_b$ for

some a, b; (b) the conflict is created by $m_a^{(1)}, m_b^{(2)}$ for some a, b; and (c) the conflict is created by $m_a^{(2)}, m_b^{(2)}$ for some a, b. We shall bound each of these possibilities separately.

In case (a), the number of $m_c^{(1)}, c \in \{1, \ldots, s\} \setminus \{a, b\}$ and $m_d^{(2)}, d \in \{1, \ldots, r\}$ is bounded by $|\mathcal{I}_1|^{s-2}|\mathcal{I}_2|^r$. Thus, comparing with (7.3.12), we see that it suffices to prove that the number of conflicting $m_a^{(1)}, m_b^{(1)}$ is $o(|\mathcal{I}_1|^2)$. Now, the number of conflicting $m_a^{(1)}, m_b^{(1)}$ is bounded by

$$\sum_{i\in[n]} d_i^3 = o\Big(\sum_{i\in[n]} d_i(d_i-1)\Big)^2, \tag{7.3.17}$$

where we use that $d_i = o(\sqrt{n})$, as required.

In case (b), the number of $m_c^{(1)}, c \in \{1, \ldots, s\} \setminus \{a\}$ and $m_d^{(2)}, d \in \{1, \ldots, r\} \setminus \{b\}$ is bounded by $|\mathcal{I}_1|^{s-1}|\mathcal{I}_2|^{r-1}$, while the number of conflicting $m_a^{(1)}, m_b^{(2)}$ is bounded by

$$\sum_{i\in[n]} d_i^3 \sum_{j\in[n]} d_j^2 = o\Big(\sum_{i\in[n]} d_i(d_i-1)\Big)^3, \tag{7.3.18}$$

where we again use that $d_i = o(\sqrt{n})$, as required.

In case (c), the number of $m_c^{(1)}, c \in \{1, \ldots, s\}$ and $m_d^{(2)}, d \in \{1, \ldots, r\} \setminus \{a, b\}$ is bounded by $|\mathcal{I}_1|^s|\mathcal{I}_2|^{r-2}$, while the number of conflicting $m_a^{(2)}, m_b^{(2)}$ is bounded by

$$\sum_{i\in[n]} d_i^3 \sum_{j\in[n]} d_j^2 \sum_{k\in[n]} d_k^2 = o\Big(\sum_{i\in[n]} d_i(d_i-1)\Big)^4, \tag{7.3.19}$$

where we again use that $d_i = o(\sqrt{n})$, as required. This completes the proof. □

Exercise 7.11 (Characterization moments independent Poisson variables)**.** *Show that the moments of (X, Y), where (X, Y) are independent Poisson random variables with parameters μ_X and μ_Y are identified by the relations, for $r \geq 1$,*

$$\mathbb{E}[X^r] = \mu_X \mathbb{E}[(X+1)^{r-1}], \tag{7.3.20}$$

and, for $r, s \geq 1$,

$$\mathbb{E}[X^r Y^s] = \mu_Y \mathbb{E}[X^r (Y+1)^{s-1}]. \tag{7.3.21}$$

Exercise 7.12 (Alternative proof of Proposition 7.9)**.** *Give an alternative proof of Proposition 7.9 by using Theorem 2.3(e) together with Exercise 7.11 and the fact that all joint moments of (S_n, M_n) converge to those of (S, M), where S and M are two independent Poisson random variables with means $\frac{\nu}{2}$ and $\frac{\nu^2}{4}$.*

Exercise 7.13 (Average number of triangles CM)**.** *Compute the average number of occupied triangles in $\mathrm{CM}_n(\boldsymbol{d})$.*

Exercise 7.14 (Poisson limit triangles CM)**.** *Show that the number of occupied triangles in $\mathrm{CM}_n(\boldsymbol{d})$ converges to a Poisson random variable when Condition 7.5(a)-(c) holds.*

7.4 Configuration model, uniform simple random graphs and GRGs

In this section, we shall investigate the relations between the configuration model, uniform simple random graphs with given degrees, and the generalized random graph with given weights. These results are 'folklore' in the random graph community, and allow to use the configuration model to prove results for several other models.

Proposition 7.10 (Uniform graphs with given degree sequence)**.** *For any degree sequence $(d_i)_{i\in[n]}$, and conditionally on the event $\{\mathrm{CM}_n(\boldsymbol{d})$ is a simple graph$\}$, $\mathrm{CM}_n(\boldsymbol{d})$ is a uniform simple random graph with the prescribed degree sequence.*

Proof. We recall that the graph in the configuration model is produced by a uniform matching of the corresponding configuration of half-edges. By Exercise 7.15 below, we note that, conditionally on the matching producing a simple graph, the conditional distribution of the configuration is uniform over all configurations which are such that the corresponding graph is simple:

Exercise 7.15 (A conditioned uniform variable is again uniform)**.** *Let $\mathbb{P}$ be a uniform distribution on some finite state space $\mathcal{X}$, and let U be a uniform random variable on $\mathcal{X}$. Let $\mathcal{Y} \subseteq \mathcal{X}$ be a non-empty subset of $\mathcal{X}$. Show that the conditional probability $\mathbb{P}(\cdot|U \in \mathcal{Y})$ given that U is in $\mathcal{Y}$ is the uniform distribution on $\mathcal{Y}$.*

We conclude that Proposition 7.10 is equivalent to the statement that every simple graph has an *equal* number of configurations contributing to it, which follows from Proposition 7.4. □

Exercise 7.16 (Poisson limits for self-loops, multiple edges and triangles)**.** *Assume that the fixed degree sequence $(d_i)_{i\in[n]}$ satisfies Condition 7.5(a)-(c). Let T_n denote the number of triangles in $\mathrm{CM}_n(\boldsymbol{d})$, i.e., the number of i, j, k such that $i < j < k$ and such that there are edges between i and j, between j and k and between k and i. Show that (S_n, M_n, T_n) converges to three independent Poisson random variables and compute their asymptotic parameters.*

An important consequence of Theorem 7.8 is that it allows us to compute the asymptotic number of graphs with a given degree sequence:

Corollary 7.11 (Number of graphs with given degree sequence)**.** *Assume that the degree sequence $(d_i)_{i\in[n]}$ satisfies Condition 7.5(a)-(c), and that $\ell_n = \sum_{i\in[n]} d_i$ is even. Then, the number of simple graphs with degree sequence $(d_i)_{i\in[n]}$ is equal to*

$$\mathrm{e}^{-\nu/2-\nu^2/4}\frac{(\ell_n-1)!!}{\prod_{i\in[n]} d_i!}(1+o(1)). \tag{7.4.1}$$

Proof. By Proposition 7.10, the distribution of $\mathrm{CM}_n(\boldsymbol{d})$, conditionally on $\mathrm{CM}_n(\boldsymbol{d})$ being simple, is uniform over all simple graphs with degree sequence $\boldsymbol{d} = (d_i)_{i\in[n]}$. Let $Q(\boldsymbol{d})$ denote the number of such simple graphs, and let G denote any simple random graph with degree sequence $\boldsymbol{d} = (d_i)_{i\in[n]}$. Recall from the proof of Proposition 7.4 that $N(G)$ denotes the number of configurations that give rise to G. By Proposition 7.4, we have that $N(G)$ is the same for all simple G. Recall further that the total number of configurations is given by $(\ell_n - 1)!!$. Then,

$$Q(\boldsymbol{d}) = \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ simple})\frac{(\ell_n-1)!!}{N(G)}. \tag{7.4.2}$$

By Proposition 6.15, for any simple graph G,

$$N(G) = \prod_{i\in[n]} d_i!. \tag{7.4.3}$$

Proposition 7.10 then yields the result. □

A special case of the configuration model is when all degrees are equal to some r. In this case, when we condition on the fact that the resulting graph in the configuration model to be simple, we obtain a *uniform regular random graph*. Uniform regular random graphs can

be seen as a finite approximation of a regular tree. In particular, Corollary 7.11 implies that, when nr is even, the number of regular r-ary graphs is equal to

$$\mathrm{e}^{-(r-1)/2-(r-1)^2/4}\frac{(rn-1)!!}{(r!)^n}(1+o(1)). \tag{7.4.4}$$

Exercise 7.17 (The number of r-regular graphs)**.** *Prove* (7.4.4).

Exercise 7.18 (The number of simple graphs without triangles)**.** *Assume that the fixed degree sequence $(d_i)_{i\in[n]}$ satisfies Condition 7.5(a)-(c). Compute the number of simple graphs with degree sequence $(d_i)_{i\in[n]}$ not containing any triangle. Hint: use Exercise 7.16.*

A further consequence of Theorem 7.8 is that it allows to prove a property for uniform graphs with a given degree sequence by proving it for the configuration model with that degree sequence:

Corollary 7.12 (Uniform graphs with given degree sequence and $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that $\boldsymbol{d}=(d_i)_{i\in[n]}$ satisfies Condition 7.5(a)-(c), and that $\ell_n=\sum_{i\in[n]}d_i$ is even. Then, an event $\mathcal{E}_n$ occurs with high probability for a uniform simple random graph with degrees $(d_i)_{i\in[n]}$ when it occurs with high probability for $\mathrm{CM}_n(\boldsymbol{d})$.*

Corollary 7.12 allows a simple strategy to study proporties of uniform simple random graphs with a prescribed degree sequence. Indeed, $\mathrm{CM}_n(\boldsymbol{d})$ can be constructed in a rather simple manner, which makes it easier to prove properties for $\mathrm{CM}_n(\boldsymbol{d})$ than it is for a uniform random graph with degrees $\boldsymbol{d}$. For completeness, we now prove the above statement.

Proof. Let $\mathrm{UG}_n(\boldsymbol{d})$ denote a uniform simple random graph with degrees $\boldsymbol{d}$. We need to prove that if $\lim_{n\to\infty}\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\in\mathcal{E}_n^c)=0$, then also $\lim_{n\to\infty}\mathbb{P}(\mathrm{UG}_n(\boldsymbol{d})\in\mathcal{E}_n^c)=0$. By Proposition 7.10,

$$\begin{aligned}\mathbb{P}(\mathrm{UG}_n(\boldsymbol{d})\in\mathcal{E}_n^c)&=\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\in\mathcal{E}_n^c|\mathrm{CM}_n(\boldsymbol{d})\text{ simple})\\&=\frac{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\in\mathcal{E}_n^c,\mathrm{CM}_n(\boldsymbol{d})\text{ simple})}{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\text{ simple})}\\&\le\frac{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\in\mathcal{E}_n^c)}{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\text{ simple})}.\end{aligned} \tag{7.4.5}$$

By Theorem 7.8, for which the assumptions are satisfied by the hypotheses in Corollary 7.12, $\liminf_{n\to\infty}\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\text{ simple})>0$. Moreover, $\lim_{n\to\infty}\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\in\mathcal{E}_n^c)=0$, so that $\mathbb{P}(\mathrm{UG}_n(\boldsymbol{d})\in\mathcal{E}_n^c)\to0$, as required. □

As a consequence of Proposition 7.10 and Theorem 6.14, we see that the GRG conditionally on its degrees, and $\mathrm{CM}_n(\boldsymbol{d})$ with the same degrees conditioned on producing a simple graph, have identically the same distribution. This also partially explains the popularity of the configuration model: Some results for the Erdős-Rényi random graph are more easily proved by conditioning on the degree sequence, proving the result for the configuration model, and using that the degree distribution of the Erdős-Rényi random graph is very close to a sequence of independent Poisson random variables. See Chapters 9 and 10. We shall formalize this 'folklore' result in the following theorem:

Theorem 7.13 (Relation between $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Let D_i be the degree of vertex i in $\mathrm{GRG}_n(\boldsymbol{w})$ defined in* (6.2.1)*, and let $\boldsymbol{D}=(D_i)_{i\in[n]}$. Then,*

$$\mathbb{P}(\mathrm{GRG}_n(\boldsymbol{w})=G\mid\boldsymbol{D}=\boldsymbol{d})=\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})=G\mid\mathrm{CM}_n(\boldsymbol{d})\text{ simple}). \tag{7.4.6}$$

Assume that $\boldsymbol{D} = (D_i)_{i\in[n]}$ *satisfies that Condition 7.5(a)-(c) hold in probability and that* $\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \in \mathcal{E}_n) \xrightarrow{\mathbb{P}} 1$*, where* $\mathrm{CM}_n(\boldsymbol{D})$ *denotes the configuration model with degrees equal to the (random) degrees of* $\mathrm{GRG}_n(\boldsymbol{w})$*, and* $\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \in \mathcal{E}_n)$ *is interpreted as a function of the random degrees* $\boldsymbol{D}$*. Then, by* (7.4.6)*, also* $\mathbb{P}(\mathrm{GRG}_n(\boldsymbol{w}) \in \mathcal{E}_n) \xrightarrow{\mathbb{P}} 1$.

We note that, by Theorem 6.9, in many cases, Condition 7.5(a)-(c). These properties are often easier to verify than the event $\mathcal{E}_n$ itself. We also remark that related versions of Theorem 7.13 can be stated with stronger hypotheses on the degrees. Then, the statement becomes that, when an event $\mathcal{E}_n$ occurs with high probability for $\mathrm{CM}_n(\boldsymbol{d})$ under the assumptions on the degrees, $\mathcal{E}_n$ also occurs with high probability for $\mathrm{GRG}_n(\boldsymbol{w})$.

Proof. Equation (7.4.6) follows from Theorem 6.14 and Corollary 7.12, for every simple graph G with degree sequence $\boldsymbol{d}$, as these two results imply that both $\mathrm{GRG}_n(\boldsymbol{w})$ conditionally on $\boldsymbol{D} = \boldsymbol{d}$ and $\mathrm{CM}_n(\boldsymbol{d})$ conditionally on being simple are uniform simple random graphs with degree sequence $\boldsymbol{d}$. By (7.4.6), for every event $\mathcal{E}_n$,

$$\mathbb{P}(\mathrm{GRG}_n(\boldsymbol{w}) \in \mathcal{E}_n \mid \boldsymbol{D} = \boldsymbol{d}) = \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \in \mathcal{E}_n \mid \mathrm{CM}_n(\boldsymbol{d}) \text{ simple}). \tag{7.4.7}$$

We rewrite

$$\begin{aligned}\mathbb{P}(\mathrm{GRG}_n(\boldsymbol{w}) \in \mathcal{E}_n^c) &= \mathbb{E}\Big[\mathbb{P}(\mathrm{GRG}_n(\boldsymbol{w}) \in \mathcal{E}_n^c \mid \boldsymbol{D})\Big] &(7.4.8)\\ &= \mathbb{E}\Big[\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \in \mathcal{E}_n^c \mid \mathrm{CM}_n(\boldsymbol{D}) \text{ simple})\Big]\\ &\le \mathbb{E}\Big[\Big(\frac{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \in \mathcal{E}_n^c)}{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \text{ simple})}\Big) \wedge 1\Big]. &(7.4.9)\end{aligned}$$

By assumption, $\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \in \mathcal{E}_n^c) \xrightarrow{\mathbb{P}} 0$. Further, since the degrees $\boldsymbol{D}$ satisfies Condition 7.5(a)-(c),

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \text{ simple}) \xrightarrow{\mathbb{P}} \mathrm{e}^{-\nu/2-\nu^2/4} > 0. \tag{7.4.10}$$

Therefore, by Dominated Convergence (Theorem A.10), we obtain that

$$\lim_{n\to\infty} \mathbb{E}\Big[\Big(\frac{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \in \mathcal{E}_n^c)}{\mathbb{P}(\mathrm{CM}_n(\boldsymbol{D}) \text{ simple})}\Big) \wedge 1\Big] = 0,$$

so that we conclude that $\lim_{n\to\infty} \mathbb{P}(\mathrm{GRG}_n(\boldsymbol{w}) \in \mathcal{E}_n^c) = 0$, as required. □

7.5 Configuration model with i.i.d. degrees

In this section, we apply the results of the previous sections to the configuration model with i.i.d. degrees. Indeed, we take the degrees $(D_i)_{i\ge1}$ to be an i.i.d. sequence. Since the total degree $\sum_{i\in[n]} D_i$ is with probability close to $1/2$ odd (recall Exercise 7.8), we need to make sure that the total degree is even. Therefore, by convention, we set

$$d_i = D_i + \mathbb{1}_{\{\sum_{j\in[n]} D_j \text{ odd}, i=n\}}, \tag{7.5.1}$$

and set

$$L_n = \sum_{i\in[n]} d_i = \sum_{i\in[n]} D_i + \mathbb{1}_{\{\sum_{i\in[n]} D_i \text{ odd}\}}. \tag{7.5.2}$$

Often, we shall ignore the effect of the added indicator in the definition of d_n, since it shall hardly make any difference.

We note that, similarly to the generalized random graph with i.i.d. weights, the introduction of randomness in the degrees introduces a *double randomness* in the model: firstly the randomness of the weights, and secondly, the randomness of the pairing of the edges *given* the degrees. Due to this double randomness, we need to investigate the degree sequence $(P_k^{(n)})_{k=1}^{\infty}$ defined by

$$P_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{d_i = k\}}. \tag{7.5.3}$$

If we ignore the dependence on n of d_n, then we see that $(P_k^{(n)})_{k=1}^{\infty}$ is precisely equal to the empirical distribution of the degrees, which is an i.i.d. sequence. As a result, by the Strong Law of Large Numbers, we have that

$$P_k^{(n)} \xrightarrow{a.s.} p_k \equiv \mathbb{P}(D_1 = k), \tag{7.5.4}$$

so that the empirical distribution of i.i.d. degrees is almost surely close to the probability distribution of each of the degrees. By Exercise 2.14, the above convergence also implies that $d_{\mathrm{TV}}(P^{(n)}, p) \xrightarrow{a.s.} 0$, where $p = (p_k)_{k \geq 1}$ and $P^{(n)} = (P_k^{(n)})_{k \geq 1}$.

The main results are the following:

Theorem 7.14 (Degree sequence of erased configuration model with i.i.d. degrees)**.** *Let $(D_i)_{i \in [n]}$ be an i.i.d. sequence of finite mean random variables with $\mathbb{P}(D \geq 1) = 1$. The degree sequence of the erased configuration model $(P_k^{(\mathrm{er})})_{k \geq 1}$ with degrees $(D_i)_{i \in [n]}$ converges to $(p_k)_{k \geq 1}$. More precisely,*

$$\mathbb{P}\Big(\sum_{k=1}^{\infty} |P_k^{(\mathrm{er})} - p_k| \geq \varepsilon\Big) \to 0. \tag{7.5.5}$$

Proof. By Exercise 7.9, when $\mathbb{E}[D] < \infty$, Condition 7.5(a)-(b) hold, where the convergence is in probability. As a result, Theorem 7.14 follows directly from Theorem 7.6. □

We next investigate the probability of obtaining a simple graph in $\mathrm{CM}_n(\boldsymbol{D})$:

Theorem 7.15 (Probability of simplicity in $\mathrm{CM}_n(\boldsymbol{D})$)**.** *Let $(D_i)_{i \geq 1}$ be an i.i.d. sequence of random variables with $\mathrm{Var}(D) < \infty$ and $\mathbb{P}(D \geq 1) = 1$. Then, the probability that $\mathrm{CM}_n(\boldsymbol{D})$ is simple is asymptotically equal to $\mathrm{e}^{-\nu/2-\nu^2/4}$, where $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$.*

Proof. By Exercise 7.9, when $\mathbb{E}[D] < \infty$, Condition 7.5(a)-(b) hold, where the convergence is in probability. As a result, Theorem 7.15 follows directly from Theorem 7.8. □

We finally investigate the case where the mean is *infinite*, with the aim to produce a random graph with power-law degrees with an exponent $\tau \in (1, 2)$. In this case, the graph topology is rather different, as the majority of edges is in fact multiple, and self-loops from vertices with high degrees are abundant. As a result, the erased configuration model has rather different degrees compared to those in the multigraph. Therefore, in order to produce a more realistic graph, we need to perform some sort of a truncation procedure. We start by investigating the case where we condition the degrees to be bounded above by some $a_n = o(n)$, which, in effect reduces the number of self-loops significantly.

Theorem 7.16 (Degree sequence of erased configuration model with i.i.d. conditioned infinite mean degrees)**.** *Let $(D_i^{(n)})_{i \in [n]}$ be i.i.d. copies of the random variable D conditioned on $D \leq a_n$. Then, for every $a_n = o(n)$, the empirical degree distribution of the erased configuration model $(P_k^{(\mathrm{er})})_{k=1}^{\infty}$ with degrees $(D_i^{(n)})_{i \in [n]}$ converges to $(p_k)_{k \geq 1}$, where $p_k = \mathbb{P}(D = k)$. More precisely,*

$$\mathbb{P}\Big(\sum_{k=1}^{\infty} |P_k^{(\mathrm{er})} - p_k| \geq \varepsilon\Big) \to 0. \tag{7.5.6}$$

Theorem 7.16 is similar in spirit to Theorem 6.12 for the generalized random graph, and is left as an exercise:

Exercise 7.19 (Proof of Theorem 7.16)**.** *Adapt the proof of Theorem 7.14 to prove Theorem 7.16.*

□

We continue by studying the erased configuration model with infinite mean degrees in the unconditioned case. We assume that there exists a slowly varying function $x \mapsto L(x)$ such that

$$1 - F(x) = x^{1-\tau} L(x), \tag{7.5.7}$$

where $F(x) = \mathbb{P}(D \leq x)$ and where $\tau \in (1, 2)$. We now investigate the degree sequence in the configuration model with infinite mean degrees, where we do not condition the degrees to be at most a_n. We shall make substantial use of Theorem 2.28. In order to describe the result, we need a few definitions. We define the (random) probability distribution $P = (P_i)_{i \geq 1}$ as follows. Let, as in Theorem 2.28, $(E_i)_{i \geq 1}$ be i.i.d. exponential random variables with parameter 1, and define $\Gamma_i = \sum_{j=1}^{i} E_j$. Let $(D_i)_{i \geq 1}$ be an i.i.d. sequence of random variables with distribution function F_D in (7.5.7), and let $D_{(n:n)} \geq D_{(n-1:n)} \geq \cdots \geq D_{(1:n)}$ be the order statistics of $(D_i)_{i \in [n]}$. We recall from Theorem 2.28 that there exists a sequence u_n, with $u_n n^{-1/(\tau-1)}$ slowly varying, such that

$$u_n^{-1} \left(L_n, \{D_{(i)}\}_{i=1}^{\infty} \right) \xrightarrow{d} \left(\sum_{j \geq 1} \Gamma_j^{-1/(\tau-1)}, (\Gamma_i^{-1/(\tau-1)})_{i \geq 1} \right). \tag{7.5.8}$$

We abbreviate $\eta = \sum_{j \geq 1} \Gamma_j^{-1/(\tau-1)}$ and $\xi_i = \Gamma_i^{-1/(\tau-1)}$, and let

$$P_i = \xi_i / \eta, \tag{7.5.9}$$

so that, by (7.5.8),

$$\sum_{i=1}^{\infty} P_i = 1. \tag{7.5.10}$$

However, the P_i are all random variables, so that $P = (P_i)_{i \geq 1}$ is a random probability distribution. We further write $M_{P,k}$ for a multinomial distribution with parameters k and probabilities $P = (P_i)_{i \geq 1}$, and U_{P,D_k} is the number of distinct outcomes of the random variable M_{P,D_k}, where D_k is independent of $P = (P_i)_{i \geq 1}$ and the multinomial trials.

Theorem 7.17 (Degree sequence of erased configuration model with i.i.d. infinite mean degrees)**.** *Let $(D_i)_{i \geq 1}$ be i.i.d. copies of a random variable D_1 having distribution function F satisfying (7.5.7). Fix $k \in \mathbb{N}$. The degree of vertex k in the erased configuration model with degrees $(D_i)_{i \in [n]}$ converges in distribution to the random variable U_{P,D_k}, where $P = (P_i)_{i \geq 1}$ is given by (7.5.9), and the random variables D_k and $P = (P_i)_{i \geq 1}$ are independent.*

Theorem 7.17 is similar in spirit to Theorem 6.13 for the generalized random graph.

Proof. We fix vertex k, and note that its degree is given by D_k. With high probability, we have that $D_k \leq \log n$, so that D_k is not one of the largest order statistics. Therefore, D_k is independent of $(\eta, \xi_1, \xi_2, \dots)$. The vertex k now has D_k half-edges, which need to be connected to other half-edges. The probability that any half-edge is connected to the vertex with the j^{th} largest degree is asymptotic to

$$P_j^{(n)} = D_{(n-j+1:n)} / L_n, \tag{7.5.11}$$

where, by Theorem 2.28 (see also (7.5.8)),

$$(P_j^{(n)})_{j\geq 1} \xrightarrow{d} (\xi_j/\eta)_{j\geq 1}. \tag{7.5.12}$$

Moreover, the vertices to which the D_k half-edges are connected are close to being independent, when $D_k \leq \log n$. As a result, the D_k half-edges of vertex k are paired to D_k vertices, and the number of edges of vertex k that are paired to the vertex with the i^{th} largest degree are asymptotically given by the i^{th} coordinate of $M_{P^{(n)},D_k}$. The random variable $M_{P^{(n)},D_k}$ converges in distribution to M_{P,D_k}. We note that in the erased configuration model, the degree of the vertex k is equal to the number of *distinct* vertices to which k is connected, which is therefore equal to the number of distinct outcomes of the random variable M_{P,D_k}, which, by definition, is equal to U_{P,D_k}. □

We next investigate the properties of the degree distribution, to obtain an equivalent result as in Exercise 6.19.

Theorem 7.18 (Power law with exponent 2 for erased configuration model with infinite mean degrees)**.** *Let the distribution function F of D_k satisfy* (7.5.7) *with $L(x) = 1$. Then, the asymptotic degree of vertex k, which is given by U_{P,D_k} satisfies that*

$$\mathbb{P}(U_{P,D_k} \geq x) \leq x^{-1}. \tag{7.5.13}$$

The result in Theorem 7.18 is similar in spirit to Exercise 6.19. It would be of interest to prove a precise identity here as well.

Proof. We give a sketch of proof only. We condition on $D_k = \lceil x^b \rceil$, for some $b \geq 1$. Then, in order that $U_{P,D_k} \geq x$, at least $x/2$ values larger than $x/2$ need to be chosen. By (2.6.17), we have that the probability that value k is chosen, for some large value k, is close to $k^{-1/(\tau-1)}/\eta$. Therefore, the probability that a value at least k is chosen is close to

$$k^{-1/(\tau-1)+1}/\eta = k^{(\tau-2)/(\tau-1)}/\eta. \tag{7.5.14}$$

Moreover, conditionally on $D_k = \lceil x^b \rceil$, the number of values larger than $x/2$ that are chosen is equal to a Binomial random variable with $\lceil x^b \rceil$ trials and success probability

$$q_x = x^{(\tau-2)/(\tau-1)}/\eta. \tag{7.5.15}$$

Therefore, conditionally on $D_k = \lceil x^b \rceil$, and using Theorem 2.18, the probability that at least $x/2$ values larger than $x/2$ are chosen is negligible when, for some sufficiently large $C > 0$,

$$|x^b q_x - \frac{x}{2}| \geq C \log x \sqrt{x^b q_x}. \tag{7.5.16}$$

Equations (7.5.15) and (7.5.16) above imply that $b = 1 - (\tau-2)/(\tau-1) = 1/(\tau-1)$. As a result, we obtain that

$$\mathbb{P}(U_{P,D_k} \geq x) \leq \mathbb{P}(D_k \geq \lceil x^b \rceil) \leq x^{-b(\tau-1)} = x^{-1}. \tag{7.5.17}$$

□

7.6 Notes and discussion

Notes on Section 7.1. The configuration model has a long history. It was introduced in [40] to study uniform random graphs with a given degree sequence (see also [44, Section 2.4]). The introduction was inspired by, and generalized the results in, the work of Bender

and Canfield [25]. The original work allowed for a careful computation of the number of graphs with prescribed degrees, using a probabilistic argument. This is the probabilistic method at its best, and also explains the emphasis on the study of the probability for the graph to be simple. It was further studied in [169, 170], where it was investigated when the resulting graph has a giant component. We shall further comment on these results in Chapter 10 below.

Notes on Section 7.2. The result in Theorem 7.6 can be found in [130]. The term erased configuration model is first used in [58, Section 2.1].

Notes on Section 7.4. Corollary 7.12 implies that the uniform simple random graph model is *contiguous* to the configuration model, in the sense that events with vanishing probability for the configuration model also have vanishing probability for the uniform simple random graph model with the same degree sequence. See [131] for a discussion of contiguity of random graphs. Theorem 7.13 implies that the generalized random graph conditioned on having degree sequence $\boldsymbol{d}$ is contiguous to the configuration model with that degree sequence, whenever the degree sequence satisfies Condition 7.5(a)-(c).

Notes on Section 7.5. A version of Theorem 7.14 can be found in [58]. Results on the erased configuration model as in Theorems 7.17-7.18 have appeared in [32], where first passage percolation on $\mathrm{CM}_n(\boldsymbol{D})$ was studied with infinite mean degrees, both for the erased as well as for the original configuration model, and it is shown that the behavior in the two models is completely different.

Chapter 8
Preferential attachment models

The generalized random graph model and the configuration model described in Chapters 6 and 7, respectively, are *static* models, i.e., the size of the graph is *fixed*, and we have not modeled the *growth* of the graph. There is a large body of work investigating *dynamic* models for complex networks, often in the context of the World-Wide Web. In various forms, such models have been shown to lead to power-law degree sequences, and, thus, they offer a possible *explanation* for the occurrence of power-law degree sequences in random graphs. The existence of power-law degree sequences in various real networks is quite striking, and models offering a convincing explanation can teach us about the mechanisms which give rise to their scale-free nature.

A possible and convincing explanation for the occurrence of power-law degree sequences is offered by the *preferential attachment paradigm*. In the preferential attachment model, vertices are added sequentially with a number of edges connected to them. These edges are attached to a receiving vertex with a probability proportional to the degree of the receiving vertex at that time, thus favoring vertices with large degrees. For this model, it is shown that the number of vertices with degree k decays proportionally to k^{-3} [50], and this result is a special case of the more general result that we shall describe in this chapter.

The idea behind preferential attachment is simple. In an evolving graph, i.e., a graph that evolves in time, the newly added vertices are connected to the already existing vertices. In an Erdős-Rényi random graph, which can also be formulated as an evolving graph, where edges are added and removed, these edges would be connected to each individual with equal probability.

Exercise 8.1 (A dynamic formulation of $\mathrm{ER}_n(p)$)**.** *Give a dynamical model for the Erdős-Rényi random graph, where at each time n we add a single individual, and where at time n the graph is equal to* $\mathrm{ER}_n(p)$*. See also the dynamic description of the Norros-Reittu model on Page 138.*

Now think of the newly added vertex as a new individual in a social population, which we model as a graph by letting the individuals be the vertices and the edges be the acquaintance relations. Is it then realistic that the edges connect to each already present individual with equal probability, or is the newcomer more likely to get to know socially active individuals, who already know many people? If the latter is true, then we should forget about equal probabilities for receiving ends of the edges of the newcomer, and introduce a bias in his/her connections towards more social individuals. Phrased in a mathematical way, it should be more likely that the edges be connected to vertices that already have a high degree. A possible model for such a growing graph was proposed by Barabási and Albert [20], and has incited an enormous research effort since.

Strictly speaking, Barabási and Albert in [20] were not the first to propose such a model, and we shall start by referring to the old literature on the subject. Yule [228] was the first to propose a growing model where preferential attachment is present, in the context of the evolution of species. He derives the power law distribution that we shall also find in this chapter. Simon [203] provides a more modern version of the preferential attachment model, as he puts it

> "*Because Yule's paper predates the modern theory of stochastic processes, his derivation was necessarily more involved than the one we shall employ here.*"

The stochastic model of Simon is formulated in the context of the occurrence of words in large pieces of text (as in [229]), and is based on two assumptions, namely (i) that the

probability that the $(k+1)^{\text{st}}$ word is a word that has already appeared exactly i times is proportional to the number of occurrences of words that have occurred exactly i times, and (ii) that there is a constant probability that the $(k+1)^{\text{st}}$ word is a new word. Together, these two assumptions give rise to frequency distributions of words that obey a power law, with a power-law exponent that is a simple function of the probability of adding a new vertex. We shall see a similar effect occurring in this chapter. A second place where the model studied by Simon and Yule can be found is in work by Champernowne [62], in the context of income distributions in populations.

In [20], Barabási and Albert describe the preferential attachment graph informally as follows:

> "*To incorporate the growing character of the network, starting with a small number* (m_0) *of vertices, at every time step we add a new vertex with* $m(\leq m_0)$ *edges that link the new vertex to* m *different vertices already present in the system. To incorporate preferential attachment, we assume that the probability* Π *that a new vertex will be connected to a vertex* i *depends on the connectivity* k_i *of that vertex, so that* $\Pi(k_i) = k_i / \sum_j k_j$*. After* t *time steps, the model leads to a random network with* $t + m_0$ *vertices and* mt *edges.*"

This description of the model is informal, but it must have been given precise meaning in [20] (since, in particular, Barabási and Albert present simulations of the model predicting a power-law degree sequence with exponent close to $\tau = 3$). The model description does not explain how the first edge is connected (note that at time $t = 1$, there are no edges, so the first edge can not be attached according to the degrees of the existing vertices), and does not give the dependencies between the m edges added at time t. We are left wondering whether these edges are independent, whether we allow for self-loops, whether we should update the degrees after each attachment of a single edge, etc. In fact, each of these choices has, by now, been considered in the literature, and the results, in particular the occurrence of power laws and the power-law exponent, do not depend sensitively on the respective choices. See Section 8.7 for an extensive overview of the literature on preferential attachment models.

The first to investigate the model rigorously, were Bollobás, Riordan, Spencer and Tusnady [50]. They complain heavily about the lack of a formal definition in [20], arguing that

> "*The description of the random graph process quoted above (i.e, in [20], edt.) is rather imprecise. First, as the degrees are initially zero, it is not clear how the process is started. More seriously, the expected number of edges linking a new vertex* v *to earlier vertices is* $\sum_i \Pi(k_i) = 1$*, rather than* m*. Also, when choosing in one go a set* S *of* m *earlier vertices as the neighbors of* v*, the distribution of* S *is* not *specified by giving the marginal probability that each vertex lies in* S*.*"

One could say that these differences in formulations form the heart of much confusion between mathematicians and theoretical physicists. To resolve these problems, choices had to be made, and these choices were, according to [50], made first in [48], by specifying the initial graph to consist of a vertex with m self-loops, and that the degrees will be updated in the process of attaching the m edges. This model will be described in full detail in Section 8.1 below.

This chapter is organized as follows. In Section 8.1, we introduce the model. In Section 8.2, we investigate how the degrees of fixed vertices evolve as the graph grows. In Section 8.3, we investigate the degree sequences in preferential attachment models. The main result is Theorem 8.2, which states that the preferential attachment model has a power-law degree sequence. The proof of Theorem 8.2 consists of two key steps, which are formulated and proved in Sections 8.4 and 8.5, respectively. In Section 8.6, we investigate the maximal

degree in a preferential attachment model. In Section 8.7, we also discuss many related preferential attachment models. We close this chapter with notes and discussion in Section 8.8.

8.1 Introduction to the model

In this chapter, we prove that the preferential attachment model has a power-law degree sequence. We start by introducing the model. The model we investigate produces a *graph sequence* which we denote by $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$, which for every t yields a graph of t vertices and mt edges for some $m = 1, 2, \ldots$ We start by defining the model for $m = 1$. In this case, $\mathrm{PA}_{1,\delta}(1)$ consists of a single vertex with a single self-loop. We denote the vertices of $\mathrm{PA}_t(1,\delta)$ by $\{v_1^{(1)}, \ldots, v_t^{(1)}\}$. We denote the degree of vertex $v_i^{(1)}$ in $\mathrm{PA}_t(1,\delta)$ by $D_i(t)$, where a self-loop increases the degree by 2.

Then, conditionally on $\mathrm{PA}_t(1,\delta)$, the growth rule to obtain $\mathrm{PA}_{t+1}(1,\delta)$ is as follows. We add a single vertex $v_{t+1}^{(1)}$ having a single edge. This edge is connected to a second end point, which is equal to $v_{t+1}^{(1)}$ with probability $(1+\delta)/(t(2+\delta)+(1+\delta))$, and to a vertex $v_i^{(1)} \in \mathrm{PA}_t(1,\delta)$ with probability $(D_i(t)+\delta)/(t(2+\delta)+(1+\delta))$, where $\delta \geq -1$ is a parameter of the model. Thus,

$$\mathbb{P}\big(v_{t+1}^{(1)} \to v_i^{(1)} \big| \mathrm{PA}_t(1,\delta)\big) = \begin{cases} \frac{1+\delta}{t(2+\delta)+(1+\delta)} & \text{for } i = t+1, \\ \frac{D_i(t)+\delta}{t(2+\delta)+(1+\delta)} & \text{for } i \in [t]. \end{cases} \tag{8.1.1}$$

Exercise 8.2 (Non-negativity of $D_i(t)+\delta$)**.** *Verify that $D_i(t) \geq 1$ for all i and t, so that $D_i(t)+\delta \geq 0$ for all $\delta \geq -1$.*

Exercise 8.3 (Attachment probabilities sum up to one)**.** *Verify that the probabilities in* (8.1.1) *sum up to one.*

The model with $m > 1$ is defined in terms of the model for $m = 1$ as follows. We start with $\mathrm{PA}_{mt}(1,\delta/m)$, and denote the vertices in $\mathrm{PA}_{mt}(1,\delta/m)$ by $v_1^{(1)}, \ldots, v_{mt}^{(1)}$. Then we identify $v_1^{(1)}, \ldots, v_m^{(1)}$ in $\mathrm{PA}_{mt}(1,\delta/m)$ to be $v_1^{(m)}$ in $\mathrm{PA}_{m,\delta}(t)$, and $v_{m+1}^{(1)}, \ldots, v_{2m}^{(1)}$ in $\mathrm{PA}_{mt}(1,\delta/m)$ to be $v_2^{(m)}$ in $\mathrm{PA}_{m,\delta}(t)$, and, more generally, $v_{(j-1)m+1}^{(1)}, \ldots, v_{jm}^{(1)}$ in $\mathrm{PA}_{mt}(1,\delta/m)$ to be $v_j^{(m)}$ in $\mathrm{PA}_{m,\delta}(t)$. This defines the model for general $m \geq 1$. The above identification procedure is sometimes called the *collapsing of vertices*. We note that $\mathrm{PA}_{m,\delta}(t)$ is a multigraph with precisely t vertices and mt edges, so that the total degree is equal to $2mt$.

Exercise 8.4 (Total degree)**.** *Prove that the total degree of $\mathrm{PA}_{m,\delta}(t)$ equals $2mt$.*

In order to explain the description of $\mathrm{PA}_{m,\delta}(t)$ in terms of $\mathrm{PA}_{mt}(1,\delta/m)$, we note that an edge in $\mathrm{PA}_{mt}(1,\delta/m)$ is attached to vertex $v_k^{(1)}$ with probability proportional to the weight of vertex $v_k^{(1)}$, where the weight is equal to the degree of vertex $v_k^{(1)}$ plus δ/m. Now, vertices $v_{(j-1)m+1}^{(1)}, \ldots, v_{jm}^{(1)}$ in $\mathrm{PA}_{mt}(1,\delta/m)$ are identified or collapsed to vertex $v_j^{(m)}$ in $\mathrm{PA}_{m,\delta}(t)$. Thus, an edge in $\mathrm{PA}_{m,\delta}(t)$ is attached to vertex $v_j^{(m)}$ with probability proportional to the total weight of the vertices $v_{(j-1)m+1}^{(1)}, \ldots, v_{jm}^{(1)}$. Since the sum of the degrees of the vertices $v_{(j-1)m+1}^{(1)}, \ldots, v_{jm}^{(1)}$ is equal to the degree of vertex $v_j^{(m)}$, this probability is proportional to the degree of vertex $v_j^{(m)}$ in $\mathrm{PA}_{m,\delta}(t)$ plus δ. We note that in the above construction and for $m \geq 2$, the degrees are updated after each edge is attached. This is what we refer to as *intermediate updating of the degrees*.

The important feature of the model is that edges are more likely to be connected to vertices with large degrees, thus making the degrees even larger. This effect is called

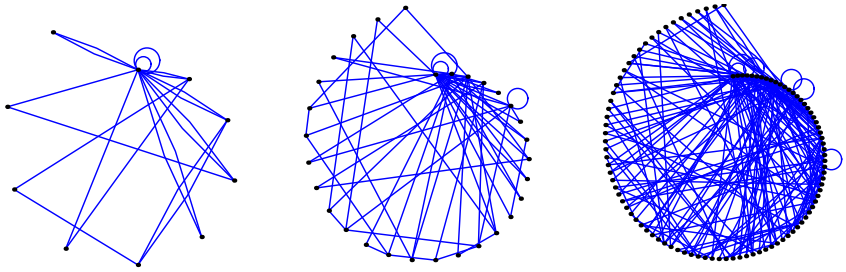

Figure 8.1: Preferential attachment random graph with $m = 2$ and $\delta = 0$ of sizes 10, 30 and 100.

preferential attachment. Preferential attachment may explain *why* there are quite large degrees. Therefore, the preferential attachment model is sometimes called the *Rich-get-Richer model*. It is quite natural to believe in preferential attachment in many real networks. For example, one is more likely to get to know a person who already knows many people, making preferential attachment not unlikely in social networks. However, the precise form of preferential attachment in (8.1.1) is only one possible example.

The above model is a slight variation of models that have appeared in the literature. The model with $\delta = 0$ is the *Barabási-Albert model*, which has received substantial attention in the literature and which was first formally defined in [48]. We have added the extra parameter δ to make the model more general.

The definition of $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ in terms of $\{\mathrm{PA}_t(1,\delta/m)\}_{t=1}^{\infty}$ is quite convenient. However, we can also equivalently define the model for $m \geq 2$ directly. We start with $\mathrm{PA}_1(m,\delta)$ consisting of a single vertex with m self-loops. To construct $\mathrm{PA}_{t+1}(m,\delta)$ from $\mathrm{PA}_t(m,\delta)$, we add a single vertex with m edges attached to it. These edges are attached *sequentially* with intermediate updating of the degrees as follows. The e^{th} edge is connected to vertex $v_i^{(m)}$, for $i \in [t]$, with probability proportional to $(D_i(e-1,t)+\delta)$, where, for $e = 1,\ldots,m$, $D_i(e,t)$ is the degree of vertex i after the e^{th} edge is attached, and to vertex $v_{t+1}^{(m)}$ with probability proportional to $(D_{t+1}(e-1,t)+e\delta/m)$, with the convention that $D_{t+1}(0,t) = 1$. This alternative definition makes it perfectly clear how the choices missing in [20] are made. Indeed, the degrees are updated during the process of attaching the edges, and the initial graph at time 1 consists of a single vertex with m self-loops. Naturally, the edges could also be attached sequentially by a different rule, for example by attaching the edges independently according to the distribution for the first edge. Also, one has the choice to allow for self-loops or not. See Figure 8.1 for a realization of $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ for $m = 2$ and $\delta = 0$, and Figure 8.2 for a realization of $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ for $m = 2$ and $\delta = -1$.

Exercise 8.5 (Collapsing vs. growth of the PA model)**.** *Prove that the alternative definition of* $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ *is indeed equal to the one obtained by collapsing* m *consecutive vertices in* $\{\mathrm{PA}_t(1,\delta/m)\}_{t=1}^{\infty}$.

Exercise 8.6 (Graph topology for $\delta = -1$)**.** *Show that when* $\delta = -1$*, the graph* $\mathrm{PA}_t(1,\delta)$ *consists of a self-loop at vertex* $v_1^{(1)}$*, and each other vertex is connected to* $v_1^{(1)}$ *with precisely one edge. What is the implication of this result for* $m > 1$*?*

In some cases, it will be convenient to consider a slight variation on the above model where, for $m = 1$, self-loops do not occur. We shall denote this variation by $\{\mathrm{PA}_t^{(b)}(m,\delta)\}_{t\geq 2}$ and sometimes refer to this model by *model (b)*. To define $\mathrm{PA}_t^{(b)}(1,\delta)$, we let $\mathrm{PA}_2^{(b)}(1,\delta)$

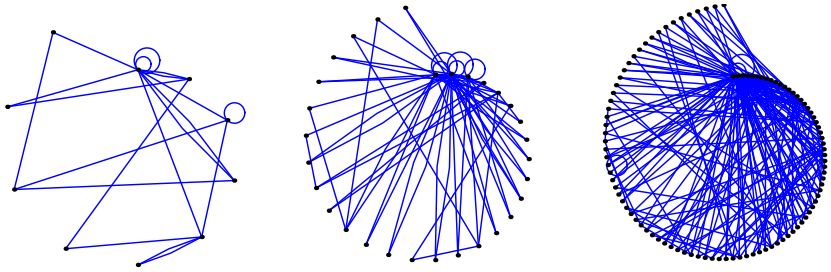

Figure 8.2: Preferential attachment random graph with $m = 2$ and $\delta = -1$ of sizes 10, 30 and 100.

consist of two vertices $v_1^{(1)}$ and $v_2^{(1)}$ with two edges between them, and we replace the growth rule in (8.1.1) by the rule that, for all $i \in [t]$,

$$\mathbb{P}\big(v_{t+1}^{(1)} \to v_i^{(1)} \big| \mathrm{PA}_t^{(b)}(1,\delta)\big) = \frac{D_i(t)+\delta}{t(2+\delta)}. \tag{8.1.2}$$

The advantage of this model is that it leads to a *connected graph*. We again define the model with $m \geq 2$ and $\delta > -m$ in terms of $\{\mathrm{PA}_t^{(b)}(1,\delta/m)\}_{t=2}^{\infty}$ as below Exercise 8.3. We also note that the differences between $\{\mathrm{PA}_t(m,\delta)\}_{t\geq 1}$ and $\{\mathrm{PA}_t^{(b)}(m,\delta)\}_{t\geq 2}$ are minor, since the probability of a self-loop in $\mathrm{PA}_t(m,\delta)$ is quite small when t is large. Thus, most of the results we shall prove in this chapter for $\{\mathrm{PA}_t(m,\delta)\}_{t\geq 1}$ shall also apply to $\{\mathrm{PA}_t^{(b)}(m,\delta)\}_{t\geq 2}$, but we shall not state these extensions explicitly.

Interestingly, the above model with $\delta \geq 0$ can be viewed as an interpolation between the models with $\delta = 0$ and $\delta = \infty$. We show this for $m = 1$, the statement for $m \geq 2$ can again be seen by collapsing the vertices. We again let the graph at time 2 consist of two vertices with two edges between them. We fix $\alpha \in [0,1]$. Then, we first draw a random variable X_{t+1} taking values 0 with probability α and $X_{t+1} = 1$ with probability $1-\alpha$. The random variables $\{X_t\}_{t=1}^{\infty}$ are independent. When $X_{t+1} = 0$, then we attach the $(t+1)^{\text{st}}$ edge to a *uniform* vertex in $[t]$. When $X_{t+1} = 1$, then we attach the $(t+1)^{\text{st}}$ edge to vertex $i \in [t]$ with probability $D_i(t)/(2t)$. We denote this model by $\{\mathrm{PA}_t^{(b')}(1,\alpha)\}_{t=1}^{\infty}$. When $\alpha \geq 0$ is chosen appropriately, then this is precisely the above preferential attachment model:

Exercise 8.7 (Alternative formulation of $\mathrm{PA}_t(1,\delta)$)**.** *For $\alpha = \frac{\delta}{2+\delta}$, the law of $\{\mathrm{PA}_t^{(b')}(1,\alpha)\}_{t=2}^{\infty}$ is equal to the one of $\{\mathrm{PA}_t(1,\delta)\}_{t=1}^{\infty}$.*

Exercise 8.8 (Degrees grow to infinity a.s.)**.** *Fix $m = 1$. Prove that $D_i(t) \xrightarrow{a.s.} \infty$. Hint: use that, with $\{I_t\}_{t=i}^{\infty}$ a sequence of independent Bernoulli random variables with $\mathbb{P}(I_t = 1) = (1+\delta)/(t(2+\delta)+1+\delta)$, we have that $\sum_{s=i}^{t} I_s \preceq D_i(t)$. What does this imply for $m > 1$?*

8.2 Degrees of fixed vertices

We start by investigating the degrees of given vertices. To formulate the results, we define the *Gamma-function* $t \mapsto \Gamma(t)$ for $t > 0$ by

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx. \tag{8.2.1}$$

We also make use of the recursion formula

$$\Gamma(t+1) = t\Gamma(t). \tag{8.2.2}$$

Exercise 8.9 (Recursion formula for the Gamma function)**.** *Prove* (8.2.2) *using partial integration, and also prove that* $\Gamma(n) = (n-1)!$ *for* $n = 1, 2, \ldots$.

The main result in this section is the following:

Theorem 8.1 (Degrees of fixed vertices)**.** *Fix* $m = 1$ *and* $\delta > -1$*. Then,* $D_i(t)/t^{\frac{1}{2+\delta}}$ *converges almost surely to a random variable* ξ_i *as* $t \to \infty$*, and*

$$\mathbb{E}[D_i(t) + \delta] = (1+\delta)\frac{\Gamma(t+1)\Gamma(i - \frac{1}{2+\delta})}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(i)}. \tag{8.2.3}$$

In Section 8.6, we shall considerably extend the result in Theorem 8.1. For example, we shall also prove the almost sure convergence of *maximal* degree.

Proof. Fix $m = 1$. We compute that

$$\begin{aligned}
\mathbb{E}[D_i(t+1) + \delta | D_i(t)] &= D_i(t) + \delta + \mathbb{E}[D_i(t+1) - D_i(t) | D_i(t)] \\
&= D_i(t) + \delta + \frac{D_i(t) + \delta}{(2+\delta)t + 1 + \delta} \\
&= (D_i(t) + \delta)\frac{(2+\delta)t + 2 + \delta}{(2+\delta)t + 1 + \delta} \\
&= (D_i(t) + \delta)\frac{(2+\delta)(t+1)}{(2+\delta)t + 1 + \delta}.
\end{aligned} \tag{8.2.4}$$

Using also that

$$\begin{aligned}
\mathbb{E}[D_i(i) + \delta] &= 1 + \delta + \frac{1+\delta}{(2+\delta)(i-1) + 1 + \delta} = (1+\delta)\frac{(2+\delta)(i-1) + 2 + \delta}{(2+\delta)(i-1) + 1 + \delta} \\
&= (1+\delta)\frac{(2+\delta)i}{(2+\delta)(i-1) + 1 + \delta},
\end{aligned} \tag{8.2.5}$$

we obtain that

$$M_i(t) = \frac{D_i(t) + \delta}{1+\delta} \prod_{s=i-1}^{t-1} \frac{(2+\delta)s + 1 + \delta}{(2+\delta)(s+1)} \tag{8.2.6}$$

is a non-negative martingale with mean 1. As a consequence of the martingale convergence theorem (Theorem 2.21), as $t \to \infty$, $M_i(t)$ converges almost surely to a limiting random variable ξ_i.

We compute that

$$\prod_{s=i-1}^{t-1} \frac{(2+\delta)s + 1 + \delta}{(2+\delta)s + 2 + \delta} = \prod_{s=i-1}^{t-1} \frac{s + \frac{1+\delta}{2+\delta}}{s+1} = \frac{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(i)}{\Gamma(t+1)\Gamma(i - \frac{1}{2+\delta})}. \tag{8.2.7}$$

It is not hard to see that, using Stirling's formula,

$$\frac{\Gamma(t+a)}{\Gamma(t)} = t^a(1 + O(1/t)). \tag{8.2.8}$$

Therefore, we have that $D_i(t)/t^{\frac{1}{2+\delta}}$ converges in distribution to a random variable M_i having expected value $(1+\delta)\frac{\Gamma(i-\frac{1}{2+\delta})}{\Gamma(i)}$. In particular, the degrees of the first i vertices at time t is at most of order $t^{\frac{1}{2+\delta}}$. Note, however, that we do not yet know whether $\mathbb{P}(\xi_i = 0) = 0$ or not!

Exercise 8.10 (Asymptotics for ratio $\Gamma(t+a)/\Gamma(t)$)**.** *Prove* (8.2.8)*, using that [109, 8.327]*

$$e^{-t}t^{t+\frac{1}{2}}\sqrt{2\pi} \le \Gamma(t+1) \le e^{-t}t^{t+\frac{1}{2}}\sqrt{2\pi}e^{\frac{1}{12t}}. \tag{8.2.9}$$

□

Note that we can extend the above result to the case when $m \ge 1$, by using the relation between $\mathrm{PA}_{m,\delta}(t)$ and $\mathrm{PA}_{mt}(1, \delta/m)$. This implies in particular that

$$\mathbb{E}_m^{\delta}[D_i(t)] = \sum_{s=1}^{m} \mathbb{E}_1^{\delta/m}[D_{m(i-1)+s}(mt)], \tag{8.2.10}$$

where we have added a subscript m and a superscript δ to denote the values of m and δ involved.

Exercise 8.11 (Mean degree for $m \ge 2$)**.** *Prove* (8.2.10) *and use it to compute* $\mathbb{E}_m^{\delta}[D_i(t)]$.

Exercise 8.12 (A.s. limit of degrees for $m \ge 2$)**.** *Prove that, for* $m \ge 2$ *and any* $i \ge 1$, $D_i(t)(mt)^{-1/(2+\delta/m)} \xrightarrow{a.s.} \xi_i'$, *where*

$$\xi_i' = \sum_{j=(i-1)m+1}^{mi} \xi_j, \tag{8.2.11}$$

and ξ_j *is the almost sure limit of* $D_j(t)$ *in* $\{\mathrm{PA}_t(1, \delta/m)\}_{t=1}^{\infty}$.

Exercise 8.13 (Mean degree for model (b))**.** *Prove that for* $\mathrm{PA}_t^{(b)}(1, \delta)$, (8.2.3) *is adapted to*

$$\mathbb{E}[D_i(t) + \delta] = (1+\delta)\frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(i)}{\Gamma(t)\Gamma(i+\frac{1}{2+\delta})}. \tag{8.2.12}$$

We close this section by giving a heuristic explanation for the occurrence of a power-law degree sequence in preferential attachment models. Theorem 8.1 in conjunction with Exercise 8.12 implies that there exists an a_m such that, for i, t large, and any $m \ge 1$,

$$\mathbb{E}[D_i(t)] \sim a_m \Big(\frac{t}{i}\Big)^{1/(2+\delta/m)}. \tag{8.2.13}$$

When the graph indeed has a power-law degree sequence, then the number of vertices with degrees at least k will be close to $ctk^{-(\tau-1)}$ for some $\tau > 1$ and some $c > 0$. The number of vertices with degree at least k at time t is equal to $N_{\ge k}(t) = \sum_{i=1}^{t} \mathbb{1}_{\{D_i(t)\ge k\}}$. Now, assume that in the above formula, we are allowed to replace $\mathbb{1}_{\{D_i(t)\ge k\}}$ by $\mathbb{1}_{\{\mathbb{E}[D_i(t)]\ge k\}}$ (there is a big leap of faith here). Then we would obtain that

$$\begin{aligned} N_{\ge k}(t) &\sim \sum_{i=1}^{t} \mathbb{1}_{\{\mathbb{E}[D_i(t)]\ge k\}} \sim \sum_{i=1}^{t} \mathbb{1}_{\{a_m\left(\frac{t}{i}\right)^{1/(2+\delta/m)}\ge k\}} \\ &= \sum_{i=1}^{t} \mathbb{1}_{\{i\le t a_m^{2+\delta/m} k^{-(2+\delta/m)}\}} = t a_m^{2+\delta/m} k^{-(2+\delta/m)}, \end{aligned} \tag{8.2.14}$$

so that we obtain a power-law with exponent $\tau - 1 = 2 + \delta/m$, so that $\tau = 3 + \delta/m$. The above heuristic shall be made precise in the following section, but the proof will be quite a bit more subtle than the above heuristic!

8.3 Degree sequences of preferential attachment models

The main result establishes the scale-free nature of preferential attachment graphs. In order to state the result, we need some notation. We write

$$P_k(t) = \frac{1}{t}\sum_{i=1}^{t} \mathbb{1}_{\{D_i(t)=k\}} \tag{8.3.1}$$

for the (random) proportion of vertices with degree k at time t. For $m \geq 1$ and $\delta > -m$, we define $\{p_k\}_{k=0}^{\infty}$ to be the probability distribution given by $p_k = 0$ for $k = 0, \ldots, m-1$ and, for $k \geq m$,

$$p_k = (2 + \frac{\delta}{m})\frac{\Gamma(k+\delta)\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)\Gamma(k+3+\delta+\frac{\delta}{m})} \tag{8.3.2}$$

For $m = 1$, (8.3.2) reduces to

$$p_k = (2+\delta)\frac{\Gamma(k+\delta)\Gamma(3+2\delta)}{\Gamma(k+3+2\delta)\Gamma(1+\delta)}. \tag{8.3.3}$$

Also, when $\delta = 0$ and $k \geq m$, (8.3.2) simplifies to

$$p_k = \frac{2\Gamma(k)\Gamma(m+2)}{\Gamma(k+3)\Gamma(m)} = \frac{2m(m+1)}{k(k+1)(k+2)}. \tag{8.3.4}$$

We start by proving that $\{p_k\}_{k=1}^{\infty}$ is a *probability distribution*. For this, we note that, by (8.2.2),

$$\frac{\Gamma(k+a)}{\Gamma(k+b)} = \frac{1}{b-a-1}\Big(\frac{\Gamma(k+a)}{\Gamma(k-1+b)} - \frac{\Gamma(k+1+a)}{\Gamma(k+b)}\Big). \tag{8.3.5}$$

Applying (8.3.5) to $a = \delta$, $b = 3+\delta+\frac{\delta}{m}$, we obtain that, for $k \geq m$,

$$p_k = \frac{\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)}\Big(\frac{\Gamma(k+\delta)}{\Gamma(k+2+\delta+\frac{\delta}{m})} - \frac{\Gamma(k+1+\delta)}{\Gamma(k+3+\delta+\frac{\delta}{m})}\Big). \tag{8.3.6}$$

Using that $p_k = 0$ for $k < m$, and by a telescoping sum identity,

$$\sum_{k\geq 1} p_k = \sum_{k\geq m} p_k = \frac{\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)}\frac{\Gamma(m+\delta)}{\Gamma(m+2+\delta+\frac{\delta}{m})} = 1. \tag{8.3.7}$$

Thus, since also $p_k \geq 0$, we obtain that $\{p_k\}_{k=1}^{\infty}$ indeed is a probability distribution. We shall see that $\{p_k\}_{k=1}^{\infty}$ arises as the limiting degree distribution for $\mathrm{PA}_{m,\delta}(t)$:

Theorem 8.2 (Degree sequence in preferential attachment model)**.** *Fix $\delta > -m$ and $m \geq 1$. Then, there exists a constant $C = C(m,\delta) > 0$ such that, as $t \to \infty$,*

$$\mathbb{P}\Big(\max_k |P_k(t) - p_k| \geq C\sqrt{\frac{\log t}{t}}\Big) = o(1). \tag{8.3.8}$$

Theorem 8.2 identifies the asymptotic degree sequence of $\mathrm{PA}_{m,\delta}(t)$ as $\{p_k\}_{k=1}^{\infty}$. We next show that, for k large, p_k is close to a power-law distribution. For this, we first note that from (8.3.2) and (8.2.8), as $k \to \infty$,

$$p_k = c_{m,\delta}k^{-\tau}(1 + O(\frac{1}{k})), \tag{8.3.9}$$

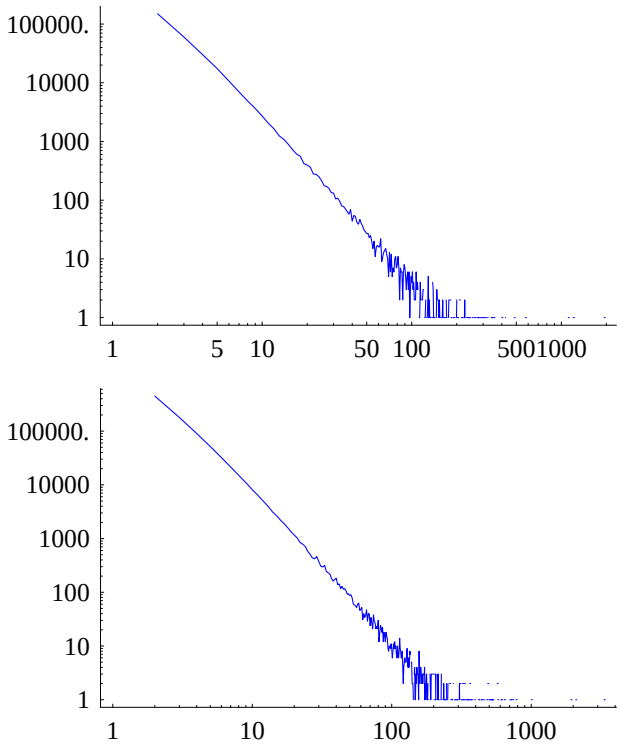

Figure 8.3: The degree sequences of a preferential attachment random graph with $m = 2, \delta = 0$ of sizes 300,000 and 1,000,000 in log-log scale.

where

$$\tau = 3 + \frac{\delta}{m} > 2, \tag{8.3.10}$$

and

$$c_{m,\delta} = \frac{(2 + \frac{\delta}{m})\Gamma(m + 2 + \delta + \frac{\delta}{m})}{\Gamma(m + \delta)}. \tag{8.3.11}$$

Therefore, by Theorem 8.2 and (8.3.9), the asymptotic degree sequence of $\mathrm{PA}_{m,\delta}(t)$ is close to a power law with exponent $\tau = 3 + \delta/m$. We note that any exponent $\tau > 2$ is possible by choosing $\delta > -m$ and $m \geq 1$ appropriately. The power-law degree sequence can clearly be observed in a simulation, see Figure 8.3, where a realization of the degree sequence of $\mathrm{PA}_{m,\delta}(t)$ is shown for $m = 2$, $\delta = 0$ and $t = 300,000$ and $t = 1,000,000$.

The important feature of the preferential attachment model is that, unlike the configuration model and the generalized random graph, the power law in $\mathrm{PA}_{m,\delta}(t)$ is *explained* by giving a model for the growth of the graph that produces power-law degrees. Therefore, preferential attachment offers a convincing explanation as to *why* power-law degree sequences occur. As Barabási puts it [19]

> "*...the scale-free topology is evidence of organizing principles acting at each stage of the network formation. (...) No matter how large and complex a*

network becomes, as long as preferential attachment and growth are present it will maintain its hub-dominated scale-free topology.".

Many more possible explanations have been given for *why* power laws occur in real networks, and many adaptations of the above simple preferential attachment model have been studied in the literature, all giving rise to power-law degrees. See Section 8.7 for an overview of the literature.

The remainder of this chapter shall be primarily devoted to the proof of Theorem 8.2, which is divided into two main parts. In Section 8.4, we prove that the degree sequence is concentrated around its mean, and in Section 8.5, we identify the mean of the degree sequence. In the course of the proof, we also prove results related to Theorem 8.2.

Exercise 8.14 (The degree of a uniform vertex)**.** *Prove that Theorem 8.2 implies that the degree at time t of a uniform vertex in $[t]$ converges in probability to a random variable with probability mass function $\{p_k\}_{k=1}^{\infty}$.*

Exercise 8.15 (Degree sequence uniform recursive tree [126])**.** *In a uniform recursive tree we attach each vertex to a uniformly chosen old vertex. This can be seen as the case where $m = 1$ and $\delta = \infty$ of $\{\mathrm{PA}_t^{(b)}(m, \delta)\}_{t\geq 2}$. Show that Theorem 8.2 remains true, but now with $p_k = 2^{-(k+1)}$.*

8.4 Concentration of the degree sequence

In this section, we prove that the (random) degree sequence is sufficiently concentrated around its expected degree sequence. We use a martingale argument which first appeared in [50], and has been used in basically all subsequent works proving power-law degree sequences for preferential attachment models. The argument is very pretty and general, and we spend some time explaining the details of it.

We start by stating the main result in this section. In its statement, we use the notation

$$N_k(t) = \sum_{i=1}^{t} \mathbb{1}_{\{D_i(t)=k\}} = tP_k(t) \tag{8.4.1}$$

for the total number of vertices with degree k at time t.

Proposition 8.3 (Concentration of the degrees)**.** *Fix $\delta \geq -m$ and $m \geq 1$. Then, for any $C > m\sqrt{8}$, as $t \to \infty$,*

$$\mathbb{P}\Big(\max_k |N_k(t) - \mathbb{E}[N_k(t)]| \geq C\sqrt{t \log t}\Big) = o(1). \tag{8.4.2}$$

We note that Theorem 8.2 predicts that $N_k(t) \approx tp_k$. Thus, at least for k for which p_k is not too small, i.e., $tp_k \gg \sqrt{t \log t}$, Proposition 8.3 suggests that the number of vertices with degree equal to k is very close to its expected value. Needless to say, in order to prove Theorem 8.2, we still need to investigate $\mathbb{E}[N_k(t)]$, and prove that it is quite close to tp_k. This is the second main ingredient in the proof of Theorem 8.2 and is formulated in Proposition 8.4. We first prove Proposition 8.3.

Proof. We start by reducing the proof. First of all, $N_k(t) = 0$ when $k > m(t + 1)$.

Therefore,

$$\begin{aligned}\mathbb{P}\Big(\max_k |N_k(t) - \mathbb{E}[N_k(t)]| \geq C\sqrt{t\log t}\Big) &= \mathbb{P}\Big(\max_{k\leq m(t+1)} |N_k(t) - \mathbb{E}[N_k(t)]| \geq C\sqrt{t\log t}\Big)\\ &\leq \sum_{k=1}^{m(t+1)} \mathbb{P}\Big(|N_k(t) - \mathbb{E}[N_k(t)]| \geq C\sqrt{t\log t}\Big).\end{aligned} \tag{8.4.3}$$

We shall prove that for any $C > m\sqrt{8}$, uniformly in $k \leq m(t+1)$,

$$\mathbb{P}\Big(|N_k(t) - \mathbb{E}[N_k(t)]| \geq C\sqrt{t\log t}\Big) = o(t^{-1}), \tag{8.4.4}$$

which would complete the proof of Proposition 8.3.

For $n = 0, \ldots, t$, we denote by

$$M_n = \mathbb{E}\big[N_k(t)|\mathrm{PA}_n(m,\delta)\big] \tag{8.4.5}$$

the conditional expected number of vertices with degree k at time t, conditionally on the graph $\mathrm{PA}_n(m,\delta)$ at time $n \in \{0,\ldots,t\}$. We shall show that $\{M_n\}_{n=0}^t$ is a martingale.

Firstly, since $N_k(t)$ is bounded by the total number of vertices at time t, we have $N_k(t) \leq t$, so that

$$\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = \mathbb{E}[N_k(t)] \leq t < \infty. \tag{8.4.6}$$

Secondly, by the tower property of conditional expectations, and the fact that $\mathrm{PA}_n(m,\delta)$ can be deduced from $\mathrm{PA}_{m,\delta}(n+1)$, we have that, for all $n \leq t-1$,

$$\begin{aligned}\mathbb{E}[M_{n+1}|\mathrm{PA}_n(m,\delta)] &= \mathbb{E}\Big[\mathbb{E}\big[N_k(t)|\mathrm{PA}_{m,\delta}(n+1)\big]\Big|\mathrm{PA}_n(m,\delta)\Big]\\ &= \mathbb{E}\big[N_k(t)|\mathrm{PA}_n(m,\delta)\big] = M_n,\end{aligned} \tag{8.4.7}$$

so that $\{M_n\}_{n=0}^t$ satisfies the conditional expectation requirement for a martingale. In fact, $\{M_n\}_{n=0}^t$ is a so-called *Doob martingale* (see also Exercise 2.22).

Therefore, $\{M_n\}_{n=0}^t$ also satisfies the moment condition for martingales. We conclude that $\{M_n\}_{n=0}^t$ is a martingale process with respect to $\{\mathrm{PA}_n(m,\delta)\}_{n=0}^t$. This is the first main ingredient of the martingale proof of (8.4.4).

For the second ingredient, we note that M_0 is identified as

$$M_0 = \mathbb{E}\big[N_k(t)|\mathrm{PA}_{m,\delta}(0)\big] = \mathbb{E}[N_k(t)], \tag{8.4.8}$$

Since $\mathrm{PA}_{m,\delta}(0)$ is the empty graph. Furthermore, M_t is trivially identified as

$$M_t = \mathbb{E}\big[N_k(t)|\mathrm{PA}_t(m,\delta)\big] = N_k(t), \tag{8.4.9}$$

since one can determine $N_k(t)$ from $\mathrm{PA}_t(m,\delta)$. Therefore, we have that

$$N_k(t) - \mathbb{E}[N_k(t)] = M_t - M_0. \tag{8.4.10}$$

This completes the second key ingredient in the martingale proof of (8.4.4).

The third key ingredient is the Azuma-Hoeffding inequality, Theorem 2.23. For this, we need to investigate the support of $|M_n - M_{n-1}|$. We claim that, for all $n \in [t]$, a.s.,

$$|M_n - M_{n-1}| \leq 2m. \tag{8.4.11}$$

In order to prove this, we note that

$$M_n = \mathbb{E}[N_k(t)|\mathrm{PA}_n(m,\delta)] = \sum_{i=1}^{t} \mathbb{P}(D_i(t) = k|\mathrm{PA}_n(m,\delta)), \tag{8.4.12}$$

and, similarly,

$$M_{n-1} = \sum_{i=1}^{t} \mathbb{P}(D_i(t) = k|\mathrm{PA}_{n-1}(m,\delta)), \tag{8.4.13}$$

so that

$$M_n - M_{n-1} = \sum_{i=1}^{t} \mathbb{P}(D_i(t) = k|\mathrm{PA}_n(m,\delta)) - \mathbb{P}(D_i(t) = k|\mathrm{PA}_{n-1}(m,\delta)). \tag{8.4.14}$$

Thus, we need to investigate the influence of the extra information contained in $\mathrm{PA}_n(m,\delta)$ compared to the information contained in $\mathrm{PA}_{n-1}(m,\delta)$. For any $s = 1,\ldots,t$, conditioning on $\mathrm{PA}_s(m,\delta)$ is the *same* as conditioning to which vertices the first sm edges are attached. Thus, in $\mathrm{PA}_{n-1}(m,\delta)$, we know where the edges of the vertices $v_1^{(m)},\ldots,v_{n-1}^{(m)}$ are attached to. In $\mathrm{PA}_n(m,\delta)$, we have the additional information of where the m edges originating from the vertex $v_n^{(m)}$ are attached to. These m edges effect the degrees of at most m other vertices, namely, the receiving ends of these m edges.

For the conditional expectations given $\mathrm{PA}_s(m,\delta)$, we need to take the expectation with respect to all possible ways of attaching the remaining edges originating from the vertices $v_{s+1}^{(m)},\ldots,v_t^{(m)}$. As explained above, only the distribution of the degrees of the vertices in $\mathrm{PA}_t(m,\delta)$ to which the m edges originating from $v_n^{(m)}$ are connected are effected by the knowledge of $\mathrm{PA}_n(m,\delta)$ compared to $\mathrm{PA}_{n-1}(m,\delta)$. This number of vertices is at most m, so that the distribution of the degrees of at most $2m$ vertices is different in the law of $\mathrm{PA}_t(m,\delta)$ conditionally on $\mathrm{PA}_{n-1}(m,\delta)$ compared to the law of $\mathrm{PA}_t(m,\delta)$ conditionally on $\mathrm{PA}_n(m,\delta)$. This implies (8.4.11).

The Azuma-Hoeffding's Inequality (Theorem 2.23) then yields that, for any $a > 0$,

$$\mathbb{P}\Big(|N_k(t) - \mathbb{E}[N_k(t)]| \geq a\Big) \leq 2e^{-\frac{a^2}{8m^2t}}. \tag{8.4.15}$$

Taking $a = C\sqrt{t\log t}$ for any C with $C^2 > 8m^2$ then proves that

$$\mathbb{P}\Big(|N_k(t) - \mathbb{E}[N_k(t)]| \geq C\sqrt{t\log t}\Big) \leq 2e^{-(\log t)\frac{C^2}{8m^2}} = o(t^{-1}). \tag{8.4.16}$$

This completes the proof of (8.4.4), and thus of Proposition 8.3. □

The above proof is rather general, and can also be used to prove concentration around the mean of other graph properties that are related to the degrees. An example is the following. Denote by

$$N_{\geq k}(t) = \sum_{l=k}^{\infty} N_l(t) \tag{8.4.17}$$

the total number of vertices with degrees at least k. Then we can also prove that $N_{\geq k}(t)$ concentrates. Indeed, for $C > \sqrt{8m}$,

$$\mathbb{P}\Big(|N_{\geq k}(t) - \mathbb{E}[N_{\geq k}(t)]| \geq C\sqrt{t\log t}\Big) = o(t^{-1}). \tag{8.4.18}$$

The proof uses the same ingredients as given above for $N_{\geq k}(t)$, where now we can make use of the martingale

$$M_n' = \mathbb{E}[N_{\geq k}(t)|\mathrm{PA}_n(m,\delta)]. \tag{8.4.19}$$

Exercise 8.16 (Concentration of the number of vertex of degree at least k)**.** *Prove* (8.4.18) *by adapting the proof of Proposition 8.3.*

8.5 Expected degree sequence

The main result in this section investigates the expected number of vertices with degree equal to k. We denote the expected number of vertices of degree k in $\mathrm{PA}_t(m,\delta)$ by

$$\bar{N}_k(t) = \mathbb{E}\big[tP_k(t)\big]. \tag{8.5.1}$$

The main aim is to prove that $\bar{N}_k(t)$ is close to $p_k t$, where p_k is defined in (8.3.3). This is the content of the following proposition:

Proposition 8.4 (Expected degree sequence)**.** *Fix $\delta > -m$ and $m \geq 1$. Then, there exists a constant $C = C(\delta, m)$ such that for all $t \geq 1$ and all $k \in \mathbb{N}$,*

$$|\bar{N}_k(t) - p_k t| \leq C. \tag{8.5.2}$$

The proof of Proposition 8.4 is split into two separate cases. We first prove the claim for $m = 1$ in Section 8.5.1, and extend the proof to $m > 1$ in Section 8.5.2.

Exercise 8.17 (The total degree of high degree vertices)**.** *Use Propositions 8.4 and 8.3 to prove that for $l = l(t) \to \infty$ as $t \to \infty$ such that $tl^{2-\tau} \geq K\sqrt{t \log t}$ for some $K > 0$ sufficiently large, there exists a constant $B > 0$ such that with probability exceeding $1 - o(t^{-1})$, for all such l,*

$$\sum_{i:D_i(t)\geq l} D_i(t) \geq Btl^{2-\tau}. \tag{8.5.3}$$

Show further that, with probability exceeding $1 - o(t^{-1})$, for all such l,

$$N_{\geq l}(t) \gg \sqrt{t}. \tag{8.5.4}$$

8.5.1 Expected degree sequence for $m = 1$

In this section, we study the expected degree sequence when $m = 1$. We adapt the argument in [45]. We start by writing

$$\mathbb{E}\big[N_k(t+1)|\mathrm{PA}_t(1,\delta)\big] = N_k(t) + \mathbb{E}[N_k(t+1) - N_k(t)|\mathrm{PA}_t(1,\delta)]. \tag{8.5.5}$$

Conditionally on $\mathrm{PA}_t(1,\delta)$, there are four ways how $N_k(t+1) - N_k(t)$ can be unequal to zero:

(a) The end vertex of the (unique) edge incident to vertex $v_{t+1}^{(1)}$ had degree $k-1$, so that its degree is increased to k, which happens with probability $\frac{k-1+\delta}{t(2+\delta)+(1+\delta)}$. Note that there are $N_{k-1}(t)$ end vertices with degree $k-1$ at time t;

(b) The end vertex of the (unique) edge incident to vertex $v_{t+1}^{(1)}$ had degree k, so that its degree is increased to $k+1$, which happens with probability $\frac{k+\delta}{t(2+\delta)+(1+\delta)}$. Note that there are $N_k(t)$ end vertices with degree k at time t;

(c) The degree of vertex $v_{t+1}^{(1)}$ is one, so that $N_1(t)$ is increased by one, when the end vertex of the (unique) edge incident to vertex $v_{t+1}^{(1)}$ is not $v_{t+1}^{(1)}$, which happens with probability $1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}$;

(d) The degree of vertex $v_{t+1}^{(1)}$ is equal to two, so that $N_2(t)$ is increased by one, when the end vertex of the (unique) edge incident to vertex $v_{t+1}^{(1)}$ is equal to $v_{t+1}^{(1)}$, which happens with probability $\frac{1+\delta}{t(2+\delta)+(1+\delta)}$.

The changes in the degree sequence in cases (a) and (b) arise due to the attachment of the edge (thus, the degree of one of the vertices $v_1^{(1)}, \ldots, v_t^{(1)}$ is changed), whereas in cases (c) and (d) we determine the degree of the added vertex $v_{t+1}^{(1)}$.

Taking all these cases into account, we arrive at the key identity

$$\begin{aligned}\mathbb{E}\big[N_k(t+1) - N_k(t) | \mathrm{PA}_t(1,\delta)\big] &= \frac{k-1+\delta}{t(2+\delta)+(1+\delta)} N_{k-1}(t) \\ &\quad - \frac{k+\delta}{t(2+\delta)+(1+\delta)} N_k(t) \\ &\quad + \mathbb{1}_{\{k=1\}} \Big(1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}\Big) \\ &\quad + \mathbb{1}_{\{k=2\}} \frac{1+\delta}{t(2+\delta)+(1+\delta)}. \end{aligned} \tag{8.5.6}$$

Here, $k \geq 1$, and for $k = 0$, by convention, we define

$$N_0(t) = 0. \tag{8.5.7}$$

By taking the expectation on both sides of (8.5.6), obtain

$$\begin{aligned}\mathbb{E}[N_k(t+1)] &= \mathbb{E}[N_k(t)] + \mathbb{E}[N_k(t+1) - N_k(t)] \\ &= \mathbb{E}[N_k(t)] + \mathbb{E}\big[\mathbb{E}[N_k(t+1) - N_k(t) | \mathrm{PA}_t(1,\delta)]\big]. \end{aligned} \tag{8.5.8}$$

Now using (8.5.6) gives us the explicit recurrence relation that, for $k \geq 1$,

$$\begin{aligned}\bar{N}_k(t+1) &= \bar{N}_k(t) + \frac{k-1+\delta}{t(2+\delta)+(1+\delta)} \bar{N}_{k-1}(t) \\ &\quad - \frac{k+\delta}{t(2+\delta)+(1+\delta)} \bar{N}_k(t) \\ &\quad + \mathbb{1}_{\{k=1\}} \big(1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}\big) \\ &\quad + \mathbb{1}_{\{k=2\}} \frac{1+\delta}{t(2+\delta)+(1+\delta)}. \end{aligned} \tag{8.5.9}$$

Equation (8.5.9) will the the key to the proof of Proposition 8.4 for $m = 1$. We start by explaining its relation to (8.3.3). Indeed, when $\bar{N}_k(t) \approx t p_k$, then one might expect that $\bar{N}_k(t+1) - \bar{N}_k(t) \approx p_k$. Substituting these approximations into (8.5.9), and approximating $t/(t(2+\delta)+(1+\delta)) \approx 1/(2+\delta)$ and $\frac{1+\delta}{t(2+\delta)+(1+\delta)} \approx 0$, we arrive at the fact that p_k must satisfy the recurrence relation, for $k \geq 1$,

$$p_k = \frac{k-1+\delta}{2+\delta} p_{k-1} - \frac{k+\delta}{2+\delta} p_k + \mathbb{1}_{\{k=1\}}, \tag{8.5.10}$$

where we define $p_0 = 0$. We now show that the unique solution to (8.5.10) is (8.3.3). We can rewrite

$$p_k = \frac{k-1+\delta}{k+2+2\delta} p_{k-1} + \frac{2+\delta}{k+2+2\delta} \mathbb{1}_{\{k=1\}}. \tag{8.5.11}$$

When $k = 1$, using that $p_0 = 0$, we obtain

$$p_1 = \frac{2+\delta}{3+2\delta}. \tag{8.5.12}$$

On the other hand, when $k > 1$, we arrive at

$$p_k = \frac{k-1+\delta}{k+2+2\delta} p_{k-1}. \tag{8.5.13}$$

Therefore, using (8.2.2) repeatedly,

$$\begin{aligned} p_k &= \frac{\Gamma(k+\delta)\Gamma(4+2\delta)}{\Gamma(k+3+2\delta)\Gamma(1+\delta)} p_1 = \frac{(2+\delta)\Gamma(k+\delta)\Gamma(4+2\delta)}{(3+2\delta)\Gamma(k+3+2\delta)\Gamma(1+\delta)} \\ &= \frac{(2+\delta)\Gamma(k+\delta)\Gamma(3+2\delta)}{\Gamma(k+3+2\delta)\Gamma(1+\delta)}, \end{aligned} \tag{8.5.14}$$

and we see that the unique solution of (8.5.10) is p_k in (8.3.3).

The next step is to use (8.5.9) and (8.5.10) to prove Proposition 8.4 for $m = 1$. To this end, we define

$$\varepsilon_k(t) = \bar{N}_k(t) - tp_k. \tag{8.5.15}$$

Then, in order to prove Proposition 8.4 for $m = 1$, we are left to prove that there exists a constant $C = C(\delta)$ such that

$$\max_k |\varepsilon_k(t)| \le C. \tag{8.5.16}$$

The value of C will be determined in the course of the proof.

Now we deviate from the proof in [45]. In [45], induction in k was performed. Instead, we use induction in t. First of all, we note that we can rewrite (8.5.10) as

$$\begin{aligned} (t+1)p_k &= tp_k + p_k \\ &= tp_k + \frac{k-1+\delta}{2+\delta} p_{k-1} - \frac{k+\delta}{2+\delta} p_k + \mathbb{1}_{\{k=1\}} \\ &= tp_k + \frac{k-1+\delta}{t(2+\delta)+(1+\delta)} tp_{k-1} - \frac{k+\delta}{t(2+\delta)+(1+\delta)} tp_k + \mathbb{1}_{\{k=1\}} \\ &\quad + \Big(\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big)(k-1+\delta)p_{k-1} \\ &\quad - \Big(\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big)(k+\delta)p_k. \end{aligned} \tag{8.5.17}$$

We abbreviate

$$\kappa_k(t) = -\Big(\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big)\Big((k+\delta)p_k - (k-1+\delta)p_{k-1}\Big), \tag{8.5.18}$$

$$\gamma_k(t) = \frac{1+\delta}{t(2+\delta)+(1+\delta)}\big(\mathbb{1}_{\{k=2\}} - \mathbb{1}_{\{k=1\}}\big). \tag{8.5.19}$$

Then, (8.5.9) and (8.5.17) can be combined to yield that

$$\varepsilon_k(t+1) = \Big(1 - \frac{k+\delta}{t(2+\delta)+(1+\delta)}\Big)\varepsilon_k(t) + \frac{k-1+\delta}{t(2+\delta)+(1+\delta)}\varepsilon_{k-1}(t) + \kappa_k(t) + \gamma_k(t). \tag{8.5.20}$$

We prove the bounds on $\varepsilon_k(t)$ in (8.5.16) by induction on $t \ge 1$. We start by initializing the induction hypothesis. When $t = 1$, we have that $\mathrm{PA}_{1,\delta}(1)$ consists of a vertex with a single self-loop. Thus,

$$\bar{N}_k(1) = \mathbb{1}_{\{k=2\}}. \tag{8.5.21}$$

Therefore, since also $p_k \leq 1$, we arrive at the estimate that, uniformly in $k \geq 1$,

$$|\varepsilon_k(1)| = |\bar{N}_k(1) - p_k| \leq \max\{\bar{N}_k(1), p_k\} \leq 1. \tag{8.5.22}$$

We have initialized the induction hypothesis for $t = 1$ in (8.5.16) for any $C \geq 1$.

We next advance the induction hypothesis. We start with $k = 1$. In this case, we have that $\varepsilon_0(t) = N_0(t) - p_0 = 0$ by convention, so that (8.5.20) reduces to

$$\varepsilon_1(t+1) = \Big(1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}\Big)\varepsilon_1(t) + \kappa_1(t) + \gamma_1(t). \tag{8.5.23}$$

We note that

$$1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)} \geq 0, \tag{8.5.24}$$

so that

$$|\varepsilon_1(t+1)| \leq \Big(1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}\Big)|\varepsilon_1(t)| + |\kappa_1(t)| + |\gamma_1(t)|. \tag{8.5.25}$$

Using the explicit forms in (8.5.18) and (8.5.19), it is not hard to see that there are universal constants $C_\kappa = C_\kappa(\delta)$ and $C_\gamma = C_\gamma(\delta)$ such that, uniformly in $k \geq 1$,

$$|\kappa_k(t)| \leq C_\kappa (t+1)^{-1}, \qquad |\gamma_k(t)| \leq C_\gamma (t+1)^{-1}. \tag{8.5.26}$$

Exercise 8.18 (Formulas for $C\gamma$ and $C\kappa$)**.** *Show that $C_\gamma = 1$ does the job, and $C_\kappa = \sup_{k\geq 1}(k+\delta)p_k = (1+\delta)p_1 = \frac{(1+\delta)(2+\delta)}{3+2\delta}$.*

Using the induction hypothesis (8.5.16), as well as (8.5.26), we arrive at

$$|\varepsilon_1(t+1)| \leq C\Big(1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}\Big) + (C_\kappa + C_\gamma)(t+1)^{-1}. \tag{8.5.27}$$

Next, we use that $t(2+\delta)+(1+\delta) \leq (t+1)(2+\delta)$, so that

$$|\varepsilon_1(t+1)| \leq C - (t+1)^{-1}\Big(C\frac{1+\delta}{2+\delta} - (C_\kappa + C_\gamma)\Big) \leq C, \tag{8.5.28}$$

whenever

$$C \geq \frac{2+\delta}{1+\delta}(C_\kappa + C_\gamma). \tag{8.5.29}$$

This advances the induction hypothesis for $k = 1$.

We now extend the argument to $k \geq 2$. We again use (8.5.20). We note that

$$1 - \frac{k+\delta}{t(2+\delta)+(1+\delta)} \geq 0 \tag{8.5.30}$$

as long as

$$k \leq t(2+\delta) + 1. \tag{8.5.31}$$

We will assume (8.5.31) for the time being, and deal with $k \geq t(2+\delta)+2$ later.

By (8.5.20) and (8.5.31), we obtain that, for $k \geq 2$ and $\delta > -1$, so that $k - 1 + \delta \geq 0$,

$$|\varepsilon_k(t+1)| \leq \Big(1 - \frac{k+\delta}{t(2+\delta)+(1+\delta)}\Big)|\varepsilon_k(t)| + \frac{k-1+\delta}{t(2+\delta)+(1+\delta)}|\varepsilon_{k-1}(t)| + |\kappa_k(t)| + |\gamma_k(t)|. \tag{8.5.32}$$

Again using the induction hypothesis (8.5.16), as well as (8.5.26), we arrive at

$$\begin{aligned}|\varepsilon_k(t+1)| &\leq C\Big(1-\frac{k+\delta}{t(2+\delta)+(1+\delta)}\Big)+C\frac{k-1+\delta}{t(2+\delta)+(1+\delta)}+(C_\kappa+C_\gamma)(t+1)^{-1}\\ &= C\Big(1-\frac{1}{t(2+\delta)+(1+\delta)}\Big)+(C_\kappa+C_\gamma)(t+1)^{-1}. \end{aligned} \tag{8.5.33}$$

As before,

$$t(2+\delta)+(1+\delta)\leq (t+1)(2+\delta), \tag{8.5.34}$$

so that

$$|\varepsilon_k(t+1)|\leq C-(t+1)^{-1}\Big(\frac{C}{2+\delta}-(C_\kappa+C_\gamma)\Big)\leq C, \tag{8.5.35}$$

whenever

$$C\geq (2+\delta)(C_\kappa+C_\gamma). \tag{8.5.36}$$

Finally, we deal with the case that $k\geq t(2+\delta)+2$. Note that $k\geq t(2+\delta)+2>t+2$ when $\delta>-1$. Since the maximal degree of $\mathrm{PA}_t(1,\delta)$ is $t+2$ (which happens precisely when all edges are connected to the initial vertex), we have that $\bar{N}_k(t+1)=0$ for $k\geq t(2+\delta)+2$. Therefore, for $k\geq t(2+\delta)+2$,

$$|\varepsilon_k(t+1)|=(t+1)p_k. \tag{8.5.37}$$

By (8.3.9) and (8.3.10), uniformly for $k\geq t(2+\delta)+2\geq t+2$ for $\delta\geq -1$, there exists a $C_p=C_p(\delta)$ such that

$$p_k\leq C_p(t+1)^{-(3+\delta)}. \tag{8.5.38}$$

For $\delta>-1$, and again uniformly for $k\geq t+2$,

$$(t+1)p_k\leq C_p(t+1)^{-(2+\delta)}\leq C_p. \tag{8.5.39}$$

Therefore, if $C\geq C_p$, then also the claim follows for $k\geq t(2+\delta)+2$. Comparing to (8.5.29) and (8.5.29), we choose

$$C=\max\Big\{(2+\delta)(C_\kappa+C_\gamma),\frac{(2+\delta)(C_\kappa+C_\gamma)}{1+\delta},C_p\Big\}. \tag{8.5.40}$$

This advances the induction hypothesis for $k\geq 2$, and completes the proof of Proposition 8.4 when $m=1$ and $\delta>-1$. □

8.5.2 Expected degree sequence for $m>1^*$

In this section, we prove Proposition 8.4 for $m>1$. We adapt the argument in Section 8.5.1 above. In Section 8.5.1, we have been rather explicit in the derivation of the recursion relation in (8.5.9), which in turn gives the explicit recursion relation on the errors $\varepsilon_k(t)$ in (8.5.20). In this section, we make the derivation more abstract, since the explicit derivations become too involved when $m>1$. The current argument is rather flexible, and can, e.g., be extended to different preferential attachment models.

We make use of the fact that to go from $\mathrm{PA}_t(m,\delta)$ to $\mathrm{PA}_{t+1}(m,\delta)$, we add precisely m edges in a preferential way. This process can be described in terms of certain operators. For a sequence of numbers $Q=\{Q_k\}_{k=1}^\infty$, we define the operator $T_{t+1}\colon \mathbb{R}^\infty\mapsto\mathbb{R}^\infty$ by

$$(T_{t+1}Q)_k=\Big(1-\frac{k+\delta}{t(2+\delta')+(1+\delta')}\Big)Q_k+\frac{k-1+\delta}{t(2+\delta')+(1+\delta')}Q_{k-1}, \tag{8.5.41}$$

where we recall that $\delta' = \delta/m$. Then, writing $\bar{N}(t) = \{\bar{N}_k(t)\}_{k=1}^{\infty}$, we can rewrite (8.5.9) when $m = 1$ so that $\delta' = \delta$,

$$\bar{N}_k(t+1) = (T_{t+1}\bar{N}(t))_k + \mathbb{1}_{\{k=1\}}\Big(1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}\Big) + \mathbb{1}_{\{k=2\}}\frac{1+\delta}{t(2+\delta)+(1+\delta)}. \tag{8.5.42}$$

Thus, as remarked above (8.5.6), the operator T_{t+1} describes the effect to the degree sequence of a single addition of the $(t+1)^{\text{st}}$ edge, apart from the degree of the newly added vertex. The latter degree is equal to 1 with probability $1 - \frac{1+\delta}{t(2+\delta)+(1+\delta)}$, and equal to 2 with probability $\frac{1+\delta}{t(2+\delta)+(1+\delta)}$. This explains the origin of each of the terms appearing in (8.5.9).

In the case when $m > 1$, every vertex has m edges that are each connected in a preferential way. Therefore, we need to investigate the effect of attaching m edges in sequel. Due to the fact that we update the degrees after attaching an edge, the effect of attaching the $(j+1)^{\text{st}}$ edge is described by applying the operator T_j to $\bar{N}(j)$. When we add the edges incident to the t^{th} vertex, this corresponds to attaching the edges $m(t-1)+1, \ldots, mt$ in sequel with intermediate updating. The effect on the degrees of vertices $v_1, \ldots, v_t$ is described precisely by applying first T_{mt+1} to describe the effect of the addition of the first edge, followed by T_{mt+2} to describe the effect of the addition of the second edge, etc. Therefore, the recurrence relation of the expected number of vertices with degree k is changed to

$$\bar{N}_k(t+1) = (T_{t+1}^{(m)}\bar{N}(t))_k + \alpha_k(t), \tag{8.5.43}$$

where

$$T_{t+1}^{(m)} = T_{m(t+1)} \circ \cdots \circ T_{mt+1}, \tag{8.5.44}$$

and where, for $k = m, \ldots, 2m$, we have that $\alpha_k(t)$ is equal to the probability that the degree of the $(t+1)^{\text{st}}$ added vertex is precisely equal to n. Indeed, when t changes to $t+1$, then the graph grows by one vertex. Its degree is equal to n with probability $q_n(t)$, so that the contribution of this vertex is equal to $\alpha_k(t)$. On the other hand, the edges that are connected from the $(t+1)^{\text{st}}$ vertex also change the degrees of the other vertices. The expected number of vertices with degree k among vertices $v_1, \ldots, v_t$ is precisely given by $(T_{t+1}^{(m)}\bar{N}(t))_k$. Thus, the operator $T_{t+1}^{(m)}$ describes the effect to the degrees of vertices $v_1, \ldots, v_t$ of the attachment of the edges emanating from vertex v_{t+1}. This explains (8.5.43).

When t grows large, then the probability distribution $k \mapsto \alpha_k(t)$ is such that $\alpha_m(t)$ is very close to 1, while $\alpha_k(t)$ is close to zero when $k > m$. Indeed, for $k > m$, at least one of the m edges should be connected to its brother half-edge, so that

$$\sum_{k=m+1}^{2m} \alpha_k(t) \leq \frac{m^2(1+\delta)}{mt(2+\delta')+(1+\delta')}. \tag{8.5.45}$$

We define

$$\gamma_k(t) = \alpha_k(t) - \mathbb{1}_{\{k=m\}}, \tag{8.5.46}$$

then we obtain from (8.5.45) that there exists a constant $C_\gamma = C_\gamma(\delta, m)$ such that

$$|\gamma_k(t)| \leq C_\gamma(t+1)^{-1}. \tag{8.5.47}$$

The bound in (8.5.47) replaces the bound on $|\gamma_k(t)|$ for $m = 1$ in (8.5.26).

Denote the operator $S^{(m)}$ on sequences of numbers $Q = \{Q_k\}_{k=1}^{\infty}$ by

$$(S^{(m)}Q)_k = m\frac{k-1+\delta}{2m+\delta}Q_{k-1} - m\frac{k+\delta}{2m+\delta}Q_k. \tag{8.5.48}$$

Then, for $m = 1$, we have that (8.5.10) is equivalent to

$$p_k = (S^{(1)}p)_k + \mathbb{1}_{\{k=1\}}. \tag{8.5.49}$$

For $m > 1$, we replace the above recursion on p by $p_k = 0$ for $k < m$ and, for $k \geq m$,

$$p_k = (S^{(m)}p)_k + \mathbb{1}_{\{k=m\}}. \tag{8.5.50}$$

Again, we can explicitly solve for $p = \{p_k\}_{k=1}^{\infty}$. The solution is given in the following lemma:

Lemma 8.5 (Solution recursion for $m > 1$)**.** *Fix $\delta > -1$ and $m \geq 1$. Then, the solution to* (8.5.50) *is given by* (8.3.2)*.*

Proof. We start by noting that $p_k = 0$ for $k < m$, and identify p_m as

$$p_m = -m\frac{m+\delta}{2m+\delta}p_m + 1, \tag{8.5.51}$$

so that

$$p_m = \frac{2m+\delta}{m(m+\delta)+2m+\delta} = \frac{2+\frac{\delta}{m}}{(m+\delta)+2+\frac{\delta}{m}}. \tag{8.5.52}$$

For $k > m$, the recursion relation in (8.5.50) becomes

$$p_k = \frac{m(k-1+\delta)}{m(k+\delta)+2m+\delta}p_{k-1} = \frac{k-1+\delta}{k+\delta+2+\frac{\delta}{m}}p_{k-1}. \tag{8.5.53}$$

As a result, we obtain that, again repeatedly using (8.2.2),

$$\begin{aligned} p_k &= \frac{\Gamma(k+\delta)\Gamma(m+3+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)\Gamma(k+3+\delta+\frac{\delta}{m})}p_m \\ &= \frac{\Gamma(k+\delta)\Gamma(m+3+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)\Gamma(k+3+\delta+\frac{\delta}{m})}\frac{(2+\frac{\delta}{m})}{(m+\delta+2+\frac{\delta}{m})} \\ &= \frac{(2+\frac{\delta}{m})\Gamma(k+\delta)\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)\Gamma(k+3+\delta+\frac{\delta}{m})}. \end{aligned} \tag{8.5.54}$$

□

Similarly to (8.5.17), we can rewrite (8.5.50) as

$$\begin{aligned} (t+1)p_k &= tp_k + p_k = tp_k + (S^{(m)}p)_k + \mathbb{1}_{\{k=m\}} \\ &= (T_{t+1}^{(m)}tp)_k + \mathbb{1}_{\{k=m\}} - \kappa_k(t), \end{aligned} \tag{8.5.55}$$

where, writing I for the identity operator,

$$\kappa_k(t) = -\big(\big[S^{(m)} + t(I - T_{t+1}^{(m)})\big]p\big)_k. \tag{8.5.56}$$

While (8.5.56) is not very explicit, a similar argument as the ones leading to (8.5.26) can be used to deduce an identical bound. That is the content of the following lemma:

Lemma 8.6 (A bound on $\kappa_k(t)$)**.** *Fix $\delta \geq -1$ and $m \geq 1$. Then there exists a constant $C_\kappa = C_\kappa(\delta, m)$ such that*

$$|\kappa_k(t)| \leq C_\kappa(t+1)^{-1}. \tag{8.5.57}$$

We defer the proof of Lemma 8.6 to the end of this section, and continue with the proof of Proposition 8.4 for $m > 1$.

We define, for $k \geq m$,

$$\varepsilon_k(t) = \bar{N}_k(t) - tp_k. \tag{8.5.58}$$

Subtracting (8.5.55) from (8.5.43) and writing $\varepsilon(t) = \{\varepsilon_k(t)\}_{k=1}^{\infty}$ leads to

$$\varepsilon_k(t+1) = (T_{t+1}^{(m)}\varepsilon(t))_k + \kappa_k(t) + \gamma_k(t). \tag{8.5.59}$$

In order to study the recurrence relation (8.5.59) in more detail, we investigate the properties of the operator $T_t^{(m)}$. To state the result, we introduce some notation. We let $Q = \{Q_k\}_{k=1}^{\infty}$ be a sequence of real numbers, and we let $\mathcal{Q} = \mathbb{R}^{\infty}$ denote the set of all such sequences. For $Q \in \mathcal{Q}$, we define the supremum-norm to be

$$\|Q\|_{\infty} = \sup_{k=1}^{\infty} |Q_k|. \tag{8.5.60}$$

Thus, in functional analytic terms, we consider the ℓ^{∞} norm on $\mathcal{Q} = \mathbb{R}^{\infty}$.

Furthermore, we let $\mathcal{Q}_m(t) \subseteq \mathcal{Q}$ be the subset of sequences for which $Q_k = 0$ for $k > m(t+1)$, i.e.,

$$\mathcal{Q}_m(t) = \{Q \in \mathcal{Q} : Q_k = 0 \ \ \forall k > m(t+1)\}. \tag{8.5.61}$$

Clearly, $\bar{N}(t) \in \mathcal{Q}_m(t)$.

We regard $T_{t+1}^{(m)}$ in (8.5.44) as an operator on $\mathcal{Q}$. We now derive its functional analytic properties:

Lemma 8.7 (A contraction property)**.** *Fix $\delta \geq -1$ and $m \geq 1$. Then $T_{t+1}^{(m)}$ maps $\mathcal{Q}_m(t)$ into $\mathcal{Q}_m(t+1)$ and, for every $Q \in \mathcal{Q}_m(t)$,*

$$\|T_{t+1}^{(m)}Q\|_{\infty} \leq \Big(1 - \frac{1}{t(2m+\delta) + (m+\delta)}\Big)\|Q\|_{\infty}. \tag{8.5.62}$$

Lemma 8.7 implies that $T_{t+1}^{(m)}$ acts as a *contraction* on elements of $\mathcal{Q}_m(t)$. Using Lemmas 8.6 and 8.7, as well as (8.5.47) allows us to complete the proof of Proposition 8.4:

Proof of Proposition 8.4. We use (8.5.59). We define the sequence $\varepsilon'(t) = \{\varepsilon'_k(t)\}_{k=1}^{\infty}$ by

$$\varepsilon'_k(t) = \varepsilon_k(t)\mathbb{1}_{\{k \leq m(t+1)\}}. \tag{8.5.63}$$

Then, by construction, $\varepsilon'(t) \in \mathcal{Q}_m(t)$. Therefore, by Lemma 8.7,

$$\begin{aligned}\|\varepsilon(t+1)\|_{\infty} &\leq \|T_{t+1}^{(m)}\varepsilon'(t)\|_{\infty} + \|\varepsilon'(t+1) - \varepsilon(t+1)\|_{\infty} + \|\kappa(t)\|_{\infty} + \|\gamma(t)\|_{\infty} \\ &\leq \Big(1 - \frac{1}{(2m+\delta) + (m+\delta)}\Big)\|\varepsilon'(t)\|_{\infty} \\ &\quad + \|\varepsilon'(t+1) - \varepsilon(t+1)\|_{\infty} + \|\kappa(t)\|_{\infty} + \|\gamma(t)\|_{\infty}. \end{aligned} \tag{8.5.64}$$

Equation (8.5.47) is equivalent to the statement that

$$\|\gamma(t)\|_{\infty} \leq \frac{C_{\gamma}}{t+1}. \tag{8.5.65}$$

Lemma 8.6 implies that

$$\|\kappa(t)\|_{\infty} \leq \frac{C_{\kappa}}{t+1}. \tag{8.5.66}$$

It is not hard to see that

$$\|\varepsilon'(t+1) - \varepsilon(t+1)\|_\infty \leq C_{\varepsilon'}(t+1)^{-(\tau-1)}, \tag{8.5.67}$$

where $\tau > 2$ is defined in (8.3.10). See (8.5.38)–(8.5.39) for the analogous proof for $m = 1$.

Therefore,

$$\|\varepsilon(t+1)\|_\infty \leq \Big(1 - \frac{1}{t(2m+\delta) + (m+\delta)}\Big)\|\varepsilon(t)\|_\infty + \frac{(C_\gamma + C_\kappa + C_{\varepsilon'})}{t+2}. \tag{8.5.68}$$

Using further that, for $m \geq 1$ and $\delta > -m$,

$$t(2m+\delta) + (m+\delta) \leq (2m+\delta)(t+1) \tag{8.5.69}$$

we arrive at

$$\|\varepsilon(t+1)\|_\infty \leq \Big(1 - \frac{1}{(t+1)(2m+\delta)}\Big)\|\varepsilon(t)\|_\infty + \frac{(C_\gamma + C_\kappa + C_{\varepsilon'})}{t+1}. \tag{8.5.70}$$

Now we can advance the induction hypothesis

$$\|\varepsilon(t)\|_\infty \leq C. \tag{8.5.71}$$

For some $C > 0$ sufficiently large, this statement trivially holds for $t = 1$. To advance it, we use (8.5.70), to see that

$$\|\varepsilon(t+1)\|_\infty \leq \Big(1 - \frac{1}{(2m+\delta)(t+1)}\Big)C + \frac{(C_\gamma + C_\kappa + C_{\varepsilon'})}{t+1} \leq C, \tag{8.5.72}$$

whenever

$$C \geq (2m+\delta)(C_\gamma + C_\kappa + C_{\varepsilon'}). \tag{8.5.73}$$

This advances the induction hypothesis, and completes the proof that $|\bar{N}_k(t) - p_k t| \leq C$ for $m \geq 2$. □

Proof of Lemmas 8.6 and 8.7. We first prove Lemma 8.7, and then Lemma 8.6.

Proof of Lemma 8.7. We recall that

$$T_{t+1}^{(m)} = T_{m(t+1)} \circ \cdots \circ T_{mt+1}, \tag{8.5.74}$$

Thus, the fact that $T_{t+1}^{(m)}$ maps $\mathcal{Q}_m(t)$ into $\mathcal{Q}_m(t+1)$ follows from the fact that T_{t+1} maps $\mathcal{Q}_1(t)$ into $\mathcal{Q}_1(t+1)$. This proves the first claim in Lemma 8.7.

To prove that the contraction property of $T_{t+1}^{(m)}$ in (8.5.62), we shall first prove that, for all $Q \in \mathcal{Q}_1(mt+a-1)$, $a = 1, \ldots, m$, $\delta > -m$ and $\delta' = \delta/m > -1$, we have

$$\|(T_{mt+a}Q)\|_\infty \leq \Big(1 - \frac{1}{t(2+\delta) + (1+\delta)}\Big)\|Q\|_\infty. \tag{8.5.75}$$

For this, we recall from (8.5.41) that

$$(T_{mt+a}Q)_k = \Big(1 - \frac{k+\delta}{(mt+a)(2+\delta') + (1+\delta')}\Big)Q_k + \frac{k-1+\delta}{(mt+a)(2+\delta') + (1+\delta')}Q_{k-1}. \tag{8.5.76}$$

When $Q \in \mathcal{Q}_1(mt+a)$, then, for all k for which $Q_k \neq 0$,

$$1 - \frac{k+\delta}{(mt+a-1)(2+\delta')+(1+\delta')} \in [0,1], \tag{8.5.77}$$

and, for $k \geq 2$, also

$$\frac{k-1+\delta}{(mt+a-1)(2+\delta')+(1+\delta')} \in [0,1]. \tag{8.5.78}$$

As a consequence, we have that

$$\begin{aligned}\|T_{mt+a}Q\|_\infty &\leq \sup_k \Big[\Big(1 - \frac{k+\delta}{(mt+a-1)(2+\delta')+(1+\delta')}\Big)\|Q\|_\infty \\ &\qquad + \frac{k-1+\delta}{(mt+a-1)(2+\delta')+(1+\delta')}\|Q\|_\infty\Big] \\ &= \Big(1 - \frac{1}{(mt+a-1)(2+\delta')+(1+\delta')}\Big)\|Q\|_\infty. \end{aligned}\tag{8.5.79}$$

Now, by (8.5.79), the application of T_{mt+a} to an element Q of $\mathcal{Q}_1(mt+a-1)$ reduces its norm. By (8.5.74), we therefore conclude that, for every $Q \in \mathcal{Q}_m(t)$,

$$\begin{aligned}\|T^{(m)}_{t+1}Q\|_\infty \leq \|T_{mt+1}Q\|_\infty &\leq \Big(1 - \frac{1}{mt(2+\delta')+(1+\delta')}\Big)\|Q\|_\infty \\ &= \Big(1 - \frac{1}{t(2m+\delta)+(m+\delta)}\Big)\|Q\|_\infty, \end{aligned}\tag{8.5.80}$$

since $\delta' = \delta/m$. This completes the proof of Lemma 8.7. □

Proof of Lemma 8.6. We recall

$$\kappa_k(t) = \Big(\big[S^{(m)} + t(I - T^{(m)}_{t+1})\big]p\Big)_k. \tag{8.5.81}$$

We start with

$$T^{(m)}_{t+1} = T_{m(t+1)} \circ \cdots \circ T_{mt+1} = \big(I + (T_{m(t+1)} - I)\big) \circ \cdots \circ \big(I + (T_{mt+1} - I)\big). \tag{8.5.82}$$

Clearly,

$$\big((T_{t+1} - I)Q\big)_k = -\frac{k+\delta}{t(2+\delta')+(1+\delta')}Q_k + \frac{k-1+\delta}{t(2+\delta')+(1+\delta')}Q_{k-1}. \tag{8.5.83}$$

When $\sup_k k|Q_k| \leq K$, then there exists a constant C such that

$$\sup_k \Big|\big((T_{t+1} - I)Q\big)_k\Big| \leq \frac{C}{t+1}. \tag{8.5.84}$$

Moreover, when $\sup_k k^2|Q_k| \leq K$, then there exists a constant $C = C_K$ such that, when $u, v \geq t$,

$$\sup_k \big|((T_{u+1} - I) \circ (T_{v+1} - I)Q)_k\big| \leq \frac{C}{(t+1)^2}. \tag{8.5.85}$$

We expand out the brackets in (8.5.82), and note that, by (8.5.85) and the fact that the operators T_u are contractions that the terms where we have at least two factors $T_u - I$ lead to error terms. More precisely, we conclude that, when $\sup_k k^2|Q_k| \leq K$,

$$(T^{(m)}_{t+1}Q)_k = Q_k + \sum_{a=1}^{m} \big((T_{mt+a} - I)Q\big)_k + E_k(t,Q), \tag{8.5.86}$$

where, uniformly in k and Q for which $\sup_k k^2|Q_k| \leq K$,

$$|E_k(t,Q)| \leq \frac{C_K}{(t+1)^2}. \tag{8.5.87}$$

As a result, we obtain that

$$\big((I - T_{t+1}^{(m)})Q\big)_k = -\sum_{a=1}^{m} \big((T_{mt+a} - I)Q\big)_k - E_k(t,Q). \tag{8.5.88}$$

Furthermore, for every $a = 1, \ldots, m$,

$$\big((T_{mt+a} - I)Q\big)_k = \frac{1}{mt}(S^{(m)}Q)_k + F_{k,a}(t,Q), \tag{8.5.89}$$

where, uniformly in k, Q for which $\sup_k k|Q_k| \leq K$ and $a = 1, \ldots, m$,

$$|F_{k,a}(t,Q)| \leq \frac{C'_K}{(t+1)^2}. \tag{8.5.90}$$

Therefore, we also obtain that

$$\sum_{a=1}^{m} \big((T_{mt+a} - I)Q\big)_k = \frac{1}{t}(S^{(m)}Q)_k + F_k(t,Q), \tag{8.5.91}$$

where

$$F_k(t,Q) = \sum_{a=1}^{m} F_{k,a}(t,Q). \tag{8.5.92}$$

We summarize from (8.5.88) and (8.5.91) that

$$\big([S^{(m)} + t(I - T_{t+1}^{(m)})]Q\big)_k = -tF_k(t,Q) - tE_k(t,Q), \tag{8.5.93}$$

so that

$$\kappa_k(t) = \big([S^{(m)} + t(I - T_{t+1}^{(m)})]p\big)_k = -tF_k(t,p) - tE_k(t,p). \tag{8.5.94}$$

We note that by (8.3.9) and (8.3.10), p satisfies that

$$\sup_k k^2 p_k \leq C_p, \tag{8.5.95}$$

so that we conclude that

$$\begin{aligned} \|\kappa(t)\|_\infty &= \sup_k \Big|\big([S^{(m)} + t(I - T_{t+1}^{(m)})p]\big)_k\Big| \leq \sup_k t\big(|E_k(t,p)| + |F_k(t,p)|\big) \\ &\leq \frac{t(C_K + C'_K)}{(t+1)^2} \leq \frac{C_K + C'_K}{t+1}. \end{aligned} \tag{8.5.96}$$

□

8.5.3 Degree sequence: completion proof of Theorem 8.2

We only prove the result for $m = 1$, the proof for $m > 1$ being identical. By Proposition 8.4, we obtain that

$$\max_k |\mathbb{E}[N_k(t)] - p_k t| \leq C. \tag{8.5.97}$$

Therefore, by Proposition 8.3 we obtain

$$\mathbb{P}\Big(\max_k |N_k(t) - p_k t| \geq C(1 + \sqrt{t \log t})\Big) = o(1), \tag{8.5.98}$$

which, since $P_k(t) = N_k(t)/t$, implies that

$$\mathbb{P}\Big(\max_k |P_k(t) - p_k| \geq \frac{C}{t}(1 + \sqrt{t \log t})\Big) = o(1). \tag{8.5.99}$$

Equation (8.5.99) in turn implies Theorem 8.2. □

8.6 Maximal degree in preferential attachment models

In this section, we shall investigate the maximal degree and the clustering of the graph $\mathrm{PA}_{m,\delta}(t)$. In order to state the results on the maximal degree, we denote

$$M_t = \max_{i=1}^{t} D_i(t). \tag{8.6.1}$$

The main result on the maximal degree is the following theorem:

Theorem 8.8 (Maximal degree of $\mathrm{PA}_{m,\delta}(t)$)**.** *Fix $m \geq 1$ and $\delta > -m$. Then,*

$$M_t t^{-\frac{1}{\tau-1}} \xrightarrow{a.s.} \mu, \tag{8.6.2}$$

with $\mathbb{P}(\mu = 0) = 0$.

Below, we shall be able to compute all moments of the limiting random variables ξ_i of $D_i(t)t^{-1/(2+\delta)}$. We do not recognize these moments as the moments of a continuous random variable.

Exercise 8.19 ([85])**.** *Fix $m = 1$ and $\delta > -1$. Then, prove that for all $t \geq i$*

$$\mathbb{P}(D_i(t) = j) \leq C_j \frac{\Gamma(t)\Gamma(i + \frac{1+\delta}{2+\delta})}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(i)}, \tag{8.6.3}$$

where $C_1 = 1$ and

$$C_j = \frac{j - 1 + \delta}{j - 1} C_{j-1}. \tag{8.6.4}$$

Mori [172] studied various martingales related to degrees, and used them to prove that the maximal degree of $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ converges a.s. We shall reproduce his argument here, applied to a slightly different model. See also [88, Section 4.3]. We fix $m = 1$ for the time being, and extend the results to $m \geq 2$ at the end of this section.

In [172], the graph at time 1 consists of two vertices, 0 and 1, connected by a single edge. In the attachment scheme, no self-loops are created, so that the resulting graph is a tree. The proof generalizes easily to other initial configurations and attachment rules, and we shall adapt the argument here to the usual preferential attachment model in which self-loops *do* occur and $\mathrm{PA}_1(1, \delta)$ consists of one vertex with a single self-loop. At the t^{th}

step, a new vertex is added and connected to an existing vertex. A vertex of degree k is chosen with probability $(k+\delta)/n(t)$ where $\delta > -1$ and $n(t) = t(2+\delta) + 1 + \delta$ is the sum of the weights for the random graph with t edges and t vertices.

Let $X_j(t) = D_j(t) + \delta$ be the weight of vertex j at time t, let $\Delta_j(t+1) = X_j(t+1) - X_j(t)$. If $j \le t$, then

$$\mathbb{P}\left(\Delta_j(t+1) = 1 | \mathrm{PA}_t(1,\delta)\right) = X_j(t)/n(t). \tag{8.6.5}$$

From this, we get

$$\mathbb{E}\left(X_j(t+1) | \mathrm{PA}_t(1,\delta)\right) = X_j(t)\left(1 + \frac{1}{n(t)}\right) \tag{8.6.6}$$

so $c_t X_j(t)$ will be a martingale if and only if $c_t/c_{t+1} = n(t)/(1+n(t))$. Anticipating the definition of a larger collection of martingales we let

$$c_k(t) = \frac{\Gamma(t + \frac{1+\delta}{2+\delta})}{\Gamma(t + \frac{k+1+\delta}{2+\delta})}, \qquad t \ge 1, k \ge 0, \tag{8.6.7}$$

For fixed $k \ge 0$, by (8.2.8),

$$c_k(t) = t^{-k/(2+\delta)}(1 + o(1)) \quad \text{as } t \to \infty \tag{8.6.8}$$

Using the recursion $\Gamma(r) = (r-1)\Gamma(r-1)$ we have

$$\frac{c_k(t+1)}{c_k(t)} = \frac{t + \frac{1+\delta}{2+\delta}}{t + \frac{k+1+\delta}{2+\delta}} = \frac{n(t)}{n(t)+k}. \tag{8.6.9}$$

In particular, it follows that $c_1(t)X_j(t)$ is a martingale for $t \ge j$. Being a positive martingale it will converge a.s. to a random variable ξ_j, as discussed in full detail in Theorem 8.1. To study the joint distribution of the $X_j(t)$ we make use of a whole class of martingales. We first introduce some notation. For $a, b > -1$ with $a - b > -1$, where a, b are not necessarily integers, we write

$$\binom{a}{b} = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}. \tag{8.6.10}$$

The restriction on a, b is such that the arguments of the Gamma-function are all strictly positive. Then the following proposition identifies a whole class of useful martingales related to the degrees of the vertices:

Proposition 8.9 (A rich class of degree martingales)**.** *Let $r \ge 0$ be a non-negative integer, $k_1, k_2, \ldots, k_r > -\max\{1, 1+\delta\}$, and $0 \le j_1 < \ldots < j_r$ be non-negative integers. Then, with $k = \sum_i k_i$,*

$$Z_{\vec{j},\vec{k}}(t) = c_k(t) \prod_{i=1}^{r} \binom{X_{j_i}(t) + k_i - 1}{k_i} \tag{8.6.11}$$

is a martingale for $t \ge \max\{j_r, 1\}$.

The restriction $k_i > -\max\{1, 1+\delta\}$ is to satisfy the restrictions $a, b, a-b > -1$ in (8.6.10), since $X_j(t) \ge 1 + \delta$. Since $\delta > -1$, this means that Proposition 8.9 also holds for certain $k_i < 0$.

Exercise 8.20 (Martingale mean)**.** *Use Proposition 8.9 to show that, for all $t \ge \max\{j_r, 1\}$,*

$$\mathbb{E}[Z_{\vec{j},\vec{k}}(t)] = \prod_{i=1}^{r} \frac{c_{K_i}(j_i)}{c_{K_{i-1}}(j_i)} \binom{k_i + \delta}{k_i}, \tag{8.6.12}$$

where $K_i = \sum_{a=1}^{i} k_a$.

Proof. By considering the two cases $\Delta_j(t) = 0$ or $\Delta_j(t) = 1$, and using (8.6.10) and $\Gamma(r) = (r-1)\Gamma(r-1)$, it is easy to check that, for all k,

$$\begin{aligned}\binom{X_j(t+1)+k-1}{k} &= \binom{X_j(t)+k-1}{k}\frac{\Gamma(X_j(t+1)+k)}{\Gamma(X_j(t)+k)}\\ &= \binom{X_j(t)+k-1}{k}\left(1+\frac{k\Delta_j(t)}{X_j(t)}\right). \end{aligned} \tag{8.6.13}$$

At most one $X_j(t)$ can change, so that

$$\prod_{i=1}^{r}\left(1+\frac{k_i\Delta_{j_i}(t)}{X_{j_i}(t)}\right) = \left(1+\sum_{i=1}^{r}\frac{k_i\Delta_{j_i}(t)}{X_{j_i}(t)}\right). \tag{8.6.14}$$

Together, (8.6.13) and (8.6.14) imply that

$$\prod_{i=1}^{r}\binom{X_{j_i}(t+1)+k_i-1}{k_i} = \left(1+\sum_{i=1}^{r}\frac{k_i\Delta_{j_i}(t)}{X_{j_i}(t)}\right)\prod_{i=1}^{r}\binom{X_{j_i}(t)+k_i-1}{k_i}. \tag{8.6.15}$$

Since $\mathbb{P}\big(\Delta_j(t+1)=1|\mathrm{PA}_t(1,\delta)\big) = X_j(t)/n(t)$, using the definition of $Z_{\vec{j},\vec{k}}(t)$ and taking expected value,

$$\mathbb{E}\left(Z_{\vec{j},\vec{k}}(t+1)|\mathrm{PA}_t(1,\delta)\right) = Z_{\vec{j},\vec{k}}(t)\cdot\frac{c_k(t+1)}{c_k(t)}\left(1+\frac{\sum_i k_i}{n(t)}\right) = Z_{\vec{j},\vec{k}}(t), \tag{8.6.16}$$

where $k = \sum_i k_i$ and the last equality follows from (8.6.9). □

Being a non-negative martingale, $Z_{\vec{j},\vec{k}}(t)$ converges. From the form of the martingale, the convergence result for the factors, and the asymptotics for the normalizing constants in (8.6.8), the limit must be $\prod_{i=1}^{r}\xi_i^{k_i}/\Gamma(k_i+1)$, where we recall that ξ_i is the almost sure limit of $D_i(t)t^{-1/(2+\delta)}$. Here we make use of (8.2.8), which implies that

$$\binom{X_j(t)+k-1}{k} = X_j(t)^k(1+O(1/X_j(t))), \tag{8.6.17}$$

together with the fact that $D_i(t) \xrightarrow{a.s.} \infty$ (see Exercise 8.8).

Our next step is to check that the martingale converges in L^1. To do this we begin by observing that (8.6.8) implies $c_m(t)^2/c_{2m}(t) \to 1$ and we have

$$\binom{x+k-1}{k}^2 = \left(\frac{\Gamma(x+k)}{\Gamma(x)\Gamma(k+1)}\right)^2 = \frac{\Gamma(x+k)}{\Gamma(x)}\frac{\Gamma(x+k)}{\Gamma(x)\Gamma(k+1)^2}. \tag{8.6.18}$$

Now we use that $x \mapsto \Gamma(x+k)/\Gamma(x)$ is increasing for $k \geq 0$, so that

$$\binom{x+k-1}{k}^2 \leq \frac{\Gamma(x+2k)}{\Gamma(x+k)}\frac{\Gamma(x+k)}{\Gamma(x)\Gamma(k+1)^2} = \binom{x+2k-1}{2k}\cdot\binom{2k}{k}. \tag{8.6.19}$$

From this it follows that

$$Z_{\vec{j},\vec{k}}(t)^2 \leq C_{\vec{k}}Z_{\vec{j},2\vec{k}}(t), \tag{8.6.20}$$

where

$$C_{\vec{k}} = \prod_{i=1}^{r} \binom{2k_i}{k_i}. \tag{8.6.21}$$

Therefore, $Z_{\vec{j},\vec{k}}(t)$ is an L^2-bounded martingale, and hence converge in L^1.

Taking $r = 1$ we have, for all $j \geq 1$ integer and $k \in \mathbb{R}$ with $k \geq 0$,

$$\mathbb{E}[\xi_j^k/\Gamma(k+1)] = \lim_{t\to\infty} \mathbb{E}[Z_{j,k}(t)] = \mathbb{E}[Z_{j,k}(j)] = c_k(j)\binom{k+\delta}{k}. \tag{8.6.22}$$

Recalling that $c_k(j) = \frac{\Gamma(j+\frac{1+\delta}{2+\delta})}{\Gamma(j+\frac{k+1+\delta}{2+\delta})}$, we thus arrive at the fact that, for all j non-negative integers, and all k non-negative,

$$\mathbb{E}[\xi_j^k] = \frac{\Gamma(j+\frac{1+\delta}{2+\delta})}{\Gamma(j+\frac{k+1+\delta}{2+\delta})}\frac{\Gamma(k+1+\delta)}{\Gamma(1+\delta)}. \tag{8.6.23}$$

It is, as far as we know, unknown which random variable has these moments, but we *can* see that the above moments identify the distribution:

Exercise 8.21 (Uniqueness of limit)**.** *Prove that the moments in* (8.6.23) *identify the distribution of ξ_j uniquely. Prove also that $\mathbb{P}(\xi_j > x) > 0$ for every $x > 0$, so that ξ_j has unbounded support.*

Exercise 8.22 (A.s. limit of $D_j(t)$ in terms of limit $D_1(t)$)**.** *Show that ξ_j has the same distribution as*

$$\xi_1 \prod_{k=1}^{j} B_k, \tag{8.6.24}$$

where B_k has a Beta$(1,(2+\delta)k-1)$-distribution.

Exercise 8.23 (Martingales for alternative construction PA model [172])**.** *Prove that when the graph at time 0 is given by two vertices with a single edge between them, and we do not allow for self-loops, then* (8.6.22) *remains valid when we instead define*

$$c_k(t) = \frac{\Gamma(t+\frac{\delta}{2+\delta})}{\Gamma(t+\frac{k+\delta}{2+\delta})} \quad t \geq 1, k \geq 0. \tag{8.6.25}$$

We complete this discussion by showing that $\mathbb{P}(\xi_j = 0) = 0$ for all $j \geq 1$. For this, we use (8.2.8), which implies that, for $k > -\max\{1, 1+\delta\}$,

$$\limsup_{t\to\infty} \mathbb{E}\Big[\Big(\frac{X_j(t)}{t^{1/(2+\delta)}}\Big)^k\Big] \leq A_k \limsup_{t\to\infty} \mathbb{E}[Z_{j,k}(t)] < \infty. \tag{8.6.26}$$

Since $\delta > -1$, we have $-1-\delta < 0$, so that the a negative moment of $X_j(t)/t^{1/(2+\delta)}$ remains uniformly bounded. This implies that $\mathbb{P}(\xi_j = 0) = 0$. Indeed, we use that $X_j(t)/t^{1/(2+\delta)} \xrightarrow{a.s.} \xi_j$, which implies that $X_j(t)/t^{1/(2+\delta)} \xrightarrow{d} \xi_j$, so that, using the Markov property (Theorem 2.14), for every $\varepsilon > 0$ and $k \in (-\max\{1, 1+\delta\}, 0)$,

$$\mathbb{P}(\xi_j \leq \varepsilon) = \limsup_{t\to\infty} \mathbb{P}\Big(X_j(t)/t^{1/(2+\delta)} \leq \varepsilon\Big) \leq \limsup_{t\to\infty} \varepsilon^{-k}\mathbb{E}\Big[\Big(\frac{X_j(t)}{t^{1/(2+\delta)}}\Big)^k\Big] = O(\varepsilon^{-k}). \tag{8.6.27}$$

Letting $\varepsilon \downarrow 0$, we obtain that $\mathbb{P}(\xi_j = 0) = 0$.

We next move on to study the maximal degree M_t. Let M_t denote the maximal degree in our random graph after t steps, and, for $t \geq j$, let

$$M_j(t) = \max_{0 \leq i \leq j} Z_{i,1}(t). \tag{8.6.28}$$

Note that $M_t(t) = c_1(t)(M_t + \delta)$. We shall now prove that $M_t(t) \xrightarrow{a.s.} \sup_{j=1}^{\infty} \xi_j$:

Proof of Theorem 8.8 for $m = 1$. We start by proving Theorem 8.8 for $m = 1$. Being a maximum of martingales, $\{M_t(t)\}_{t=1}^{\infty}$ is a non-negative submartingale. Therefore, $M_t(t) \xrightarrow{a.s.} \mu$ for some limiting random variable μ, and we are left to prove that $\mu = \sup_{j \geq 0} \xi_j$.

Since $Z_{j,1}(t)^k$ is a submartingale for every $k \geq 1$, and $Z_{j,1}(t)^k$ converges in L^1 to ξ_j^k, we further have that

$$\mathbb{E}[Z_{j,1}(t)^k] \leq \mathbb{E}[\xi_j^k]. \tag{8.6.29}$$

Then, using the trivial inequality

$$M_t(t)^k = \max_{0 \leq i \leq t} Z_{i,1}(t)^k \leq \sum_{j=0}^{t} Z_{j,1}(t)^k, \tag{8.6.30}$$

and (8.6.29), we obtain

$$\mathbb{E}[M_t(t)^k] \leq \sum_{j=0}^{t} \mathbb{E}[Z_{j,1}(t)^k] \leq \sum_{j=0}^{\infty} \mathbb{E}[\xi_j^k] = \Gamma(k+1) \binom{k+\delta}{k} \sum_{j=0}^{\infty} c_k(j), \tag{8.6.31}$$

which is finite by (8.6.8) if $k > 2 + \delta$. Thus $M_t(t)$ is bounded in L^k for every integer $k > 2 + \delta$, and hence bounded and convergent in L^p for any $p \geq 1$. Therefore, to prove that $\mu = \sup_{j \geq 0} \xi_j$, we are left to prove that $M_t(t)$ converges to $\sup_{j \geq 0} \xi_j$ in L^k for some k.

Let $k > 2 + \delta$ be fixed. Then, by a similar inequality as in (8.6.30),

$$\mathbb{E}\big[(M_t(t) - M_j(t))^k\big] \leq \sum_{i=j+1}^{t} \mathbb{E}[Z_{i,1}(t)^k] \tag{8.6.32}$$

Since $M_j(t)$ is a *finite* maximum of martingales, it is again a non-negative submartingale which each converge almost surely and in L^k for any $k > 2 + \delta$, its almost sure limit is equal to $\max_{0 \leq i \leq j} \xi_i = \mu_j$, Therefore, the limit of the left-hand side of (8.6.32) is

$$\mathbb{E}\Big[\Big(\lim_{t \to \infty} t^{-1/(2+\delta)} M_t - \mu_j \Big)^k \Big] \tag{8.6.33}$$

while the right-hand side of (8.6.32) increases to (compare to (8.6.29))

$$\sum_{i=j+1}^{\infty} \mathbb{E}[\xi_i^k] = k! \binom{k+\delta}{k} \sum_{i=j+1}^{\infty} c_k(i), \tag{8.6.34}$$

which is small if j is large by (8.6.8). Recall that $t^{-1/(2+\delta)} M_t \xrightarrow{a.s.} \mu$. Therefore, we obtain that

$$\lim_{j \to \infty} \mathbb{E}\Big[(\mu - \mu_j)^k \Big] = 0. \tag{8.6.35}$$

Hence $\lim_{t\to\infty} t^{-1/(2+\delta)} M_t = \mu$ as claimed.

When $m \geq 2$, then the above can be used as well. Indeed, in this case, we have that by Exercise 8.12, $D_i(t)(mt)^{-1/(2+\delta/m)} \xrightarrow{a.s.} \xi_i'$, where

$$\xi_i' = \sum_{j=(i-1)m+1}^{mi} \xi_j, \tag{8.6.36}$$

and ξ_j is the almost sure limit of $D_j(t)$ in $\{\mathrm{PA}_{1,\delta/m}(t)\}_{t=1}^{\infty}$. This implies that $M_t \xrightarrow{a.s.} \mu = \sup_{j=1}^{\infty} \xi_j'$. We omit the details. □

Since $\mathbb{P}(\xi_1 = 0) = 0$, we have that $\mathbb{P}(\mu = 0) = \mathbb{P}(\sup_{j=1}^{\infty} \xi_j = 0) \leq \mathbb{P}(\xi_1 = 0) = 0$. Thus, we see that M_t really is of order $t^{1/(2+\delta)}$, and is not smaller.

8.7 Related preferential attachment models

There are numerous related preferential attachment models in the literature. Here we discuss a few of them:

A directed preferential attachment model. In [45], a directed preferential attachment model is investigated, and it is proved that the degrees obey a power law similar to the one in Theorem 8.2. We first describe the model. Let G_0 be any fixed initial directed graph with t_0 edges. Fix some non-negative parameters $\alpha, \beta, \gamma, \delta_{\mathrm{in}}$ and δ_{out}, where $\alpha + \beta + \gamma = 1$.

We next define $G(t)$. In order to do so, we say that we choose a vertex *according to* $f_i(t)$ when we choose vertex i with probability

$$\frac{f_i(t)}{\sum_j f_j(t)}. \tag{8.7.1}$$

Thus, the probability that we choose a vertex i is proportional to the value of the function $f_i(t)$. Also, we denote the in-degree of vertex i in $G(t)$ by $D_{\mathrm{in},i}(t)$, and the out-degree of vertex i in $G(t)$ by $D_{\mathrm{out},i}(t)$.

We let $G(t_0) = G_0$, where t_0 is chosen appropriately, as we will indicate below. For $t \geq t_0$, we form $G(t+1)$ from $G(t)$ according to the following growth rules:

(A) With probability α, we add a new vertex v together with an edge from v to an existing vertex which is chosen according to $D_{\mathrm{in},i}(t) + \delta_{\mathrm{in}}$.

(B) With probability β, we add an edge between the existing vertices v and w, where v and w are chosen independently, v according to $D_{\mathrm{in},i}(t) + \delta_{\mathrm{in}}$ and w according to $D_{\mathrm{out},i}(t) + \delta_{\mathrm{out}}$.

(C) With probability γ, we add a vertex w and an edge from an existing vertex v to w according to $D_{\mathrm{out},i}(t) + \delta_{\mathrm{out}}$.

The above growth rule produces a graph process $\{G(t)\}_{t\geq t_0}$ where $G(t)$ has precisely t edges. The number of vertices in $G(t)$ is denoted by $T(t)$, where $T(t) \sim \mathrm{Bin}(t, \alpha + \gamma)$.

It is not hard to see that if $\alpha\delta_{\mathrm{in}} + \gamma = 0$, then all vertices outside of G_0 will have in-degree zero, while if $\gamma = 1$ all vertices outside of G_0 will have in-degree one. Similar trivial graph processes arise when $\gamma\delta_{\mathrm{out}} + \alpha = 0$ or $\alpha = 1$.

Exercise 8.24 (Special cases directed PA model)**.** *Prove that if $\alpha\delta_{\mathrm{in}} + \gamma = 0$, then all vertices outside of G_0 will have in-degree zero, while if $\gamma = 1$ all vertices outside of G_0 will have in-degree one.*

We exclude the above cases. Then, [45] show that both the in-degree and the out degree of the graph converge, in the sense that we will explain now. Denote by $X_i(t)$ the in-degree sequence of $G(t)$, so that

$$X_k(t) = \sum_{v \in G(t)} \mathbb{1}_{\{D_{\mathrm{in},v}(t)=k\}}, \tag{8.7.2}$$

and, similarly, let $Y_i(t)$ be the out-degree sequence of $G(t)$, so that

$$Y_k(t) = \sum_{v \in G(t)} \mathbb{1}_{\{D_{\mathrm{out},v}(t)=k\}}. \tag{8.7.3}$$

Denote

$$\tau_{\mathrm{in}} = 1 + \frac{1+\delta_{\mathrm{in}}(\alpha+\beta)}{\alpha+\beta}, \qquad \tau_{\mathrm{out}} = 1 + \frac{1+\delta_{\mathrm{out}}(\gamma+\beta)}{\gamma+\beta}. \tag{8.7.4}$$

Then [45, Theorem 3.1] shows that there exist probability distributions $p = \{p_k\}_{k=0}^{\infty}$ and $q = \{q_k\}_{k=0}^{\infty}$ such that with high probability

$$X_k(t) - p_k t = o(t), \qquad Y_k(t) - q_k t = o(t), \tag{8.7.5}$$

while, for $k \to \infty$,

$$p_k = C_{\mathrm{in}} k^{-\tau_{\mathrm{in}}}(1+o(1)), \qquad q_k = C_{\mathrm{out}} k^{-\tau_{\mathrm{out}}}(1+o(1)). \tag{8.7.6}$$

In fact, the probability distributions p and q are determined explicitly, as in (8.3.2) above, and p and q have a similar shape as p in (8.3.2). Also, since $\delta_{\mathrm{in}}, \delta_{\mathrm{out}} \geq 0$, and $\alpha+\beta, \gamma+\beta \leq 1$, we again have that $\tau_{\mathrm{in}}, \tau_{\mathrm{out}} \in (2, \infty)$. In [45], there is also a result on the joint distribution of the in- and out-degrees of $G(t)$, which we shall not state here.

The proof in [45] is similar to the one chosen here. Again the proof is split into a concentration result as in Proposition 8.3, and a determination of the expected empirical degree sequence in Proposition 8.4. In fact, the proof Proposition 8.4 is adapted after the proof in [45], which also writes down the recurrence relation in (8.5.20), but analyses it in a different way, by performing induction on k, rather than on t as we do in Sections 8.5.1 and 8.5.2. As a result, the result proved in Proposition 8.4 is slightly stronger. A related result on a directed preferential attachment model can be found in [60]. In this model, the preferential attachment probabilities only depend on the in-degrees, rather than on the total degree, and power-law in-degrees are proved.

A general preferential attachment model. A quite general version of preferential attachment models is presented in [76]. In this paper, an undirected graph process is defined. At time 0, there is a single initial vertex v_0. Then, to go from $G(t)$ to $G(t+1)$, either a new vertex can be added or a number of edges between existing vertices. The first case is called NEW, the second OLD. With probability α, we choose to apply the procedure OLD, and with probability $1-\alpha$ we apply the procedure NEW.

In the procedure NEW, we add a single vertex, and let $f = \{f_i\}_{i=1}^{\infty}$ be such that f_i is the probability that the new vertex generates i edges. With probability β, the end vertices of these edges are chosen *uniformly* among the vertices, and, with probability $1-\beta$, the end vertices of the added edges are chosen proportionally to the degree.

In the procedure OLD, we choose a single old vertex. With probability δ, this vertex is chosen uniformly, and with probability $1-\delta$, it is chosen with probability proportionally to the degree. We let $g = \{g_i\}_{i=1}^{\infty}$ be such that g_i is the probability that the old vertex generates i edges. With probability γ, the end vertices of these edges are chosen *uniformly* among the vertices, and, with probability $1-\gamma$, the end vertices of the added edges are chosen proportionally to the degree.

The main result in [76] states that the empirical degree distribution converges to a probability distribution which obeys a power law with a certain exponent τ which depends on the parameters of the model. More precisely, a result such as in Theorem 8.2 is proved, at least for $k \leq t^{1/21}$. Also, a version of Proposition 8.4 is proved, where the error term $\mathbb{E}[P_k(t)] - tp_k$ is proved to be at most $Mt^{1/2} \log t$. For this result, some technical conditions need to be made on the first moment of f, as well as on the distribution g. The result is nice, because it is quite general. The precise bounds are a bit weaker than the ones presented here.

Interestingly, also the maximal degree is investigated, and it is shown that the maximal degree is of order $\Theta(t^{1/(\tau-1)})$ as one would expect. This result is proved as long as $\tau < 3$. [1] Finally, the results close to those that we present here are given in [4]. In fact, the error bound in Proposition 8.4 is proved there for $m = 1$ for several models. The result for $m > 1$ is, however, not contained there.

Non-linear preferential attachment. There is also work on preferential attachment models where the probability of connecting to a vertex with degree k depends in a non-linear way on k. In [152], the attachment probabilities have been chosen proportional to k^γ for some γ. The linear case was non-rigorously investigated in [151], and the cases where $\gamma \neq 1$ in [152]. As one can expect, the results depend dramatically in the choice of γ. When $\gamma < 1$, the degree sequence is predicted to have a power law with a certain stretched exponential cut-off. Indeed, the number of vertices with degree k at time t is predicted to be roughly equal to $t\alpha_k$, where

$$\alpha_k = \frac{\mu}{k^\gamma} \prod_{j=1}^{k} \frac{1}{1 + \frac{\mu}{j^\gamma}}, \tag{8.7.7}$$

and where μ satisfies the implicit equation that $\sum_k \alpha_k = 1$. When $\gamma > 1$, then [151] predicts that there is a single vertex that is connected to nearly all the other vertices. In more detail, when $\gamma \in (1 + \frac{1}{m+1}, 1 + \frac{1}{m})$, it is predicted that there are only finitely many vertices that receive more than $m+1$ links, while there are, asymptotically, infinitely many vertices that receive at least m links. This was proved rigorously in [184].

In [197], random trees with possibly non-linear preferential attachment are studied by relating them to continuous-time branching processes and using properties of such branching processes. Their analysis can be seen as a way to make the heuristic in Section 1.3.2 precise. To explain their results, let w_i be the *weight* of a vertex of degree i. The random tree evolves, conditionally on the tree at time t, by attaching the $(t+1)^{\text{st}}$ vertex to vertex i with probability proportional to $w_{D_i(t)-1}$. Let λ^* be the solution, if it exists, of the equation

$$1 = \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{w_i}{w_i + \lambda}. \tag{8.7.8}$$

Then, it is proved in [197] that the degree distribution converges to $p_w = \{p_w(k)\}_{k=1}^{\infty}$, where[2]

$$p_w(k) = \frac{\lambda^*}{w_k + \lambda^*} \prod_{i=0}^{k} \frac{w_i}{w_i + \lambda^*}. \tag{8.7.9}$$

[1] On [76, Page 318], it is mentioned that when the power law holds with power law exponent τ, that this suggests that the maximal degree should grow like $t^{1/\tau}$. However, when the degrees are independent and identically distributed with a power law exponent equal to τ, then the maximal degree should grow like $\Theta(t^{1/(\tau-1)})$, which is precisely what is proved in [76, Theorems 2 and 5].

[2] The notion of degree used in [197] is slightly different since [197] makes use of the in-degree only. For trees, we have that the degree is the in-degree plus 1, which explains the apparent difference in the formulas.

For linear preferential attachment models where $w_i = i + 1 + \delta$, we have that $\lambda^* = \delta$, so that (8.7.9) reduces to (8.3.3):

Exercise 8.25 (The affine preferential attachment case)**.** *Prove that, when* $\lambda^* = \delta$ *and* $w_i = i + 1 + \delta$, (8.7.9) *reduces to* (8.3.3)*.*

Interestingly, in [197] not only the degree of a uniformly chosen vertex is studied, but also its neighborhood. We refrain from describing these results here. These analyses are extended beyond the tree case in [31].

Preferential attachment with fitness. The models studied in [35, 36, 101] include preferential attachment models with *random fitness*. In general, in such models, the vertex v_i which is added at time i is given a random fitness (ζ_i, η_i). The later vertex v_t at time $t > i$ connects to vertex v_i with a conditional probability which is proportional to $\zeta_i D_i(t) + \eta_i$. The variable ζ_i is called the *multiplicative fitness*, and η_i is the *additive fitness*. The case of additive fitness only was introduced in [101], the case of multiplicative fitness was introduced in [35, 36] and studied further in [54]. Bhamidi [31] finds the exact degree distribution both for the additive and multiplicative models.

Preferential attachment and power-law exponents in $(1, 2)$**.** In all models, and similarly to Theorem 8.2, the power law exponents τ are limited to the range $(2, \infty)$. It would be of interest to find simple examples where the power law exponent can lie in the interval $(1, 2)$. A possible solution to this is presented in [82], where a preferential attachment model is presented in which a *random* number of edges can be added which is, unlike [76], not bounded. In this case, when the number of edges obeys a power law, then there is a cross-over between a preferential attachment power law and the power law from the edges, the one with the smallest exponent winning. Unfortunately, the case where the weights have degrees with power-law exponent in $(1, 2)$ is not entirely analyzed. The conjecture in [82] in this case is partially proved by Bhamidi in [31, Theorem 40].

Universal techniques to study preferential attachment models. In [31], Bhamidi investigates various preferential attachment models using universal techniques from continuous-time branching processes (see [9] and the works by Jagers and Nerman [124, 125, 175]) to prove powerful results for preferential attachment graphs. Models that can be treated within this general methodology include fitness models [35, 36, 101], competition-induced preferential attachment models [28, 29], linear preferential attachment models as studied in this chapter, but also sublinear preferential attachment models and preferential attachment models with a cut-off. Bhamidi is able to prove results for (1) the degree distribution of the graph; (2) the maximal degree; (3) the degree of the initial root; (4) the local neighborhoods of vertices; (5) the height of various preferential attachment trees; and (6) properties of percolation on the graph, where we erase the edges independently and with equal probability.

8.8 Notes and discussion

Notes on Section 8.1. There are various ways of modeling the Rich-get-Richer or preferential attachment phenomenon, and in these notes, we shall describe some related models. The most general model is studied in [76], the main result being that the degrees obey a power law. A model where the added edges are conditionally independent given the degrees is given in [140]. A directed preferential attachment model is presented in [27].

Notes on Section 8.2. The degrees of fixed vertices plays a crucial role in the analysis of preferential attachment models, see e.g. [50]. In [213], several moments of the degrees are computed for the Albert-Barabási model, including the result in Theorem 8.1 and several extensions.

Notes on Section 8.3. Most papers on specific preferential attachment models prove that the degree sequences obey a power law. We shall refer in more detail to the various papers on the topic when we discuss the various different ways of proving Proposition 8.4. General results in this direction can be found for example in [31].

Notes on Section 8.4. The proof of Theorem 8.2 relies on two key propositions, namely, Propositions 8.3 and 8.4. Proposition 8.3 is a key ingredient in the investigation of the degrees in preferential attachment models, and is used in many related results for other models. The first version, as far as we know, of this proof is in [50].

Notes on Section 8.5. The proof of the expected empirical degree sequence in Proposition 8.4 is new, and proves a stronger result than the one for $\delta = 0$ appearing in [50]. The proof of Proposition 8.4 is also quite flexible. For example, instead of the growth rule in (8.1.1), we could attach the m edges of the newly added vertex $v_{t+1}^{(m)}$ each independently and with equal probability to a vertex $i \in [t]$ with probability proportional to $D_i(t) + \delta$. More precisely, this means that, for $t \geq 3$,

$$\mathbb{P}\big(v_{t+1}^{(m)} \to v_i^{(1)} \big| \mathrm{PA}_t(m, \delta)\big) = \frac{D_i(t) + \delta}{t(2m + \delta)} \qquad \text{for } i \in [t], \tag{8.8.1}$$

and, conditionally on $\mathrm{PA}_t(m, \delta)$, the attachment of the edges are *independent*. We can define $\mathrm{PA}_2(m, \delta)$ to consist of 2 vertices connected by m edges.

It is not hard to see that the proof of Proposition 8.3 applies verbatim:

Exercise 8.26 (Adaptation concentration degree sequence)**.** *Adapt the proof of Proposition 8.3 showing the concentration of the degrees to the preferential attachment model defined in* (8.8.1)*.*

It is not hard to see that also the proof of Proposition 8.4 applies by making the obvious changes. In fact, the limiting degree sequence remains unaltered. A second slightly different model, in which edges are added independently without intermediate updating, is studied by Jordan in [137].

The original proof in [50] of the asymptotics of the expected empirical degree sequence for $\delta = 0$ makes use of an interesting relation between this model and so-called *n-pairings*. An n-pairing is a partition of the set $\{1, \ldots, 2n\}$ into pairs. We can think about the pairs as being points on the x-axis, and the pairs as chords joining them. This allows us to speak of the left- and right-endpoints of the pairs.

The link between an n-pairing and the preferential attachment model with $\delta = 0$ and $m = 1$ is obtained as follows. We start from the left, and merge all left-endpoints up to and including the first right endpoint into the single vertex v_1. Then, we merge all further left-endpoints up to the next right endpoint into vertex v_2, etc. For the edges, we replace each pair by a directed edge from the vertex corresponding to its right endpoint to the vertex corresponding to its left endpoint. Then, as noted in [48], the resulting graph has the same distribution as $G_1(t)$. The proof in [50] then uses explicit computations to prove that for $k \leq t^{1/15}$,

$$\mathbb{E}[N_k(t)] = tp_k(1 + o(1)). \tag{8.8.2}$$

The advantage of the current proof is that the restriction on k in $k \leq t^{1/15}$ is absent, that the error term in (8.8.2) is bounded uniformly by a constant, and that the proof applies to $\delta = 0$ and $\delta \neq 0$.

The approach of Hagberg and Wiuf in [112] is closest to ours. In it, the authors assume that the model is a preferential attachment model, where the expected number of vertices of degree k in the graph at time $t+1$, conditionally on the graph at time t solves

$$\mathbb{E}[N_k(t+1)|N(t)] = (1 - \frac{a_k}{t})N_k(t) - \frac{a_{k-1}}{t}N_{k-1}(t) + c_k, \tag{8.8.3}$$

where $N_k(t)$ is the number of vertices of degree k at time t, $N(t) = \{N_k(t)\}_{k=0}^{\infty}$ and it is assumed that $a_{-1} = 0$, and where $c_k \geq 0$ and $a_k \geq a_{k-1}$. Also, it is assumed that $|N_k(t) - N_k(t-1)|$ is uniformly bounded. This is almost true for the model considered in this chapter. Finally, $\{N(t)\}_{t=0}^{\infty}$ is assumed to be a Markov process, starting at some time t_0 in a configuration $N(t_0)$. Then, with

$$\alpha_k = \sum_{j=0}^{k} \frac{c_j}{1+a_j} \prod_{i=j+1}^{\infty} \frac{a_{i-1}}{1+a_i}, \tag{8.8.4}$$

it is shown that $N_t(k)/t$ converges to α_k.

Exercise 8.27 (Monotonicity error [112])**.** *Show that*

$$\max_{j=1}^{k} |\mathbb{E}[N_t(j)] - \alpha_j t| \tag{8.8.5}$$

is non-increasing.

Notes on Section 8.6. The beautiful martingale description in Proposition 8.9 is due to Mori [172] (see also [173]). We largely follow the presentation in [88, Section 4.3], adapting it to the setting of preferential attachment models in Section 8.1. The fact that Proposition 8.9 also holds for *non-integer* k_i is, as far as we know, new. This is relevant, since it identifies *all* moments of the limiting random variables ξ_j, which might prove useful in order to identify their distribution, which, however, has not been done yet.

Intermezzo: Back to real networks II...

In the previous sections, we have discussed various models having flexible degree sequences. The generalized random graph and the configuration model give us static flexible models for random graphs with various degree sequences. Preferential attachment models give us a convincing explanation of the abundance of power-law degree sequences in various applications. In Chapters 6–8, we have focussed on the properties of the *degrees* of such graphs. However, we have noted in Chapter 1 that many real networks not only have degree sequences that are rather different from the ones of the Erdős-Rényi random graph, also many examples are *small worlds* and have a *giant connected component*.

In the Chapters 9–11, we shall return to the models discussed in Chapters 6–8, and focus on their critical behavior as well as on the distances in these random graph models. Interestingly, a large chunk of the non-rigorous physics literature suggests that the behavior in various *different* random graph models can be described by only a *few* essential parameters. The key parameter of each of these models in the power-law degree exponent, and the physics literature predicts that the behavior in random graph models with similar degree sequences is similar. This is an example of the notion of *universality*, a notion which is central in statistical physics. Despite its importance, there are only few example of universality that can be rigorously proved. In Chapters 9–11, we shall investigate the level of universality present in random graph models.

Chapter 9
Inhomogeneous random graphs revisited

In this chapter, we discuss the phase transition and distances in inhomogeneous random graphs, which can be viewed as generalizations of the generalized random graph investigated in Chapter 6. For technical reasons, we primarily work with the Norros-Reittu model $\mathrm{NR}_n(\boldsymbol{w})$, all our results also apply to the generalized random graph $\mathrm{GRG}_n(\boldsymbol{w})$ and the Chung-Lu model $\mathrm{CL}_n(\boldsymbol{w})$.

This chapter is organized as follows. In Section 9.1, we investigate the connectivity structure of generalized random graphs, by investigating the phase transition of its largest connected component and its small-world properties. The results in Section 9.1 are all proved. In Section 9.2, we give the proofs of the lower bounds on the distances, in Sections 9.3– 9.4, we prove the corresponding upper bounds. The proof of the phase transition for generalized random graphs is deferred to Chapter 10, where the phase transition is proved for the configuration model. In Section 9.5, we discuss the general setting of inhomogeneous random graphs from [47] with its detailed and deep results. In Section 9.6, we state some recent related results, and in Section 9.7, we close this chapter with notes and discussion.

9.1 Connectivity structure of generalized random graphs

9.1.1 The phase transition in generalized random graphs

In this section, we investigate the size of the largest connected component in generalized random graphs.

Example 9.1 (Population of two types (Cont.))**.** Recall Example 6.1, where we assumed that two distinct types of vertices are present. The first type has on average m_1 neighbors, the second type m_2, where $m_1 \neq m_2$. We have modeled this with a $\mathrm{GRG}_n(\boldsymbol{w})$ where n_1 vertices have weight m_1 and n_2 vertices have weight m_2, and write $n = n_1 + n_2$. Assume that $n_1/n \to p$. Under what conditions does $\mathrm{GRG}_n(\boldsymbol{w})$ have a giant component? When we compare with Theorem 4.8, we may suspect that a giant component exists whenever a uniform vertex has on average more than 1 neighbor. The latter is true when

$$pm_1 + (1-p)m_2 > 1. \tag{9.1.1}$$

In this section, we examine whether this intuition is correct.

Exercise 9.1 (Average degree in two populations)**.** Show that the average degree is close to $pm_1 + (1-p)m_2$ in the setting of Example 9.1.

In order to state the main result in this section, we recall the limiting degree distribution in the generalized random graph given in (6.3.2)

$$p_k = \mathbb{E}\big[\mathrm{e}^{-W}\frac{W^k}{k!}\big], \qquad k \geq 0. \tag{9.1.2}$$

Then, Theorem 6.9 shows that the degree sequence in $\mathrm{GRG}_n(\boldsymbol{w})$ is close to $(p_k)_{k\geq 0}$. The main result is as follows:

Theorem 9.2 (Phase transition in generalized random graphs)**.** *Suppose that Condition 6.4(a)-(b) hold and consider the random graph* $\mathrm{NR}_n(\boldsymbol{w})$*, letting* $n \to \infty$*. Let* $\mathcal{C}_{\max}$ *and* $\mathcal{C}_{(2)}$ *be the largest and second largest components of* $\mathrm{NR}_n(\boldsymbol{w})$*.*

(a) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$*, then there exist* $\xi \in (0,1), \zeta \in (0,1)$ *such that*

$$\begin{aligned} |\mathcal{C}_{\max}|/n &\xrightarrow{\mathbb{P}} \zeta, \\ v_k(\mathcal{C}_{\max})/n &\xrightarrow{\mathbb{P}} p_k(1-\xi^k), \text{ for every } k \geq 0, \\ |E(\mathcal{C}_{\max})|/n &\xrightarrow{\mathbb{P}} \tfrac{1}{2}\mathbb{E}[W](1-\xi^2). \end{aligned}$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ *and* $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$.

(b) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \leq 1$*, then* $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ *and* $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.

The above results apply to $\mathrm{GRG}_n(\boldsymbol{w})$ *and* $\mathrm{CL}_n(\boldsymbol{w})$ *under the same conditions.*

The proof of Theorem 9.2 is deferred to Section 10.1.2 in Chapter 10, where a similar result is proved for the configuration model. By the strong relation between the configuration model and the generalized random graph, this result can be seen to imply Theorem 9.2.

Exercise 9.2 (The phase transition for two populations)**.** *Show that the* $\zeta > 0$ *precisely when* $[pm_1^2 + (1-p)m_2^2]/[pm_1 + (1-p)m_2] > 1$ *in the setting of Example 9.1. Find an example of* p, m_1, m_2 *where the average degree is less than one, yet there exists a giant component.*

Reformulation in terms of branching processes. We start by reformulating the results in Theorem 9.2 in terms of *branching processes*. We can interpret ξ as the extinction probability of a branching process, and ζ as the survival probability of a related two-stage branching process.

We start by introducing two-stage branching processes with a mixed Poisson offspring. We define the branching process $(\mathcal{Z}_l)_{l\geq 0}$ as starting from $\mathcal{Z}_0 = 1$, where in the first generation the offspring distribution is equal to $(p_k)_{k\geq 0}$ given in (9.1.2), whereas in the second and further generations the offspring is chosen in accordance to

$$g_k = \frac{(k+1)p_{k+1}}{\mathbb{E}[W]} = \frac{1}{\mathbb{E}[W]}\mathbb{E}\big[\mathrm{e}^{-W}\frac{W^{k+1}}{k!}\big], \qquad k \geq 0, \tag{9.1.3}$$

where W is the mixing distribution. The expected number of offspring in the second and further generations is given by

$$\sum_{k=1}^{\infty} k g_k = \frac{1}{\mathbb{E}[W]}\mathbb{E}\big[\sum_{k=1}^{\infty} k(k+1)\mathrm{e}^{-W}\frac{W^{k+1}}{(k+1)!}\big] = \mathbb{E}[W^2]/\mathbb{E}[W] = \nu. \tag{9.1.4}$$

In particular, a branching process with mixed Poisson offspring distribution is supercritical when $\mathbb{E}[W^2] > \mathbb{E}[W]$. In Section 9.3.1, we relate the neighborhoods of $\mathrm{NR}_n(\boldsymbol{W})$ to an ordinary branching process having asymptotic expected offspring equal to $\nu = \mathbb{E}[W^2]/\mathbb{E}[W]$. Therefore, $\nu > 1$ is equivalent to the branching process being supercritical.

We now heuristically explain this relation to branching processes by describing the exploration of a vertex V chosen uniformly from the vertex set $[n]$. By Theorem 6.9, the probability that its degree is k converges to p_k, for every $k \geq 1$. This explains the offspring of the root of our branching process approximation.

To describe the offspring of the direct neighbors of the root, we need to examine the forward degree of a uniform neighbor of the root. Here, by forward degree, we mean the number of vertices unequal to the root to which a neighbor of the root is connected. Intuitively, this forward degree is not much different from the degree minus one of a vertex contained in a *uniform edge*. We apply Theorem 6.5, which states that the total number

of edges in $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$ is well approximated by its mean ℓ_n. We also use that the probability that an edge is occupied is close to $w_i w_j/\ell_n$, so that the probability $g_k^{(n)}$ that the forward degree equals k is close to

$$\begin{aligned} g_k^{(n)} &\approx \frac{1}{\ell_n} \sum_{i,j\in[n]} \mathbb{P}(ij \text{ occupied}, D_j = k+1) \\ &= \frac{1}{\ell_n} \sum_{i,j\in[n]} \frac{w_i w_j}{\ell_n} \mathbb{P}(X_j = k) \\ &= \frac{1}{\ell_n} \sum_{j\in[n]} w_j \mathbb{P}(X_j = k), \end{aligned} \tag{9.1.5}$$

where $X_j = \sum_{s\in[n]:\, s\neq i,j} I_{sj}$, where (I_{sj}) are independent Bernoulli variables with $\mathbb{P}(I_{sj} = 1) = w_s w_j/\ell_n$. By Theorem 2.9, X_j is close to a Poisson random variable with mean $\sum_{s\in[n]} w_s w_j/\ell_n = w_j$, so that

$$g_k^{(n)} \approx \frac{1}{\ell_n} \sum_{j\in[n]} w_j \mathbb{P}(\mathrm{Poi}(w_j) = k) = \mathbb{P}(\mathrm{Poi}(W_n^*) = k), \tag{9.1.6}$$

where W_n^* is the size-biased distribution of W_n. Here, for a non-negative random variable X with $\mathbb{E}[X] > 0$, we let X^* denote its *size-biased distribution* given by

$$\mathbb{P}(X^* \leq x) = \frac{\mathbb{E}[X \mathbb{1}_{\{X\leq x\}}]}{\mathbb{E}[X]}. \tag{9.1.7}$$

When Condition 6.4(a)-(b) holds, we have that $W_n^* \xrightarrow{d} W^*$, which is the size-biased distribution of W. We arrive at

$$g_k^{(n)} \approx \mathbb{P}(\mathrm{Poi}(W^*) = k) = \frac{1}{\mathbb{E}[W]} \mathbb{E}\big[\mathrm{e}^{-W} \frac{W^{k+1}}{k!}\big] = g_k, \tag{9.1.8}$$

which explains (9.1.3). It turns out that the steps in the above heuristic explanation can be made precise. See Section 9.3.1, where we relate the neighborhood of a uniform vertex in $\mathrm{NR}_n(\boldsymbol{w})$ to a mixed Poisson random variable.

The above heuristically argues that the number of vertices unequal to the root connected to any neighbor of the root has asymptotic law $(g_k)_{k\geq 0}$. However, every time we find a vertex in the cluster of the root, the number of *available* vertices decreases, in a similar way as the depletion-of-points effect in the exploration of clusters for the Erdős-Rényi random graph $\mathrm{ER}_n(\lambda/n)$. Since the number of vertices is originally n, we can grow the cluster of the root for a long time before we note this effect.

As a result, the degrees of different vertices in the exploration process is close to being i.i.d., leading to a branching process approximation. In terms of this branching process, we can interpret ζ in Theorem 9.2 as the survival probability of the above two-stage branching process, so that ζ satisfies

$$\zeta = \sum_{k=1}^{\infty} p_k (1 - \xi^k), \tag{9.1.9}$$

where ξ is the extinction probability of the branching process with offspring distribution $(g_k)_{k\geq 0}$. Clearly, $\xi = 1$ precisely when

$$\nu = \sum_{k\geq 0} k g_k \leq 1, \tag{9.1.10}$$

which explains the condition on ν in Theorem 9.2(a). Further, by Theorem 6.9, there are approximately np_k vertices with degree k. Assuming approximate independence of each of these k neighbors, each of them survives with probability $1-\xi$, so that the probability that at least one survives equals $1-\xi^k$. When one of the neighbors survive, the vertex itself will be part of the giant component, which explains why $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1-\xi^k)$. Finally, an edge consists of two half-edges, and an edge is part of the giant component precisely when one of the vertices incident to it is, which occurs with probability $1-\xi^2$. There are in total approximately $\ell_n/2 \approx n\mathbb{E}[W]/2$ edges, which explains why $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[W](1-\xi^2)$. Therefore, all results in Theorem 9.2 have a simple explanation in terms of the branching process approximation of the connected component of a uniform vertex in $[n]$ for $\mathrm{NR}_n(\boldsymbol{w})$.

Exercise 9.3 (Degree sequence of giant component)**.** *Show that the proportion of vertices of the giant component $\mathcal{C}_{\max}$ having degree k is close to $p_k(1-\xi^k)/\zeta$.*

Exercise 9.4 (Degree sequence of complement of giant component)**.** *Show that when $\xi < 1$, the proportion of vertices outside the giant component $\mathcal{C}_{\max}$ having degree k is close to $p_k\xi^k/(1-\zeta)$. Conclude that the degree sequence of the complement of the giant component never satisfies a power law. Can you give an intuitive explanation for this?*

We close this section by discussing the consequences of the phase transition for the attack vulnerability of $\mathrm{CL}_n(\boldsymbol{w})$:

Attack vulnerability of $\mathrm{CL}_n(\boldsymbol{w})$. Suppose an adversary attacks a network by removing some of its vertices. A *clever* adversary would remove the vertices in a clever way, this is often referred to as a *deliberate* attack. On the other hand, the vertices might also be exposed to random failures, which is often referred to as a *random* attack. The results as stated above do not specifically apply to these settings, but do have intuitive consequences. We model a deliberate attack as the removal of a proportion of the vertices with highest weights, whereas a random attack is modeled by random removal of the vertices with a given probability. One of the aims is to quantify the effect of such attacks, and in particular the difference in random and deliberate attacks. We shall denote the proportion of removed vertices by p. We shall always assume that $\nu > 1$, so that a giant component exists, and we investigate under what conditions on p and the graph $\mathrm{CL}_n(\boldsymbol{w})$, the giant component remains to exist.

We start by addressing the case of random attack for the $\mathrm{CL}_n(\boldsymbol{w})$ model under Condition 6.4(a)-(c), where $\mathbb{E}[W^2] < \infty$. One of the difficulties of the above set-up is that we remove *vertices* rather than *edges*, so that the resulting graph is no longer an IRG. In percolation jargon, we deal with site percolation rather than with edge percolation. We start by relating the obtained graph to an IRG.

Note that when we *explore* a cluster of a vertex after an attack, then the vertex may not have been affected by the attack, which has probability p. After this, in the exploration, we shall always inspect an edge between a vertex which is unaffected by the attack and a vertex of which we do not yet know whether it has been attacked or not. As a result, for random attacks, the probability that it is affected is precisely equal to p. Therefore, it is similar to the random graph where p_{ij} is replaced with $(1-p) \times p_{ij}$. For a branching process, this identification is exact, and we have that $\zeta_{\kappa,p} = (1-p)\zeta_{(1-p)\kappa}$, where $\zeta_{\kappa,p}$ denotes the survival probability of the branching process where each individual is killed with probability p independently of all other randomness. For $\mathrm{CL}_n(\boldsymbol{w})$, this equality is only asymptotic. In the case where $\mathbb{E}[W^2] < \infty$, so that $\nu < \infty$, this means that there exists a critical value $p_c = 1 - 1/\nu$, such that if $p < p_c$, the $\mathrm{CL}_n(\boldsymbol{w})$ where vertices are removed with probability p, the giant component persists, while if $p > p_c$, then the giant component is destroyed. Thus, when $\mathbb{E}[W^2] < \infty$, the $\mathrm{CL}_n(\boldsymbol{w})$ is *sensitive to random attacks*. When $\mathbb{E}[W^2] = \infty$, on the other hand, $\nu = \infty$, so that the giant component persists for *every*

$p \in [0,1)$, and the graph is called *robust to random attacks*. Here we must note that the *size* of the giant component does decrease, since $\zeta_{\kappa,p} < p\zeta_\kappa$!

For a deliberate attack, we remove the proportion p of vertices with highest weight. This means that $\boldsymbol{w}$ is replaced with $\boldsymbol{w}(p)$, which is equal to $w_i(p) = w_i \mathbb{1}_{\{i>np\}}$, and we denote the resulting edge probabilities by

$$p_{ij}(p) = \max\{1, w_i(p)w_j(p)/\ell_n\}. \tag{9.1.11}$$

In this case, the resulting graph on $[n] \setminus [np]$ is again a Chung-Lu model, for which ν is replaced with $\nu(p)$ given by

$$\nu(p) = \mathbb{E}[[1-F]^{-1}(U)^2 \mathbb{1}_{\{U>p\}}]/\mathbb{E}[W], \tag{9.1.12}$$

where U is uniform on $[0,1]$. Now, for any distribution function F, $\mathbb{E}[[1-F]^{-1}(U)^2 \mathbb{1}_{\{U>p\}}] < \infty$, so that, for p sufficiently close to 1, $\nu(p) < 1$. Thus, the $\mathrm{CL}_n(\boldsymbol{w})$ model is always sensitive to deliberate attacks.

Exercise 9.5 (Finiteness of $\nu(p)$)**.** *Prove that $\nu(p) < \infty$ for every $p \in (0,1]$ and any distribution function F.*

Exercise 9.6 (Connectivity of uniformly chosen vertices)**.** *Suppose we draw two vertices uniformly at random from $[n]$. Prove that Theorem 9.2 implies that the probability that the vertices are connected converges to ζ^2.*

9.1.2 The small-world phenomenon in generalized random graphs

In this section, we discuss typical distances in $\mathrm{NR}_n(\boldsymbol{w})$. We define H_n to be the graph distance between two vertices chosen uniformly at random from $[n]$, where the graph distance between two vertices is the minimal number of edges in all paths connecting the vertices. It is possible that no path connecting the vertices exists, in which case, we define $H_n = +\infty$. By Theorem 9.2, $\mathbb{P}(H - n = +\infty) \to 1 - \zeta^2 > 0$, since $\zeta < 1$ (see Exercise 9.6). In particular, when $\zeta = 0$, which is equivalent to $\nu \le 1$, $\mathbb{P}(H - n = +\infty) \to 1$. Therefore, in our main results, we shall condition on $H_n < \infty$.

Distances in inhomogeneous random graphs with finite variance weights. We start by investigating the behavior of H_n for $\mathrm{NR}_n(\boldsymbol{w})$ in the case where the weights have finite variance:

Theorem 9.3 (Typical distances in $\mathrm{NR}_n(\boldsymbol{w})$ for finite-variance weights)**.** *In the Norros-Reittu model $\mathrm{NR}_n(\boldsymbol{w})$, where the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfy Condition 6.4(a)-(c) and where $\nu > 1$, conditionally on $H_n < \infty$,*

$$H_n / \log n \xrightarrow{\mathbb{P}} 1/\log \nu. \tag{9.1.13}$$

The same result applies, under the same conditions, to $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$.

We give a complete proof of Theorem 9.3 in Sections 9.2-9.4 below. The intuition behind Theorem 9.3 is as follows. In Section 9.1.1, we have argued that the neighborhood of a uniform vertex in $\mathrm{NR}_n(\boldsymbol{w})$ is well-approximated by a two-stage branching process, where the second and all later generations have offspring distribution $(g_k)_{k\ge0}$ in (9.1.3). When $\nu = \sum_{k\ge0} k g_k < \infty$, then the number of vertices at distance k is close to $M\nu^k$, where M is the martingale limit of $\mathcal{Z}_k/\nu^k$. To know what H_n is, we need to grow the neighborhoods from the first uniform vertex until we find the second uniform vertex. The latter happens with reasonable probability when $\mathcal{Z}_k \approx n$, which suggests that the relevant k is such that $\nu^k \approx n$, so that $k \approx \log_\nu n$.

While the above heuristic is quite convincing, the argument is fatally flawed. Indeed, as argued in Section 9.1.1, the neighborhoods of a uniform vertex are well-approximated by a branching process as long as the number of vertices found is much smaller than n. When the number of vertices found becomes of order n, the depletion-of-points effect has already started to kick in. Therefore, the above approach is doomed to fail. Our proof instead, is divided in a lower and and upper bound on the typical distance H_n. For the proof of the lower bound in Section 9.2.1, we show that the *expected* number of paths of k edges between two uniform vertices is approximately ν^k/ℓ_n, so that such a path **whp**does not exist when $k \leq (1-\varepsilon)\log_\nu n$. For the proof of the upper bound in Section 9.4, we use a second moment method to show that, conditionally on the two uniformly chosen vertices being in the giant component, **whp**there exists a path of $(1+\varepsilon)\log_\nu n$ edges.

Exercise 9.7 (Typical distances in $\mathrm{ER}_n(\lambda/n)$)**.** *Prove that* $H_n/\log n \xrightarrow{\mathbb{P}} 1/\log\lambda$ *in* $\mathrm{ER}_n(\lambda/n)$.

Theorem 9.3 leaves open what happens when $\nu = \infty$. We can use Theorem 9.3 to show that $H_n = o_{\mathbb{P}}(\log n)$:

Exercise 9.8 (Typical distances when $\nu = \infty$)**.** *Prove that* $H_n/\log n \xrightarrow{\mathbb{P}} 0$ *when* $\nu = \infty$.

Distances in inhomogeneous random graphs with infinite variance weights. In this section, we study typical distances in the Norros-Reittu random graph $\mathrm{NR}_n(\boldsymbol{w})$, in the case where the degrees obey a power-law with degree exponent τ satisfying that $\tau \in (2,3)$. In this case, $\nu = \infty$, so that $H_n = o_{\mathbb{P}}(\log n)$ (recall Exercise 9.8).

Many of our arguments also apply to the generalized random graph $\mathrm{GRG}_n(\boldsymbol{w})$ and the Chung-Lu model $\mathrm{CL}_n(\boldsymbol{w})$. In this section, we discuss the setting where the weights $\boldsymbol{w}$ are heavy tailed. Recall that $F_n(x)$ denotes the proportion of vertices i for which $w_i \leq x$. Then, we assume that there exists a $\tau \in (2,3)$ such that for all $\delta > 0$, there exists $c_1 = c_1(\delta)$ and $c_2 = c_2(\delta)$ such that, uniformly in n,

$$c_1 x^{-(\tau-1+\delta)} \leq [1-F_n](x) \leq c_2 x^{-(\tau-1-\delta)}, \tag{9.1.14}$$

where the upper bound is expected to hold for every $x \geq 1$, while the lower bound is only required to hold for $1 \leq x \leq n^\alpha$ for some $\alpha > 1/2$.

The assumption in (9.1.14) is what we need precisely, and it states that $[1-F_n](x)$ obeys power-law bounds for appropriate values of x. Note that the lower bound in (9.1.14) cannot be valid for *all* x, since $F_n(x) > 0$ implies that $F_n(x) \geq 1/n$, so that the lower and upper bound in (9.1.14) are contradicting when $x \gg n^{1/(\tau-1)}$. Thus, the lower bound can hold only for $x = O(n^{1/(\tau-1)})$. When $\tau \in (2,3)$, we have that $1/(\tau-1) \in (1/2,1)$, and we only need the lower bound to hold for $x \leq n^\alpha$ for *some* $\alpha \in (1/2,1)$.

We now give simpler conditions for (9.1.14) in special cases:

Exercise 9.9 (Power-law tails in key example of deterministic weights)**.** *Let* $\boldsymbol{w}$ *be defined as in* (6.1.10)*, and assume that* F *satisfies*

$$1 - F(x) = x^{-(\tau-1)}L(x), \tag{9.1.15}$$

where the exponent satisfies $\tau \in (2,3)$*, and where* $x \mapsto L(x)$ *is slowly varying. Prove that* (9.1.14) *holds.*

Exercise 9.10 (Power-law tails for i.i.d. weights)**.** *For i.i.d. weights* $\boldsymbol{w} = (w_i)_{i\in[n]}$ *with distribution* F *satisfying that* (9.1.15) *with* $\tau \in (2,3)$*, and where* $x \mapsto L(x)$ *is slowly varying. Prove that* (9.1.14) *holds with probability converging to 1.*

Our main result is as follows:

Theorem 9.4 (Typical distances in $\mathrm{NR}_n(\boldsymbol{w})$ for $\tau \in (2,3)$)**.** *Fix the Norros-Reittu model* $\mathrm{NR}_n(\boldsymbol{w})$*, where the weights* $\boldsymbol{w} = (w_i)_{i\in[n]}$ *satisfy Condition 6.4(a)-(b) and* (9.1.14)*. Then, conditionally on* $H_n < \infty$*,*

$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|}. \tag{9.1.16}$$

The same results apply, under the same conditions, to $\mathrm{CL}_n(\boldsymbol{w})$ *and* $\mathrm{GRG}_n(\boldsymbol{w})$*.*

Theorem 9.4 implies that $\mathrm{NR}_n(\boldsymbol{w})$ with $\boldsymbol{w}$ as in (6.1.10), for $\tau \in (2,3)$, is an ultra-small world when (9.2.14) is satisfied.

The main tool to study distances in $\mathrm{NR}_n(\boldsymbol{w})$ is a comparison to branching processes, which is particularly pretty for $\mathrm{NR}_n(\boldsymbol{w})$. In the next two sections, we prove Theorems 9.3–9.4. When $\tau > 3$, then the branching process approximation has *finite mean*, and we can make use of the martingale limit results of the number of individuals in generation k as $k \to \infty$. When $\tau \in (2,3)$, on the other hand, the branching process has infinite mean. In this case, the number of individuals in generation k, conditionally on survival, grows *superexponentially*, which explains why distances grow doubly logarithmically. See Section 10.6, where this is explained in more detail in the context of the configuration model. The superexponential growth implies that a path between two vertices typically passes through vertices with growing weights as we move away from the two vertices. Thus, starting from the first vertex $V_1 \in [n]$, the path connecting V_1 to V_2 uses vertices that first grow until the midpoint of the path is reached, and then decrease again to reach V_2. This can be understood by noting that the probability that vertex with weight w is not connected to any vertex with weight larger than $y > w$ in $\mathrm{NR}_n(\boldsymbol{w})$ is

$$\mathrm{e}^{-\sum_{i:\, w_i > y} w w_i/\ell_n} = \mathrm{e}^{-w[1-F_n^*](y)}, \tag{9.1.17}$$

where $F_n^*(y) = \sum_{i:\, w_i \le y} w_i/\ell_n$ is the distribution function of W_n^* introduced in (9.1.6). When (9.1.14) holds, it follows that $[1-F_n^*](y)$ is close to $y^{-(\tau-2)}$, the size-biasing increasing the power by one. Therefore, the probability that vertex with weight w is not connected to any vertex with weight larger than $y > w$ in $\mathrm{NR}_n(\boldsymbol{w})$ is approximately $\mathrm{e}^{-wy^{-(\tau-2)}}$. Take w large, then this probability is small when $y \gg w^{1/(\tau-2)}$. Thus, a vertex of weight w is **whp** connected to a vertex of weight $w^{1/(\tau-2)}$. Since $1/(\tau-2) > 1$ when $\tau \in (2,3)$, we obtain that vertices with large weights w are **whp** connected to vertices with weight at least $w^{1/(\tau-2)}$.

The proof of Theorems 9.3–9.4 are organized as follows. In Section 9.2, we prove the lower bounds on the typical distance in $\mathrm{NR}_n(\boldsymbol{w})$, both when $\tau > 3$ and when $\tau \in (2,3)$. In Section 9.3, we describe a stochastic domination of the neighborhoods in $\mathrm{NR}_n(\boldsymbol{w})$ in terms of a marked Poisson branching process and prove the $\log\log n$ upper bound for $\tau \in (2,3)$. In Section 9.4, we investigate the number of paths between sets of vertices in $\mathrm{NR}_n(\boldsymbol{w})$, and use this to prove the $\log n$ upper bound when $\tau > 3$. In each of our proofs, we formulate the precise results as separate theorems, and prove them under conditions that are slightly weaker than those in Theorems 9.3–9.4.

9.2 Lower bounds on typical distances

In this section, we prove lower bounds on typical graph distances. In Section 9.2.1, we prove the lower bound in Theorem 9.32(i) in the setting of Theorem 9.3.

9.2.1 Logarithmic lower bound graph distance in finite variance case

In this section, we prove a logarithmic lower bound on the graph distance in $\mathrm{NR}_n(\boldsymbol{w})$. The main result is as follows:

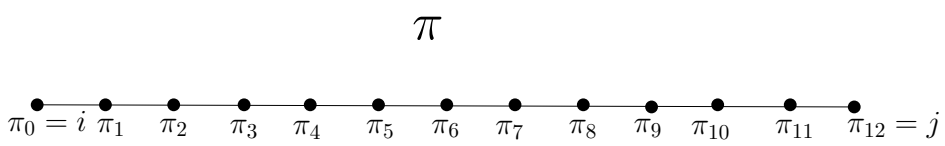

Figure 9.1: A 12-step self-avoiding path connecting vertices i and j.

Theorem 9.5 (Logarithmic lower bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Assume that*

$$\limsup_{n\to\infty} \nu_n = \nu, \tag{9.2.1}$$

where $\nu \in (1,\infty)$ *and*

$$\nu_n = \mathbb{E}[W_n^2]/\mathbb{E}[W_n] = \sum_{i\in[n]} w_i^2 / \sum_{i\in[n]} w_i. \tag{9.2.2}$$

Then, for any $\varepsilon > 0$*,*

$$\mathbb{P}(H_n \le (1-\varepsilon)\log_\nu n) = o(1). \tag{9.2.3}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ *and* $\mathrm{GRG}_n(\boldsymbol{w})$ *under the same conditions.*

Proof. We recall that H_n is the graph distance in $\mathrm{NR}_n(\boldsymbol{w})$ between two vertices drawn uniformly at random from $[n]$. The idea behind the proof of Theorem 9.5 is that it is quite unlikely that a path exists that is much shorter than $\log_\nu n$ edges. In order to show this, we use a first moment bound and show that the *expected* number of occupied paths connecting the two vertices chosen uniformly at random from $[n]$ having length at most k is $o(1)$. We now fill in the details.

We denote the graph distance between $i,j \in [n]$ in $\mathrm{NR}_n(\boldsymbol{w})$ by $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j)$ and abbreviate $k_n = \lceil (1-\varepsilon)\log_\nu n \rceil$. Then, conditioning on the uniform vertices chosen and Boole's inequality gives

$$\begin{aligned}\mathbb{P}(H_n \le k_n) &= \frac{1}{n^2}\sum_{i,j\in[n]} \mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n) \\ &= \frac{1}{n^2}\sum_{i,j\in[n]}\sum_{k=0}^{k_n} \mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) = k).\end{aligned} \tag{9.2.4}$$

We make use of *path counting techniques.* A path $\pi = (\pi_0,\dots,\pi_k)$ of length k between vertices i and j is a sequence of vertices connecting $\pi_0 = i$ to $\pi_k = j$. We call a path π *self-avoiding* when it visits every vertex at most once, i.e., $\pi_i \neq \pi_j$ for every $i \neq j$. Let $\mathcal{P}_k(i,j)$ denote the set of k-step self-avoiding paths between vertices i and j. See Figure 9.2.1 for an example of a 12-step self-avoiding path between i and j.

When $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) = k$, there must be path of length k such that all edges (π_l,π_{l+1}) are occupied in $\mathrm{NR}_n(\boldsymbol{w})$, for $l = 0,\dots,k-1$. The probability in $\mathrm{NR}_n(\boldsymbol{w})$ that the edge (π_l,π_{l+1}) is occupied is equal to

$$1 - \mathrm{e}^{-w_{\pi_l} w_{\pi_{l+1}}/\ell_n} \le w_{\pi_l} w_{\pi_{l+1}}/\ell_n. \tag{9.2.5}$$

For $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$, an identical upper bound holds.

We say that π is *occupied* when all edges in π are occupied in $\mathrm{NR}_n(\boldsymbol{w})$. Then, by Boole's inequality,

$$\mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j)=k) \le \mathbb{P}(\exists \pi \in \mathcal{P}_k(i,j)\colon \pi \text{ occupied}) \le \sum_{\pi\in\mathcal{P}_k(i,j)} \mathbb{P}(\pi \text{ occupied}). \quad (9.2.6)$$

For any path $\pi \in \mathcal{P}_k(i,j)$,

$$\begin{aligned}\mathbb{P}(\pi \text{ occupied}) &= \prod_{s=0}^{k-1} \mathbb{P}((\pi_l,\pi_{l+1}) \text{ occupied}) \le \prod_{l=0}^{k-1} w_{\pi_l} w_{\pi_{l+1}}/\ell_n \qquad (9.2.7)\\ &= \frac{w_{\pi_0} w_{\pi_k}}{\ell_n} \prod_{l=1}^{k} w_{\pi_l}^2/\ell_n = \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k} w_{\pi_l}^2/\ell_n.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j)=k) &\le \frac{w_i w_j}{\ell_n} \sum_{\pi\in\mathcal{P}_k(i,j)} \prod_{l=1}^{k} \frac{w_{\pi_l}^2}{\ell_n} \qquad (9.2.8)\\ &= \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k} \Big(\sum_{\pi_l\in[n]} \frac{w_{\pi_l}^2}{\ell_n}\Big) = \frac{w_i w_j}{\ell_n} \nu_n^k,\end{aligned}$$

where ν_n is defined in (9.2.2), so that

$$\mathbb{P}(H_n \le k_n) \le \frac{1}{n^2} \sum_{i,j\in[n]} \sum_{k=0}^{k_n} \frac{w_i w_j}{\ell_n} \nu_n^k = \frac{\ell_n}{n^2} \sum_{k=0}^{k_n} \nu_n^k = \frac{\ell_n}{n^2} \frac{\nu_n^{k_n+1}-1}{\nu_n - 1}. \quad (9.2.9)$$

By (9.2.1), $\limsup_{n\to\infty} \nu_n = \nu \in (1,\infty)$, so that, for n large enough, $\nu_n \ge (\nu-\delta) > 1$, while $\ell_n/n = \mathbb{E}[W_n] \to \mathbb{E}[W] < \infty$. Thus, since $\nu \mapsto (\nu^{k+1}-1)/(\nu-1)$ is increasing for every integer $k \ge 0$,

$$\mathbb{P}(H_n \le k_n) \le O((\nu-\delta)^{k_n}/n) = o(1), \quad (9.2.10)$$

when $\delta = \delta(\varepsilon) > 0$ is chosen such that $(1-\varepsilon)/\log(\nu-\delta) < 1$, and since $k_n = \lceil (1-\varepsilon)\log_\nu n \rceil$. □

The condition (9.2.1) is slightly weaker than Condition 6.4, which is assumed in Theorem 9.3:

Exercise 9.11 (Conditions (9.2.1) and Condition 6.4)**.** *Show that when there is precisely one vertex with weight $w_1 = \sqrt{n}$, whereas $w_i = \lambda > 1$, then (9.2.1) holds, but Condition 6.4 does not. Argue that the upper bound derived in Theorem 9.5 is not sharp, since the vertex 1 can occur at most once in a self-avoiding path.*

Exercise 9.12 (Lower bound on fluctuations)**.** *Adapt the proof of Theorem 9.5 to show that for every ε, we can find a constant $K = K(\varepsilon) > 0$ such that*

$$\mathbb{P}\big(H_n \le \frac{\log n}{\log \nu_n} - K\big) \le \varepsilon. \quad (9.2.11)$$

Conclude that if $\log \nu_n = \log \nu + o(1/\log n)$, then the same statement holds with $\log_\nu n$ replacing $\frac{\log n}{\log \nu_n}$.

We close this section by extending the above result to settings where ν_n is not necessarily bounded:

Corollary 9.6 (Lower bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$ for $\tau = 3$)**.** *Let ν_n be given in* (9.2.2)*. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\Big(H_n \le (1-\varepsilon)\log_{\nu_n} n\Big) = o(1). \tag{9.2.12}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ under the same conditions.

The proof of Corollary 9.6 is left as an exercise:

Exercise 9.13 (Proof Corollary 9.6)**.** *Adapt the proof to Theorem 9.5 to prove Corollary 9.6.*

Exercise 9.14 (Lower bound on typical distances for $\tau = 3$)**.** *Let $w_i = c\sqrt{(n/i)}$, so that $\tau = 3$. Prove that $\nu_n/\log n \to c$. Use Corollary 9.6 to obtain that for any $\varepsilon > 0$,*

$$\mathbb{P}\Big(H_n \le (1-\varepsilon)\frac{\log n}{\log\log n}\Big) = o(1). \tag{9.2.13}$$

Exercise 9.15 (Lower bound on typical distances for $\tau \in (2,3)$)**.** *Let $w_i = c/i^{1/(\tau-1)}$ with $\tau \in (2,3)$. Prove that there exists a constant $c' > 0$ such that $\nu_n \ge c' n^{(3-\tau)/(\tau-1)}$. Show that Corollary 9.6 implies that $H_n \ge (\tau-1)/(\tau-3)$ in this case. How useful is this bound?*

9.2.2 A $\log\log$ lower bound on typical distances in the infinite variance case

In this section, we prove a log log-lower bound on the average distances of $\mathrm{NR}_n(\boldsymbol{w})$ for $\tau \in (2,3)$. The main result we prove is the following theorem:

Theorem 9.7 (Loglog lower bound on average distances in $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Suppose that the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfy Condition 6.4(a) and that there exists a $\tau \in (2,3)$ and c_2 such that, for all $x \ge 1$,*

$$[1-F_n](x) \le c_2 x^{-(\tau-1)}, \tag{9.2.14}$$

Then, for every $\varepsilon > 0$,

$$\mathbb{P}\Big(H_n \le (1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|}\Big) = o(1). \tag{9.2.15}$$

We follow the proof of Theorem 9.5 as closely as possible. The problem with that proof is that, under the condition in (9.2.14), ν_n is too large. Indeed, Exercise 9.15 shows that the lower bound obtained in Corollary 9.6 is a constant, which is not very useful. What fails is that there are too many vertices with too high weight. However, it is quite unlikely that a vertex with a high weight is chosen. Indeed, as argued in (9.1.17), when starting from a vertex with weight w, say, the probability that it is directly connected to a vertex having weight a_n is at most

$$\sum_{j:\, w_j \ge y} \frac{w w_j}{\ell_n} = w[1-F_n^*](y), \tag{9.2.16}$$

which is small when y is too large. On the other hand, the main contribution to ν_n comes from vertices having maximal weight of the order $n^{1/(\tau-1)}$. This problem is resolved by a suitable truncation argument on the weights of the vertices in the occupied paths, which effectively removes these high-weight vertices. Therefore, instead of obtaining $\nu_n = \sum_{s\in[n]} w_s^2/\ell_n$, we obtain a partial sum of this restricted to vertices having a relatively small

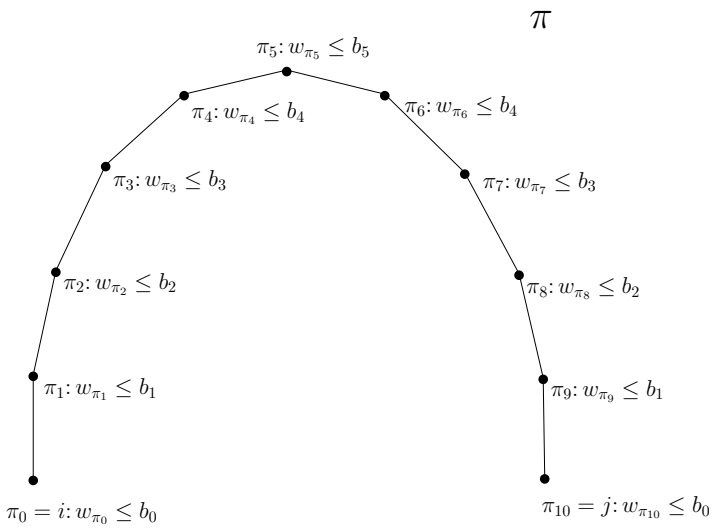

Figure 9.2: A 10-step good path connecting i and j and the upper bounds on the weight of its vertices. The height of a vertex is high for vertices with large weights.

weight. Effectively, this means that we split the space of all paths into *good paths*, i.e., paths that avoid vertices with too large weight, and *bad paths*, which are paths that jump to vertices with too high weight.

We now present the details for this argument. We again start from

$$\mathbb{P}(H_n \le k_n) = \frac{1}{n^2} \sum_{i,j\in[n]} \mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n). \tag{9.2.17}$$

When $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n$, there exists an occupied path $\pi \in \mathcal{P}_k(i,j)$ for some $k \le k_n$.

We fix an *increasing* sequence of numbers $(b_l)_{l=0}^{\infty}$ that serve as truncation values for the weights of vertices along our occupied path. We determine the precise values of $(b_l)_{l=0}^{\infty}$ below. We say that a path $\pi \in \mathcal{P}_k(i,j)$ is *good* when $w_{\pi_l} \le b_l \wedge b_{k-l}$ for every $l = 0, \dots, k$, and bad otherwise. The condition $w_{\pi_l} \le b_l \wedge b_{k-l}$ for every $l = 0, \dots, k$ is equivalent to the statement that $w_{\pi_l} \le b_l$ for $l \le \lceil k/2 \rceil$, while $w_{\pi_l} \le b_{k-l}$ for $\lceil k/2 \rceil < l \le k$. Thus, b_l provides an upper bound on the weight of the l^{th} vertex and the $(k-l)^{\text{th}}$ vertex of the occupied path, ensuring that the weights occurring in the occupied path can not be too large. See Figure 9.2.2 for a description of a good path and the bounds on the weight of its vertices.

Let $\mathcal{GP}_k(i,j)$ be the set of good paths in $\mathcal{P}_k(i,j)$. Let

$$\mathcal{E}_k(i,j) = \{\exists \pi \in \mathcal{GP}_k(i,j) \colon \pi \text{ occupied}\} \tag{9.2.18}$$

denote the event that there exists a good path of length k.

When $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n$, but there does not exist a $k \le k_n$ and a good occupied path $\pi \in \mathcal{GP}_k(i,j)$, then either there exists an $l \le \lceil k/2 \rceil$ such that $w_{\pi_s} \le b_s$ for every $s < l$, while $w_{\pi_l} > b_l$, or there exists an $l \le \lceil k/2 \rceil$ such that $w_{\pi_{k-s}} \le b_{k-s}$ for every $s < l$, while $w_{\pi_{k-l}} > b_{k-l}$. Let $\mathcal{P}_k(i) = \cup_{l\in[n]} \mathcal{P}_k(i,l)$ denote the set of all paths of length k from i, and let

$$\mathcal{BP}_k(i) = \{\pi \in \mathcal{P}_k(i) \colon w_{\pi_l} > b_l, w_{\pi_s} \le b_s \forall s < l\} \tag{9.2.19}$$

denote the set of *bad paths* of length k, i.e., those $\pi \in \mathcal{P}_k(i)$ that are not in $\mathcal{GP}_k(i,\pi_k)$. Let $\mathcal{F}_l(i)$ be the event that there exists a bad path of length l starting from i, i.e.,

$$\mathcal{F}_l(i) = \{\exists \pi \in \mathcal{BP}_l(i) \colon \pi \text{ occupied}\}. \tag{9.2.20}$$

Then, since $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n$ implies that there either is a good path or a bad path,

$$\{\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n\} \subseteq \bigcup_{k \le k_n} \Big(\mathcal{F}_k(i) \cup \mathcal{F}_k(j) \cup \mathcal{E}_k(i,j)\Big), \tag{9.2.21}$$

so that, by Boole's inequality,

$$\mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n) \le \sum_{k=0}^{k_n} \big[\mathbb{P}(\mathcal{F}_k(i)) + \mathbb{P}(\mathcal{F}_k(j)) + \mathbb{P}(\mathcal{E}_k(i,j))\big]. \tag{9.2.22}$$

In order to estimate the probabilities $\mathbb{P}(\mathcal{F}_k(i))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$, we introduce some notation. For $b \ge 0$, let

$$\nu_n(b) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i^2 \mathbb{1}_{\{w_i \le b\}}, \tag{9.2.23}$$

be the restriction of ν_n to vertices with weights at most b, and let

$$F_n^*(x) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i \mathbb{1}_{\{w_i \le x\}} \tag{9.2.24}$$

be the distribution function of W_n^*, the size-biased version of W_n. The following lemma gives bounds on $\mathbb{P}(\mathcal{F}_k(i))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$:

Lemma 9.8 (Truncated path probabilities)**.** *For every $k \ge 1$, $(b_l)_{l \ge 0}$ with $b_l \ge 0$ and $l \mapsto b_l$ non-decreasing,*

$$\mathbb{P}(\mathcal{F}_k(i)) \le w_i [1 - F_n^*](b_k) \prod_{l=1}^{k-1} \nu_n(b_l), \tag{9.2.25}$$

and

$$\mathbb{P}(\mathcal{E}_k(i,j)) \le \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}). \tag{9.2.26}$$

When $b_l = \infty$ for each l, the bound in (9.2.26) equals that obtained in (9.2.8).

Proof. We start by proving (9.2.25). By Boole's inequality,

$$\mathbb{P}(\mathcal{F}_k(i)) = \mathbb{P}(\exists \pi \in \mathcal{BP}_l(i) \colon \pi \text{ occupied}) \le \sum_{\pi \in \mathcal{BP}_l(i)} \mathbb{P}(\pi \text{ occupied}). \tag{9.2.27}$$

By (9.2.7), (9.2.23) and (9.2.24),

$$\begin{aligned} \mathbb{P}(\mathcal{F}_k(i)) &\le \sum_{\pi \in \mathcal{BP}_l(i)} \frac{w_i w_{\pi_k}}{\ell_n} \prod_{l=1}^{k-1} w_{\pi_l}^2/\ell_n \qquad (9.2.28) \\ &\le w_i \sum_{\pi_k \colon w_{\pi_k} \ge b_k} \frac{w_{\pi_k}}{\ell_n} \times \prod_{l=1}^{k-1} \sum_{\pi_l \colon w_{\pi_l} \le b_l} w_{\pi_l}^2/\ell_n \\ &= w_i [1 - F_n^*](b_k) \prod_{l=1}^{k} \nu_n(b_l). \end{aligned}$$

The proof of (9.2.26) is similar. Indeed, by (9.2.7),

$$\mathbb{P}(\mathcal{E}_k(i,j)) \leq \sum_{\pi\in\mathcal{GP}_k(i,j)} \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} w_{\pi_l}^2/\ell_n \leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu(b_l \wedge b_{k-l}). \tag{9.2.29}$$

Now follow the steps in the proof of (9.2.25). □

We continue by proving upper bounds on $[1-F_n^*](x)$ and $\nu_n(b)$:

Lemma 9.9 (Bounds on sums)**.** *Suppose that the weights $\boldsymbol{w}=(w_i)_{i\in[n]}$ satisfy Condition 6.4(a) and that there exist $\tau\in(2,3)$ and c_2 such that, for all $x\geq 1$,*

$$[1-F_n](x) \leq c_2 x^{-(\tau-1)}. \tag{9.2.30}$$

Then, there exists a constant $c_2^*>0$ such that, for all $x\geq 1$,

$$[1-F_n^*](x) \leq c_2^* x^{-(\tau-2)}, \tag{9.2.31}$$

and there exists a $c_\nu>0$ such that for all $b\geq 1$,

$$\nu_n(b) \leq c_\nu b^{3-\tau}. \tag{9.2.32}$$

Proof. For (9.2.32), we bound

$$\nu_n(b) = \mathbb{E}[W_n^* \mathbb{1}_{\{W_n^*\leq b\}}] \leq = \int_0^b [1-F_n^*](x)dx \leq c_2^* \int_0^b x^{-(\tau-2)}dx = \frac{c_2^*}{3-\tau} b^{3-\tau}, \tag{9.2.33}$$

which implies (9.2.32) for $c_\nu = c_2^*/(3-\tau)$. Write out

$$[1-F_n^*](x) = \frac{1}{\ell_n} \sum_{i\in[n]} w_i \mathbb{1}_{\{w_i>x\}} = \frac{\mathbb{E}[W_n \mathbb{1}_{\{W_n>x\}}]}{\mathbb{E}[W_n]}. \tag{9.2.34}$$

Now use the fact that, for any non-negative random variable X,

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X>x)dx. \tag{9.2.35}$$

Applying this to $X = W_n \mathbb{1}_{\{W_n>x\}}$ yields

$$[1-F_n^*](x) = \frac{1}{\mathbb{E}[W_n]} \int_x^\infty [1-F_n(y)]dy. \tag{9.2.36}$$

By (9.2.30),

$$[1-F_n^*](x) \leq \frac{1}{\mathbb{E}[W_n]} \int_x^\infty c_2 y^{-(\tau-1)}dy = \frac{c_2}{(\tau-2)\mathbb{E}[W_n]} x^{-(\tau-2)} \leq c_2^* x^{-(\tau-2)}, \tag{9.2.37}$$

when $c_2^* = c_2/[(\tau-2)(\mathbb{E}[W]-\varepsilon)$ and n is sufficiently large. □

With Lemmas 9.8 and 9.9 at hand, we are ready to choose $(b_l)_{l\geq 0}$ and to complete the proof of Theorem 9.7:

Proof of Theorem 9.7. Take $k_n = 2(1-\varepsilon)\log\log n/|\log(\tau-2)|$. By (9.2.17) and (9.2.21),

$$\mathbb{P}(H_n \leq k_n) \leq \frac{1}{n} + \sum_{k=1}^{k_n} \big[\frac{2}{n} \sum_{i\in[n]} \mathbb{P}(\mathcal{F}_k(i)) + \frac{1}{n^2} \sum_{i,j\in[n]:\, i\neq j} \mathbb{P}(\mathcal{E}_k(i,j))\big],$$

where the contribution $1/n$ is due to $i=j$ for which $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,i)=0$. We use Lemmas 9.8 and 9.9 to provide bounds on $\mathbb{P}(\mathcal{F}_k(i))$, $\mathbb{P}(\mathcal{F}_k(j))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$. These bounds are quite similar.

We first describe how we choose the truncation values $(b_l)_{l=0}^{\infty}$ so that $[1-F_n^*](b_k)$ is so small that $\mathbb{P}(\mathcal{F}_k(i))$ is small, and, for this choice of $(b_l)_{l=0}^{\infty}$, we show that $\mathbb{P}(\mathcal{E}_k(i,j))$ is small. Intuitively, this means that it is quite unlikely that i or j is connected to a vertex at distance k with too high weight, i.e., having weight at least b_k. At the same time, it is also unlikely that there is a path $\pi\in\mathcal{P}_k(i,j)$ whose weights are all small, i.e., for which $w_{\pi_k}\le b_k$ for every $k\le k_n$, because k_n is to small.

By Lemma 9.8, we wish to choose b_k so that $\mathbb{P}(\mathcal{F}_k(i))=[1-F_n^*](b_k)\prod_{l=0}^{k-1}\nu_n(b_l)$ is small. Below (9.1.17), it is argued that $b_k\approx b_{k-1}^{1/(\tau-2)}$. In order to make this probability small, we will take b_k somewhat larger. We now present the details. We take $\delta\in(0,\tau-2)$ sufficiently small and let

$$a=1/(\tau-2-\delta)>1. \tag{9.2.38}$$

Take $b_0=\mathrm{e}^A$ for some constant $A\ge 0$ sufficiently large and define $(b_l)_{l\ge 0}$ recursively by

$$b_l=b_{l-1}^a, \qquad \text{so that} \qquad b_l=b_0^{a^l}=\mathrm{e}^{A(\tau-2-\delta)^{-l}}. \tag{9.2.39}$$

We start from (9.2.22). By Lemma 9.8, we obtain an upper bound on $\mathbb{P}(\mathcal{F}_k(i))$ in terms of factors $\nu_n(b_l)$ and $[1-F_n^*](b_k)$, which are bounded in Lemma 9.9. We start by applying the bound on $\nu_n(b_l)$ to obtain

$$\begin{aligned}\prod_{l=1}^{k-1}\nu_n(b_l)&\le\prod_{l=1}^{k-1}c_\nu b_l^{\tau-3}=c_\nu^k\mathrm{e}^{K(3-\tau)\sum_{l=1}^{k-1}a^l}\\&\le c_\nu^{k-1}\mathrm{e}^{K(3-\tau)a^k/(a-1)}=c_2^*w_ic_\nu^{k-1}b_k^{(3-\tau)/(a-1)}.\end{aligned} \tag{9.2.40}$$

Combining (9.2.40) with the bound on $[1-F_n^*](b_k)$ in Lemma 9.9 yields

$$\mathbb{P}(\mathcal{F}_k(i))\le c_2^*w_ic_\nu^kb_k^{-(\tau-2)+(3-\tau)/(a-1)}. \tag{9.2.41}$$

Since $3-\tau+\delta<1$ when $\tau\in(2,3)$ and $\delta\in(0,\tau-2)$,

$$\begin{aligned}(\tau-2)-(3-\tau)/(a-1)&=(\tau-2)-(3-\tau)(\tau-2-\delta)/(3-\tau+\delta)\\&=\delta/(3-\tau+\delta)>\delta,\end{aligned} \tag{9.2.42}$$

so that

$$\mathbb{P}(\mathcal{F}_k(i))\le c_2^*w_ic_\nu^kb_k^{-\delta}. \tag{9.2.43}$$

As a result, for each $\delta>0$

$$\frac{1}{n}\sum_{i\in[n]}\sum_{k=0}^{k_n}\mathbb{P}(\mathcal{F}_k(i))\le c_2^*\frac{1}{n}\sum_{i\in[n]}w_i\mathbb{1}_{\{w_i>K\}}+\frac{1}{n}\sum_{i\in[n]}c_2^*w_i\sum_{k\ge 1}c_\nu^kb_k^{-\delta}\le\varepsilon, \tag{9.2.44}$$

when we take $A=A(\delta,\varepsilon)$ sufficiently large.

Similarly, since $b_l\ge 1$, by (9.2.40),

$$\mathbb{P}(\mathcal{E}_k(i,j))\le\frac{w_iw_j}{\ell_n}\prod_{l=1}^{k-1}\nu_n(b_l\wedge b_{k-l})\le\frac{w_iw_j}{\ell_n}c_\nu^{k-1}b_{\lceil k/2\rceil}^{2(3-\tau)/(a-1)}, \tag{9.2.45}$$

so that, using further that $l \mapsto b_l$ is increasing,

$$\sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i,j\in[n]} \mathbb{P}(\mathcal{E}_k(i,j)) \le \frac{1}{n^2} \sum_{k=1}^{k_n} \sum_{i,j\in[n]} \frac{w_i w_j}{\ell_n} c_\nu^{k-1} b_{\lceil k/2\rceil}^{2(3-\tau)/(a-1)} \tag{9.2.46}$$

$$\le \frac{\ell_n}{n^2} k_n c_\nu^{k_n-1} b_{\lceil k_n/2\rceil}^{2(3-\tau)/(a-1)}.$$

Recall that $k \le k_n = 2(1-\varepsilon)\log\log n/|\log(\tau-2)|$. Take $\delta = \delta(\varepsilon) > 0$ so small that $(\tau-2-\delta)^{-(k_n+1)/2} \le (\log n)^{1-\varepsilon/4}$. Then,

$$b_{\lceil k_n/2\rceil} \le \mathrm{e}^{A(\tau-2-\delta)^{-(k_n+1)/2}} \le \mathrm{e}^{A(\log n)^{1-\varepsilon/4}}, \tag{9.2.47}$$

and we conclude that

$$\sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i,j\in[n]} \mathbb{P}(\mathcal{E}_k(i,j)) \le \frac{\ell_n}{n^2} k_n c_\nu^{k_n} \exp\big(2A(3-\tau)(\log n)^{1-\varepsilon/4}\big) = o(1), \tag{9.2.48}$$

since $k_n = O(\log\log n)$ and $\ell_n/n^2 = \Theta(1/n)$. This completes the proof of Theorem 9.7. □

Exercise 9.16 (Lower bound on fluctuations*). *Adapt the proof of Theorem 9.7 to show that for every ε, we can find a constant $K = K(\varepsilon) > 0$ such that*

$$\mathbb{P}\big(H_n \le \frac{2\log\log n}{|\log(\tau-2)|} - K\big) \le \varepsilon. \tag{9.2.49}$$

Hint: choose $b_k = Lb_{k-1}^{1/(\tau-2)}$, where the constant $L >$ is chosen sufficiently large.

9.3 Branching process comparisons and the log log upper bound

In this section, we prove the log log upper bound on typical graph distances in the infinite variance case. In Section 9.3.1, we start by comparing the neighborhoods of vertices in $\mathrm{NR}_n(\boldsymbol{w})$ to branching processes, a technique that is crucial in the derivation of all our lower bounds. In Section 9.3.2, we use this comparison to prove the log log upper bound on typical distances when $\tau \in (2,3)$.

9.3.1 Comparison to branching processes

In this section, we describe a beautiful comparison of the neighborhoods of a uniformly chosen vertex in inhomogeneous random graphs, such as the generalized random graph, the Chung-Lu model and the Norros-Reittu model, and a marked branching process. This comparison is particularly pretty when considering the Norros-Reittu model, where there is an explicit stochastic domination result of these neighborhoods are bounded by a so-called *two-stage* branching process with a mixed Poisson offspring distribution. [1]

Stochastic domination of clusters by a branching process. We shall dominate the cluster of a vertex in the Norros-Reittu model by the total progeny of a two-stage branching processes with mixed Poisson offspring. This domination is such that we also control the difference, and makes the heuristic argument below Theorem 9.2 precise.

We now describe the cluster exploration of a uniformly chosen vertex $V \in [n]$. Define the *mark distribution* to be the random variable M with distribution

$$\mathbb{P}(M = m) = w_m/\ell_n, \qquad m \in [n]. \tag{9.3.1}$$

Let $(X_w)_w$ be a collection of independent random variables, where

[1] In [102], the two-stage branching process is called a delayed branching process.

(a) the number of children of the root $X_\varnothing$ has a mixed Poisson distribution with random parameter $w_{M_\varnothing}$, where $M_\varnothing$ is uniformly chosen in $[n]$;

(b) X_w has a mixed Poisson distribution with random parameter w_{M_w}, where $(M_w)_{w\neq\varnothing}$ are i.i.d. random marks with distribution (9.3.1) independently of $M_\varnothing$.

We call $(X_w, M_w)_w$ a *marked mixed-Poisson branching process* (MMPBP).

Clearly, $w_V = w_{M_\varnothing}$ has distribution W_n defined in (6.1.13), while the distribution of w_{M_w} for each w with $|w| \geq 1$ is i.i.d. with distribution w_M given by

$$\mathbb{P}(w_M \leq x) = \sum_{m=1}^{n} \mathbb{1}_{\{w_m \leq x\}} \mathbb{P}(M = m) = \frac{1}{\ell_n} \sum_{m=1}^{n} w_m \mathbb{1}_{\{w_m \leq x\}} = \mathbb{P}(W_n^* \leq x) = F_n^*(x), \tag{9.3.2}$$

where W_n^* is the *size-biased distribution* of W_n and F_n^* is defined in (9.2.24).

When we are only interested in *numbers of individuals*, then we obtain a two-stage branching process since the random variables $(X_w)_w$ are independent, and the random variables $(X_w)_{w\neq\varnothing}$ are i.i.d. However, in the sequel, we make explicit use of the marks $(M_w)_{w\neq\varnothing}$, as the complete information $(X_w, M_w)_w$ gives us a way to retrieve the cluster of the vertex $M_\varnothing$, something that would not be possible on the basis of $(X_w)_w$ only.

In order to define the cluster exploration in $\mathrm{NR}_n(\boldsymbol{w})$, we introduce a *thinning* that guarantees that we only inspect a vertex once. We think of M_w as being the vertex label in $\mathrm{NR}_n(\boldsymbol{w})$ of the tree vertex w, and $X_w = \mathrm{Poi}(w_{M_w})$ as its *potential number of children.* These potential children effectively become children when their marks correspond to vertices in $\mathrm{NR}_n(\boldsymbol{w})$ that have not yet appeared. The thinning ensures this. To describe the thinning, we set $\varnothing$ unthinned, and, for w with $w \neq \varnothing$, we thin w when either (i) one of the tree vertices on the (unique) path between the root $\varnothing$ and w has been thinned, or (ii) when $M_w = M_{w'}$ for some unthinned vertex $w' < w$. We now make the connection between the thinned marked mixed Poisson branching process and the cluster exploration precise:

Proposition 9.10 (Clusters as thinned marked branching processes)**.** *The cluster of a uniformly chosen vertex $\mathcal{C}(V)$ is equal in distribution to $\{M_w \colon w \text{ unthinned}\}$, the marks of unthinned vertices encountered in the marked mixed Poisson branching process up to the end of the exploration. Similarly, the set of vertices at graph distance k from V has the same distribution as*

$$\Big(\{M_w \colon w \text{ unthinned}, |w| = k\}\Big)_{k\geq 0}. \tag{9.3.3}$$

Proof. We prove the two statements simultaneously. By construction, the distribution of V is the same as that of $M_\varnothing$, the mark of the root of the marked mixed Poisson branching process. We continue by proving that the direct neighbors of the root $\varnothing$ agree in both constructions. In $\mathrm{NR}_n(\boldsymbol{w})$, the direct neighbors are equal to $\{j \in [n] \setminus \{l\} \colon I_{lj} = 1\}$, where $(I_{lj})_{j\in[n]\setminus\{l\}}$ are independent $\mathrm{Be}(p_{lj})$ random variables with $p_{lj} = 1 - \mathrm{e}^{-w_l w_k/\ell_n}$.

We now prove that the same is true for the marked mixed Poisson branching process. Conditionally on $M_\varnothing = l$, the root has a $\mathrm{Poi}(w_l)$ number of children, where these $\mathrm{Poi}(w_l)$ offspring receive i.i.d. marks. We make use of the following fundamental property of the Poisson distribution:

Lemma 9.11 (A Poisson number of multinomial trials)**.** *Let X have a Poisson distribution with parameter λ. Perform X multinomial trials, where the i^{th} outcome appears with probability p_i for some probabilities $(p_i)_{i=1}^k$. Let $(X_i)_{i=1}^k$, where X_i denotes the total number of outcomes i. Then $(X_i)_{i=1}^k$ is a sequence of* independent *Poisson random variables with parameters $(\lambda p_i)_{i=1}^k$.*

Proof. Let $(x_i)_{i=1}^k$ denote a sequence of non-negative integers, denote $x = \sum_{i=1}^k x_i$ and compute

$$\begin{aligned}\mathbb{P}((X_i)_{i=1}^k = (x_i)_{i=1}^k) &= \mathbb{P}(X = x)\mathbb{P}((X_i)_{i=1}^k = (x_i)_{i=1}^k \mid X = x) \\ &= \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}\binom{x}{x_1, x_2, \ldots, x_k} p_1^{x_1}\cdots p_k^{x_k} = \prod_{i=1}^k \mathrm{e}^{-\lambda x_i}\frac{\lambda^{x_i}}{(x_i)!}.\end{aligned} \tag{9.3.4}$$

□

By Lemma 9.11, the random vector $(X_{\varnothing,j})_{j\in[n]}$, where $X_{\varnothing,j}$ is the number of offspring of the root that receive mark j, is a vector of *independent* Poisson random variables with parameters $w_l w_j/\ell_n$. Due to the thinning, a mark occurs precisely when $X_{\varnothing,j} \geq 1$. Therefore, the mark j occurs, independently for all $j \in [n]$, with probability $1 - \mathrm{e}^{-w_l w_j/\ell_n} = p_{jk}^{(\mathrm{NR})}$. This proves that the set of marks of children of the root in the MMPBD has the same distribution as the set of neighbors of the chosen vertex in $\mathrm{NR}_n(\boldsymbol{w})$.

Next, we look at the number of new elements of $\mathcal{C}(V)$ neighboring the vertex which has received word w. First, condition on $M_w = l$, and assume that w is not thinned. Conditionally on $M_w = l$, the number of children of w in the MMPBP has distribution $\mathrm{Poi}(w_l)$. Each of these $\mathrm{Poi}(w_l)$ children receives an i.i.d. mark. Let $X_{w,j}$ denote the number of children of w that receive mark j.

By Lemma 9.11, $(X_{w,j})_{j\in[n]}$ is again a vector of independent Poisson random variables with parameters $w_l w_j/\ell_n$. Due to the thinning, a mark appears within the offspring of individual w precisely when $X_{w,j} \geq 1$, and these events are independent. In particular, for each j that has not appeared as the mark of an unthinned vertex, the probability that it occurs equals $1 - \mathrm{e}^{-w_j w_k/\ell_n} = p_{jk}^{(\mathrm{NR})}$, as required. □

The law of the branching process. For given weights $(w_i)_{i\in[n]}$, we now describe the distribution of the marked mixed Poisson branching process (MMPBP). Since the marks are mutually *independent*, the marked Poisson process is a branching process if we ignore the information about the marks. The offspring distribution $p^{(n)}$ of Z_1, i.e., the first generation of $(Z_l)_{l\geq 0}$, is given by

$$p_k^{(n)} = \mathbb{P}\big(\mathrm{Poi}(w_V) = k\big) = \sum_{i\in[n]} \mathbb{P}\big(\mathrm{Poi}(w_i) = k \big| V = i\big)\mathbb{P}(V = i) = \frac{1}{n}\sum_{i\in[n]} \mathrm{e}^{-w_i}\frac{w_i^k}{k!}, \tag{9.3.5}$$

for $k \geq 0$, so that $p^{(n)}$ is a mixed Poisson distribution with mixing distribution W_n, i.e.,

$$p_k^{(n)} = \mathbb{E}\big[\mathrm{e}^{-W_n}\frac{W_n^k}{k!}\big]. \tag{9.3.6}$$

Recall that individuals in the second and further generations have a random mark distributed as an independent copy of M given by (9.3.1). Hence, if we denote the offspring distribution of the second and further generations by $g^{(n)}$, then we obtain, for all $k \geq 0$,

$$\begin{aligned}g_k^{(n)} &= \mathbb{P}\big(\mathrm{Poi}(w_M) = k\big) = \sum_{i\in[n]} \mathbb{P}\big(\mathrm{Poi}(w_i) = k \big| M = i\big)\mathbb{P}(M = i) \\ &= \frac{1}{\ell_n}\sum_{i\in[n]} \mathrm{e}^{-w_i}\frac{w_i^{k+1}}{k!}.\end{aligned} \tag{9.3.7}$$

We recognize the offspring distribution in (9.3.7) as a mixed Poisson distribution, where the mixing random variable now has distribution W_n^*:

$$g_k^{(n)} = \mathbb{E}\big[\mathrm{e}^{-W_n^*}\frac{(W_n^*)^k}{k!}\big]. \tag{9.3.8}$$

Together, (9.3.5) and (9.3.7) identify the distribution $(Z_l)_{l\geq 0}$ as a two-stage branching process, where (1) the first generation has a mixed Poisson distribution with mixing random variable w_V, where $V \in [n]$ is chosen uniformly at random, and where (2) the offspring distribution of the second and further generations has a mixed Poisson distribution with mixing random variable w_M, where $M \in [n]$ has distribution given by (9.3.1). This yields a stochastic upper bound on the neighborhoods of a uniformly chosen $V \in [n]$ in the Norros-Reittu model $\mathrm{NR}_n(\boldsymbol{w})$. In the case where $\boldsymbol{w}$ is *constant*, the above gives an interesting direct upper bound of $|\mathcal{C}(V)|$ in terms of a Poisson branching process:

Exercise 9.17 (Erdős-Rényi random graph)**.** *Prove that* $\mathrm{NR}_n(\boldsymbol{w}) = \mathrm{ER}_n(\lambda/n)$ *when* $\boldsymbol{w}$ *is constant with* $w_i = -n\log(1-\lambda/n)$.

Exercise 9.18 (Erdős-Rényi random graph (Cont.))**.** *Show that Exercise 9.17 together with Proposition 9.10 imply that* $|\mathcal{C}(V)| \preceq T^*$, *where* T^* *is the total progeny of a Poisson branching process with mean* $-n\log(1-\lambda/n)$ *offspring.*

The limiting two-stage branching process. In the previous section, we have described the neighborhood of a uniform in terms of a (thinned) two-stage branching process. This leads to the definitions of the probability mass functions $p^{(n)}$ and $g^{(n)}$ in (9.3.5) and (9.3.7). We now study what happens when $n \to \infty$.

Recall the two-stage branching processes with a mixed Poisson offspring introduced in Section 9.1.1. We proceed by proving that (9.1.2) and (9.1.3) are the limits of (9.3.5) and (9.3.7):

Lemma 9.12 (Weak convergence of MMPBP)**.** *Let the weights* $\boldsymbol{w} = (w_i)_{i\in[n]}$ *satisfy Condition 6.4(a). Then, for every* $k \geq 0$,

$$\lim_{n\to\infty} p_k^{(n)} = p_k. \tag{9.3.9}$$

When, instead, the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ *satisfy Condition 6.4(a)-(b) with* $\mathbb{E}[W] < \infty$, *where* W *has distribution function* F, *then, for all* $k \geq 0$,

$$\lim_{n\to\infty} g_k^{(n)} = g_k. \tag{9.3.10}$$

Proof. By Condition 6.4(a), $W_n \xrightarrow{d} W$, where W has distribution function F. By (9.3.6), and the fact that, for each $k \geq 0$, the function $w \mapsto \mathrm{e}^{-w}w^k/k!$ is a bounded continuous function, we obtain from the continuous mapping theorem that $p_k^{(n)} \to p_k$.

For (9.3.10), we note that

$$g_k^{(n)} = \frac{(k+1)p_{k+1}^{(n)}}{\mathbb{E}[W_n]}. \tag{9.3.11}$$

By Condition 6.4(b), $\mathbb{E}[W_n] \to \mathbb{E}[W]$. By (9.3.9), also $p_{k+1}^{(n)} \to p_{k+1}$, so that, indeed, $g_k^{(n)} \to \frac{(k+1)p_{k+1}}{\mathbb{E}[W]} = g_k$. □

Branching process notation. We now define the generation sizes of our branching process. We define $Z_0 = 1$, and, for $m \geq 1$,

$$Z_m = \#\{w\colon |w| = m\}. \tag{9.3.12}$$

Since each individual in generation m has precisely one parent in generation $m-1$, we can alternatively write

$$Z_m = \sum_{w\colon |w|=m-1} X_w. \tag{9.3.13}$$

We see that $(Z_m)_{m\geq 0}$ has the same distribution as the generation sizes of a so-called *two-stage mixed Poisson branching process*, in which the root has $Z_1 \sim \mathrm{Poi}(w_V)$ children, where V is chosen uniformly in $[n]$, and all other individuals have offspring distribution given by $\mathrm{Poi}(w_M)$. Throughout the remainder of this chapter, we use $(Z_l)_{l\geq 0}$ denote the n-dependent two-stage marked mixed-Poisson branching process (MMPBP) with offspring distributions $p^{(n)}$ in (9.3.6) for the root, and offspring distributions $g^{(n)}$ in (9.3.7) for all later generations. We also let $(\mathcal{Z}_l)_{l\geq 0}$ denote the limiting two-stage MMPBP with offspring distributions p in (9.1.2) for the root, and offspring distributions g in (9.1.3) for all later generations. We write $(\widetilde{Z}_m)_{m\geq 0}$ for the *thinned* MMPBP by $\widetilde{Z}_0 = 1$ and

$$\widetilde{Z}_m = \sum_{w\colon |w|=m} (1 - \mathbb{1}_{\{w \text{ thinned}\}}). \tag{9.3.14}$$

Thus, $\widetilde{Z}_m$ denotes the number of unthinned words w of generation $|w| = m$. Then, obviously, by the coupling between the two processes and the fact that the thinning only removes vertices,

$$\widetilde{Z}_m \leq Z_m \qquad \text{a.s.} \tag{9.3.15}$$

We are interested in *typical distances*, which is the graph distance between the vertices V_1 and V_2, where $V_1, V_2 \in [n]$ are chosen uniformly at random. Therefore, we sometimes use a superscript (1) or (2), and write, e.g., $\widetilde{Z}_m^{(1)}$, when we consider the constructions from vertex V_1 and V_2, respectively. A straightforward adaptation of Proposition 9.10 implies that, on the event that $H_n \geq 2m$,

$$\widetilde{Z}_m^{(1)} \leq Z_m^{(1)}, \quad \text{and} \quad \widetilde{Z}_m^{(2)} \leq Z_m^{(2)} \qquad \text{a.s.} \tag{9.3.16}$$

Indeed, when $H_n \geq 2m$, and conditionally on $\mathcal{N}_m(V_1)$, in the graph exploration from V_2, we must avoid the vertices in $\mathcal{N}_m(V_1)$, as well as thin the vertices in $(\mathcal{N}_l(V_2))_{l=0}^m$. This explains (9.3.16).

Corollary 9.13 (Weak convergence of neighborhoods of two vertices)**.** *Let the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfy Condition 6.4(a)-(b). Then, for each $m \geq 1$,*

$$\mathbb{P}(H_n \leq 2m) = o(1), \tag{9.3.17}$$

and

$$(\widetilde{Z}_m^{(1)}, \widetilde{Z}_m^{(2)}) \xrightarrow{d} (\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)}), \tag{9.3.18}$$

where $(\mathcal{Z}_l^{(1)}, \mathcal{Z}_l^{(2)})_{l\geq 0}$ are two independent limiting two-stage marked mixed-Poisson branching processes.

Proof. We start by proving that $\widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$. By Lemma 9.12, $Z_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$, and, by (9.3.16), $\widetilde{Z}_m^{(1)} \leq Z_m^{(1)}$. Further, by Proposition 9.10, $\widetilde{Z}_m^{(1)} = Z_m^{(1)}$ unless one of the vertices in $Z_m^{(1)}$ has been thinned. Since $Z_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$, which is a bounded random variable, **whp** we draw at most $Z_m^{(1)} \leq a_n$ marks for any $a_n \to \infty$. The probability that within these a_n draws, the mark i appears *twice*, thus causing the second individual to be thinned, is at most

$$\Big(\frac{w_i}{\ell_n}\Big)^2 \binom{a_n}{2} \leq \frac{(a_n w_i)^2}{\ell_n^2}, \tag{9.3.19}$$

since there at most $\binom{a_n}{2}$ pairs of times that mark i can be drawn. By Boole's inequality, the probability that an individual from the first a_n individuals is thinned is therefore at most

$$\frac{a_n^2}{\ell_n^2} \sum_{i\in[n]} w_i^2. \tag{9.3.20}$$

By Exercise 6.3, Condition 6.4(a) and (b) imply that $\max_{i\in[n]} w_i = o(n)$, so that the probability that an individual from the first a_n individuals is thinned is at most $o(1)a_n^2$. Choosing $a_n \to \infty$ sufficiently slowly proves that $\mathbb{P}(\widetilde{Z}_m^{(1)} \neq Z_m^{(1)}) = o(1)$, completing the proof that $\widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$.

Further, let $\mathcal{N}_{\le m}(V_1)$ denote the vertices in $\mathrm{NR}_n(\boldsymbol{w})$ that are at distance at most m from V_1. Similarly as in the proof of $\widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$, we can also show that $\mathcal{N}_{\le m}(V_1)$ converges in distribution to the set of marks in the MMPBP up to generation m.

We continue show that $H_n > 2m$ occurs **whp**. We know that $\widetilde{Z}_{2m}^{(1)} \xrightarrow{d} \mathcal{Z}_{2m}^{(1)}$, which is a finite random variable. Now, $H_n \le 2m$ occurs precisely when V_2 is one of the $\mathcal{Z}_{2m}^{(1)}$ individuals, which, conditionally on $\mathcal{Z}_{2m}^{(1)}$, has probability $\mathcal{Z}_{2m}^{(1)}/n = o_{\mathbb{P}}(1)$. Therefore, $\mathbb{P}(H_n \le 2m) = o(1)$, as required. We condition on $\mathcal{N}_{\le m}(V_1)$ and on $H_n > 2m$. Then, the cluster of V_2 up to graph distance m is again an inhomogeneous random graph on $[n] \setminus \mathcal{N}_{\le m}(V_1)$ with edge probabilities given by $p_{ij} = 1 - \mathrm{e}^{-w_i w_j/\ell_n}$. Since $|\mathcal{N}_{\le m}(V_1)| = \widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(2)}$, which is a finite random variable, it follows that, conditionally on $\mathcal{N}_{\le m}(V_1)$, the number of vertices in $[n]\setminus\mathcal{N}_{\le m}(V_1)$ at distance m from V_2 converges in distribution to $\mathcal{Z}_m^{(2)}$. Since this is true *conditionally on* $\mathcal{N}_{\le m}(V_1)$, and the limit is independent of $\mathcal{N}_{\le m}(V_1)$, also $\mathcal{Z}_m^{(1)}$ and $\mathcal{Z}_m^{(2)}$) are independent. This completes the proof. □

Exercise 9.19 (The diameter tends to infinity)**.** *Let the weights satisfy Condition 6.4(a)-(b). Use* (10.4.22) *to show that the diameter of* $\mathrm{NR}_n(\boldsymbol{w})$ *tends to infinity in probability.*

9.3.2 A $\log\log$ upper bound for $\tau \in (2,3)$

In this section, we prove a $\log\log n$ upper bound on the typical distance H_n using the comparison to branching processes obtained in the previous section. Throughout this section, we assume that there exist $\tau \in (2,3)$, $\alpha > 1/2$ and c_1 such that, uniformly in n and $x \le n^\alpha$,

$$[1 - F_n](x) \ge c_1 x^{-(\tau-1)}. \tag{9.3.21}$$

The bound in (9.3.21) corresponds to the lower bound in (9.1.14). The main result in this section is the following theorem:

Theorem 9.14 (A $\log\log$ upper bound on typical distance for $\tau \in (2,3)$)**.** *Suppose that empirical distribution function F_n of the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfies Condition 6.4(a)-(b) and* (9.3.21)*. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(H_n \le \frac{2(1+\varepsilon)\log\log n}{|\log(\tau-2)|} \mid H_n < \infty\Big) = 1. \tag{9.3.22}$$

Proof. The proof Theorem 9.14 is organized as follows. We start by showing that the giant-weight vertices, i.e., the vertices with extremely high weight, are all connected to one another. Thus, the giant-weight vertices form a complete graph. This is often referred to as a *clique* in the random graph community. In the second step, we show that connections from a vertex to the set of giant weight vertices occur at distance at most $\frac{(1+\varepsilon)\log\log n}{|\log(\tau-2)|}$. The latter is only true when the vertex is in the giant component, a fact we need to carefully into account. In the final step, we complete the proof of Theorem 9.14. We now start by defining the set of giant-weight vertices.

The giant-weight vertices form a clique. Recall the definition of $\alpha > 1/2$ in (9.3.21). Let

$$\mathrm{Giant}_n = \{i\colon w_i \ge n^\alpha\} \tag{9.3.23}$$

denote the set of vertices with giant weights. Let $A \subseteq [n]$. We say that A forms a *clique* when the edges a_1a_2 are occupied for all $a_1, a_2 \in A$. We continue by proving that, **whp**, Giant_n forms a clique:

Lemma 9.15 (High-weight vertices form clique)**.** *Under the conditions of Theorem 9.14,*

$$\mathbb{P}(\mathrm{Giant}_n \textit{ does not form clique}) \leq n^2 \mathrm{e}^{-n^{2\alpha}/\ell_n} \tag{9.3.24}$$

Proof. Let $g_1, g_2 \in \mathrm{Giant}_n$, so that $w_{g_1}, w_{g_2} \geq n^\alpha$. There are at most $|\mathrm{Giant}_n|^2 \leq n^2$ pairs of vertices in Giant_n, so that

$$\mathbb{P}(\mathrm{Giant}_n \text{ does not form clique}) \leq n^2 \max_{g_1,g_2 \in \mathrm{Giant}_n} \mathbb{P}(g_1g_2 \text{ vacant}). \tag{9.3.25}$$

The edge g_1g_2 is vacant with probability

$$\mathbb{P}(g_1g_2 \text{ vacant}) = \mathrm{e}^{-w_{g_1}w_{g_2}/\ell_n} \leq \mathrm{e}^{-n^{2\alpha}/\ell_n}, \tag{9.3.26}$$

since $w_g \geq n^\alpha$ for every $g \in \mathrm{Giant}_n$. Multiplying out gives the result. □

Connections to Giant_n occur at $\log\log n$ distances. We next show that vertices that survive up to distance m have a high probability of connecting to Giant_n using a path of at most $(1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}$ edges:

Proposition 9.16 (Connecting to Giant_n)**.** *Let $i \in [n]$ be such that $w_i > 1$. Under the conditions of Theorem 9.14, there exist $c, c_1^* > 0$ and $\eta > 0$ such that*

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) \geq (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}\Big) \leq c\mathrm{e}^{-c_1^* w_i^\eta}. \tag{9.3.27}$$

Consequently, with $\mathcal{N}_m(i)$ denoting the vertices at graph distance m from i, $\mathcal{N}_{\leq m}(i)$ the vertices at graph distance at most m from i, and $\mathcal{W}_m(i) = \sum_{k\in\mathcal{N}_m(i)} w_k$ denoting the weight of vertices in $\mathcal{N}_m(i)$,

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(\mathcal{N}_m(i), \mathrm{Giant}_n) \geq (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|} \mid \mathcal{N}_{\leq m}(i)\Big) \leq c\mathrm{e}^{-c_1^* \mathcal{W}_m(i)^\eta}. \tag{9.3.28}$$

Proof. We start by proving (9.3.27). The bound in (9.3.27) is trivial unless w_i is large. We let $x_0 = i$, and define, recursively,

$$x_\ell = \max\{j \in [n] \colon x_{\ell-1}j \text{ occupied}\}. \tag{9.3.29}$$

Thus, x_ℓ is maximal weight neighbor of $x_{\ell-1}$. We stop the above recursion when $w_{x_\ell} \geq n^\alpha$, since then $x_\ell \in \mathrm{Giant}_n$. Recall the heuristic below (9.1.17), which shows that a vertex with weight w is **whp** connected to a vertex with weight $w^{1/(\tau-2)}$. We now make this precise.

We take $a = 1/(\tau-2+\delta)$, where we choose $\delta > 0$ so small that $a > 1$. By (9.2.24),

$$\mathbb{P}(w_{x_{\ell+1}} < w_{x_\ell}^a \mid (x_s)_{s\leq\ell}) = \mathrm{e}^{-w_{x_\ell}\sum_{l\colon w_l \geq w_{x_\ell}^a} w_l/\ell_n} = \mathrm{e}^{-w_{x_\ell}[1-F_n^*](w_{x_\ell}^a)}. \tag{9.3.30}$$

We split the argument depending on whether $w_{x_\ell}^a \leq n^\alpha$ or not. Firstly, when $w_{x_\ell}^a \leq n^\alpha$, by (9.3.21) and uniformly for $x \leq n^\alpha$,

$$[1-F_n^*](x) \geq \frac{xn}{\ell_n}[1-F_n](x) \geq c_1^* x^{-(\tau-2)}, \tag{9.3.31}$$

where, for n large enough, we can take $c_1^* = c_1/(2\mathbb{E}[W])$. Therefore,

$$\mathbb{P}(w_{x_{\ell+1}} < w_{x_\ell}^a \mid (x_s)_{s\le\ell}) \le \mathrm{e}^{-c_1^* w_{x_\ell}^{1-(\tau-2)a}} \le \mathrm{e}^{-c_1^* w_{x_\ell}^{\delta}}, \tag{9.3.32}$$

since $a = 1/(\tau-2+\delta) > 1$ so that $1-(\tau-2)a = a\delta > \delta$.

Secondly, when $w_{x_\ell}^a > n^\alpha$, but $w_{x_\ell} < n^\alpha$, we can use (9.3.31) for $x = n^\alpha$ to obtain

$$\mathbb{P}(w_{x_{\ell+1}} < n^\alpha \mid (x_s)_{s\le\ell}) \le \mathrm{e}^{-c_1^* w_{x_\ell} n^{\alpha(\tau-2)}} \le \mathrm{e}^{-c_1^* n^{\alpha[1-(\tau-2)]/a}} \le \mathrm{e}^{-c_1^* n^{\alpha\delta/a}}. \tag{9.3.33}$$

Therefore, in both cases, and with $\eta = \alpha\delta/a$,

$$\mathbb{P}(w_{x_{\ell+1}} < (n^\alpha \wedge w_{x_\ell}^a) \mid (x_s)_{s\le\ell})\mathrm{e}^{-c_1^* w_{x_\ell}^\eta}. \tag{9.3.34}$$

As a result, when x_ℓ is such that w_{x_ℓ} is quite large, **whp**, $w_{x_{\ell+1}} \ge w_{x_\ell}$. This produces, **whp**, a short path to Giant_n. We now investigate the properties of this path.

Let the recursion stop at some integer time k. The key observation is that when this occurs, we must have that $w_{x_{\ell+1}} > w_{x_\ell}^a$ for each $\ell \le k-1$ where k is such that $w_{x_{k-1}} \in [n^{\alpha/a}, n^\alpha]$, and at the same time $w_{x_k} \ge n^\alpha$. Then, the following holds:

(1) $w_{x_\ell} \ge w_{x_0}^{a^\ell} = w_i^{a^\ell}$ for every $\ell \le k-1$,

(2) $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) \le k$.

By (1), $w_{x_{k-1}} \ge w_i^{a^{k-1}}$, and $w_{x_{k-1}} \in [n^{\alpha/a}, n^\alpha]$. Therefore, $w_i^{a^{k-1}} \le n^\alpha$, which, in turn, implies that

$$a^{k-1} \le \alpha \log n, \qquad \text{or} \qquad k-1 \le (\log\log n + \log\alpha)(\log a). \tag{9.3.35}$$

Let $k_n = (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}$. By (1) and (2), when $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) > k_n$ occurs, then there must exist an $\ell \le k_n$ such that $w_{x_{\ell+1}} \le n^\alpha \wedge w_{x_\ell}^a$. We conclude that

$$\begin{aligned}
\mathbb{P}\Big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) \ge k_n\Big) &\le \sum_{\ell=0}^{k_n} \mathbb{P}(w_{x_{\ell+1}} \le w_{x_\ell}^a) \\
&\le \sum_{\ell=0}^{k_n} \mathbb{E}[\mathbb{P}(w_{x_{\ell+1}} \le w_{x_\ell}^a \mid (x_s)_{s\le\ell})] \\
&\le \sum_{\ell=0}^{k_n} \mathbb{E}[\mathrm{e}^{-c_1^* w_{x_\ell}^\eta}] \le \sum_{\ell=0}^{k_n} \mathrm{e}^{-c_1^* w_i^{\delta a^\ell}} \le c\mathrm{e}^{-c_1^* w_i^\eta}.
\end{aligned} \tag{9.3.36}$$

The proof of (9.3.28) is similar, by conditioning on $\mathcal{N}_m(i)$ and by noting that we can interpret $\mathcal{N}_m(i)$ as a single vertex having weight $\mathcal{W}_m(i) = \sum_{k\in\mathcal{N}_m(i)} w_k$. □

Completion of the proof of Theorem 9.14. To prove the upper bound in Theorem 9.14, for $\varepsilon \in (0,1)$, we take

$$k_n = (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}, \tag{9.3.37}$$

so that it suffices to show, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(H_n \le 2k_n \mid H_n < \infty) = 1. \tag{9.3.38}$$

Since

$$\mathbb{P}(H_n \le 2k_n \mid H_n < \infty) = \frac{\mathbb{P}(H_n \le 2k_n)}{\mathbb{P}(H_n < \infty)}, \tag{9.3.39}$$

this follows from the bounds

$$\liminf_{n\to\infty} \mathbb{P}(H_n < \infty) \le \zeta^2, \tag{9.3.40}$$

$$\limsup_{n\to\infty} \mathbb{P}(H_n \le 2k_n) \ge \zeta^2, \tag{9.3.41}$$

with $\zeta > 0$ the survival probability of the underlying branching process approximation to the neighborhood shells of $\mathrm{NR}_n(\boldsymbol{w})$. For (9.3.40), we split, for some $m \ge 1$,

$$\mathbb{P}(H_n < \infty) \le \mathbb{P}(H_n \le 2m) + \mathbb{P}(\widetilde{Z}_m^{(1)} > 0, \widetilde{Z}_m^{(2)} > 0, H_n > 2m). \tag{9.3.42}$$

By (10.4.22) in Corollary 9.13, $\mathbb{P}(H_n \le 2m) = o(1)$, and, by (9.3.18) in Corollary 9.13,

$$\lim_{n\to\infty} \mathbb{P}(\widetilde{Z}_m^{(1)} > 0, \widetilde{Z}_m^{(2)} > 0, H_n > 2m) = \mathbb{P}(\mathcal{Z}_m > 0)^2, \tag{9.3.43}$$

which converges to ζ^2 when $m \to \infty$. This proves (9.3.40).

To prove (9.3.41), we fix $m \ge 1$ and write

$$\begin{aligned}\mathbb{P}(2m < H_n \le 2k_n) &\ge \mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_i, \mathrm{Giant}_n) \le k_n, i = 1, 2, H_n > 2m\big) \\ &\ge \mathbb{P}(\widetilde{Z}_m^{(1)} > 0, \widetilde{Z}_m^{(2)} > 0, H_n > 2m) \\ &\quad - 2\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) < k_n, \widetilde{Z}_m^{(1)} > 0\big).\end{aligned} \tag{9.3.44}$$

By (9.3.43), the first term converges to ζ_m^2, which in turn converges to ζ^2 when $m \to \infty$.

For the second term, we condition on $\mathcal{N}_{\le m}(V_1), \mathcal{N}_{\le m}(V_2)$, and use that $\widetilde{Z}_m^{(1)}$ is measurable w.r.t. $\mathcal{N}_{\le m}(V_1)$ to obtain

$$\begin{aligned}&\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) < k_n, \widetilde{Z}_m^{(1)} > 0\big) \\ &\quad = \mathbb{E}\Big[\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) > k_n \mid \mathcal{N}_{\le m}(V_1)\big) \mathbb{1}_{\{\widetilde{Z}_m^{(1)} > 0\}}\Big].\end{aligned} \tag{9.3.45}$$

By Proposition 9.16,

$$\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) > k_n \mid \mathcal{N}_{\le m}(V_1)\big) \le c\mathrm{e}^{-c_1^* \mathcal{W}_m(V_1)^\eta}. \tag{9.3.46}$$

By Lemma 9.12 and the fact that $(w_k)_{k\in\mathcal{N}_m(V_1)}$ are i.i.d. copies of the random variable W_n, conditionally on being distinct and different from $(w_l)_{l\in\mathcal{N}_{\le m-1}(V_1)}$,

$$\mathcal{W}_m(V_1) \xrightarrow{d} \sum_{i=1}^{\mathcal{Z}_m^{(1)}} W(i), \tag{9.3.47}$$

where $(W(i))_{i\ge1}$ are i.i.d. random variables with distribution function F. Therefore,

$$\mathcal{W}_m(V_1) \xrightarrow{\mathbb{P}} \infty \tag{9.3.48}$$

when first $n \to \infty$ followed by $m \to \infty$, and we use that $\mathcal{Z}_m^{(1)} \xrightarrow{\mathbb{P}} \infty$ since $\mathcal{Z}_m^{(1)} > 0$. As a result,

$$\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) > k_n \mid \mathcal{N}_{\le m}(V_1)\big) \mathbb{1}_{\{\widetilde{Z}_m^{(1)} > 0\}} \xrightarrow{\mathbb{P}} 0, \tag{9.3.49}$$

which by Lebesgues Dominated Convergence Theorem (Theorem A.9) implies that

$$\mathbb{E}\Big[\mathrm{e}^{-c_1^* \mathcal{W}_m(V_1)^\eta} \mathbb{1}_{\{\widetilde{Z}_m^{(1)} > 0\}}\Big] \to 0, \tag{9.3.50}$$

when first $n \to \infty$ followed by $m \to \infty$. This proves (9.3.41), and thus completes the proof of the upper bound in Theorem 9.4. □

9.4 Path counting and the log upper bound for $\tau > 3$

9.4.1 Path counting techniques

In this section, we study path counting techniques in the context of inhomogeneous random graphs (IRGs). We generalize the setting somewhat, and consider an IRG on the vertices $\mathcal{I}$ with edge probabilities $p_{ij} = u_i u_j$, for some weights $(u_i)_{i\in\mathcal{I}}$. We obtain $\mathrm{CL}_n(\boldsymbol{w})$ by taking $u_i = w_i/\sqrt{\ell_n}$ and $\mathcal{I} = [n]$. Since the $\mathrm{NR}_n(\boldsymbol{w})$ random graph is closely related to $\mathrm{CL}_n(\boldsymbol{w})$, this suffices for our purposes.

For $a, b \in \mathcal{I}$ and $k \geq 1$, let

$$N_k(a,b) = \#\{\pi \in \mathcal{P}_k(a,b)\colon \pi \text{ occupied}\} \tag{9.4.1}$$

denote the number of paths of length k between the vertices a and b. Let

$$n_k(a,b) = \mathbb{E}[N_k(a,b)] \tag{9.4.2}$$

denote the expected number of occupied paths of length k connecting a and b. Define

$$\bar{n}_k(a,b) = u_a u_b \Big(\sum_{i\in\mathcal{I}\setminus\{a,b\}} u_i^2\Big)^{k-1}, \qquad \underline{n}_k(a,b) = u_a u_b \Big(\sum_{i\in\mathcal{I}_{a,b,k}} u_i^2\Big)^{k-1}, \tag{9.4.3}$$

where $\mathcal{I}_{a,b,k}$ is the subset of $\mathcal{I}$ in which a and b, as well as the $k+2$ indexes with highest weights have been removed. In Section 9.2, we have implicitly proved an upper bound on $\mathbb{E}[N_k(a,b)]$ of the form

$$n_k(a,b) \leq \bar{n}_k(a,b). \tag{9.4.4}$$

Exercise 9.20 (Upper bound on the expected number of paths)**.** *Prove* (9.4.4) *for an inhomogeneous random graph with vertex set $\mathcal{I}$ and with edge probabilities $p_{ij} = u_i u_j$ for every $i, j \in \mathcal{I}$.*

Let

$$\nu_{\mathcal{I}} = \sum_{i\in\mathcal{I}} u_i^2, \qquad \gamma_{\mathcal{I}} = \sum_{i\in\mathcal{I}} u_i^3 \tag{9.4.5}$$

denote the sums of squares and third powers of $(u_i)_{i\in\mathcal{I}}$, respectively. Our aim is to show that **whp** paths of length k exist between the vertices a and b. We do this by applying a second moment method on $N_k(a,b)$, for which we need a lower bound on $\mathbb{E}[N_k(a,b)]$ and an upper bound on $\mathrm{Var}(N_k(a,b))$, which are such that $\mathrm{Var}(N_k(a,b)) = o(\mathbb{E}[N_k(a,b)]^2)$ (recall Theorem 2.15. We prove lower bounds on $\mathbb{E}[N_k(a,b)]$ and upper bounds on $\mathrm{Var}(N_k(a,b))$ in the following proposition:

Proposition 9.17 (Variance of numbers of paths)**.** *For any $k \geq 1$, $a, b \in \mathcal{I}$ and $(u_i)_{i\in\mathcal{I}}$,*

$$\mathbb{E}[N_k(a,b)] \geq \underline{n}_k(a,b), \tag{9.4.6}$$

while, assuming that $\nu_{\mathcal{I}} > 1$,

$$\begin{aligned}\mathrm{Var}(N_k(a,b)) \leq n_k(a,b) + \bar{n}_k(a,b)^2\Big(&\frac{\gamma_{\mathcal{I}}\nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}}-1}\Big(\frac{1}{u_a}+\frac{1}{u_b}\Big) + \frac{\gamma_{\mathcal{I}}^2\nu_{\mathcal{I}}}{u_a u_b(\nu_{\mathcal{I}}-1)^2} \\ &+ k\Big(1+\frac{\gamma_{\mathcal{I}}}{u_a\nu_{\mathcal{I}}}\Big)\Big(1+\frac{\gamma_{\mathcal{I}}}{u_b\nu_{\mathcal{I}}}\Big)\frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}}-1}\big(\mathrm{e}^{2k^3\gamma_{\mathcal{I}}^2/\nu_{\mathcal{I}}^3}-1\big)\Big).\end{aligned} \tag{9.4.7}$$

We apply Proposition 9.17 in cases where $\mathbb{E}[N_k(a,b)] = n_k(a,b) \to \infty$, by taking $\mathcal{I}$ is a large subset of $[n]$ and $u_i = w_i/\sqrt{\ell_n}$. In this case, $\nu_{\mathcal{I}} \approx \nu_n \approx \nu > 1$. In our applications of Proposition 9.17, the ratio $\bar{n}_k(a,b)/\underline{n}_k(a,b)$ will be bounded, and $k^3\gamma_{\mathcal{I}}^2/\nu_{\mathcal{I}}^3 = o(1)$, so that the last term is an error term. The starting and end vertices $a, b \in \mathcal{I}$ will correspond to a union of vertices in $[n]$ of quite large size. As a result, $\gamma_{\mathcal{I}}/u_a$ and $\gamma_{\mathcal{I}}/u_a$ are typically small, so that

$$\mathrm{Var}(N_k(a,b)) \approx \frac{\gamma_{\mathcal{I}}\nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}}-1}\big(\frac{1}{u_a}+\frac{1}{u_b}\big) \tag{9.4.8}$$

is small. The choice of a, b and $\mathcal{I}$ is quite delicate, which explains why we formulate Proposition 9.17 in such generality.

Proof. We note that $N_k(a,b)$ is a sum of indicators

$$N_k(a,b) = \sum_{\pi\in\mathcal{P}_k(a,b)} \mathbb{1}_{\{\pi \text{ occupied}\}}. \tag{9.4.9}$$

As a result,

$$\begin{aligned}\mathbb{E}[N_k(a,b)] &= \sum_{\pi\in\mathcal{P}_k(a,b)} \mathbb{P}(\pi \text{ occupied}) = \sum_{\pi\in\mathcal{P}_k(a,b)}\prod_{l=0}^{k} u_{\pi_l}u_{\pi_{l+1}} \\ &= u_{\pi_0}u_{\pi_k}\sum_{\pi\in\mathcal{P}_k(a,b)}\prod_{l=1}^{k-1}u_{\pi_l}^2.\end{aligned} \tag{9.4.10}$$

For $\pi \in \mathcal{P}_k(a,b)$, $\pi_0 = a, \pi_k = b$. Further,

$$\sum_{\pi\in\mathcal{P}_k(a,b)}\prod_{l=1}^{k-1}u_{\pi_l}^2 = \sum_{i_1,\ldots,i_{k-1}\in\mathcal{I}\setminus\{a,b\}}^{*}\prod_{l=1}^{k-1}u_{\pi_l}^2, \tag{9.4.11}$$

where we recall that $\sum^*_{i_1,\ldots,i_r\in\mathcal{I}}$ denotes a sum over *distinct* indices. Each sum over i_j yields a factor that is at least $\sum_{i\in\mathcal{I}_{a,b,k}} u_i^2$, which proves (10.4.19).

To compute $\mathrm{Var}(N_k(a,b))$, we again start from (9.4.9), which yields

$$\mathrm{Var}(N_k(a,b)) = \sum_{\pi,\rho\in\mathcal{P}_k(a,b)}\big[\mathbb{P}(\pi,\rho \text{ occupied}) - \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied})\big]. \tag{9.4.12}$$

For π, ρ, we denote by $\pi\cap\rho$ the edges the paths π and ρ have in common. The occupation statuses of π and ρ are independent precisely when $\pi\cap\rho = \varnothing$, so that

$$\mathrm{Var}(N_k(a,b)) \le \sum_{\substack{\pi,\,\rho\,\in\,\mathcal{P}_k(a,\,b)\\ \pi\,\cap\,\rho\,\neq\,\varnothing}} \mathbb{P}(\pi,\rho \text{ occupied}). \tag{9.4.13}$$

Define $\rho\setminus\pi$ to be the edges in ρ that are *not* part of π, so that

$$\begin{aligned}\mathbb{P}(\pi,\rho \text{ occupied}) &= \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied} \mid \pi \text{ occupied}) \\ &= \prod_{l=0}^{k} u_{\pi_l}u_{\pi_{l+1}}\prod_{e\in\rho\setminus\pi} u_{\bar{e}}u_{\underline{e}},\end{aligned} \tag{9.4.14}$$

where, for an edge $e = \{x, y\}$, we write $\bar{e} = x, \underline{e} = y$. When $\pi = \rho$, then

$$\mathbb{P}(\pi, \rho \text{ occupied}) = \mathbb{P}(\pi \text{ occupied}), \tag{9.4.15}$$

and this contributes $n_k(a, b)$ to $\mathrm{Var}(N_k(a, b))$. From now on, we consider $\pi \neq \rho$.

The probability in (9.4.14) needs to be summed over all possible pairs of paths (π, ρ) with $\pi \neq \rho$ that share at least one edge. In order to do this effectively, we start by introducing some notation.

Let $l = |\pi \cap \rho|$ denote the number of edges in $\pi \cap \rho$, so that $l \geq 1$ precisely when $\pi \cap \rho \neq \varnothing$. Since $\pi \neq \rho$, $l \leq k - 1$. When $\pi \neq \rho$, we have that $l \leq k - 1$, and since π and ρ are self-avoiding paths between a and b, l cannot be equal to $k - 1$, so that we consider $l \leq k - 2$ from now on. Let $k - l = |\rho \setminus \pi|$ be the number of edges in ρ that are not part of π.

Let m denote the number of connected subpaths in $\rho \setminus \pi$, so that $m \geq 1$ whenever $\pi \neq \rho$. Since $\pi_0 = \rho_0 = a$ and $\pi_k = \rho_k = b$, these subpaths start and end in vertices along the path π. We can view these subpaths as excursions of the path ρ from the walk π. By construction, between two excursions, there is at least one edge that π and ρ have in common.

Fix m. We define $\mathrm{Shape}(\pi, \rho)$, the *shape* of the pair (π, ρ), by

$$\mathrm{Shape}(\pi, \rho) = (\vec{x}_{m+1}, \vec{s}_m, \vec{t}_m, \vec{o}_{m+1}, \vec{r}_{m+1}), \tag{9.4.16}$$

where

(1) $\vec{x}_{m+1} \in \mathbb{N}_0^{m+1}$ and $x_j \geq 0$ is the length of the subpath in $\rho \cap \pi$ in between the $(j-1)^{\text{st}}$ and j^{th} subpath of $\pi \setminus \rho$. Here x_1 is the number of common edges in the subpath of $\rho \cap \pi$ that contains a, while x_{m+1} is the number of common edges in the subpath of $\rho \cap \pi$ that contains b, so that $x_1 \geq 0$ and $x_{m+1} \geq 0$. For $j \in \{2, \ldots, m\}$, $x_j \geq 1$;

(2) $\vec{s}_m \in \mathbb{N}^m$ and $s_j \geq 1$ is the number of edges in the j^{th} subpath of $\pi \setminus \rho$;

(3) $\vec{t}_m \in \mathbb{N}^m$ and $t_j \geq 1$ is the number of edges in the j^{th} subpath of $\rho \setminus \pi$;

(4) $\vec{o}_{m+1} \in [m+1]^{m+1}$ and o_j is the order of the j^{th} common subpath in $\rho \cap \pi$ of the path π in ρ, i.e., $o_2 = 5$ means that the second subpath that π has in common with ρ is the 5^{th} subpath that ρ has in common with π. Note that $o_1 = 1$ and $o_{m+1} = m + 1$, since π and ρ start and end in a and b, respectively;

(5) $\vec{r}_{m+1} \in \{0, 1\}^{m+1}$ is such that r_j describes the direction in which the j^{th} common subpath in $\rho \cap \pi$ of the path π is traversed by ρ, with $r_j = 1$ when the direction is the same for π and ρ and 0 otherwise. Thus, $r_1 = r_{m+1} = 1$.

The information in $\mathrm{Shape}(\pi, \rho)$ is precisely that necessary to piece together the topology of the two paths, except for the information of the *vertices* involved in π and ρ. See Figure 9.4.1 for an example of a pair of paths (π, ρ) and its corresponding shape.

With $l = |\pi \cap \rho|$, we have

$$\sum_{j=1}^{m+1} x_j = l, \qquad \sum_{j=1}^{m} s_j = \sum_{j=1}^{m} t_j = k - l. \tag{9.4.17}$$

Let $\mathrm{Shape}_{m,l}$ denote the set of shapes corresponding to pairs of paths (π, ρ) with m excursions and l common edges, so that (9.4.17) hold. Then,

$$\mathrm{Var}(N_k(a, b)) \leq \bar{n}_k(a, b) + \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \sum_{\sigma \in \mathrm{Shape}_{m,l}} \sum_{\substack{\pi, \rho \in \mathcal{P}_k(a,b) \\ \mathrm{Shape}(\pi, \rho) = \sigma}} \mathbb{P}(\pi, \rho \text{ occupied}). \tag{9.4.18}$$

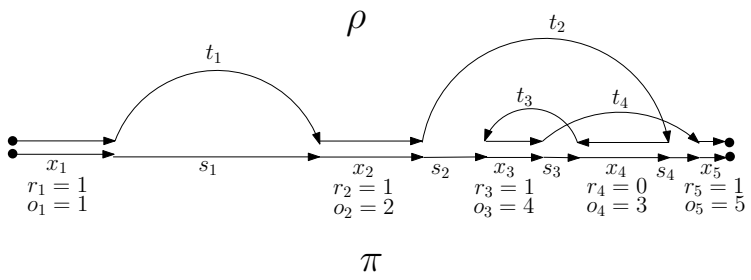

Figure 9.3: An example of a pair of paths (π, ρ) and its corresponding shape.

When $\mathrm{Shape}(\pi, \rho) = \sigma$ for some $\sigma \in \mathrm{Shape}_{m,l}$, and since π and ρ both start and end in a and b, the union of paths $\pi \cup \rho$ visits $k + 1 + l - m$ distinct vertices. The vertex a is in $1 + \delta_{x_1,0}$ edges, and b in $1 + \delta_{x_{m+1},0}$ edges. Of the other $k - 1$ vertices in π, precisely $2m - \delta_{x_1,0} - \delta_{x_{m+1},0}$ are part of of three edges, and $k - 1 - 2m + \delta_{x_1,0} + \delta_{x_{m+1},0}$ are part of in two edges. The remaining $k - l - m$ vertices in ρ that are not part of π are part of precisely 2 edges. By construction, the $k + 1$ vertices of both π and ρ are disjoint, but the remaining $k - l - m$ vertices in ρ may intersect those of π. Therefore, denoting $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$,

$$\mathbb{P}(\pi, \rho \text{ occupied}) = u_a^{1+a_1} u_b^{1+a_{m+1}} \prod_{s=1}^{2m-a_1-a_{m+1}} u_{v_s}^3 \prod_{t=2m-a_1-a_{m+1}+1}^{2(k-1)-l-m} u_{v_t}^2, \tag{9.4.19}$$

where $\{(v_1, \ldots, v_{k+1+l-m})\} \in \mathcal{I}^{k-1+l-m}$.

For a fixed $\sigma \in \mathrm{Shape}_{m,l}$ now bound the sum over $\pi, \rho \in \mathcal{P}_k(a, b)$ such that $\mathrm{Shape}(\pi, \rho) = \sigma$ from above by summing (9.4.19) over *all* $\{(v_1, \ldots, v_{k-1+l-m})\} \in \mathcal{I}^{k-1+l-m}$, to obtain for any $\sigma \in \mathrm{Shape}_{m,l}$,

$$\begin{aligned} &\sum_{\substack{\pi,\, \rho \in \mathcal{P}_k(a,b) \\ \mathrm{Shape}(\pi,\rho) = \sigma}} \mathbb{P}(\pi, \rho \text{ occupied}) \qquad (9.4.20) \\ &\leq u_a u_b \gamma_{\mathcal{I}}^{2m} \nu_{\mathcal{I}}^{2k-1-3m-l} \big(\frac{u_a \nu_{\mathcal{I}}}{\gamma_{\mathcal{I}}}\big)^{\delta_{x_1,0}} \big(\frac{u_b \nu_{\mathcal{I}}}{\gamma_{\mathcal{I}}}\big)^{\delta_{x_{m+1},0}} \\ &= \bar{n}_k(a,b)^2 \gamma_{\mathcal{I}}^{2(m+1)} \nu_{\mathcal{I}}^{-3(m-1)-l} \big(\frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\big)^{1-\delta_{x_1,0}} \big(\frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\big)^{1-\delta_{x_{m+1},0}}. \end{aligned}$$

Therefore, we arrive at

$$\begin{aligned} \mathrm{Var}(N_k(a,b)) \leq n_k(a,b) + \bar{n}_k(a,b)^2 \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \gamma_{\mathcal{I}}^{2(m-1)} \nu_{\mathcal{I}}^{-3(m-1)-l} \qquad (9.4.21) \\ \times \sum_{\sigma \in \mathrm{Shape}_{m,l}} \big(\frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\big)^{1-\delta_{x_1,0}} \big(\frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\big)^{1-\delta_{x_{m+1},0}}. \end{aligned}$$

We continue to compute the number of shapes in the following lemma:

Lemma 9.18 (The number of shapes)**.** *Fix $m \geq 1$ and $l \leq k - 2$. For $m = 1$, the number of shapes in* $\mathrm{Shape}_{m,l}$ *fixed $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$ equals l when $a_1 = a_{m+1} = 0$, 1*

when $a_1 + a_{m+1} = 1$ *and* 0 *when* $a_1 = a_{m+1} = 1$. *For* $m \geq 2$, *the number of shapes in* $\mathrm{Shape}_{m,l}$ *fixed* $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$ *is bounded by*

$$2^{m-1}(m-1)!\binom{k-l-1}{m-1}^2\binom{l}{m-a_1-a_{m+1}}. \tag{9.4.22}$$

Proof. Since $r_1 = r_{m+1} = 1$, there are 2^{m-1} directions in which the common parts can be traversed. Since there are m distinct parts, there are $m+1$ common parts. The first part contains vertex a, the last part contains vertex b. Thus, there are $(m-1)!$ orders $\vec{o}_{m+1}$ of the common parts when we have fixed the directions the paths can be traversed.

In counting the number of $\vec{x}_{m+1}, \vec{s}_m, \vec{t}_m$, we repeatedly use the fact that there are $\binom{a-1}{b-1}$ possible sequences $(y_1, \ldots, y_b) \in \mathbb{N}_0^b$ such that $\sum_{j=1}^b y_j = a$. This can be seen by representing a as a sequence of a ones, separated by $a-1$ zeros. We draw b zeros, which we can do in $\binom{a-1}{b-1}$ possible ways. Then, we note that a sequence $(y_1, \ldots, y_b) \in \mathbb{N}_0^b$ such that $\sum_{j=1}^b y_j = a$ can be obtained uniquely by letting y_i be the number of ones in between the $(i-1)^{\text{st}}$ and i^{th} chosen zero. Similarly, there are $\binom{a+b-1}{b-1}$ possible sequences $(y_1, \ldots, y_b) \in \mathbb{N}^b$ such that $\sum_{j=1}^b y_j = a$, since we can apply the previous equality to $(y_1+1, \ldots, y_b+1) \in \mathbb{N}^b$.

Using the above, we continue to count the number of shapes. The number of $(s_1, \ldots, s_m) \in \mathbb{N}^m$ such that $s_j \geq 1$ and $\sum_{j=1}^m s_j = k-l$ equals

$$\binom{k-l-1}{m-1}. \tag{9.4.23}$$

The same applies to $(t_1, \ldots, t_m) \in \mathbb{N}^m$ such that $t_j \geq 1$ and $\sum_{j=1}^m t_j = l$. In counting the number of possible $\vec{x}_{m+1}$ such that $\sum_{j=1}^{m+1} x_j = l$, we need to count their numbers separately for $x_1 = 0$ and $x_1 \geq 1$, and for $x_{m+1} = 0$ and $x_{m+1} \geq 1$. When $m = 1$, the number is zero when $x_1 = x_2 = 0$, since $x_1 = x_2 = 0$ implies that the paths share no edges. Denote $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$, and suppose that $m - a_1 - a_{m+1} \geq 0$. Then, there are

$$\binom{l}{m-a_1-a_{m+1}} \tag{9.4.24}$$

possible choice of $\vec{x}_{m+1}$ with fixed $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$. The claim follows by multiplying these bounds on the number of choices for $\vec{r}_{m+1}$, $\vec{o}_{m+1}$, $\vec{s}_m$, $\vec{t}_m$ and $\vec{x}_{m+1}$. □

We complete the proof of Proposition 9.17. By (9.4.21) and applying Lemma 9.18, it suffices to sum

$$\begin{aligned} &2^{m-1}(m-1)!\binom{k-l-1}{m-1}^2\binom{l}{m-a_1-a_{m+1}} \\ &\quad\times \big(\frac{2\gamma_{\mathcal{I}}^2}{\nu_{\mathcal{I}}^3}\big)^{m-1}\nu_{\mathcal{I}}^{-l}\big(\frac{\gamma_{\mathcal{I}}}{u_a\nu_{\mathcal{I}}}\big)^{1-a_1}\big(\frac{\gamma_{\mathcal{I}}}{u_b\nu_{\mathcal{I}}}\big)^{1-a_{m+1}} \end{aligned} \tag{9.4.25}$$

over $l \in [k-2]$, $m \in [k-l]$ and $a_1, a_{m+1} \in \{0,1\}$, where, by convention, $\binom{l}{-1} = 0$.

We start with $m = 1$, for which we obtain that the sum of (9.4.25) over the other

variables equals

$$\gamma_{\mathcal{I}}\big(\frac{1}{u_a}+\frac{1}{u_b}\big)\sum_{l=1}^{\infty}\nu_{\mathcal{I}}^{-(l-1)}+\frac{\gamma_{\mathcal{I}}^2}{u_a u_b \nu_{\mathcal{I}}}\sum_{l=1}^{\infty} l\nu_{\mathcal{I}}^{-(l-1)}$$
$$=\frac{\gamma_{\mathcal{I}}\nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}}-1}\big(\frac{1}{u_a}+\frac{1}{u_b}\big)+\frac{\gamma_{\mathcal{I}}^2\nu_{\mathcal{I}}}{u_a u_b(\nu_{\mathcal{I}}-1)^2}, \tag{9.4.26}$$

where we use that, for $a \in [0,1)$,

$$\sum_{l=0}^{\infty} a^{-l} = a/(1-a), \qquad \sum_{l=0}^{\infty} l a^{-(l-1)} = a^2/(1-a)^2. \tag{9.4.27}$$

The terms in (9.4.26) are the first two terms appearing on the right-hand side of (10.4.20).

This leaves us to bound the contribution when $m \geq 2$. We continue by bounding

$$\binom{k-l-1}{m-1}(m-1)! = \frac{1}{(m-1)!}\Big(\frac{(k-l-1)!}{(k-l-m)!}\Big)^2 \leq \frac{k^{2(m-1)}}{(m-1)!}, \tag{9.4.28}$$

and, using that $\binom{a}{b} \leq a^b/b!$ and $l \leq k$,

$$\binom{l}{m-a_1-a_{m+1}} \leq \frac{l^{m-a_1-a_{m+1}}}{(m-a_1-a_{m+1})!} \leq k^m. \tag{9.4.29}$$

Therefore, the number of shapes in $\mathrm{Shape}_{m,l}$ is, for each $l \geq 1$ and $m \geq 2$, bounded by

$$2^{m-1}\frac{k^{2(m-1)}}{(m-1)!}k^m = k\frac{(2k^3)^{m-1}}{(m-1)!}. \tag{9.4.30}$$

Since the above is independent of l, we can start by summing (9.4.25) over $l \geq 1$, and over $a_1, a_{m+1} \in \{0,1\}$ to obtain a bound of the form

$$k\big(1+\frac{\gamma_{\mathcal{I}}}{u_a\nu_{\mathcal{I}}}\big)\big(1+\frac{\gamma_{\mathcal{I}}}{u_b\nu_{\mathcal{I}}}\big)\frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}}-1}\sum_{m\geq 2}\frac{(2k^3)^{m-1}}{(m-1)!}\Big(\frac{\gamma_{\mathcal{I}}^2}{\nu_{\mathcal{I}}^3}\Big)^{m-1} \tag{9.4.31}$$
$$= k\big(1+\frac{\gamma_{\mathcal{I}}}{u_a\nu_{\mathcal{I}}}\big)\big(1+\frac{\gamma_{\mathcal{I}}}{u_b\nu_{\mathcal{I}}}\big)\frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}}-1}\big(\mathrm{e}^{2k^3\gamma_{\mathcal{I}}^2/\nu_{\mathcal{I}}^3}-1\big).$$

The term in (9.4.31) is the last term appearing on the right-hand side of (10.4.20). Summing the bounds in (9.4.26) and (9.4.31) proves (10.4.20). □

Exercise 9.21 (Variance of two paths)**.** *Prove that* $\mathrm{Var}(N_k(a,b)) \leq \mathbb{E}[N_k(a,b)]$ *for* $k = 2$.

Exercise 9.22 (Variance of three paths)**.** *Compute* $\mathrm{Var}(N_3(a,b))$ *explicitly, and compare it to the bound in* (10.4.20).

Exercise 9.23 (Variance on paths for $\mathrm{ER}_n(\lambda/n)$)**.** *Let* $A, B \subseteq [n]$*, and let* $N_k(A,B)$ *denote the number of paths of length* k *connecting* A *to* B *(where a path connecting* A *and* B *avoids* A *and* B *except for the starting and end point). Show that for* $k \leq K\log n$*,*

$$\mathbb{E}[N_k(A,B)] = \lambda^k|A||B|\big(1-\frac{|A|+|B|}{n}\big)^k(1+o(1)). \tag{9.4.32}$$

Use Proposition 9.17 to bound the variance of $N_k(A,B)$*, and prove that*

$$N_k(A,B)/\mathbb{E}[N_k(A,B)] \xrightarrow{\mathbb{P}} 1 \tag{9.4.33}$$

when $|A|, |B| \to \infty$ *with* $|A|+|B| = o(n/k)$.

9.4.2 Logarithmic upper bound on typical distances in finite variance case

In this section, we prove that two uniformly chosen vertices that are conditioned to be connected are with high probability within distance $(1+\varepsilon)\log_\nu n$, as formulated in the following theorem:

Theorem 9.19 (Logarithmic upper bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Assume that Condition 6.4(a)-(c) hold, where $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \in (1,\infty)$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}(H_n \le (1+\varepsilon)\log_\nu n \mid H_n < \infty) = 1 + o(1). \tag{9.4.34}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ under the same conditions.

Organization of the proof of Theorem 9.19. We prove Theorem 9.19 by combining the *branching process comparison* to a *second moment method* using Proposition 9.17 on the number of paths of a given length. More precisely, we fix $m \ge 1$ large, and recall that $\mathcal{N}_{\le m}(V_1)$ and $\mathcal{N}_{\le m}(V_2)$ denote the vertices at distance at most m from V_1 and V_2 respectively, and let $\mathcal{N}_m(V_1)$ and $\mathcal{N}_m(V_2)$ denote the vertices at distance precisely equal to m. We condition on $\mathcal{N}_{\le m}(V_1)$ and $\mathcal{N}_{\le m}(V_2)$ such that $\mathcal{N}_m(V_1) \neq \varnothing$ and $\mathcal{N}_m(V_2) \neq \varnothing$. By Corollary 9.13, the probabilities of the latter event is close to ζ_m^2, where $\zeta_m = \mathbb{P}(\mathcal{Z}_m^{(1)} > 0)$ is the probability that the branching process survives to generation m. Then, $\zeta_m \to \zeta$ when $m \to \infty$, and, conditionally on $\mathcal{Z}_m^{(1)} > 0$, $\mathcal{Z}_m^{(1)} \ge M$ **whp**, for any M and as $m \to \infty$. This explains the branching process approximation. We take $u_i = w_i/\sqrt{\ell_n}$,

$$a = \mathcal{N}_m(V_1), \qquad b = \mathcal{N}_m(V_2), \tag{9.4.35}$$

so that

$$u_a = \frac{1}{\sqrt{\ell_n}} \sum_{i \in \mathcal{N}_m(V_1)} w_i = \mathcal{W}_m(V_1)/\sqrt{\ell_n}, \qquad b = \frac{1}{\sqrt{\ell_n}} \sum_{i \in \mathcal{N}_m(V_2)} w_i = \mathcal{W}_m(V_2)/\sqrt{\ell_n}. \tag{9.4.36}$$

We formalize the above ideas in the following lemma:

Lemma 9.20 (Branching process approximation)**.** *As $n \to \infty$,*

$$(\mathcal{W}_m(V_1), \mathcal{W}_m(V_2)) \xrightarrow{d} \Big(\sum_{j=1}^{\mathcal{Z}_m^{(1)}} W^{(1)}(j), \sum_{j=1}^{\mathcal{Z}_m^{(2)}} W^{(2)}(j) \Big), \tag{9.4.37}$$

where $(\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)})$ are the generation sizes of two independent branching processes, and $(W^{(1)}(j))_{j\ge1}$ and $(W^{(2)}(j))_{j\ge1}$ are two independent sequences of i.i.d. random variables with distribution F.

Proof. By Corollary 9.13, $\widetilde{Z}_m^{(1)} = |\mathcal{N}_m(V_1)|$ and $\widetilde{Z}_m^{(2)} = |\mathcal{N}_m(V_2)|$ jointly converge in distribution to $(\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)})$, which are independent generation sizes of MMPBPs. Each of the individuals in $\mathcal{N}_m(V_1)$ and $\mathcal{N}_m(V_2)$ receives a mark M_i, and its weight is w_{M_i}. By Proposition 9.10, these marks are i.i.d. random variables conditioned to be distinct (if they were not distinct, then the corresponding individuals would have been thinned). Since, **whp**, the proof of Corollary 9.13 shows that no vertex is thinned, **whp** we can ignore the thinning. Then, $\mathcal{W}_m(V_1) = \sum_{j=1}^{\widetilde{Z}_m^{(1)}} W_n(j)$, where $(W_n(j))_{j\ge1}$ are i.i.d. copies of W_n. By Condition 6.4(a), $W_n \xrightarrow{d} W$, so that $\mathcal{W}_m(V_1) \xrightarrow{d} \sum_{j=1}^{\mathcal{Z}_m^{(1)}} W^{(1)}(j)$. The joint convergence follows in a similar fashion. □

Second moment method and path counting. Fix $k = k_n = (1+\varepsilon)\log_\nu n - 2m$. We next present the details of the second moment method that shows that **whp**, on the event that $\mathcal{N}_m(V_1) \neq \varnothing$ and $\mathcal{N}_m(V_2) \neq \varnothing$, there exist a path of length $k_n - 2m$ connecting $\mathcal{N}_m(V_1)$ and $\mathcal{N}_m(V_2)$. This ensures that, on the event that $\mathcal{N}_m(V_1) \neq \varnothing$ and $\mathcal{N}_m(V_2) \neq \varnothing$, the event $H_n \leq k_n - 2m$ occurs **whp**. For this, we take $u_i = w_i/\sqrt{\ell_n}$. We fix $K \geq 1$ sufficiently large and take

$$\mathcal{I} = \{i \in [n] \colon w_i \leq K\} \setminus (\mathcal{N}_{\leq m}(V_1) \cup \mathcal{N}_{\leq m}(V_2)). \tag{9.4.38}$$

We investigate the constants appearing in Proposition 9.17 in the following lemma:

Lemma 9.21 (Parameters in path counting)**.** *Conditionally on $\mathcal{N}_{\leq m}(V_1)$ and $\mathcal{N}_{\leq m}(V_2)$, and with $a = \mathcal{N}_m(V_1), b = \mathcal{N}_m(V_2)$, for $k = (1+\varepsilon)\log_\nu n$,*

$$\lim_{n\to\infty} n_k(a,b) = \infty, \qquad \lim_{n\to\infty} \frac{\bar{n}_k(a,b)}{\underline{n}_k(a,b)} = 1, \tag{9.4.39}$$

and

$$\frac{\mathrm{Var}(N_k(a,b))}{\mathbb{E}[N_k(a,b)]^2} \leq \frac{K\nu^2}{\nu-1}\Big(\frac{1}{\sqrt{\ell_n}u_a} + \frac{1}{\sqrt{\ell_n}u_b}\Big) + \frac{K^2\nu^2}{(\nu-1)\ell_n u_a u_b} + o_{\mathbb{P}}(1). \tag{9.4.40}$$

Proof. By (9.4.3),

$$\underline{n}_k(a,b) = u_a u_b \nu_{\mathcal{I}_{a,b}}^{k-1}, \tag{9.4.41}$$

and

$$\frac{\bar{n}_k(a,b)}{\underline{n}_k(a,b)} = \big(\nu_{\mathcal{I}_{a,b}}/\nu_{\mathcal{I}_{a,b,k}}\big)^{k-1}. \tag{9.4.42}$$

We start by investigating $\nu_{\mathcal{I}}$. Denote

$$\nu(K) = \frac{\mathbb{E}[W^2 \mathbb{1}_{\{W\leq K\}}]}{\mathbb{E}[W]}. \tag{9.4.43}$$

Then, by (9.4.38) and the fact that $\mathcal{N}_{\leq m}(V_1)$ and $\mathcal{N}_{\leq m}(V_2)$ contain a finite number of vertices,

$$\lim_{n\to\infty} \nu_{\mathcal{I}} = \nu(K). \tag{9.4.44}$$

The same applies to $\nu_{\mathcal{I}_{a,b}}$ and $\nu_{\mathcal{I}_{a,b,k}}$. Then, with $K > 0$ chosen so large that $\nu(K) \geq \nu - \varepsilon/2$ and with $k = (1+\varepsilon)\log_\nu n$,

$$\nu_{\mathcal{I}_{a,b}}^{k-1} = \mathcal{W}_m(V_1)\mathcal{W}_m(V_2)\frac{n}{\ell_n} n^{(1+\varepsilon)\log \nu_{\mathcal{I}_{a,b}}/\log\nu - 1} \to \infty, \tag{9.4.45}$$

where K and n are so large that $(1+\varepsilon)\nu_{\mathcal{I}}/\nu > 1$. This proves the first property in (9.4.39).

To prove the second property in (9.4.39), we note that the set $\mathcal{I}_{a,b,k}$ is obtained from $\mathcal{I}_{a,b}$ by removing the k vertices with highest weight. Since $w_i \leq K$ for all $i \in \mathcal{I}$ (recall (9.4.38)), $\nu_{\mathcal{I}_{a,b}} \leq \nu_{\mathcal{I}_{a,b,k}} + kK/\ell_n$. Since $k \leq A\log n$, we therefore arrive at

$$\frac{\bar{n}_k(a,b)}{\underline{n}_k(a,b)} \leq \big(1 + kK/(\ell_n \nu_{\mathcal{I}_{a,b,k}})\big)^{k-1} = \mathrm{e}^{k^2 K/(\ell_n \nu_{\mathcal{I}_{a,b,k}})} \to 1, \tag{9.4.46}$$

as required.

To prove (9.4.40), we rely on Proposition 9.17. We have already shown that $n_k(a,b) = \mathbb{E}[N_k(a,b)] \to \infty$, so that the first term on the right-hand side of (10.4.20) is $o(\mathbb{E}[N_k(a,b)]^2)$. Further, by (9.4.38),

$$\gamma_{\mathcal{I}} \leq \nu_{\mathcal{I}}(\max_{i\in\mathcal{I}} u_i) \leq \frac{\nu_{\mathcal{I}} K}{\sqrt{\ell_n}}, \tag{9.4.47}$$

so that, for $k \le A \log n$ with $A > 1$ fixed,

$$(1 + \frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}})(1 + \frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}) k (\mathrm{e}^{2k^3 \gamma_{\mathcal{I}}^2/\nu_{\mathcal{I}}^3} - 1) = o_{\mathbb{P}}(1). \tag{9.4.48}$$

Substituting these bounds into (9.4.40) and using (9.4.39) yields the claim. □

Completing of the proof of Theorem 9.19. Now we are are ready to complete the proof of Theorem 9.19. We must show that

$$\mathbb{P}(k_n < H_n < \infty) = o(1). \tag{9.4.49}$$

Indeed, then $\mathbb{P}(H_n > k_n \mid H_n < \infty) = o(1)$ since, $\mathbb{P}(H_n < \infty) \to \zeta^2 > 0$ by Theorem 9.2. We rewrite

$$\begin{aligned} \mathbb{P}(k_n < H_n < \infty) &= \mathbb{P}(k_n < H_n < \infty, \mathcal{N}_m(V_1) \neq \varnothing, \mathcal{N}_m(V_2) \neq \varnothing) \\ &\le \mathbb{P}(N_{k_n - 2m}(\mathcal{N}_m(V_1), \mathcal{N}_m(V_2)) = 0, \mathcal{N}_m(V_1) \neq \varnothing, \mathcal{N}_m(V_2) \neq \varnothing). \end{aligned} \tag{9.4.50}$$

Recall that $k = k_n = (1 + \varepsilon) \log_\nu n$. By the Chebychev inequality (Theorem 2.15), and given $\mathcal{N}_{\le m}(V_1), \mathcal{N}_{\le m}(V_2)$, the conditional probability on $\{H_n > k_n\}$ is at most

$$\frac{\mathrm{Var}(N_{k-2m}(a,b))}{\mathbb{E}[N_{k-2m}(a,b)]^2} \le \frac{K\nu^2}{\nu - 1}\big(\frac{1}{\sqrt{\ell_n} u_a} + \frac{1}{\sqrt{\ell_n} u_b}\big) + \frac{K^2 \nu^2}{(\nu - 1)\ell_n u_a u_b} + o_{\mathbb{P}}(1). \tag{9.4.51}$$

When $\mathcal{N}_m(V_1) \neq \varnothing$ and $\mathcal{N}_m(V_2) \neq \varnothing$, by (9.4.36),

$$\frac{1}{\sqrt{\ell_n} u_a} + \frac{1}{\sqrt{\ell_n} u_b} \xrightarrow{\mathbb{P}} \Big(\sum_{j=1}^{\mathcal{Z}_m^{(1)}} W^{(1)}(j) \Big)^{-1} + \Big(\sum_{j=1}^{\mathcal{Z}_m^{(2)}} W^{(2)}(j) \Big)^{-1} \xrightarrow{\mathbb{P}} 0, \tag{9.4.52}$$

when $m \to \infty$. Therefore,

$$\mathbb{P}\Big(N_{k-2m}(a,b) = 0 \mid \mathcal{N}_m(V_1) \neq \varnothing, \mathcal{N}_m(V_2) \neq \varnothing \Big) \xrightarrow{\mathbb{P}} 0, \tag{9.4.53}$$

and, by Lebesgues Dominated Convergence Theorem (Theorem A.9),

$$\mathbb{P}(H_n > k_n, \mathcal{N}_m(V_1) \neq \varnothing, \mathcal{N}_m(V_2) \neq \varnothing) \to 0, \tag{9.4.54}$$

which completes the proof. □

We close this section by describing what happens when $\tau = 3$, and there are no slowly varying functions:

Distances for the critical case $\tau = 3$ When $\tau = 3$, w_i is approximately $c(n/i)^{1/2}$. It turns our that this changes the distances only by a doubly logarithmic factor:

Theorem 9.22 (Logarithmic upper bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Assume that Condition 6.4(a)-(b) hold, and that there exists constants $c_2 > c_1 > 0$ and $\alpha > 0$ such that for all $x \le n^\alpha$,*

$$[1 - F_n](x) \ge c_1/x^2, \tag{9.4.55}$$

and for all $x \ge 0$,

$$[1 - F_n](x) \le c_2/x^2. \tag{9.4.56}$$

Then, conditionally on $H_n < \infty$,

$$\frac{H_n \log\log n}{\log n} \xrightarrow{\mathbb{P}} 1. \tag{9.4.57}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ under the same conditions.

The lower bound in Theorem 9.22 is already stated in Corollary 9.6. The upper bound can be proved using the path counting techniques in Proposition 9.17 and adaptations of it. We now sketch the proof.

Let $\eta > 0$ and let

$$\alpha_n = \mathrm{e}^{\nu_n^{1-\eta}}. \tag{9.4.58}$$

Define the *core* of $\mathrm{NR}_n(\boldsymbol{w})$ to be

$$\mathrm{Core}_n = \{i\colon w_i \geq \alpha_n\}. \tag{9.4.59}$$

The proof of Theorem 9.22 follows from the following two propositions:

Proposition 9.23 (Typical distances in the core)**.** *Under the conditions of Theorem 9.22, let $V_1', V_2' \in \mathrm{Core}_n$ be chosen with probability proportional to their weight, i.e.,*

$$\mathbb{P}(V_i' = j) = \frac{w_j}{\sum_{v\in\mathrm{Core}_n} w_v}, \tag{9.4.60}$$

and let H_n' be the graph distance between V_1', V_2' in Core_n. Then, for any $\varepsilon > 0$, there exists an $\eta > 0$ such that

$$\mathbb{P}\big(H_n' \leq \frac{(1+\varepsilon)\log n}{\log\log n}\big) \to 1. \tag{9.4.61}$$

Proposition 9.24 (From the periphery to the core)**.** *Under the conditions of Theorem 9.22, let V_1, V_2 be two vertices chosen uniformly at random from $[n]$. Then, for any $\eta > 0$,*

$$\mathbb{P}(d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Core}_n) \leq \nu_n^{1-\eta}, d_{\mathrm{NR}_n(\boldsymbol{w})}(V_2, \mathrm{Core}_n) \leq \nu_n^{1-\eta}) \to \zeta^2. \tag{9.4.62}$$

To see that Propositions 9.23–9.24 imply Theorem 9.22, we note that

$$\begin{aligned} H_n &= d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, V_2) \\ &\leq d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Core}_n) + d_{\mathrm{NR}_n(\boldsymbol{w})}(V_2, \mathrm{Core}_n) + d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1', V_2'), \end{aligned} \tag{9.4.63}$$

where $V_1', V_2' \in \mathrm{Core}_n$ are the vertices in Core_n found first in the breadth-first search. Then, by Proposition 9.10, $V_1', V_2' \in \mathrm{Core}_n$ are chosen with probability proportional to their weight. Therefore, when n is so large that $\nu_n^{1-\eta} \leq \varepsilon \log n/(4\log\log n)$,

$$\begin{aligned} &\mathbb{P}(H_n \leq (1+\varepsilon)\frac{\log n}{\log\log n}) \\ &\quad\geq \mathbb{P}\big(d_{\mathrm{NR}_n(\boldsymbol{w})}(V_i, \mathrm{Core}_n) \leq \nu_n^{1-\eta}, i = 1,2, d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1', V_2') \leq (1+\varepsilon/2)\frac{\log n}{\log\log n}\big). \end{aligned} \tag{9.4.64}$$

By Proposition 9.24, the probability of the first event converges to ζ^2, and by Proposition 9.24, the probability of the second event converges to 1. We conclude that

$$\mathbb{P}\big(H_n \leq (1+\varepsilon)\frac{\log n}{\log\log n}\big) \to \zeta^2.$$

Since also $\mathbb{P}(H_n < \infty) \to \zeta^2$, this completes the proof. □

The proofs of Propositions 9.23–9.24 follow from path counting techniques similar to the ones carried out above. We now sketch their proofs, starting with Proposition 9.23. For Proposition 9.23, we take

$$a = V_1', \qquad b = V_2', \qquad \mathcal{I} = \{w_i\colon w_i \in [K, \sqrt{\alpha_n}]\}. \tag{9.4.65}$$

The whole point is that there exists a constant $c > 0$ such that

$$\nu_{\mathcal{I}} \geq c \log \alpha_n = c\nu_n^{1-\eta}, \tag{9.4.66}$$

while $u_a \geq \alpha_n/\sqrt{\ell_n}$, $u_b \geq \alpha_n/\sqrt{\ell_n}$, so that

$$\mathbb{E}[N_k(a,b)] \approx \alpha_n^2 c^k \nu_n^{k(1-\eta)}/\ell_n \to \infty \tag{9.4.67}$$

for $k = \log n/((1-\eta)\log \nu_n) \leq (1+\varepsilon/2)\log n/\log \nu_n$ when η is such that $1/(1-\eta) \leq 1+\varepsilon/2$. Further,

$$\gamma_{\mathcal{I}} \leq \sqrt{\alpha_n}/\sqrt{\ell_n}, \tag{9.4.68}$$

so that $\mathrm{Var}(N_k(a,b))/\mathbb{E}[N_k(a,b)]^2 \to 0$ by Proposition 9.17.

Exercise 9.24 (ν_n bound for $\tau = 3$)**.** *Prove that* (9.4.55) *and* (9.4.56) *imply that* $\nu_{\mathcal{I}} \geq c \log \alpha_n$ *by using*

$$\frac{1}{n}\sum_{i\in\mathcal{I}} w_i^2 = \mathbb{E}[W_n^2 \mathbb{1}_{\{W_n \in [K, \sqrt{\alpha_n}]\}}] = 2\int_K^{\sqrt{\alpha_n}} x[F_n(\sqrt{\alpha_n}) - F_n(x)]dx. \tag{9.4.69}$$

Exercise 9.25 (Expected number of paths within Core_n diverges)**.** *Prove that*

$$\mathbb{E}[N_k(a,b)] \to \infty$$

for $a = V_1'$, $b = V_2'$ *and* $k = \log n/((1-\eta)\log \nu_n)$.

Exercise 9.26 (Concentration of number of paths within Core_n)**.** *Prove that*

$$\mathrm{Var}(N_k(a,b))/\mathbb{E}[N_k(a,b)]^2 \to 0$$

for $a = V_1'$, $b = V_2'$ *and* $k = \log n/((1-\eta)\log \nu_n)$.

For Proposition 9.23, we again condition on $\mathcal{N}_m(V_1) \neq \varnothing, \mathcal{N}_m(V_2) \neq \varnothing$, the probability of which converges to ζ^2 when first $n \to \infty$ followed by $m \to \infty$. Then, we perform a second moment method on the number of paths between $\mathcal{N}_m(V_i)$ and Core_n. For this, we take $k = \nu_n^{1-\eta}$ and

$$a = \mathcal{N}_m(V_1), \qquad b = \mathrm{Core}_n, \qquad \mathcal{I} = \{i\colon w_i \leq K\} \setminus (\mathcal{N}_{\leq m}(V_1) \cup \mathcal{N}_{\leq m}(V_2)). \tag{9.4.70}$$

Then we follow the proof in (9.4.51)–(9.4.54) to show that

$$\mathbb{P}(d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Core}_n) > \nu_n^{1-\eta}, \mathcal{N}_m(V_1) \neq \varnothing, \mathcal{N}_m(V_2) \neq \varnothing) \to 0, \tag{9.4.71}$$

as required. Note, for this, that, conditionally on $\mathcal{N}_{\leq m}(V_1), \mathcal{N}_{\leq m}(V_2)$

$$\mathbb{E}[N_k(a,b)] \approx \nu(K)^k \mathcal{W}_m(V_1) \frac{1}{\ell_n} \sum_{i\in\mathrm{Core}_n} w_i, \tag{9.4.72}$$

where $\nu(K) \to \infty$ as $K \to \infty$, and where, by (9.4.55),

$$\frac{1}{n}\sum_{i\in\mathrm{Core}_n} w_i \geq \alpha_n[1 - F_n](\alpha_n) \geq c/\alpha_n. \tag{9.4.73}$$

Therefore, $\mathbb{E}[N_k(a,b)] \to \infty$ as soon as $k \geq 2\log \alpha_n/\log \nu(K)$, which is satisfied for K sufficiently large and $k = \nu_n^{1-\eta}$.

Exercise 9.27 (Completion proof Proposition 9.23)**.** *Complete the proof of Proposition 9.23 by adapting the arguments in* (9.4.51)–(9.4.54).

9.5 General inhomogeneous random graphs

In this section, we introduce the general setting of inhomogeneous random graphs. The inhomogeneous random graph is a generalization of the Erdős-Rényi random graph $\mathrm{ER}_n(p)$ as well as the inhomogeneous random graphs studied in Chapter 6. We start by motivating its choice, which is inspired by Example 6.1.

Example 9.25 (Population of two types (Cont.))**.** Suppose that we have a complex network in which there are n_1 vertices of type 1 and n_2 of type 2. Type 1 individuals have on average m_1 neighbors, type 2 individuals m_2, where $m_1 \neq m_2$. Further, suppose that the probability that a type 1 individual is a friend of a type 2 individual is quite different from the probability that a type 1 individual is a friend of a type 1 individual.

In the model proposed in Example 6.3, the probability that a type i individual is a friend of a type j individual (where $i, j, \in \{1, 2\}$) is equal to $m_i m_j/(\ell_n + m_i m_j)$, where $\ell_n = n_1 m_1 + n_2 m_2$. Approximating this probability by $m_i m_j/\ell_n$, we see that the probability that a type 1 individual is friend of a type 2 individual is quite related to the probability that a type 1 individual is friend of a type 1 individual. Indeed, take two type 1 and two type 2 individuals. Then, the probability that the type 1 individuals are friends and the type 2 individuals are friends is almost the same as the probability that first type 1 individual is friend with the first type 2 individual, and the second type 1 individual is friend of the second type 2 individual. This relation seems quite artificial, and is in many practical situations unwanted. The problem originates in the *product structure* of the edge probabilities in the generalized random graph.

We assume that our individuals have types which are in a certain type space $\mathcal{S}$. When there are individuals of just 2 types, as in Example 9.25, then it suffices to take $\mathcal{S} = \{1, 2\}$. However, the model allows for rather general sets of types of the individuals, both finite as well as (countably or uncountably) infinite. An example of an uncountably infinite type space could be types related to the *ages* of the individuals in the population. We therefore also need to know how many individuals there are of a given type. This is described in terms of a *measure* μ_n, where, for $A \subseteq \mathcal{S}$, $\mu_n(A)$ denotes the proportion of individuals having a type in A.

In our general model, instead of vertex weights, the edge probabilities are moderated by a *kernel* $\kappa\colon \mathcal{S}^2 \to [0, \infty)$. The probability that two individuals of types x_1 and x_2 are friends is approximately $\kappa(x_1, x_2)/n$. Since there are many choices for κ, we arrive at a rather flexible model, where individuals have types and connection probabilities are related to the types of the individuals involved.

We start by making the above definitions formal, by defining what our ground space is and what a kernel is:

Definition 9.26 (Kernel)**.** *(i) A ground space is a pair $(\mathcal{S}, \mu)$, where $\mathcal{S}$ is a separable metric space and μ is a Borel probability measure on $\mathcal{S}$.*

(ii) A vertex space $\mathcal{V}$ is a triple $(\mathcal{S}, \mu, (\boldsymbol{x}_n)_{n\geq 1})$, where $(\mathcal{S}, \mu)$ is a ground space and, for each $n \geq 1$, $\boldsymbol{x}_n$ is a random sequence $(x_1, x_2, \ldots, x_n)$ of n points of $\mathcal{S}$, such that

$$\mu_n(A) = \#\{i\colon x_i \in A\}/n \to \mu(A) \tag{9.5.1}$$

for every μ-continuity set $A \subseteq \mathcal{S}$. The convergence in (9.5.1) *is denoted by $\mu_n \xrightarrow{p} \mu$.*

(iii) A kernel κ is a symmetric non-negative (Borel) measurable function on $\mathcal{S}^2$. By a kernel on a vertex space $(\mathcal{S}, \mu, (\boldsymbol{x}_n)_{n\geq 1})$ we mean a kernel on $(\mathcal{S}, \mu)$.

Before defining the precise random graph model, we state the necessary conditions on our kernels. We write $E(G)$ for the number of edges in a graph G. Note that

$$\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}))] = \sum_{i<j} p_{ij}, \tag{9.5.2}$$

so that our model has bounded degree in expectation precisely when $\frac{1}{n}\sum_{i<j} p_{ij}$ remains bounded. In our applications, we wish that the average degree per vertex in fact *converges*. This explains the main conditions we pose on the kernel κ

Definition 9.27 (Graphical and irreducible kernels)**.** *(i) A kernel κ is* graphical *if the following conditions hold:*

(a) κ is continuous a.e. on $\mathcal{S}^2$;

(b)

$$\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) < \infty; \tag{9.5.3}$$

(c)

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))] \to \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy). \tag{9.5.4}$$

Similarly, a sequence (κ_n) of kernels is called graphical with limit κ *when*

$$y_n \to y \quad \text{and} \quad z_n \to z \quad \text{imply that} \quad \kappa_n(y_n,z_n) \to \kappa(y,z), \tag{9.5.5}$$

where κ satisfies conditions (a) and (b) above, and

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)))] \to \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy). \tag{9.5.6}$$

(ii) A kernel κ is called reducible *if*

$$\exists A \subseteq \mathcal{S} \quad \text{with} \quad 0<\mu(A)<1 \quad \text{such that} \quad \kappa = 0 \quad \text{a.e. on} \quad A\times(\mathcal{S}\backslash A);$$

otherwise κ is irreducible.

We now discuss the above definitions. The assumptions in (9.5.3), (9.5.4), (9.5.6) imply that the expected number of edges is proportional to n, and that the proportionality constant is precisely equal to $\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy)$. Thus, in the terminology of Chapter 1, $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is *sparse*.

Roughly speaking, κ is reducible if the vertex set of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ can be split in two parts so that the probability of an edge from one part to the other is zero, and irreducible otherwise. When κ is reducible, then the random graph splits into two independent random graphs on the two disjoint subsets A and $\mathcal{S}\setminus A$. Therefore, we could have equally started with each of them separately, explaining why the notion of irreducibility is quite natural.

In many cases, we shall take $\mathcal{S}=[0,1]$, $x_i = i/n$ and μ the Lebesgue-measure on $[0,1]$. Then, clearly, (9.5.1) is satisfied. In fact, in [129], it is shown that we can always restrict to $\mathcal{S}=[0,1]$ by suitably adapting the other choices of our model. However, for notational purposes, it is more convenient to work with general $\mathcal{S}$. For example, where $\mathcal{S}=\{1\}$ is just a single type, the model reduces to the Erdős-Rényi random graph, and in the setting where $\mathcal{S}=[0,1]$, this is slightly more cumbersome:

Exercise 9.28 (Erdős-Rényi random graph)**.** *Show that when $\mathcal{S}=[0,1]$ and $p_{ij} = \kappa(i/n, j/n)/n$ with $\kappa\colon [0,1]^2 \to [0,\infty)$ being continuous, then the model is the Erdős-Rényi random graph with edge probability λ/n precisely when $\kappa(x,y)=\lambda$. Is this also true when $\kappa\colon [0,1]^2 \to [0,\infty)$ is not continuous?*

Now we come to the definition of our random graph. Given a kernel κ, for $n \in \mathbb{N}$, we let $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ be the random graph on $[n]$, each possible edge ij, $i,j \in [n]$, is present with probability

$$p_{ij}(\kappa) = p_{ij} = \frac{1}{n}[\kappa(x_i, x_j) \wedge n], \tag{9.5.7}$$

and the events that different edges are present are independent. Similarly, $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is defined with κ_n replacing κ in (9.5.7).

For $\mathrm{CL}_n(\boldsymbol{w})$ with $\boldsymbol{w} = (w_i)_{i\in[n]}$ as in (6.1.10), we take $\mathcal{S} = [0,1], x_i = i/n$ and

$$\kappa_n(x,y) = [1-F]^{-1}(x)[1-F]^{-1}(y)n/\ell_n. \tag{9.5.8}$$

For $\mathrm{CL}_n(\boldsymbol{w})$ with $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfying Condition 6.4, instead, we take $\mathcal{S} = [0,1]$, $x_i = i/n$ and

$$\kappa_n(i/n, j/n) = w_i w_j/\mathbb{E}[W_n]. \tag{9.5.9}$$

We next study the Chung-Lu random graph:

Exercise 9.29 (The Chung-Lu model)**.** *Prove that when κ is given by*

$$\kappa(x,y) = [1-F]^{-1}(x)[1-F]^{-1}(y)/\mathbb{E}[W], \tag{9.5.10}$$

then κ is graphical precisely when $\mathbb{E}[W] < \infty$, where W has distribution function F. Further, κ is always irreducible.

Exercise 9.30 (The Chung-Lu model repeated)**.** *Let $\tilde{w}_i = [1-F]^{-1}(i/n)\sqrt{n\mathbb{E}[W]/\ell_n}$ and $w_i = [1-F]^{-1}(i/n)$ as in (6.1.10). Then $\mathrm{CL}_n(\tilde{\boldsymbol{w}})$ and $\mathrm{CL}_n(\boldsymbol{w})$ are asymptotically equivalent whenever $(\frac{\mathbb{E}[W_n]}{\ell_n} - 1)^2 = o(n)$.*

In [47], also the choices

$$p_{ij}^{(\mathrm{NR})}(\kappa_n) = 1 - \mathrm{e}^{-\kappa_n(x_i,x_j)/n}, \tag{9.5.11}$$

or

$$p_{ij}^{(\mathrm{GRG})}(\kappa_n) = p_{ij} = \frac{\kappa(x_i,x_j)}{n + \kappa(x_i,x_j)} \tag{9.5.12}$$

are investigated. All results in [47] remain valid for the choices in (9.5.11) and (9.5.12). When

$$\sum_{i,j\in[n]} \kappa_n(x_i,x_j)^3 = o(n^{3/2}), \tag{9.5.13}$$

this follows immediately from Theorem 6.17:

Exercise 9.31 (Asymptotic equivalence for general IRGs)**.** *Prove that the random graphs $\mathrm{IRG}_n(\boldsymbol{p})$ with p_{ij} as in (9.5.7) is asymptotically equivalent to $\mathrm{IRG}_n(\boldsymbol{p})$ with $p_{ij} = p_{ij}^{(\mathrm{NR})}(\kappa_n)$ and to $\mathrm{IRG}_n(\boldsymbol{p})$ with $p_{ij} = p_{ij}^{(\mathrm{GRG})}(\kappa_n)$ when (9.5.13) holds.*

In the next section, we discuss some examples of inhomogeneous random graphs.

9.5.1 Examples of inhomogeneous random graphs

The Erdős-Rényi random graph. If $\mathcal{S}$ is general and $\kappa(x,y) = \lambda$ for every $x,y \in \mathcal{S}$, then the edge probabilities p_{ij} given by (9.5.7) are all equal to λ/n (for $n > \lambda$). Then $\mathrm{IRG}_n(\boldsymbol{p}(\kappa)) = \mathrm{ER}_n(\lambda/n)$. The simplest choice here is to take $\mathcal{S} = \{1\}$.

The homogeneous bipartite random graph. Let n be even, ket $\mathcal{S} = \{0, 1\}$, let $x_i = 0$ for $i \in [n/2]$ and $x_i = 1$ for $i \in [n] \setminus [n/2]$. Further, let κ be defined by $\kappa(x, y) = 0$ when $x \neq y$ and $\kappa(x, y) = \lambda$ when $x = y$. Then $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is the random bipartite graph with $n/2$ vertices in each class, where each possible edge between classes is present with probability λ/n, independently of the other edges.

Exercise 9.32 (Definitions 9.26-9.27 for homogeneous bipartite graph)**.** *Prove that Definitions 9.26-9.27 hold for the homogeneous bipartite graph.*

The finite-type case. Fix $r \geq 2$ and suppose we have a graph with r different types of vertices. Let $\mathcal{S} = \{1, \dots, r\}$. Let n_i denote the number of vertices of type i, and let $\mu_n(i) = n_i/n$. Let $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ be the random graph where two vertices of types i and j, respectively, joined by an edge with probability $n^{-1}\kappa(i, j)$ (for $n \geq \max \kappa$). Then κ is equivalent to an $r \times r$ matrix, and the random graph $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ has vertices of r different types (or colors). The finite-types case has been studied by Söderberg [205, 206, 207, 208]. We conclude that our general IRG covers the cases of a finite (or even countably infinite) number of types.

Exercise 9.33 (Homogeneous bipartite graph)**.** *Prove that the homogeneous bipartite random graph is a special case of the finite-types case.*

Exercise 9.34 (Irreducibility for the finite-types case)**.** *Prove that, in the finite-type case, irreducibility follows when there exists an m such that the m^{th} power of the matrix $(\kappa(i, j))_{i,j\in[r]}$ contains no zeros.*

Exercise 9.35 (Graphical limit in the finite-types case)**.** *Prove that, in the finite-type case,* (9.5.1) *holds precisely when*

$$\lim_{n\to\infty} n_i/n = p_i. \tag{9.5.14}$$

9.5.2 Degree sequence of IRG

We next turn to the degrees of the vertices of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. As we shall see, the degree of a vertex of a given type x is asymptotically Poisson with a mean

$$\lambda(x) = \int_{\mathcal{S}} \kappa(x, y)\mu(dy) \tag{9.5.15}$$

that depends on x. This leads to a mixed Poisson distribution for the degree D of a (uniformly chosen) random vertex of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. We recall that N_k denotes the number of vertices of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ with degree k.

Theorem 9.28 (The degree sequence of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$)**.** *Let (κ_n) be a graphical sequence of kernels with limit κ. For any fixed $k \geq 0$,*

$$N_k/n \xrightarrow{\mathbb{P}} \int_{\mathcal{S}} \frac{\lambda(x)^k}{k!} \mathrm{e}^{-\lambda(x)} \mu(dx) \tag{9.5.16}$$

where $x \mapsto \lambda(x)$ is defined by

$$\lambda(x) = \int_{\mathcal{S}} \kappa(x, y)\mu(dy). \tag{9.5.17}$$

Equivalently,

$$N_k/n \xrightarrow{\mathbb{P}} \mathbb{P}(\Xi = k), \tag{9.5.18}$$

where Ξ has the mixed Poisson distribution with distribution W_λ given by

$$\mathbb{P}(W_\lambda \le x) = \int_0^x \lambda(y)\mu(dy). \tag{9.5.19}$$

In the remainder of this section, we shall give a proof of Theorem 9.28. We start by proving Theorem 9.28 for the finite-types case, which is substantially easier. After this, we give a proof in the general case, for which we shall need to prove results on approximations of sequences of graphical kernels.

Proof of Theorem 9.28 in the finite-types case. Consider first the finite-type case. Take a vertex v of type i, let D_v be its degree, and let $D_{v,j}$ be the number of edges from v to vertices of type $j \in [r]$. Then, clearly, $D_v = \sum_j D_{v,j}$.

Recall that, in the finite-types case, the edge probability between vertices of types i and j is denoted by $(\kappa(i,j) \wedge n)/n$. Assume that $n \ge \max \kappa$. The random variables $(D_{v,j})_{j\in[r]}$ are independent, and $D_{v,j} \sim \mathrm{Bin}(n_j - \delta_{ij}, \kappa(i,j)/n) \xrightarrow{d} \mathrm{Poi}(\mu_j \kappa(i,j))$, where n_j are the number of vertices with type j and $\mu_j = \lim_{n\to\infty} n_j/n$. Hence,

$$D_v \xrightarrow{d} \mathrm{Poi}\left(\sum_j \mu_j \kappa(i,j)\right) = \mathrm{Poi}(\lambda(i)), \tag{9.5.20}$$

where $\lambda(i) = \int \kappa(i,j) d\mu(j) = \sum_j \kappa(i,j)\mu_j$. Consequently,

$$\mathbb{P}(D_v = k) \to \mathbb{P}(\mathrm{Poi}(\lambda(i)) = k) = \frac{\lambda(i)^k}{k!} \mathrm{e}^{-\lambda(i)}. \tag{9.5.21}$$

Let $N_{k,i}$ be the number of vertices in $G^{\mathcal{V}}(n,k)$ of type i with degree k. Then, for fixed $n_1, \dots, n_r$,

$$\frac{1}{n}\mathbb{E}[N_{k,i}] = \frac{1}{n} n_i \mathbb{P}(D_v = k) \to \mu_i \mathbb{P}(\mathrm{Poi}(\lambda(i)) = k). \tag{9.5.22}$$

It is easily checked that $\mathrm{Var}(N_{k,i}) = O(n)$. Hence,

$$\frac{1}{n} N_{k,i} \xrightarrow{\mathbb{P}} \mathbb{P}(\mathrm{Poi}(\lambda(i)) = k)\mu_i = \mathbb{P}(\Xi = k), \tag{9.5.23}$$

and thus, summing over i,

$$\frac{1}{n} N_k = \sum_i \frac{1}{n} N_{k,i} \xrightarrow{\mathbb{P}} \sum_i \mathbb{P}(\mathrm{Poi}(\lambda(i)) = k)\mu_i = \mathbb{P}(\Xi = k). \tag{9.5.24}$$

This proves Theorem 9.28 in the regular finitary case. □

In order to prove Theorem 9.28 in the general case, we shall be approximating a sequence of graphical kernels (κ_n) by appropriate regular finite kernels:

Approximations of sequences of graphical kernels. Recall that $\mathcal{S}$ is a separable metric space, and that μ is a Borel measure on $\mathcal{S}$ with $0 < \mu(\mathcal{S}) < \infty$. We usually assume that $\mu(\mathcal{S}) = 1$; in this section, this makes no difference. Here the metric and topological structure of $\mathcal{S}$ will be important.

Given a sequence of finite partitions $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$, $m \ge 1$, of $\mathcal{S}$ and an $x \in \mathcal{S}$, we define $i_m(x)$ by requiring that

$$x \in A_{m,i_m(x)}. \tag{9.5.25}$$

As usual, for $A \subset \mathcal{S}$ we write $\mathrm{diam}(A)$ for $\sup\{|x-y| \colon x, y \in A\}$. By taking $\mathcal{P}_m$ as the dyadic partition into intervals of length 2^{-m} in $\mathcal{S}$, we easily see the following:

Lemma 9.29 (Approximating partition)**.** *There exists a sequence of finite partitions $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$, $m \ge 1$, of $\mathcal{S}$ such that*

(i) each A_{mi} is measurable and $\mu(\partial A_{mi}) = 0$;

(ii) for each m, $\mathcal{P}_{m+1}$ refines $\mathcal{P}_m$, i.e., each A_{mi} is a union $\bigcup_{j\in J_{mi}} A_{m+1,j}$ for some set J_{mi};

(iii) for a.e. $x \in \mathcal{S}$, $\mathrm{diam}(A_{m,i_m(x)}) \to 0$ as $m \to \infty$, where $i_m(x)$ is defined by (9.5.25)

Recall that a kernel κ is a symmetric measurable function on $\mathcal{S}\times\mathcal{S}$. Fixing a sequence of partitions with the properties described in Lemma 9.29, we can define sequences of lower and upper approximations to κ by

$$\kappa_m^-(x,y) = \inf\{\kappa(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}\}, \tag{9.5.26}$$

$$\kappa_m^+(x,y) = \sup\{\kappa(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}\}. \tag{9.5.27}$$

We thus replace κ by its infimum or supremum on each $A_{mi}\times A_{mj}$. As κ_m^+ might be $+\infty$, we shall use it only for bounded κ.

By Lemma 9.29(ii),

$$\kappa_m^- \le \kappa_{m+1}^- \quad \text{and} \quad \kappa_m^+ \ge \kappa_{m+1}^+. \tag{9.5.28}$$

Furthermore, if κ is continuous a.e. then, by Lemma 9.29(iii),

$$\kappa_m^-(x,y) \to \kappa(x,y) \quad \text{and} \quad \kappa_m^+(x,y) \to \kappa(x,y) \quad \text{for a.e. } (x,y) \in \mathcal{S}^2. \tag{9.5.29}$$

Since $k_m^- \le \kappa$, we can obviously construct our random graph so that $\mathrm{IRG}_n(\kappa_m^-) \subseteq \mathrm{IRG}_n(\kappa)$, and in the sequel we shall assume this. Similarly, we shall assume that $\mathrm{IRG}_n(\kappa_m^+) \supseteq \mathrm{IRG}_n(\kappa)$ when κ is bounded.

If (κ_n) is a graphical sequence of kernels with limit κ, we define instead

$$\kappa_m^-(x,y) := \quad \inf\{(\kappa\wedge\kappa_n)(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}, n \ge m\}, \tag{9.5.30}$$

$$\kappa_m^+(x,y) := \quad \sup\{(\kappa\vee\kappa_n)(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}, n \ge m\}. \tag{9.5.31}$$

By Lemma 9.29, we have $\kappa_m^- \le \kappa_{m+1}^-$, and from Lemma 9.29(iii) and 9.27(ii) we see that

$$\kappa_m^-(x,y) \nearrow \kappa(x,y) \quad \text{as } m\to\infty, \text{ for a.e. } (x,y) \in \mathcal{S}^2. \tag{9.5.32}$$

Moreover, when $n \ge m$, we have

$$\kappa_n \ge \kappa_m^-, \tag{9.5.33}$$

and we may assume that $\mathrm{IRG}_n(\kappa_m^-) \subseteq \mathrm{IRG}_n(\kappa_n)$. By the convergence of the sequence of kernels (κ_n), we further obtain that also the number of edges converges:

Now we are ready to complete the proof of Theorem 9.28 for general sequences of graphical kernels (κ_n). Define κ_m^- by (9.5.30). Let $\varepsilon > 0$ be given. From (9.5.6) and monotone convergence, there is an m such that

$$\iint_{\mathcal{S}^2} \kappa_m^-(x,y)\mu(dx)\mu(dy) > \iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) - \varepsilon. \tag{9.5.34}$$

For $n \ge m$, we have $\kappa_m^- \le \kappa_n$ by (9.5.33), so we may assume that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-)) \subseteq \mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. Then, by (9.5.6) and (9.5.34),

$$\begin{aligned} &\frac{1}{n}E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)) \setminus \mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-))) \\ &= \frac{1}{n}E(G(n,\kappa_n)) - \frac{1}{n}E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-))) \\ &\xrightarrow{\mathbb{P}} \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) - \frac{1}{2}\iint_{\mathcal{S}^2} \kappa_m^-(x,y)\mu(dx)\mu(dy) < \frac{\varepsilon}{2}, \end{aligned} \tag{9.5.35}$$

so that, **whp** $E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)) \setminus \mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-))) < \varepsilon n$. Let us write $N_k^{(m)}$ for the number of vertices of degree k in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-))$. It follows that **whp**

$$|N_k^{(m)} - N_k| < 2\varepsilon n \tag{9.5.36}$$

Writing $\Xi^{(m)}$ for the equivalent of Ξ defined using κ_m^- in place of κ, by the proof for the regular finitary case, $N_k^{(m)}/n \xrightarrow{\mathbb{P}} \mathbb{P}(\Xi^{(m)} = k)$. Thus, **whp**,

$$|N_k^{(m)}/n - \mathbb{P}(\Xi^{(m)} = k)| < \varepsilon. \tag{9.5.37}$$

Finally, we have $\mathbb{E}[\Xi] = \int_{\mathcal{S}} \lambda(x)\mu(dx) = \iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy)$. Since $\lambda^{(m)}(x) \le \lambda(x)$, we can assume that $\Xi^{(m)} \le \Xi$, and thus

$$\begin{aligned}\mathbb{P}(\Xi \neq \Xi^{(m)}) &= \mathbb{P}(\Xi - \Xi^{(m)} \ge 1)\\ &\le \mathbb{E}[\Xi - \Xi^{(m)}] = \iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) - \iint_{\mathcal{S}^2} \kappa_m^-(x,y)\mu(dx)\mu(dy) < \varepsilon.\end{aligned} \tag{9.5.38}$$

Combining (9.5.36), (9.5.37) and (9.5.38), we see that $|N_k/n - \mathbb{P}(\Xi = k))| < 4\varepsilon$ **whp**. □

Let Λ be the random variable $\lambda(U)$, where U is a random variable on $\mathcal{S}$ having distribution μ. Then we can also describe the mixed Poisson distribution of Ξ as $\mathrm{Poi}(\Lambda)$. Under mild conditions, the tail probabilities $\mathbb{P}(\Xi > t)$ and $\mathbb{P}(\Lambda > t)$ are similar for large t. We state this for the case of power-law tails; the result generalizes to regularly varying tails. As above, let D be the degree of a random vertex in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. Let $N_{\ge k}$ be the number of vertices with degree at least k.

Corollary 9.30 (Power-law tails for the degree sequence)**.** *Let (κ_n) be a graphical sequence of kernels with limit κ. Suppose that $\mathbb{P}(\Lambda > t) = \mu\{x\colon \lambda(x) > t\} \sim at^{-(\tau-1)}$ as $t \to \infty$, for some $a > 0$ and $\tau > 2$. Then*

$$N_{\ge k}/n \xrightarrow{\mathbb{P}} \mathbb{P}(\Xi \ge k) \sim ak^{-(\tau-1)}, \tag{9.5.39}$$

where the first limit is for k fixed and $n \to \infty$, and the second for $k \to \infty$. In particular, $\lim_{n\to\infty} \mathbb{P}(D \ge k) \sim ak^{-(\tau-1)}$ as $k \to \infty$.

Proof. It suffices to show that $\mathbb{P}(\Xi \ge k) \sim ak^{-(\tau-1)}$; the remaining conclusions then follow from Theorem 9.28. For any $\varepsilon > 0$, $\mathbb{P}(\mathrm{Poi}(\Lambda) > t|\Lambda > (1+\varepsilon)t) \to 1$ and $\mathbb{P}(\mathrm{Poi}(\Lambda) > t|\Lambda < (1-\varepsilon)t) = o(t^{-(\tau-1)})$ as $t \to \infty$. It follows that $\mathbb{P}(\Xi > t) = \mathbb{P}(\mathrm{Poi}(\Lambda) > t) \sim at^{-(\tau-1)}$ as $t \to \infty$. □

This result shows that the general IRG does include natural cases with power-law degree distributions. Naturally, we have already observed in Theorem 6.6 that this is the case for the $\mathrm{GRG}_n(\boldsymbol{w})$ when the weights sequence $\boldsymbol{w}$ is chosen appropriately.

9.5.3 Multitype branching processes

In order to study further properties of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$, we need to understand the neighborhood structure of vertices. For simplicity, let us restrict ourselves to the finite types case. Then, for a vertex of type i, the number of neighbors of type j is close to Poisson distributed with approximate mean $\kappa(i,j)$. Even when we assume independence of the neighborhood structures of different vertices, we still do not arrive at a classical branching process as discussed in Chapter 3. Instead, we can describe the neighborhood structure with a branching process in which we keep track of the type of each of the vertices. For

general κ and μ, we can even have a continuum of types. Such branching processes are called *multitype branching processes*. See e.g. [16, Chapter V] or [115, Chapter III] for more background on multitype branching processes. In this section, we shall only discuss the basics and we shall quickly go to the special case of multitype branching processes where every offspring has a Poisson distribution.

Multitype branching processes with finitely many types. Multitype branching process can be analyzed using linear algebra. In order to do so, we first introduce some notation. We first assume that we are in the finite types case, and denote the number of types by r. We let $\boldsymbol{j} = (j_1, \ldots, j_r) \in \mathbb{N}_0^r$ be a vector of non-negative integers, and denote by $p_{\boldsymbol{j}}^{(i)}$ the probability that an individual of type i gives rise to an offspring $\boldsymbol{j}$, i.e., j_1 children of type 1, j_2 children of type 2, etc. The offsprings of the different individuals are all mutually independent. Denote by $Z_{n,j}^{(i)}$ the number of individuals of type j in generation n when starting from a single particle of type i and $\mathbf{Z}_n^{(i)} = (Z_{n,1}^{(i)}, \ldots, Z_{n,j}^{(i)})$. We shall be interested in the survival or extinction of multitype branching processes, and in the growth of the generation sizes. In the multitype case, we are naturally lead to a matrix setup. We now discuss the survival versus extinction of multitype branching processes. We denote the survival probability of the multitype branching process when starting from a single individual of type i by

$$q^{(i)} = \mathbb{P}(\mathbf{Z}_n^{(i)} \neq \mathbf{0} \text{ for all } n), \tag{9.5.40}$$

and we let $\boldsymbol{q} = (q^{(1)}, \ldots, q^{(r)})$. Our first aim is to investigate when $\boldsymbol{q} = \mathbf{0}$.

Multitype branching processes and generating functions. We write $\boldsymbol{p}(\boldsymbol{j}) = (p_{\boldsymbol{j}}^{(1)}, \ldots, p_{\boldsymbol{j}}^{(r)})$ and we let

$$G^{(i)}(\boldsymbol{s}) = \sum_{\boldsymbol{j}} p_{\boldsymbol{j}}^{(i)} \prod_{a=1}^{r} s_a^{j_a} \tag{9.5.41}$$

be the joint moment generating function of the offspring of an individual of type i. We write $\mathbf{G}(\boldsymbol{s}) = (G^{(1)}(\boldsymbol{s}), \ldots, G^{(r)}(\boldsymbol{s}))$ for the vector of generating functions. We now generalize Theorem 3.1 to the multitype case. Let $\boldsymbol{q}$ satisfy $\boldsymbol{q} = \mathbf{1} - \mathbf{G}(\mathbf{1} - \boldsymbol{q})$. By convexity of $\boldsymbol{s} \mapsto \mathbf{G}(\boldsymbol{s})$, there is at most one non-zero solution to the equation $\boldsymbol{s} = \mathbf{G}(\boldsymbol{s})$ which is not equal to $\mathbf{0}$. Define

$$G_n^{(i)}(\boldsymbol{s}) = \mathbb{E}[\prod_{a=1}^{r} s_a^{Z_{n,a}^{(i)}}], \tag{9.5.42}$$

and $\mathbf{G}_n(\boldsymbol{s}) = (G_n^{(1)}(\boldsymbol{s}), \ldots, G_n^{(r)}(\boldsymbol{s}))$. Then, we have that $\mathbf{G}_{n+1}(\boldsymbol{s}) = \mathbf{G}_n(\mathbf{G}(\boldsymbol{s})) = \mathbf{G}(\mathbf{G}_n(\boldsymbol{s}))$ and $\boldsymbol{q} = \mathbf{1} - \lim_{n\to\infty} \mathbf{G}_n(\mathbf{0})$. Naturally, the extinction probability depends sensitively on the type of the ancestor of the branching process. On the other hand, under reasonable assumptions, the positivity of the survival probability is independent of the initial type. A necessary and sufficient condition for this property is that, with positive probability, a particle of type i arises as a descendent of of a particle of type j for each type i and j.

We note that when $\mathbf{G}(\boldsymbol{s}) = \mathbf{M}\boldsymbol{s}$ for some matrix $\mathbf{M}$, then each individual in the Markov chain has precisely one offspring, and we call this case *singular*. When each particle has precisely one offspring, the multitype branching process is equivalent to a Markov chain, and the process a.s. survives. Thus, in this case, there is no survival vs. extinction phase transition. We shall assume throughout the remainder that the multitype branching process is non-singular.

Survival vs. extinction and mean offspring. We continue to describe the survival versus extinction of multitype branching processes in terms of the mean offspring. Let κ_{ij} denote the expected offspring of type j of a single individual of type i, and let $\mathbf{T}_\kappa =$

$\{\kappa_{ij}\}_{i,j=1}^r$ be the matrix of offsprings. We shall assume that there exists an l such that the matrix $\mathbf{M}^l$ has only strictly positive entries. This is sometimes called *irreducibility*, as it implies that the Markov chain of the number of individuals of the various types is an irreducible Markov chain. By the Perron-Frobenius theorem, the matrix $\mathbf{M}$ has a unique largest eigenvalue $\|\mathbf{T}_\kappa\|$ with non-negative left-eigenvector $\mathbf{x}_\kappa$, and the eigenvalue $\|\mathbf{T}_\kappa\|$ can be computed as

$$\|\mathbf{T}_\kappa\| = \sup_{\mathbf{x}:\, \|\mathbf{x}\|_2 \le 1} \|\mathbf{T}_\kappa \mathbf{x}\|_2, \quad \text{where} \quad \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^r x_i^2}. \tag{9.5.43}$$

We note that

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)} | \mathbf{Z}_n^{(i)} = \mathbf{z}] = \mathbf{T}_\kappa \mathbf{z}, \tag{9.5.44}$$

so that

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)}] = \mathbf{T}_\kappa^n \mathbf{e}^{(i)}. \tag{9.5.45}$$

where $\mathbf{T}_\kappa^n$ denotes the n-fold application of the matrix $\mathbf{T}_\kappa$, and $\mathbf{e}^{(i)}$ is the vector which has on the i^{th} position a 1, and further only zeroes. The identifications in (9.5.44) and (9.5.45) have several important consequences concerning the phase transition of multitype branching processes, as we shall now discuss in more detail.

First, when $\|\mathbf{T}_\kappa\| < 1$, we have that

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)}] \le \|\mathbf{T}_\kappa\|^n \|\mathbf{e}^{(i)}\|_2, \tag{9.5.46}$$

which converges to 0 exponentially fast. Therefore, by the Markov inequality (Theorem 2.14), the multitype branching process dies out a.s. When $\|\mathbf{T}_\kappa\| > 1$, on the other hand, the sequence

$$M_n = \mathbf{x}_\kappa \mathbf{Z}_{n+1}^{(i)} \|\mathbf{T}_\kappa\|^{-n} \tag{9.5.47}$$

is a non-negative martingale, by (9.5.44) and the fact that $\mathbf{x}_\kappa$ is a left-eigenvector with eigenvalue $\|\mathbf{T}_\kappa\|$, since $\mathbf{x}_\kappa \mathbf{T}_\kappa = \|\mathbf{T}_\kappa\| \mathbf{x}_\kappa$. By the Martingale convergence theorem (Theorem 2.21), the martingale M_n converges a.s. When we further assume some further restrictions on M_n, for example that M_n has finite second moment, then we obtain that $M_n \xrightarrow{a.s.} M_\infty$ and $\mathbb{E}[M_n] \to \mathbb{E}[M_\infty]$. More precisely, there is a multitype analog of the Kesten-Stigum Theorem (Theorem 3.10). Since $\mathbb{E}[M_n] = \mathbb{E}[M_0] = \mathbf{x}_\kappa \mathbf{e}^{(i)} > 0$, we thus have that $\mathbf{Z}_{n+1}^{(i)}$ grows exponentially with a strictly positive probability, which implies that the survival probability is positive. Theorem 3.1 can be adapted to show that $\mathbf{Z}_{n+1}^{(i)} \xrightarrow{\mathbb{P}} \mathbf{0}$ when $\|\mathbf{T}_\kappa\| = 1$. See e.g. [115, Sections II.6-II.7]. We conclude that, for non-singular and irreducible multitype branching processes, we have that $\boldsymbol{q} > \mathbf{0}$ precisely when $\|\mathbf{T}_\kappa\| > 1$.

Poisson multitype branching processes. We call a multitype branching processes *Poisson* when all the number of children of each type are independent Poisson random variables. Thus, $\mathbf{Z}^{(i)} = (Z_{1,1}^{(i)}, \ldots, Z_{1,r}^{(i)})$ is a vector of independent Poisson random variables with means $(\kappa_{1,i}, \ldots, \kappa_{r,i})$. As we shall see later, Poisson multitype branching processes arise naturally when exploring a component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ starting at a vertex of type x. This is directly analogous to the use of the single-type Poisson branching process in the analysis of the Erdős-Rényi graph $\mathrm{ER}_n(\lambda/n)$ as discussed in detail in Chapters 4 and 5.

For Poisson multitype branching processes, we obtain that

$$G^{(i)}(\boldsymbol{s}) = \mathbb{E}[\prod_{a=1}^r s_a^{Z_{1,a}^{(i)}}] = \mathrm{e}^{\sum_{a=1}^r \kappa_{a,i}(s_a - 1)} = \mathrm{e}^{(\mathbf{T}_\kappa(\boldsymbol{s}-\mathbf{1}))_i}. \tag{9.5.48}$$

Thus, the vector of survival probabilities $\boldsymbol{q}$ satisfies

$$\boldsymbol{q} = \mathbf{1} - \mathrm{e}^{-\mathbf{T}_\kappa \boldsymbol{q}}. \tag{9.5.49}$$

This leads us to the investigation of eigenfunctions of non-linear operators of the form $f \mapsto 1-\mathrm{e}^{-\mathbf{T}_\kappa f}$. We now extend the above setting of finite-type Poisson multitype branching processes to the infinite type case.

Poisson multitype branching processes with infinitely many types. Let κ be a kernel. We define the Poisson multitype branching processes with kernel κ as follows. Each individual of type $x \in \mathcal{S}$ is replaced in the next generation by a set of individuals distributed as a Poisson process on $\mathcal{S}$ with intensity $\kappa(x,y)\mu(dy)$. Thus, the number of children with types in a subset $A \subseteq \mathcal{S}$ has a Poisson distribution with mean $\int_A \kappa(x,y)\mu(dy)$, and these numbers are independent for disjoint sets A and for different particles; see e.g., Kallenberg [138].

Let $\zeta_\kappa(x)$ be the survival probability of the Poisson multitype branching process with kernel κ, starting from an ancestor of type $x \in \mathcal{S}$. Set

$$\zeta_\kappa = \int_{\mathcal{S}} \zeta_\kappa(x)\mu(dx). \tag{9.5.50}$$

Again, it can be seen in a similar way as above that $\zeta\kappa > 0$ if and only if $\|\mathbf{T}_\kappa\| > 1$, where now the linear operator $\mathbf{T}_\kappa$ is defined, for $f\colon \mathcal{S} \to \mathbb{R}$,

$$(\mathbf{T}_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x,y) f(y)\mu(dy), \tag{9.5.51}$$

for any (measurable) function f such that this integral is defined (finite or $+\infty$) for a.e. $x \in \mathcal{S}$.

Note that $\mathbf{T}_\kappa f$ is defined for every $f \geq 0$, with $0 \leq \mathbf{T}_\kappa f \leq \infty$. If $\kappa \in L^1(\mathcal{S}\times\mathcal{S})$, as we shall assume throughout, then $\mathbf{T}\kappa f$ is also defined for every bounded f. In this case $\mathbf{T}_\kappa f \in L^1(\mathcal{S})$ and thus $\mathbf{T}_\kappa f$ is finite a.e.

As we shall see, the analysis of multitype branching processes with a possibly uncountable number of types is a bit more functional analytic. Similarly to the finite-type case in (9.5.43), we define

$$\|\mathbf{T}_\kappa\| = \sup\Big\{\|\mathbf{T}_\kappa f\|_2 \colon f \geq 0, \|f\|_2 \leq 1\Big\} \leq \infty. \tag{9.5.52}$$

When finite, $\|T_\kappa\|$ is the norm of $\mathbf{T}_\kappa$ as an operator on $L^2(\mathcal{S})$; it is infinite if $\mathbf{T}_\kappa$ does not define a bounded operator on L^2. The norm $\|\mathbf{T}_\kappa\|$ is at most the Hilbert-Schmidt norm of $\mathbf{T}_\kappa$:

$$\|\mathbf{T}_\kappa\| \leq \|\mathbf{T}_\kappa\|_{\mathrm{HS}} = \|\kappa\|_{L^2(\mathcal{S}\times\mathcal{S})} = \left(\iint_{\mathcal{S}^2} \kappa(x,y)^2 \mu(dx)\mu(dy)\right)^{1/2}. \tag{9.5.53}$$

We also define the non-linear operator Φ_κ by

$$\big(\Phi_\kappa f\big)(x) = 1 - \mathrm{e}^{-(\mathbf{T}_\kappa f)(x)}, \qquad x \in \mathcal{S}, \tag{9.5.54}$$

for $f \geq 0$. Note that for such f we have $0 \leq \mathbf{T}_\kappa f \leq \infty$, and thus $0 \leq \Phi_\kappa f \leq 1$. We shall characterize the survival probability $\zeta_\kappa(x)$, and thus ζ_κ, in terms of the non-linear operator Φ_κ, showing essentially that the function $x \mapsto \zeta_\kappa(x)$ is the maximal fixed point of the non-linear operator Φ_κ (recall (9.5.49)). Again, the survival probability satisfies that $\zeta_\kappa > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$, recall the finite-types case discussed in detail above.

We call a multitype branching process *supercritical* when $\|\mathbf{T}_\kappa\| > 1$, *critical* when $\|\mathbf{T}_\kappa\| < 1$, and *subcritical* when $\|\mathbf{T}_\kappa\| < 1$. Then, the above discussion can be summarized by saying that a multitype branching process survives with positive probability precisely when it is supercritical.

9.5.4 The phase transition for inhomogeneous random graphs

In this section, we discuss the phase transition in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. The main result shows that there is a giant component when the associated multitype branching process is supercritical, while otherwise there is not:

Theorem 9.31 (Giant component of IRG)**.** *Let (κ_n) be a sequence of irreducible graphical kernels with limit κ, and let $\mathcal{C}_{\max}$ denote the largest connected component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. Then,*

$$|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\kappa. \tag{9.5.55}$$

In all cases $\zeta_\kappa < 1$, while $\zeta_\kappa > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$.

Theorem 9.31 is a generalization of the law of large numbers for the largest connected component in Theorem 4.8 for $\mathrm{ER}_n(\lambda/n)$:

Exercise 9.36 (LLN for $\mathcal{C}_{\max}$ for $\mathrm{ER}_n(\lambda/n)$)**.** *Prove that, for the Erdős-Rényi random graph, Theorem 9.31 implies that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\lambda$, where ζ_λ is the survival probability of a Poisson branching process with mean λ offspring.*

We close this section by discussing a few examples of Theorem 9.31.

The bipartite random graph. We let n be even and take $\mathcal{S} = \{1, 2\}$ and

$$\kappa_n(x, y) = \kappa(x, y) = \lambda \mathbb{1}_{\{x \neq y\}}/2. \tag{9.5.56}$$

Thus, for $i < j$, the edge probabilities p_{ij} given by (9.5.7) are equal to $\lambda/(2n)$ (for $2n > \lambda$) when $i \in [n/2]$ and $j \in [n] \setminus [n/2]$.

In this case, $\|\mathbf{T}_\kappa\| = \lambda$ with corresponding eigenfunction $f(x) = 1$ for all $x \in \mathcal{S}$. Thus, Theorem 9.31 proves that there is a phase transition at $\lambda = 2$. Furthermore, the function $\zeta_\lambda(x)$ reduces to the single value $\zeta_{\lambda/2}$, which is the survival probability of a Poisson branching process with mean offspring $\lambda/2$. This is not surprising, since the degree of each vertex is $\mathrm{Bin}(n/2, \lambda/n)$, so the bipartite random graph of size n is quite closely related the Erdős-Rényi random graph of size $n/2$.

The finite-type case. The bipartite random graph can also be viewed as a random graph with two types of vertices (i.e., the vertices $[n/2]$ and $[n] \setminus [n/2]$). We now generalize the results to the finite-type case, in which we have seen that κ_n is equivalent to an $r \times r$-matrix $(\kappa_n(i, j))_{i,j \in [r]}$, where r denotes the number of types. In this case, $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ has vertices of r different types (or colors), say n_i vertices of type i, with two vertices of type i and j joined by an edge with probability $n^{-1}\kappa_n(i, j)$ (for $n \geq \max \kappa_n$). This case has been studied by Söderberg [205, 206, 208, 207], who noted Theorem 9.31 in this case.

Exercise 9.37 (Phase transition for $r = 2$)**.** *Compute ζ_κ in the case of two types, and give necessary and sufficient conditions for $\zeta_\kappa > 0$.*

Exercise 9.38 (The size of small components in the finite-types case)**.** *Prove that, in the finite-types case, when (κ_n) converges, then $\sup_{x,y,n} \kappa_n(x, y) < \infty$ holds, so that the results of Theorem 9.33 apply in the sub- and supercritical cases.*

The random graph with prescribed expected degrees. We next consider the Chung-Lu model or expected degree random graph, where κ_n is given by (9.5.9), i.e., $\kappa_n(i/n, j/n) = w_i w_j/\mathbb{E}[W_n]$.

We first assume that Condition 6.4(a)-(c) hold, so that in particular $\mathbb{E}[W^2] < \infty$, where W has distribution function F. A particular instance of this case is the choice $w_i = [1-F]^{-1}(i/n)$ in (6.1.10). In this case, the sequence (κ_n) converges to κ, where the limit κ is given by (recall (9.5.10))

$$\kappa(x,y) = \psi(x)\psi(y)/\mathbb{E}[W], \tag{9.5.57}$$

where $\psi(x) = [1-F]^{-1}(x)$ Then, we note that for each $f \geq 0$ with $\|f\|_2 = 1$,

$$(\mathbf{T}_\kappa f)(x) = \psi(x)\frac{\int_{\mathcal{S}} \psi(x)f(x)\mu(dx)}{\int_{\mathcal{S}} \psi(x)\mu(dx)}, \tag{9.5.58}$$

so that $\|\mathbf{T}_\kappa f\|_2 = \int_{\mathcal{S}} \psi(x)f(x)\mu(dx)/\int_{\mathcal{S}} \psi(x)\mu(dx)$, which is maximal when $f(x) = \psi(x)/\|\psi\|_2$. We conclude that $\|\mathbf{T}_\kappa\| = \|\psi\|_2^2/\int_{\mathcal{S}} \psi(x)\mu(dx) = \mathbb{E}[W^2]/\mathbb{E}[W]$. Thus,

$$\|\mathbf{T}_\kappa\| = \mathbb{E}[W^2]/\mathbb{E}[W], \tag{9.5.59}$$

and we recover the results in [69, 72] in the case where $\mathbb{E}[W^2] < \infty$. In the case where $\mathbb{E}[W^2] = \infty$, on the other hand, we see that $\|\mathbf{T}_\kappa f\|_2 = \infty$ for every f with $\|f\|_2 = 1$ such that $\int_{\mathcal{S}} \psi(x)f(x)\mu(dx) \neq 0$, so that $\|\mathbf{T}_\kappa\| = \infty$, so that $\mathrm{CL}_n(\boldsymbol{w})$ is always supercritical in this regime.

Exercise 9.39 (The size of small components for $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Prove that, for* $\mathrm{CL}_n(\boldsymbol{w})$ *with weights given by* (6.1.10) *and with* $1 < \nu < \infty$*, the second largest cluster has size* $|\mathcal{C}_{(2)}| = O_{\mathbb{P}}(\log n)$ *when* W *has bounded support or is a.s. bounded below by* $\varepsilon > 0$*, while if* $\nu < 1$*,* $|\mathcal{C}_{\max}| = O(\log n)$ *when* W *has bounded support. Here* W *is a random variable with distribution function* F*.*

9.5.5 Small-world effect in inhomogeneous random graphs

In this section, we consider the distances between vertices of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ where, as usual, (κ_n) is a graphical sequence of kernels with limit κ.

Let us write $\mathrm{dist}_G(i,j)$ for the graph distance between the vertices $i, j \in [n]$ in a graph G, where the graph distance is the minimum number of edges in the graph G that form a path from i to j, and, by convention, we let $\mathrm{dist}_G(i,j) = \infty$ when i, j are in different connected components.

We define the *typical graph distance* to be H_n, where $H_n = \mathrm{dist}_G(V_1, V_2)$ is the graph distance between two vertices V_1, V_2 which are chosen uniformly at random from $[n]$.

We start by discussing logarithmic asymptotics of the typical graph distance in the case where $\nu = \|\mathbf{T}_\kappa\| \in (1, \infty)$. When $\|\mathbf{T}_\kappa\| = \infty$, then our results also prove that $H_n = o_{\mathbb{P}}(\log n)$, but they do not tell us much about their exact asymptotics.

Logarithmic asymptotics of typical graph distance in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$**.** The main result on typical graph distances in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is as follows:

Theorem 9.32 (Typical distances in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$)**.** *Let* (κ_n) *be graphical sequence of kernels with limit* κ*, and with* $\nu = \|\mathbf{T}_\kappa\| \in (1, \infty)$*. Let* $\varepsilon > 0$ *be fixed. Then, for* $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$*,*

(*i*) *If* $\sup_{x,y,n} \kappa_n(x,y) < \infty$*, then*

$$\mathbb{P}(H_n \leq (1-\varepsilon)\log_\nu n) = o(1). \tag{9.5.60}$$

(ii) If κ is irreducible, then

$$\mathbb{P}(H_n \le (1+\varepsilon)\log_\nu n) = \zeta_\kappa^2 + o(1). \tag{9.5.61}$$

In the terminology of Section 1.2, Theorem 9.32(ii) implies that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is a small world when κ is irreducible and $\nu = \|\mathbf{T}_\kappa\| < \infty$. Theorem 9.32(i) shows that the graph distances are of order $\Theta(\log n)$ when $\sup_{x,y,n} \kappa_n(x,y) < \infty$, so that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is not an ultra-small world. The intuition behind Theorem 9.32 is that, by (9.5.45) and (9.5.47), a Poisson multitype branching process with kernel κ has neighborhoods that grow exponentially, i.e., the number of vertices at distance k grows like $\|\mathbf{T}_\kappa\|^k$. Thus, if we examine the distance between two vertices V_1 and V_2 chosen uniformly at random from $[n]$, then we need to explore the neighborhood of vertex V_1 up to the moment that it 'catches' vertex V_2. In this case, the neighborhood must be of size $\sim n$, so that we need that $\|\mathbf{T}_\kappa\|^k = \nu^k \sim n$, i.e., $k = k_n \sim \log_\nu n$. However, proving of such a fact is quite tricky, since there are far fewer possible further vertices to explore when the neighborhood has size $\sim n$. The proof overcomes this fact by exploring from the two vertices V_1 and V_2 simultaneously up to the first moment that these neighborhoods share a common vertex. At this moment, we have found the shortest path.

Discussion of the proof of Theorem 9.32(i). Theorem 9.32(i) is closely related to Theorem 9.5. The proof of Theorem 9.5 can be extended to $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ for certain (κ_n) by noting that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{IRG}_n(\boldsymbol{p}(\kappa))}(i,j) = k) \le \sum_{i_1,\ldots,i_{k-1}\in[n]} \prod_{l=0}^{k-1} \frac{\kappa_n(x_{i_l}, x_{i_{l+1}})}{n}, \tag{9.5.62}$$

where $i_0 = i, i_k = j$, so that

$$\mathbb{P}(H_n = k) \le \frac{1}{n^k} \sum_{i_0,i_1,\ldots,i_{k-1},i_k\in[n]} \prod_{l=0}^{k-1} \kappa_n(x_{i_l}, x_{i_{l+1}}). \tag{9.5.63}$$

If the above $(k+1)$-dimensional discrete integrals could be replaced by the continuous integral, then we would arrive at

$$\frac{1}{n}\int_{\mathcal{S}} \cdots \int_{\mathcal{S}} \prod_{l=0}^{k} \kappa(x_l, x_{l+1}) \prod_{i=0}^{k} \mu(dx_i) = \frac{1}{n}\|\mathbf{T}_\kappa^{k+1}\mathbf{1}\|_1, \tag{9.5.64}$$

which is bounded from above by $\frac{1}{n}\|\mathbf{T}_\kappa\|^{k+1}$. Repeating the bound in (9.2.10) would then prove that, when $\nu = \|\mathbf{T}_\kappa\| > 1$,

$$\mathbb{P}(H_n \le (1-\varepsilon)\log_\nu n) = o(1). \tag{9.5.65}$$

However, in the general case, it is not so easy to replace the $(k+1)$-fold discrete sum in (9.5.63) by a $(k+1)$-fold integral. In the exercises below, we show how we can carry through the above argument:

Exercise 9.40 (Logarithmic lower bound graph distances $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ of finite-types)**.** *Suppose that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is of finite type, and assume that $n_i/n \to p_i$, where n_i is the number of vertices of type i. Suppose further that the probability that a vertex of type i is directly connected to one of type j is equal to $[\kappa_{ij} \wedge n]/n$ for some matrix $(\kappa_{ij})_{i,j\in[r]}$ with largest eigenvalue $\nu = \|\mathbf{T}_\kappa\| > 1$. Prove that*

$$\mathbb{P}(H_n \le (1-\varepsilon)\log_\nu n) = o(1). \tag{9.5.66}$$

Exercise 9.41 (Extension to the setting of Theorem 9.32)**.** *Assume that the conditions in Theorem 9.32 hold. Recall the bound in* (9.5.31)*, which bounds κ_n from above by κ_m^+, which is of finite-type. Then, use the fact that $\|\mathbf{T}_{\kappa_m^+}\| \downarrow \|\mathbf{T}_\kappa\| = \nu > 1$ to conclude that $\mathbb{P}(H_n \le (1-\varepsilon)\log_\nu n) = o(1)$ holds under the conditions of Theorem 9.32.*

Note that

$$|\{\{v,w\}\colon \mathrm{dist}_G(v,w) < \infty\}| = \sum_i \binom{|\mathcal{C}_{(i)}|}{2}, \tag{9.5.67}$$

where $\{v,w\}$ denotes an unordered pair of distinct vertices of G and where $\mathcal{C}_{(i)}$ denote the connected components of G arranged in decreasing order, so that $\mathcal{C}_{(1)} = \mathcal{C}_{\max}$. Thus, by Theorem 9.31, if κ is irreducible, then

$$\mathbb{P}(H_n < \infty) = \zeta_\kappa^2 + o(1). \tag{9.5.68}$$

Thus, Theorem 9.32(i–ii) can be reformulated by saying that if $\sup_{x,y,n}\kappa_n(x,y) < \infty$ and if $\nu = \|\mathbf{T}_\kappa\| > 1$, then

$$\mathbb{P}\Big(1-\varepsilon \le \frac{H_n}{\log_\nu n} \le 1+\varepsilon \Big| H_n < \infty\Big) \to 1. \tag{9.5.69}$$

As a result, conditionally on $H_n < \infty$, we have that $H_n/\log n \xrightarrow{\mathbb{P}} 1/\log\nu$:

Exercise 9.42 (Convergence in probability of typical distance in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$)**.** *Suppose that the graphical sequence of kernels (κ_n) satisfies $\sup_{x,y,n}\kappa_n(x,y) < \infty$, where the limit κ is irreducible and $\nu = \|\mathbf{T}_\kappa\| > 1$. Prove that Theorem 9.31 together with Theorem 9.32(i–ii) imply that, conditionally on $H_n < \infty$,*

$$H_n/\log n \xrightarrow{\mathbb{P}} 1/\log\nu. \tag{9.5.70}$$

Theorem 9.32 leaves open the case when $\|\mathbf{T}_\kappa\| = \infty$, which, for example for $\mathrm{CL}_n(\boldsymbol{w})$, is the case when F has infinite second moment. [47, Theorem 3.14(iv)] states that when $\|\mathbf{T}_\kappa\| = \infty$, the typical graph distance is smaller than $\log n$. More precisely, [47, Theorem 3.14(iv)] states that if κ is irreducible and $\|\mathbf{T}_\kappa\| = \infty$, then there is a function $f(n) = o(\log n)$ such that

$$\mathbb{P}(H_n \le f(n)) = \zeta_\kappa^2 + o(1). \tag{9.5.71}$$

Exercise 9.43 (Convergence in probability of typical distance in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$)**.** *Suppose that the graphical sequence of kernels (κ_n) converges to κ, where κ is irreducible and $\|\mathbf{T}_\kappa\| = \infty$. Prove that Theorem 9.31 together with Theorem 9.32(iii) imply, conditionally on $H_n < \infty$,*

$$H_n/\log n \xrightarrow{\mathbb{P}} 0. \tag{9.5.72}$$

9.6 Related results for inhomogeneous random graphs

In this section, we discuss some related results for inhomogeneous random graphs. While we give intuition about their proofs, we shall not include them in full detail.

The largest subcritical cluster. For the classical random graph $\mathrm{ER}_n(\lambda/n)$, it is well-known that in the subcritical case for which $\lambda < 1$, the stronger bound $|\mathcal{C}_{\max}| = \Theta(\log n)$ holds (see Theorems 4.4–4.5), and that in the supercritical case for which $\lambda > 1$, $|\mathcal{C}_{(2)}| = \Theta(\log n)$. These bounds do not always hold in the general framework we are considering here, but if we add some conditions, then we can improve the estimates in Theorem 9.31 for the subcritical case to $O(\log n)$:

Theorem 9.33 (Subcritical phase and duality principle of the IRG)**.** *Consider the inhomogeneous random graph* $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$*, where* (κ_n) *is a graphical sequence of kernels with limit* κ*. Then,*

(*i*) *if* κ *is subcritical and* $\sup_{x,y,n} \kappa_n(x,y) < \infty$*, then* $|\mathcal{C}_{\max}| = O_{\mathbb{P}}(\log n)$.

(*ii*) *if* κ *is supercritical,* κ *is irreducible, and either* $\inf_{x,y,n} \kappa_n(x,y) > 0$ *or* $\sup_{x,y,n} \kappa_n(x,y) < \infty$*, then* $|\mathcal{C}_{(2)}| = O_{\mathbb{P}}(\log n)$.

When $\lim_{n\to\infty} \sup_{x,y} \kappa_n(x,y) = \infty$, the largest subcritical clusters can have rather different behavior, as we now show for the rank-1 case. Note that, by Theorem 9.31 as well as the fact that $\|\mathbf{T}_\kappa\| = \nu = \mathbb{E}[W^2]/\mathbb{E}[W]$, a rank-1 model can only be subcritical when $\mathbb{E}[W^2] < \infty$, i.e., in the case of finite variance degrees. However, when W has a power-law tail, i.e., when $\mathbb{P}(W \geq w) \sim w^{-(\tau-1)}$, then the highest weight can be much larger than $\log n$. When this is the case, then also the largest subcritical cluster is much larger than $\log n$, as proved in the following theorem:

Theorem 9.34 (Subcritical phase for rank-1 IRG)**.** *Let* $\boldsymbol{w}$ *satisfy Condition 6.4(a)-(c) with* $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] < 1$*, and, further, that there exists a constant* $c_2 > 0$ *such that*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}. \tag{9.6.1}$$

Then, for $\mathrm{NR}_n(\boldsymbol{w})$ *with* $\Delta = \max_{j\in[n]} w_j$*,*

$$|\mathcal{C}_{\max}| = \frac{\Delta}{1-\nu} + o_{\mathbb{P}}(n^{1/(\tau-1)}). \tag{9.6.2}$$

Theorem 9.34 is most interesting in the case where the limiting distribution function F in Condition 6.4 has a power-law tail. For example, for $\boldsymbol{w}$ as in (6.1.10), let F satisfy

$$[1 - F](x) = cx^{-(\tau-1)}(1 + o(1)). \tag{9.6.3}$$

Then, $\Delta = w_1 = [1-F]^{-1}(1/n) = (cn)^{1/(\tau-1)}(1+o(1))$. Therefore, $|\mathcal{C}_{\max}| = (cn)^{1/(\tau-1)}/(1-\nu) + o(n^{1/(\tau-1)}$. Thus, the largest connected component is much larger than for $\mathrm{ER}_n(\lambda/n)$ with $\lambda < 1$.

Theorem 9.34 can be intuitively understood as follows. The connected component of a typical vertex is close to a branching process, so that it is with high probability bounded since the expected value of its cluster will be close to $1/(1-\nu)$. Thus, the best way to obtain a large connected component is to start with a vertex with high weight w_i, and let all of its roughly w_i children be independent branching processes. Therefore, in expectation, each of these children is connected to another $1/(1-\nu)$ different vertices, leading to a cluster size of roughly $w_i/(1-\nu)$. This is clearly largest when $w_i = \max_{j\in[n]} w_j = \Delta$, leading to an intuitive explanation of Theorem 9.34.

Theorems 9.33 and 9.34 raise the question what the precise conditions for $|\mathcal{C}_{\max}|$ to be of order $\log n$ are. Intuitively, when $\Delta \gg \log n$, then $|\mathcal{C}_{\max}| = \Delta/(1-\nu)(1 + o_{\mathbb{P}}(1))$, whereas if $\Delta = \Theta(\log n)$, then $|\mathcal{C}_{\max}| = \Theta_{\mathbb{P}}(\log n)$ as well. In [216], it was proved that $|\mathcal{C}_{\max}|/\log n$ converges in probability to a finite constant when $\nu < 1$ and the weights are i.i.d. with distribution function F with $\mathbb{E}[\mathrm{e}^{\alpha W}] < \infty$ for some $\alpha > 0$, i.e., *exponential tails* are sufficient.

The critical behavior of rank-1 random graphs. We next discuss the effect of inhomogeneity on the size of the largest connected components in the *critical* case. As it turns out, the behavior is rather different depending on whether $\mathbb{E}[W^3] < \infty$ or not.

Theorem 9.35 (The critical behavior with finite third moments)**.** *Fix the Norros-Reittu random graph with weights $\boldsymbol{w}(t) = \boldsymbol{w}(1+tn^{(\tau-3)(\tau-1)})$. Assume that $\nu = 1$, that the weight sequence $\boldsymbol{w}$ satisfies Condition 6.4(a)-(c), and further assume that*

$$\mathbb{E}[W_n] = \mathbb{E}[W] + o(n^{-1/3}), \qquad \mathbb{E}[W_n^2] = \mathbb{E}[W^2] + o(n^{-1/3}), \qquad \mathbb{E}[W_n^3] = \mathbb{E}[W^3] + o(1) \tag{9.6.4}$$

Let $(|\mathcal{C}_{(i)}(t)|)_{i\geq 1}$ denote the clusters of $\mathrm{NR}_n(\boldsymbol{w}(t))$ with $\boldsymbol{w}(t) = (1 + tn^{-1/3})\boldsymbol{w}$, ordered in size. Then, as $n \to \infty$, for all $t \in \mathbb{R}$,

$$\big(n^{-2/3}|\mathcal{C}_{(i)}(t)|\big)_{i\geq 1} \xrightarrow{d} \big(\gamma_i^*(t)\big)_{i\geq 1}, \tag{9.6.5}$$

in the product topology, for some limiting random variables $\big(\gamma_i^*(t)\big)_{i\geq 1}$.

The limiting random variables $\big(\gamma_i^*(t)\big)_{i\geq 1}$ are, apart from a multiplication by a constant and a time-rescaling, equal to those for $\mathrm{ER}_n(\lambda/n)$ in the scaling window (see Theorem 5.4).

When $\mathbb{E}[W^{3-\varepsilon}] = \infty$ for some $\varepsilon > 0$, it turns out that the scaling of the largest critical cluster is rather different:

Theorem 9.36 (Weak convergence of the ordered critical clusters for $\tau \in (3, 4)$)**.** *Fix the Norros-Reittu random graph with weights $\boldsymbol{w}(t) = \boldsymbol{w}(1 + tn^{(\tau-3)(\tau-1)})$ defined in (6.1.10). Assume that $\nu = 1$ and that there exists a $\tau \in (3, 4)$ and $0 < c_F < \infty$ such that*

$$\lim_{x\to\infty} x^{\tau-1}[1 - F(x)] = c_F. \tag{9.6.6}$$

Let $(|\mathcal{C}_{(i)}(t)|)_{i\geq 1}$ denote the clusters of $\mathrm{NR}_n(\boldsymbol{w}(t))$, ordered in size. Then, as $n \to \infty$, for all $t \in \mathbb{R}$,

$$\big(n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{(i)}(t)|\big)_{i\geq 1} \xrightarrow{d} (\gamma_i(t))_{i\geq 1}, \tag{9.6.7}$$

in the product topology, for some non-degenerate limit $(\gamma_i(t))_{i\geq 1}$.

In this chapter, we have already seen that distances depend sensitively on the finiteness of $\mathbb{E}[W^2]$. Now we see that the critical behavior is rather different when $\mathbb{E}[W^3] < \infty$ or $\mathbb{E}[W^3] = \infty$. Interestingly, in the power-law case as described in (9.6.6), the size of the largest clusters grows like $n^{(\tau-2)/(\tau-1)}$, which is much smaller than the $n^{2/3}$ scaling when $\mathbb{E}[W^3] < \infty$. The proof of Theorems 9.35 and 9.36 also reveals that the structure of large critical clusters is quite different. When $\mathbb{E}[W^3] < \infty$, then the vertex with largest weight is in the largest connected component with vanishing probability. Therefore, the largest connected component arises by many attempts to create a large cluster, and each trial has roughly the same probability. This can be formulated as *power to the masses*. In the other hand, for weights $\boldsymbol{w}$ as in (6.1.10) for which (9.6.6) holds, the vertices with largest weight are with probability bounded away from 0 and 1 in the largest cluster, while a vertex with small weight is in the largest cluster with vanishing probability. Thus, to find the largest clusters, it suffices to explore the clusters of the high-weight vertices: *power to the wealthy!*

Fluctuations of distances in the finite variance case. We continue by studying the fluctuations of the typical graph distance when $\mathbb{E}[W^2] < \infty$. We shall impose a slightly stronger condition on the distribution function F of W, namely, that there exists a $\tau > 3$ and $c > 0$ such that

$$1 - F(x) \leq cx^{-(\tau-1)}. \tag{9.6.8}$$

Equation (9.6.8) implies that the degrees have finite variance:

Exercise 9.44 (Finite variance degrees when (9.6.8) holds)**.** *Prove that* (9.6.8) *implies that* $\mathbb{E}[W^2] < \infty$*. Use this to prove that the degrees have uniformly bounded variance when* (9.6.8) *holds.*

Theorem 9.37 (Limit law for the typical graph distance in $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Assume that* (9.6.8) *is satisfied, and let* $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$*. For* $k \geq 1$*, define* $a_k = \lfloor \log_\nu k \rfloor - \log_\nu k \in (-1, 0]$*. Then, for* $\mathrm{CL}_n(\boldsymbol{w})$ *with* $\boldsymbol{w}$ *as in* (6.1.10)*, there exist random variables* $(R_a)_{a\in(-1,0]}$ *with* $\limsup_{K\to\infty} \sup_{a\in(-1,0]} \mathbb{P}(|R_a| < K) = 1$ *such that, as* $n \to \infty$*,*

$$\mathbb{P}\big(H_n - \lfloor \log_\nu n \rfloor = k \mid H_n < \infty\big) = \mathbb{P}(R_{a_n} = k) + o(1), \qquad k \in \mathbb{Z}. \tag{9.6.9}$$

While Theorem 9.32 implies that, conditionally on $H_n < \infty$, $H_n/\log n \xrightarrow{\mathbb{P}} 1/\log \nu$, Theorem 9.37 implies that the fluctuations of H_n around $\log_\nu n$ remain uniformly bounded in probability.

The random variables $(R_a)_{a\in(-1,0]}$ can be determined in terms of the limit law in a branching process approximation of the neighborhoods of $\mathrm{CL}_n(\boldsymbol{w})$, and depend sensitively on a, which implies that although $\{H_n - \lfloor \log_\nu n \rfloor\}_{n=2}^\infty$ is a *tight* sequence of random variables, it does *not* weakly converges.

Exercise 9.45 (Tightness of centered typical graph distances in $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Prove that, under the conditions of Theorem 9.37, and conditionally on* $H_n < \infty$*, the sequence* $\{H_n - \lfloor \log_\nu n \rfloor\}_{n=2}^\infty$ *is tight.*

Exercise 9.46 (Non-convergence of centered typical graph distances in $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Prove that, under the conditions of Theorem 9.37, and conditionally on* $H_n < \infty$*, the sequence* $H_n - \lfloor \log_\nu n \rfloor$ *does not weakly converge when the distribution of* R_a *depends continuously on* a *and when there are* $a, b \in (-1, 0]$ *such that the distribution of* R_a *is not equal to the one of* R_b*.*

Exercise 9.47 (Extension Theorem 9.37 to $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Use Theorem 6.17 to prove that Theorem 9.37 holds verbatim for* $\mathrm{GRG}_n(\boldsymbol{w})$ *and* $\mathrm{NR}_n(\boldsymbol{w})$ *when* (9.6.8) *holds.*

The diameter in inhomogeneous random graphs. We shall also be interested in the *diameter* of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$, which is defined to be the maximal finite graph distance between any pair of vertices, i.e., the diameter $\mathrm{diam}(G)$ of the graph G equals

$$\mathrm{diam}(G) = \max_{u,v:\, \mathrm{dist}_G(u,v)<\infty} \mathrm{dist}_G(u, v). \tag{9.6.10}$$

We shall see that for $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$, the diameter tends to be much larger than the typical graph distances, which is due to long thin lines which are distributed as subcritical $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ with a subcritical κ by a duality principle for $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. Before we state the results, we introduce the notion of the *dual kernel*:

Definition 9.38 (Dual kernel for $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$)**.** *Let* (κ_n) *be a sequence of supercritical kernels with limit* κ*. The dual kernel is the kernel* $\widehat{\kappa}$ *defined by* $\widehat{\kappa}(x, y) = \kappa(x, y)$*, with* reference measure $d\widehat{\mu}(x) = (1 - \zeta_\kappa(x))\mu(dx)$*.*

The dual kernel shall describe the graph after the removal of the giant component. Here, the reference measure $\widehat{\mu}$ measures the size of the graph. In this case, a vertex x is in the giant component with probability $1-\zeta_\kappa(x)$, in which case it must be removed. Thus, $\widehat{\mu}$ describes the proportion of vertices of the various types which are outside the giant component. As before, we define the operator $\mathbf{T}_{\widehat{\kappa}}$ by the equality

$$(\mathbf{T}_{\widehat{\kappa}}f)(x)=\int_{\mathcal{S}}\widehat{\kappa}(x,y)f(y)d\widehat{\mu}(y)=\int_{\mathcal{S}}\kappa(x,y)f(y)[1-\zeta_\kappa(x)]\mu(dy), \tag{9.6.11}$$

and we write $\|\mathbf{T}_{\widehat{\kappa}}\|$ for

$$\|\mathbf{T}_{\widehat{\kappa}}\|=\sup\big\{\|\mathbf{T}_{\widehat{\kappa}}f\|_2\colon f\geq 0,\|f\|_{\widehat{\mu},2}=1\big\}, \tag{9.6.12}$$

where

$$\|f\|^2_{\widehat{\mu},2}=\int_{\mathcal{S}}f^2(x)\widehat{\mu}(dx). \tag{9.6.13}$$

Theorem 9.39 (The diameter of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ in the finite-types case)**.** *Let (κ_n) be a sequence of kernels with limit κ, which has finitely many types. If $0<\|\mathbf{T}_\kappa\|<1$, then*

$$\frac{\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)))}{\log n}\xrightarrow{\mathbb{P}}\frac{1}{\log\|\mathbf{T}_\kappa\|^{-1}} \tag{9.6.14}$$

as $n\to\infty$. If $\|\mathbf{T}_\kappa\|>1$ and κ irreducible, then

$$\frac{\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)))}{\log n}\xrightarrow{\mathbb{P}}\frac{2}{\log\|\mathbf{T}_{\widehat{\kappa}}\|^{-1}}+\frac{1}{\log\|\mathbf{T}_\kappa\|}, \tag{9.6.15}$$

where $\widehat{\kappa}$ is the dual kernel to κ.

If we compare Theorem 9.39 to Theorem 9.37, we see that the diameter has the same scaling as the typical graph distance, but that the limit in probability of $\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))/\log n$ is strictly larger than the one for $H_n/\log n$ conditioned on $H_n<\infty$. This effect is particularly noticeable in the case when $\tau\in(2,3)$, where $H_n/\log\log n$, conditionally on $H_n<\infty$, converges in probability to a finite limit, while $\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))/\log n$ converges to a non-zero limit. This can be explained by noticing that the diameter in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ arises due to very this lines of length of order $\log n$. Since these this lines involve only very few vertices, they will not contribute to H_n, but they do to $\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))$. This is another argument why we prefer to work with typical graph distances than with the diameter.

9.6.1 Clustered inhomogeneous random graphs

General inhomogeneous random graphs have rather low clustering. Indeed, assuming that $\kappa_n(x,y\leq n$, we can compute that the expected number of triangles in an $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is equal to

$$\mathbb{E}[\#\text{ triangles in }\mathrm{IRG}_n(\boldsymbol{p}(\kappa))]=\frac{1}{n^3}\sum_{i,j,k\in[n]}\kappa_n(x_i,x_j)\kappa_n(x_j,x_k)\kappa_n(x_k,x_i). \tag{9.6.16}$$

Under relatively weak conditions on the kernel κ_n, it follows that

$$\mathbb{E}[\#\text{ triangles in }\mathrm{IRG}_n(\boldsymbol{p}(\kappa))]\to\int_{\mathcal{S}^3}\kappa(x_1,x_2)\kappa(x_2,x_3)\kappa(x_3,x_1)\mu(dx_1)\mu(dx_2)\mu(dx_3). \tag{9.6.17}$$

Therefore, the clustering coefficient converges to zero as $1/n$. In many real-world networks, particularly in social networks, the clustering coefficient is strictly positive. In this

section, we discuss a model similar to the inhomogeneous random graph $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ that incorporates clustering.

The idea behind this model is that instead of only adding *edges* independently, we can also add other graphs on r vertices in an independent way. For example, we could study a graph where each pair of vertices is independently connected with probability λ/n, as for $\mathrm{ER}_n(\lambda/n)$, but also each collection of triples forms a triangle with probability μ/n^2, independently for all triplets and independently of the status of the edges. Here the exponent n^{-2} is chosen to as to make the expected number of triangles containing a vertex bounded.

Exercise 9.48 (Clustering in model with edges and triangles)**.** *Show that the clustering coefficient in the model where each pair of vertices is independently connected with probability λ/n, as for $\mathrm{ER}_n(\lambda/n)$ and each triples forms a triangle with probability μ/n^2, independently for all triplets and independently of the status of the edges, converges to $\mu/(\mu+\lambda^2)$.*

In social networks, also complete graphs of size four, five, etc., are present more often than in usual random graph. Therefore, we also wish to add those independently. In order to formulate this general version of the model, we start by introducing some notation.

Let $\mathcal{F}$ consist of one representative of each isomorphism class of finite connected graphs, chosen so that if $F \in \mathcal{F}$ has r vertices then $V(F) = [r] = \{1, 2, \dots, r\}$. Simple examples of such F are the complete graphs on r vertices, but also other examples are possible. Recall that $\mathcal{S}$ denotes the type space. Given $F \in \mathcal{F}$ with r vertices, let κ_F be a measurable function from $\mathcal{S}^r$ to $[0, \infty)$. The function κ_F is called the *kernel corresponding to* F. A sequence $\widetilde{\kappa} = (\kappa_F)_{F\in\mathcal{F}}$ is a *kernel family*.

Let $\widetilde{\kappa}$ be a kernel family and n an integer. We define a random graph $\mathrm{IRG}_n(\widetilde{\kappa})$ with vertex set $[n] = \{1, \dots, n\}$. First let $x_1, x_2, \dots, x_n \in \mathcal{S}$ be i.i.d. (independent and identically distributed) with the distribution μ. Given $\mathbf{x} = (x_1, \dots, x_n)$, construct $\mathrm{IRG}_n(\widetilde{\kappa})$ as follows, starting with the empty graph. For each r and each $F \in \mathcal{F}$ with $|F| = r$, and for every r-tuple of distinct vertices $(v_1, \dots, v_r) \in [n]^r$, add a copy of F on the vertices $v_1, \dots, v_r$ (with vertex i of F mapped to v_i) with probability

$$p = p(v_1, \dots, v_r; F) = \frac{\kappa_F(x_{v_1}, \dots, x_{v_r})}{n^{r-1}}, \tag{9.6.18}$$

all these choices being independent. Here, if $p > 1$, by convention we simply add a copy with probability 1. We shall often call the added copies of the various F that together form $\mathrm{IRG}_n(\widetilde{\kappa})$ *atoms* as they may be viewed as indivisible building blocks. Sometimes we refer to them as *small graphs*, although there is in general no bound on their sizes.

The reason for dividing by n^{r-1} in (9.6.18) is that we wish to consider sparse graphs; indeed, our main interest is the case when $\mathrm{IRG}_n(\widetilde{\kappa})$ has $O(n)$ edges. As it turns out, we can be slightly more general; however, when κ_F is integrable (which we shall always assume), the expected number of added copies of each graph F is $O(n)$. Note that all incompletely specified integrals are with respect to the appropriate r-fold product measure μ^r on $\mathcal{S}^r$.

There are several plausible choices for the normalization in (9.6.18). The one we have chosen means that if $\kappa_F = c$ is constant, then (asymptotically) there are on average cn copies of F in total, and each vertex is on average in rc copies of F. An alternative is to divide the expression in (9.6.18) by r; then (asymptotically) each vertex would on average be in c copies of F. Another alternative, natural when adding cliques only but less so in the general case, would be to divide by $r!$; this is equivalent to considering unordered sets of r vertices instead of ordered r-tuples. When there is only one kernel, corresponding to adding edges, this would correspond to the normalization used in [47], and in particular to that of the classical model $\mathrm{ER}_n(\lambda/n)$; the normalization we use here differs from this by a factor of 2.

In the special case where all κ_F are zero apart from κ_{K_2}, the kernel corresponding to an edge, we recover (essentially) a special case of the model of [47]; we call this the *edge-only* case, since we add only edges, not larger graphs. We write κ_2 for κ_{K_2}. Note that in the edge-only case, given $\mathbf{x}$, two vertices i and j are joined with probability

$$\frac{\kappa_2(x_i,x_j)+\kappa_2(x_j,x_j)}{n}+O\left(\frac{\big(\kappa_2(x_i,x_j)+\kappa_2(x_j,x_i)\big)^2}{n^2}\right). \tag{9.6.19}$$

The correction term will never matter, so we may as well replace κ_2 by its symmetrized version. In fact, we shall always assume that κ_F is invariant under permutations of the vertices of the graph F.

For any kernel family $\widetilde{\kappa}$, let κ_e be the corresponding *edge kernel*, defined by

$$\kappa_e(x,y)=\sum_F \sum_{ij\in E(F)} \int_{\mathcal{S}^{V(F)\setminus\{i,j\}}} \kappa_F(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_{j-1},y,x_{j+1},\dots,x_{|F|}), \tag{9.6.20}$$

where the second sum runs over all $2E(F)$ ordered pairs (i,j) with $ij\in \mathit{Edges}(F)$, and we integrate over all variables apart from x and y. Note that the sum need not always converge; since every term is positive this causes no problems: we simply allow $\kappa_e(x,y)=\infty$ for some x,y. Given x_i and x_j, the probability that i and j are joined in $G(n,\widetilde{\kappa})$ is at most $\kappa_e(x_i,x_j)/n+O(1/n^2)$. In other words, κ_e captures the edge probabilities in $G(n,\widetilde{\kappa})$, but not the correlations.

Before proceeding to deeper properties, let us note that the expected number of added copies of F is $(1+O(n^{-1}))n\int_{\mathcal{S}^{|F|}}\kappa_F$. Unsurprisingly, the actual number turns out to be concentrated about this mean. Let

$$\xi(\widetilde{\kappa})=\sum_{F\in\mathcal{F}} E(F)\int_{\mathcal{S}^{|F|}}\kappa_F=\frac{1}{2}\int_{\mathcal{S}^2}\kappa_e\le\infty \tag{9.6.21}$$

be the *asymptotic edge density* of $\widetilde{\kappa}$. Since every copy of F contributes $E(F)$ edges, the following theorem is almost obvious, provided we can ignore overlapping edges.

Theorem 9.40 (The edge density in $\mathrm{IRG}_n(\widetilde{\kappa})$)**.** *As $n\to\infty$, $\mathbb{E}[E(\mathrm{IRG}_n(\widetilde{\kappa}))/n]\to\xi(\widetilde{\kappa})\le\infty$ Moreover, $E(\mathrm{IRG}_n(\widetilde{\kappa}))/n$ converges in probability to the asymptotic edge density $\xi(\widetilde{\kappa})$. In other words, if $\xi(\widetilde{\kappa})<\infty$ then $E(\mathrm{IRG}_n(\widetilde{\kappa}))=\xi(\widetilde{\kappa})n+o_p(n)$, and if $\xi(\widetilde{\kappa})=\infty$ then $E(\mathrm{IRG}_n(\widetilde{\kappa}))>Cn$ for every constant C,* **whp**.

The main focus is the emergence of the giant component. By the *component structure* of a graph G, we mean the set of vertex sets of its components, i.e., the structure encoding only which vertices are in the same component, not the internal structure of the components themselves. When studying the component structure of $\mathrm{IRG}_n(\widetilde{\kappa})$, the model can be simplified somewhat. Recalling that the atoms $F\in\mathcal{F}$ are connected by definition, when we add an atom F to a graph G, the effect on the component structure is simply to unite all components of G that meet the vertex set of F, so only the vertex set of F matters, not its graph structure. We say that $\widetilde{\kappa}$ is a *clique kernel family* if the only non-zero kernels are those corresponding to complete graphs; the corresponding random graph model $\mathrm{IRG}_n(\widetilde{\kappa})$ is a *clique model*. For questions corresponding component structure, it suffices to study clique models. For clique kernels we write κ_r for κ_{K_r}; as above, we always assume that κ_r is symmetric, here meaning invariant under all permutations of the coordinates of $\mathcal{S}^r$. Given a general kernel family $\widetilde{\kappa}$, the corresponding (symmetrized) clique kernel family is given by $\underline{\widetilde{\kappa}}=(\kappa_r)_{r\ge2}$ with

$$\kappa_r(x_1,\dots,x_r)=\sum_{F\in\mathcal{F}:\ |F|=r}\frac{1}{r!}\sum_{\pi\in\mathfrak{S}_r}\kappa_F(x_{\pi(1)},\dots,x_{\pi(r)}), \tag{9.6.22}$$

where $\mathfrak{S}_r$ denotes the symmetric group of permutations of $[r]$. (This is consistent with our notation $\kappa_2 = \kappa_{K_r}$.) When considering the size (meaning number of vertices) of the giant component in $\mathrm{IRG}_n(\widetilde{\kappa})$, we may always replace $\widetilde{\kappa}$ by the corresponding clique kernel family.

It is often convenient to think of a clique model as a random *hypergraph*, with the cliques as the hyperedges; for this reason we call a clique kernel family a *hyperkernel*. Note that each unordered set of r vertices corresponds to $r!$ r-tuples, so the probability that we add a K_r on a given set of r vertices is $r!\kappa_r(x_{v_1}, \dots, x_{v_r})/n^{r-1}$. (More precisely, this is the expected number of K_rs added with this vertex set.)

In our analysis we also consider the linear operator T_{κ_e} defined by

$$T_{\kappa_e}(f)(x) = \int_{\mathcal{S}} \kappa_e(x,y) f(y) d\mu(y), \tag{9.6.23}$$

where κ_e is defined by (9.6.20). We need to impose some sort of integrability condition on our kernel family:

Definition 9.41. *(i) A kernel family $\widetilde{\kappa} = (\kappa_F)_{F\in\mathcal{F}}$ is integrable if*

$$\int \widetilde{\kappa} = \sum_{F\in\mathcal{F}} |F| \int_{\mathcal{S}^{|F|}} \kappa_F < \infty. \tag{9.6.24}$$

This means that the expected number of atoms containing a given vertex is bounded.

(ii) A kernel family $\widetilde{\kappa} = (\kappa_F)_{F\in\mathcal{F}}$ is edge integrable if

$$\sum_{F\in\mathcal{F}} E(F) \int_{\mathcal{S}^{|F|}} \kappa_F < \infty; \tag{9.6.25}$$

equivalently, $\xi(\kappa) < \infty$ or $\int_{\mathcal{S}^2} \kappa_e < \infty$. This means that the expected number of edges in $G(n,\widetilde{\kappa})$ is $O(n)$, see Theorem 9.40, and thus the expected degree of a given vertex is bounded.

Note that a hyperkernel (κ_r) is integrable if and only if $\sum_{r\ge 2} r \int_{\mathcal{S}^r} \kappa_r < \infty$, and edge integrable

The main results concerning the phase transition on $\mathrm{IRG}_n(\widetilde{\kappa})$ is that if $\widetilde{\kappa}$ is an integrable kernel family satisfying a certain extra assumption, then the normalized size of the giant component in $\mathrm{IRG}_n(\widetilde{\kappa})$ is simply $\zeta(\widetilde{\kappa}) + o_p(1)$. The extra assumption is an irreducibility assumption similar to Definition 9.27(ii) that essentially guarantees that the graph does not split into two pieces: we say that a symmetric kernel $\kappa_e : \mathcal{S}^2 \to [0,\infty)$ is *reducible* if

$$\exists A \subset \mathcal{S} \text{ with } 0 < \mu(A) < 1 \text{ such that } \kappa_e = 0 \text{ a.e. on } A \times (\mathcal{S}\setminus A);$$

otherwise κ_e is irreducible. Thus, κ_e is irreducible if

$$A \subseteq \mathcal{S} \text{ and } \kappa_e = 0 \text{ a.e. on } A\times(\mathcal{S}\setminus A) \text{ implies } \mu(A) = 0 \text{ or } \mu(\mathcal{S}\setminus A) = 0.$$

We are now ready to formulate the main result in this section involving the phase transition in $\mathrm{IRG}_n(\widetilde{\kappa})$. Recall that $|\mathcal{C}_{\max}|$ denotes the number of vertices in the largest connected component of the graph under consideration, and $|\mathcal{C}_{(2)}|$ the number of vertices in its second largest component.

Theorem 9.42 (The phase transition on clustered inhomogeneous random graphs)**.** *Let* $\widetilde{\kappa}' = (\kappa'_F)_{F\in\mathcal{F}}$ *be an irreducible, integrable kernel family, and let* $\widetilde{\kappa} = (\kappa_r)_{r\geq 2}$ *be the corresponding hyperkernel, given by* (9.6.22)*. Then, there exists a* $\zeta(\widetilde{\kappa}) \in [0,1)$ *such that*

$$|\mathcal{C}_{\max}| = \zeta(\widetilde{\kappa})n + o_p(n), \tag{9.6.26}$$

and $|\mathcal{C}_{(2)}| = o_p(n)$*.*

Theorem 9.42 is proved by showing that (in the hyperkernel case) the branching process that captures the 'local structure' of $\mathrm{IRG}_n(\widetilde{\kappa})$. For Theorem 9.42 to be useful we would like to know something about $\zeta(\widetilde{\kappa})$, which can be calculated from $\zeta_{\widetilde{\kappa}}$, which is in turn the largest solution to the functional equation:

$$f = 1 - \mathrm{e}^{-S_{\widetilde{\kappa}}(f)(x)}. \tag{9.6.27}$$

The question when $\zeta(\widetilde{\kappa}) > 0$ is settled in the following theorem:

Theorem 9.43 (Condition for existence giant component)**.** *Let* $\widetilde{\kappa}$ *be an integrable hyperkernel. Then,* $\zeta(\widetilde{\kappa}) > 0$ *if and only if* $||T_{\kappa_e}|| > 1$*. Furthermore, if* $\widetilde{\kappa}$ *is irreducible and* $||T_{\kappa_e}|| > 1$*, then* $\zeta_{\widetilde{\kappa}}(x)$ *is the unique non-zero solution to the functional equation* (9.6.27)*, and* $\zeta_{\widetilde{\kappa}}(x) > 0$ *holds for a.e.* x*.*

In general, $||T_{\kappa_e}||$ may be rather hard to calculate. When we suppose that each κ_r is constant, however, this can be done. Indeed, say that $\kappa_r = c_r$. Then $\kappa_e(x,y) = \sum_r r(r-1)c_r = 2\xi(\kappa)$ for all x and y, so

$$||T_{\kappa_e}|| = 2\xi(\kappa). \tag{9.6.28}$$

This is perhaps surprising: it tells us that for such uniform hyperkernels, the critical point where a giant component emerges is determined only by the total number of edges added; it does not matter what size cliques they lie in, even though, for example, the third edge in every triangle is 'wasted'. This turns out not to be true for arbitrary kernel families, where, rather each atom needs to be replaced by a clique.

9.7 Notes and discussion

Notes on Section 9.1. Theorem 9.2 is taken from [133], where the giant component is investigated for the configuration model. We explain its proof in detail in Section 10.1, where we also prove how the result for the configuration model in Theorem 10.1 can be used to prove Theorem 9.2.

Theorem 9.3 has a long history, and many versions of it have been proven in the literature. We refer the reader to [66, 68] for the Chung-Lu model, and [102] for its extensions to the Norros-Reittu model and the generalized random graph. Theorem 9.4 has also been proved in many versions, both fully as well as in partial forms, see [182], [66, 68], as well as the recent paper [84].

Notes on Section 9.2. As far as we are aware, the proof of Theorem 9.5 is new in the present context. Similar arguments have been used often though to prove lower bounds on distances in various situations.

Notes on Section 9.3. Proposition 9.10 appears first as [182, Proposition 3.1], where the connection between $\mathrm{NR}_n(\boldsymbol{w})$ and Poisson branching processes were first exploited to prove versions of Theorem 9.4.

Notes on Section 9.4. The path counting techniques in Proposition 9.17 are novel. Related proofs for the upper bound on H_n when $\nu < \infty$ often rely on branching process comparisons up to a generation $m = m_n \to \infty$.

Notes on Section 9.5. Theorem 9.28 is a special case of [47, Theorem 3.13]. Theorem 9.31 is a special case of [47, Theorem 3.1]. Earlier versions for random graphs with given expected degrees or Chung-Lu model appeared in [67, 72] (see also the monograph [70]).

The seminal paper [47] studies inhomogeneous random graph in an even more general setting, where the number of vertices in the graph need not be equal to n. In this case, the vertex space is called a *generalized vertex space*. We simplify the discussion here by assuming that the number of vertices is always equal to n. An example where the extension to a random number of vertices is crucially used is in [218], which studies an interpolation between percolation and $\mathrm{ER}_n(p)$.

In [47], there are various other results concerning the giant component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. For example, [47, Theorem 3.9] proves that the giant component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is stable in the sense that its size does not change much if we add or delete a few edges. Note that the edges added or deleted do not have to be random or independent of the existing graph, rather, they can be chosen by a adversary after inspecting the whole of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. More precisely, [47, Theorem 3.9] shows that, for small enough $\delta > 0$, the giant component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ in the supercritical regime does change by more than εn vertices if we remove any collection of δn edges.

Theorem 9.32 is a simplified version of [47, Theorem 3.14]. A first version of Theorem 9.32 was proved in [66, 68] for the expected degree random graph, in the case of *admissible* deterministic weights. We refer to [68, p. 94] for the definition of admissible degree sequences.

Theorem 9.4 for the expected degree random graph or Chung-Lu model is first proved in [66, 68], in the case of deterministic weights $w_i = c \cdot (i/n)^{-1/(\tau-1)}$, having average degree strictly greater than 1 and maximum degree m satisfying $\log m \gg \log n / \log \log n$. These restrictions were lifted in [88, Theorem 4.5.2]. Indeed, the bound on the average distance is not necessary, since, for $\tau \in (2, 3)$, $\nu = \infty$ and therefore the IRG is *always* supercritical. An upper bound as in Theorem 9.4 for the Norros-Reittu model with i.i.d. weights is proved in [182].

Notes on Section 9.6. Theorem 9.34 is [128, Corollary 4.4]. Theorem 9.35 is proved in [33], a related version with a different proof can be found in [217]. Theorem 9.36 is proved in [34]. Theorem 9.37 is proved in [102], both in the case of i.i.d. degrees as well as for deterministic weights under a mild further condition on the distribution function. Theorem 9.39 is a special case of [47, Theorem 3.16]. Even in the special case of $\mathrm{ER}_n(\lambda/n)$, it is new, and it negatively answers a question of Chung and Lu [65]. Related results for the configuration model, which also imply results for the generalized random graph, can be found in [107].

Chapter 10
Configuration model revisited

In this chapter, we state some further results on the configuration model. This chapter is organized as follows. In Section 10.1, we study when a giant component exists for the configuration model. In Section 10.2, we study the connectivity of the configuration model. In Section 10.3, we study the typical graph distance in the configuration model. In Section 10.4, we prove the distance results for the configuration model, using path counting techniques and comparisons to branching processes. In Section 10.5, we identify the diameter of the configuration model when it has infinite-variance degrees. In Section 10.6, we study infinite-mean branching processes, as these arise in the configuration model with infinite-variance degrees. In Section 10.8, we state further results in the configuration model, using among others the results in Section 10.6.

10.1 Phase transition in the configuration model

In this section, we investigate the connected components in the configuration model. We shall show that, alike for the Erdős-Rényi random graph, there is a condition for the configuration model such that with high probability a giant component exists. Again, this condition has the interpretation that an underlying branching process describing the exploration of a cluster has a strictly positive survival probability.

We start by recalling some notation from Chapter 7. We investigate the configuration model $\mathrm{CM}_n(\boldsymbol{d})$, where in most cases, the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ are assumed to satisfy Condition 7.5(a)-(b), and sometimes also Condition 7.5(c). We recall that D_n is the degree of a uniformly chosen vertex in $[n]$, i.e., $D_n = d_V$, where V is uniformly chosen from $[n]$. Equivalently,

$$\mathbb{P}(D_n = k) = n_k/n, \tag{10.1.1}$$

where n_k denotes the number of vertices of degree k. For a graph G, we write $v_k(G)$ for the number of vertices of degree k in G, and $|E(G)|$ for the number of edges. The main result concerning the size and structure of the largest connected components of $\mathrm{CM}_n(\boldsymbol{d})$ is the following:

Theorem 10.1 (Phase transition in $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Suppose that Condition 7.5(a)-(b) hold and consider the random graph $\mathrm{CM}_n(\boldsymbol{d})$, letting $n\to\infty$. Assume that $p_2 = \mathbb{P}(D=2) < 1$. Let $\mathcal{C}_{\max}$ and $\mathcal{C}_{(2)}$ be the largest and second largest components of $\mathrm{CM}_n(\boldsymbol{d})$.*

(a) If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$, then there exist $\xi\in[0,1), \zeta\in(0,1]$ such that

$$\begin{aligned} |\mathcal{C}_{\max}|/n &\xrightarrow{\mathbb{P}} \zeta, \\ v_k(\mathcal{C}_{\max})/n &\xrightarrow{\mathbb{P}} p_k(1-\xi^k), \text{ for every } k\geq 0, \\ |E(\mathcal{C}_{\max})|/n &\xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1-\xi^2). \end{aligned}$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$.

(b) If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \leq 1$, then $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.

Reformulation in terms of branching processes. We start by interpreting the results in Theorem 10.1 in terms of *branching processes*. As it turns out, we can interpret ξ as the extinction probability of a branching process, and ζ as the survival probability of a related branching process. Similarly to the cluster exploration for $\mathrm{NR}_n(\boldsymbol{w})$, we shall see that the root of the branching process plays a special role. Indeed, the offspring distribution at the root is equal to $(p_k)_{k\geq 0}$, while the offspring distribution at any other vertex in the tree is given by

$$g_k = \frac{(k+1)p_{k+1}}{\mathbb{E}[D]}. \tag{10.1.2}$$

We now heuristically explain this relation to branching processes by intuitively describing the exploration of a vertex chosen uniformly from the vertex set $[n]$. By definition, the probability that its degree is k equals $\mathbb{P}(D_n = k)$, which, by Condition 7.5(a), converges to $p_k = \mathbb{P}(D = k)$, for every $k \geq 1$. This explains the offspring of the root of our branching process approximation.

To describe the offspring of the direct neighbors of the root, we examine the degree of the vertex to which the first half-edge incident to the root is paired. We note that, by the uniform matching of half-edges, the probability that a vertex of degree k is chosen is proportional to k. Now, ignoring the fact that the root and one half-edge have already been chosen, which has an effect on the number of available or free half-edges, we have that the degree of the vertex incident to the chosen half-edge equals k with probability equal to $kp_k^{(n)}/\mathbb{E}[D_n]$, where $p_k^{(n)} = n_k/n$ is the proportion of vertices with degree k, and

$$\mathbb{E}[D_n] = \frac{1}{n}\sum_{i\in[n]} d_i = \frac{1}{n}\sum_{i\in[n]}\sum_{k=0}^{\infty} k\mathbb{1}_{\{d_i=k\}} = \sum_{k=0}^{\infty} kp_k^{(n)}, \tag{10.1.3}$$

so that $(kp_k^{(n)}/\mathbb{E}[D_n])_{k\geq 0}$ is a probability mass function. However, one of the half-edges is used up to connect to the root. Therefore, for a vertex incident to the root to have k offspring, it needs to connect its half-edge to a vertex with degree $k+1$. Therefore, the probability that the offspring of any of the direct neighbors of the root is equal to k equals

$$g_k^{(n)} = \frac{(k+1)p_{k+1}^{(n)}}{\mathbb{E}[D_n]}. \tag{10.1.4}$$

Thus, $(g_k^{(n)})_{k\geq 0}$ can be interpreted as the *forward degree of vertices in the cluster exploration.* When Condition 7.5(a)-(b) hold, then also $g_k^{(n)} \to g_k$, where $(g_k)_{k\geq 0}$ is defined in (10.1.2). As a result, we will often refer to $(g_k)_{k\geq 0}$ as the *forward degree distribution.*

The above heuristically argues that the number of vertices unequal to the root connected to any neighbor of the root has asymptotic law $(g_k)_{k\geq 0}$. However, every time we pair to half-edges, the number of *free* or *available* half-edges decreases by 2. Similarly to the depletion of points effect in the exploration of clusters for the Erdős-Rényi random graph $\mathrm{ER}_n(\lambda/n)$, the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ suffers from a *depletion of points and half-edges* effect. Thus, by iteratively connecting half-edges in a breadth-first way, the offspring distribution changes along the way, which gives potential trouble. Luckily, the number of available half-edges that we start with equals $\ell_n - 1$, which is very large when Condition 7.5(a)-(b) hold, since then $\ell_n/n = \mathbb{E}[D_n]/n \to \mathbb{E}[D] > 0$. Thus, we can pair many half-edges before we start noticing that their number decreases. As a result, the degrees of different vertices in the exploration process is close to being i.i.d., leading to a branching process approximation. In terms of this branching process, we can interpret ζ in Theorem 10.1 as the survival probability of the above two-stage branching process, so that ζ satisfies

$$\zeta = \sum_{k=1}^{\infty} p_k(1-\xi^k), \tag{10.1.5}$$

where ξ is the extinction probability of the branching process with offspring distribution $(g_k)_{k\geq 0}$, which satisfies

$$\xi = \sum_{k=0}^{\infty} g_k \xi^k. \tag{10.1.6}$$

Clearly, $\xi = 1$ precisely when

$$\nu = \sum_{k\geq 0} k g_k \leq 1. \tag{10.1.7}$$

By (10.1.2), we can rewrite

$$\nu = \frac{1}{\mathbb{E}[D]} \sum_{k\geq 0} k(k+1) p_{k+1} = \mathbb{E}[D(D-1)]/\mathbb{E}[D], \tag{10.1.8}$$

which explains the condition on ν in Theorem 10.1(a). Further, there are $n_k = np_k^{(n)} \approx np_k$ vertices with degree k. Each of the k neighbors of a vertex will survive with probability $1-\xi$, so that the probability that at least one survives equals $1-\xi^k$. When one of the neighbors survive, the vertex itself will be part of the giant component, which explains why $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1-\xi^k)$. Finally, an edge consists of two half-edges, and an edge is part of the giant component precisely when one of the vertices incident to it is, which occurs with probability $1-\xi^2$. There are in total $\ell_n/2 = n\mathbb{E}[D_n]/2 \approx n\mathbb{E}[D]/2$ edges, which explains why $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1-\xi^2)$. Therefore, all results in Theorem 10.1 have a simple explanation in terms of the branching process approximation of the connected component of a uniform vertex in $[n]$ for $\mathrm{CM}_n(\boldsymbol{d})$.

Reformulation in terms of generating functions. We next reformulate the results in terms of generating functions, which shall play a crucial role throughout our proof. Let

$$G_D(x) = \sum_{k=0}^{\infty} p_k x^k = \mathbb{E}[x^D] \tag{10.1.9}$$

be the probability generating function of the probability distribution $(p_k)_{k\geq 1}$. Recall that, for a non-negative random variable D, the random variable D^* denotes its size-biased distribution. Define further

$$G_D^*(x) = \mathbb{E}[x^{D^*}] = \sum_{k=1}^{\infty} g_k x^k = G_D'(x)/G_D'(1), \tag{10.1.10}$$

$$H(x) = \mathbb{E}[D]x\big(x - G_D^*(x)\big). \tag{10.1.11}$$

Note that $G_D^*(1) = 1$, and thus $H(0) = H(1) = 0$. Note also that

$$H'(1) = 2\mathbb{E}[D] - \sum_k k^2 p_k = \mathbb{E}[2D - D^2] = -\mathbb{E}[D(D-2)] \tag{10.1.12}$$

For further properties of $x \mapsto H(x)$, see Lemma 10.6 below. We conclude that if $\mathbb{E}[D(D-2)] = \sum_k k(k-2)p_k > 0$, then there is a unique $\xi \in (0,1)$ such that $H(\xi) = 0$, or equivalently $G_D^*(\xi) = \xi$, so that indeed ξ is the extinction probability of the branching process with offspring distribution $(g_k)_{k\geq 0}$. The functions $x \mapsto H(x)$ and $x \mapsto G_D^*(x)$ play a crucial role in our analysis of the problem.

We prove Theorem 10.1 in Section 10.1.2 below. We now remark upon the result.

The condition $\mathbb{P}(D=2)=p_2<1$ and the role of $\mathbb{P}(D=1)=p_1$. We start by discussing what the role is of $p_1=\mathbb{P}(D=1)$ and $p_2=\mathbb{P}(D=2)$. Note that when $p_1=0$, $\nu=\mathbb{E}[D(D-1)]/\mathbb{E}[D]=\sum_{k=2}^{\infty}k(k-1)p_k/\sum_{k\geq 2}kp_k\geq 1$, with strict inequality as soon as $p_k>0$ for some $k\geq 3$.

The case $p_1=0$ and $\nu=1$, i.e., $p_k=0$ for all $k\neq 0,2$, is quite exceptional. In this case, $H(x)=0$ for all x. We give three examples showing that quite different behaviors are possible. Because isolated vertices do not matter, let us assume $p_0=0$ too, and consider thus the case $p_2=1$.

One example is when $d_i=2$ for all $i\in[n]$, so we are studying a random 2-regular graph. In this case, the components are cycles and the distribution of cycle lengths in $\mathrm{CM}_n(\boldsymbol{d})$ is given by the Ewen's sampling formula ESF(1/2), see e.g. [14]. This implies that $|\mathcal{C}_{\max}|/n$ converges in distribution to a non-degenerate distribution on $[0,1]$ and not to any constant [14, Lemma 5.7]. Moreover, the same is true for $|\mathcal{C}_{(2)}|/n$ (and for $|\mathcal{C}_{(3)}|/n$,...), so in this case there are several large components. To intuitively see this result, we note that in the exploration of a cluster we start with one vertex with two half-edges. When paring a half-edge, it will connect to a vertex that again has two half-edges. Therefore, the number of half-edges to be paired is always equal to 2, up to the moment when the cycle is closed, and the cluster is completed. When there are $m=\alpha n$ half-edges left, the probability of closing up the cycle equals $1/m=1/(\alpha n)$, and, thus, the time this will take is of order n. A slight extension of this reasoning shows that the time it takes to close a cycle is nT_n, where T_n converges to a limiting non-degenerate random variable:

Exercise 10.1 (Cluster size of vertex 1 in a 2-regular graph)**.** *Let $n_2=n$, and let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that*

$$|\mathcal{C}(1)|/n\xrightarrow{d}T, \tag{10.1.13}$$

where $\mathbb{P}(T>x)=1/\sqrt{1-x}$.

A second case with $p_2=1$ is obtained by adding a small number of vertices of degree 1. More precisely, let $n_1\to\infty$, $n_1/n\to 0$, and $n_2=n-n_1$. It is then easy to see that $|\mathcal{C}_{\max}|=o_{\mathbb{P}}(n)$:

Exercise 10.2 (Cluster size in a 2-regular graph with some degree 1 vertices)**.** *Let $n_1\to\infty$ with $n_1/n\to 0$, and $n_2=n-n_1$. Let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that*

$$|\mathcal{C}(1)|/n\xrightarrow{\mathbb{P}}0. \tag{10.1.14}$$

A third case with $p_2=1$ is obtained by instead adding a small number of vertices of degree 4 (i.e., $n_4\to\infty$, $n_2=n-n_4$.) By regarding each vertex of degree 4 as two vertices of degree 2 that have merged, it is easy to see that in this case $|\mathcal{C}_{\max}|=n-o_{\mathbb{P}}(n)$, so there is a giant component containing almost everything.

We conclude that the case where $p_2=\mathbb{P}(D=2)=1$ is quite sensitive to the precise properties of the degree structure that are not captured by the limiting distribution $(p_k)_{k\geq 1}$ only. In the sequel, we shall ignore the case where $p_2=1$.

Organization of the proof of Theorem 10.1. Theorem 10.1 is proved using a clever *randomization scheme* to explore the connected components one by one. This construction is explained terms of a simple continuous-time algorithm in Section 10.1.1 below. The algorithm describes the number of vertices of given degrees that have been found, as well as the total number of unpaired half edges, at time $t>0$. It is proved that, when $n\to\infty$, these quantities all converge in probability to *deterministic functions* described in terms of the functions $x\mapsto H(x)$ and $x\mapsto G_D^*(x)$ above. In particular, the number of unpaired half-edges is given in terms of $x\mapsto H(x)$, so that the first zero of this function gives the size of the giant component. In Section 10.1.2, the algorithm is analyzed by showing that,

after a short initial period of exploring small clusters, the giant component is found, and the exploration will explore it completely.

10.1.1 Finding the largest component

The components of an arbitrary finite graph or multigraph can be found by the following standard procedure. Pick an arbitrary vertex v and determine the component of v as follows: include all the neighbors of v in an arbitrary order; then add in the neighbors of the neighbors, and so on, until no more vertices can be added. The vertices included until this moment form the component of v. If there are still vertices left in the graph, then pick any such vertex w, and repeat the above to determine the second component (the component of vertex w). Carry on in this manner until all the components have been found.

The same result can be more conveniently obtained in the following way. Regard each edge as consisting of two half-edges, each half-edge having one endpoint. We will label the vertices as sleeping or awake (= used) and the half-edges as sleeping, active or dead; the sleeping and active half-edges are also called living. We start with all vertices and half-edges sleeping. Pick a vertex and label its half-edges as active. Then take any active half-edge, say x and find its partner y in the graph; label these two half-edges as dead; further, if the endpoint of y is sleeping, label it as awake and all other half-edges there as active. Repeat as long as there is any active half-edge. When there is no active half-edge left, we have obtained the first component. Then start again with another vertex until all components are found.

We apply this algorithm to $\mathrm{CM}_n(\boldsymbol{d})$ with a given degree sequence, revealing its edges during the process. We thus observe initially only the vertex degrees and the half-edges, but not how they are joined to form edges. Hence, each time we need a partner of an half-edge, it is uniformly distributed over all other living half-edges, with the understanding that the dead half-edges are the ones that are already paired into edges. It is here that we are using the specific structure of the configuration model, which simplifies the analysis substantially.

We make the random choices of finding a partner to the edges by associating i.i.d. random maximal lifetimes τ_x to half-edge x, where τ_x has an Exp(1) distribution. We interpret these lifetimes as *clocks*, and changes in our exploration process only occur when a clock of a half-edge rings. In other words, each half-edge dies spontaneously at rate 1 (unless killed earlier). Each time we need to find the partner of a half-edge x, we then wait until the next living half-edge $\neq x$ dies and take that one. This process in continuous-time can be formulated as an *algorithm*, constructing $\mathrm{CM}_n(\boldsymbol{d})$ and exploring its components simultaneously, as follows. Recall that we start with all vertices and half-edges sleeping. The exploration is then formalized in the following steps:

Step 1 When there is no active half-edge (as in the beginning), select a sleeping vertex and declare it awake and all its half-edges active. For definiteness, we choose the vertex by choosing a half-edge uniformly at random among all sleeping half-edges. When there is no sleeping half-edge left, the process stops; the remaining sleeping vertices are all isolated and we have explored all other components.

Step 2 Pick an active half-edge (which one does not matter) and kill it, i.e., change its status to dead.

Step 3 Wait until the next half-edge dies (spontaneously, as a result of its clock ringing). This half-edge is joined to the one killed in the previous step Step 2 to form an edge of the graph. When the vertex it belongs to is sleeping, we change this vertex to awake and all other half-edges there to active. Repeat from Step 1.

The above randomized algorithm is such that components are created between the successive times Step 1 is performed, where we say that Step 1 is performed when there is no active half-edge and, as a result, a new vertex is chosen.

The vertices in the component created during one of these intervals are the vertices that are awakened during the interval. Note also that a component is completed and Step 1 is performed exactly when the number of active half-edges is 0 and a half-edge dies at a vertex where all other half-edges (if any) are dead. In the next section, we shall investigate the behavior of the key characteristics of the algorithm, such as the number of sleeping half-edges and the number of sleeping vertices of a given degree.

10.1.2 Analysis of the algorithm for $\mathrm{CM}_n(\boldsymbol{d})$

We start by introducing the key characteristics of the exploration algorithm. Let $S(t)$ and $A(t)$ be the numbers of sleeping and active half-edges, respectively, at time t, and let

$$L(t) = S(t) + A(t) \tag{10.1.15}$$

be the number of living half-edges. For definiteness, we define these random functions to be right-continuous.

Let us first look at $L(t)$. We start with ℓ_n half-edges, all sleeping and thus living, but we immediately perform Step 1 and Step 2 and kill one of them. Thus, $L(0) = \ell_n - 1$. In the sequel, as soon as a living half-edge dies, we perform Step 3 and then (instantly) either Step 2 or both Step 1 and Step 2. Since Step 1 does not change the number of living half-edges while Step 2 and Step 3 each decrease it by 1, the total result is that $L(t)$ is decreased by 2 each time one of the living half-edges dies, except when the last living one dies and the process terminates. Because of this simple dynamics of $t \mapsto S(t)$, we can give sharp asymptotics of $S(t)$ when $n \to \infty$:

Proposition 10.2 (The number of sleeping half-edges)**.** *As $n \to \infty$, for any $t_0 \geq 0$ fixed,*

$$\sup_{0 \leq t \leq t_0} |n^{-1} L(t) - \mathbb{E}[D_n] \mathrm{e}^{-2t}| \xrightarrow{\mathbb{P}} 0. \tag{10.1.16}$$

Proof. The process $t \mapsto L(t)$ satisfies $L(0) = \ell_n - 1$, and it degreases by 2 at rate $L(t)$. As a result, it is closely related to a *death process*. We study such processes in the following lemma:

Lemma 10.3 (Asymptotics of death processes)**.** *Let $d, \gamma > 0$ be given and let $(N^{(x)}(t))_{t \geq 0}$ be a Markov process such that $N^{(x)}(t) = x$ a.s., and the dynamics of $t \mapsto (N^{(x)}(t))_{t \geq 0}$ is such that when it is in position y, then it jumps down by d at rate γy. In other words, the waiting time until the next event is $\mathrm{Exp}(1/\gamma y)$ and each jump is of size d downwards. Then, for every $t_0 \geq 0$,*

$$\mathbb{E}\Big[\sup_{t \leq t_0} \big| N^{(x)}(t) - \mathrm{e}^{-\gamma d t} x \big|^2 \Big] \leq 8d(\mathrm{e}^{\gamma d t_0} - 1)x + 8d^2. \tag{10.1.17}$$

Proof. The proof follows by distinguishing several cases. First assume that $d = 1$ and that x is an integer. In this case, the process is a standard pure death process taking the values $x, x-1, x-2, \ldots, 0$, describing the number of particles alive when the particles die independently with rate $\gamma > 0$. As is well-known, and easily seen by regarding $N^{(x)}(t)$ as the sum of x independent copies of the process $N^{(1)}(t)$, the process $(\mathrm{e}^{\gamma t} N^{(x)}(t))_{t \geq 0}$, is a *martingale* starting in x. Furthermore, for every $t \geq 0$, the random variable $N^{(x)}(t)$ has a $\mathrm{Bin}(x, \mathrm{e}^{-\gamma t})$ distribution, since each of the x particles has a probability of dying before time t of $\mathrm{e}^{-\gamma t}$, and the different particles die independently. Hence, by Doob's inequality,

$$\begin{aligned} \mathbb{E}\Big[\sup_{t \leq t_0} \big| N^{(x)}(t) - \mathrm{e}^{-\gamma t} x \big|^2 \Big] &\leq \mathbb{E}\Big[\sup_{t \leq t_0} \big| \mathrm{e}^{\gamma t} N^{(x)}(t) - x \big|^2 \Big] \leq 4\mathbb{E}\Big[\big(\mathrm{e}^{\gamma t} N^{(x)}(t_0) - x \big)^2 \Big] \\ &= 4\mathrm{e}^{2\gamma t} \mathrm{Var}(N^{(x)}(t_0)) \leq 4(\mathrm{e}^{\gamma t_0} - 1)x. \end{aligned} \tag{10.1.18}$$

This proves the claim for x being integer.

Next, still assume $d = 1$, but let $x > 0$ be arbitrary. We can couple the two processes $\big(N^{(x)}(t)\big)_{t\geq 0}$ and $\big(N^{(\lfloor x\rfloor)}(t)\big)_{t\geq 0}$ with different initial values such that whenever the smaller one jumps by 1, so does the other. This coupling keeps

$$|N^{(x)}(t) - N^{(\lfloor x\rfloor)}(t)| < 1 \tag{10.1.19}$$

for all $t \geq 0$, and thus,

$$\sup_{t\leq t_0} \big|N^{(\lfloor x\rfloor)}(t) - \mathrm{e}^{-\gamma t}\lfloor x\rfloor\big| \leq \sup_{t\leq t_0} \big|N^{(x)}(t) - \mathrm{e}^{-\gamma t}x\big| + 2, \tag{10.1.20}$$

so that by (10.1.18), in turn,

$$\mathbb{E}\Big[\sup_{t\leq t_0} \big|N^{(x)}(t) - \mathrm{e}^{-\gamma t}x\big|^2\Big] \leq 8(\mathrm{e}^{\gamma t_0} - 1)x + 8. \tag{10.1.21}$$

Finally, for a general $d > 0$, we observe that $N^{(x)}(t)/d$ is a process of the same type with the parameters (γ, d, x) replaced by $(\gamma d, 1, x/d)$, and the general result follows from (10.1.21) and (10.1.18). □

The proof of Proposition 10.2 follows from Lemma 10.3 with $d = 2, x = (\ell_n - 1) = n\mathbb{E}[D_n] - 1$ and $\gamma = 1$. □

We continue by considering the sleeping half-edges. Let $V_k(t)$ be the number of sleeping vertices of degree k at time t, so that

$$S(t) = \sum_{k=1}^{\infty} kV_k(t). \tag{10.1.22}$$

Note that Step 2 does not affect sleeping half-edges, and that Step 3 implies that each sleeping vertex of degree k is eliminated (i.e., awakened) with intensity k, independently of all other vertices. There are also some sleeping vertices eliminated by Step 1, though, which complicates the dynamics of $t \mapsto V_k(t)$.

We first ignore the effect of Step 1 by letting $\widetilde{V}_k(t)$ be the number of vertices of degree k such that all its half-edges have maximal lifetimes $\tau_x > t$. Thus, none of its k half-edges would have died spontaneously up to time t, assuming they all escaped Step 1. It is reasonable to ignore the effect of Step 1 in the leading order, as we perform Step 1 until we hit the giant component, and then it takes a long time to find the entire giant component. The number of times we perform Step 1 until we find the giant component will be small, as each time we have a strictly positive probability of choosing a vertex in the giant component. Thus, intuitively, the difference between $V_k(t)$ and $\widetilde{V}_k(t)$ can be expected to be bounded.

For a given half-edge, we call the half-edges incident to the same vertex its *brother half-edges*. Let further

$$\widetilde{S}(t) = \sum_{k} k\widetilde{V}_k(t) \tag{10.1.23}$$

denote the number of half-edges whose brother half-edges have escaped spontaneous death up to time t.

Recall the functions G_D, G_D^* from (10.1.9), (10.1.10), and define

$$h(x) = x\mathbb{E}[D]G_D^*(x). \tag{10.1.24}$$

Then, we can identify the asymptotics of $(\widetilde{V}_k(t))_{t\geq 0}$ in a similar way as in Proposition 10.2:

Lemma 10.4 (The number of sleeping vertices)**.** *Assume that Condition 7.5(a)-(b) hold. Then, as $n \to \infty$, for any $t_0 \geq 0$ fixed*

$$\sup_{t \leq t_0} |n^{-1} \widetilde{V}_k(t) - p_k \mathrm{e}^{-kt}| \xrightarrow{\mathbb{P}} 0 \tag{10.1.25}$$

for every $k \geq 0$ and

$$\sup_{t \leq t_0} |n^{-1} \sum_{k=0}^{\infty} \widetilde{V}_k(t) - G_D(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0, \tag{10.1.26}$$

$$\sup_{t \leq t_0} |n^{-1} \widetilde{S}(t) - h(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0. \tag{10.1.27}$$

Proof. The statement (10.1.25) again follows from Lemma 10.3, now with $\gamma = k$, $x = n_k$ and $d = 1$. The case $k = 0$ is trivial, with $\widetilde{V}_0(t) = n_0$ for all t. We can replace $p_k^{(n)}$ by p_k by Condition 7.5(a).

By Condition 7.5(b), D_n are uniformly integrable, which means that for every $\varepsilon > 0$ there exists $K < \infty$ such that for all n, $\sum_{k>K} k n_k/n = \mathbb{E}[D_n | D_n > k] < \varepsilon$. We may further assume (or deduce from Fatou's inequality) that $\sum_{k>K} k p_k < \varepsilon$, and obtain by (10.1.25) that, **whp**,

$$\begin{aligned}
\sup_{t \leq t_0} |n^{-1} \widetilde{S}(t) - h(\mathrm{e}^{-t})| &= \sup_{t \leq t_0} \left| \sum_{k=1}^{\infty} k (n^{-1} \widetilde{V}_k(t) - p_k \mathrm{e}^{-kt}) \right| \\
&\leq \sum_{k=1}^{K} k \sup_{t \leq t_0} |n^{-1} \widetilde{V}_k(t) - p_k \mathrm{e}^{-kt}| + \sum_{k>K} k \left(\frac{n_k}{n} + p_k \right) \\
&\leq \varepsilon + \varepsilon + \varepsilon,
\end{aligned}$$

proving (10.1.27). An almost identical argument yields (10.1.26). □

Remarkably, the difference between $S(t)$ and $\widetilde{S}(t)$ is easily estimated. The following result can be viewed as the key to why this approach works. Indeed, it gives a *uniform* upper bound on the difference due to the application of Step 1:

Lemma 10.5 (Effect of Step 1)**.** *If $\Delta := \max_{i \in [n]} d_i$ is the maximum degree of $\mathrm{CM}_n(\boldsymbol{d})$, then*

$$0 \leq \widetilde{S}(t) - S(t) < \sup_{0 \leq s \leq t} (\widetilde{S}(s) - L(s)) + \Delta. \tag{10.1.28}$$

Proof. Clearly, $V_k(t) \leq \widetilde{V}_k(t)$, and thus $S(t) \leq \widetilde{S}(t)$. Furthermore, $\widetilde{S}(t) - S(t)$ increases only as a result of Step 1. Indeed, Step 1 acts to guarantee that $A(t) = L(t) - S(t) \geq 0$, and is only performed when $A(t) = 0$.

If Step 1 is performed at time t and a vertex of degree $j > 0$ is awakened, then Step 2 applies instantly and we have $A(t) = j - 1 < \Delta$, and consequently

$$\widetilde{S}(t) - S(t) = \widetilde{S}(t) - L(t) + A(t) < \widetilde{S}(t) - L(t) + \Delta. \tag{10.1.29}$$

Furthermore, $\widetilde{S}(t) - S(t)$ is never changed by Step 2 and either unchanged or decreased by Step 3. Hence, $\widetilde{S}(t) - S(t)$ does not increase until the next time Step 1 is performed. Consequently, for any time t, if s was the last time before (or equal to) t that Step 1 was performed, then $\widetilde{S}(t) - S(t) \leq \widetilde{S}(s) - S(s)$, and the result follows by (10.1.29). □

Recall that $A(t) = L(t) - S(t)$ denotes the number of awakened vertices and let

$$\widetilde{A}(t) = L(t) - \widetilde{S}(t) = A(t) - (\widetilde{S}(t) - S(t)) \tag{10.1.30}$$

denote the number of awakened vertices ignoring the effect of Step 1. Thus, $\widetilde{A}(t) \leq A(t)$.

Then, by Lemmas 10.2 and 10.4 (and (10.1.11)), we have that, for any $t_0 \geq 0$,

$$\sup_{t \leq t_0} |n^{-1}\widetilde{A}(t) - H(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0. \tag{10.1.31}$$

Lemma 10.5 can be written

$$0 \leq \widetilde{S}(t) - S(t) < -\inf_{s \leq t} \widetilde{A}(s) + \Delta \tag{10.1.32}$$

By (10.1.30) and (10.1.32), we obtain further the relation

$$\widetilde{A}(t) \leq A(t) < \widetilde{A}(t) - \inf_{s \leq t} \widetilde{A}(s) + \Delta, \tag{10.1.33}$$

which, perhaps, illuminates the relation between $A(t)$ and $\widetilde{A}(t)$. Now, the function $t \mapsto H(\mathrm{e}^{-t})$, which acts as the limit of $\widetilde{A}(t)$, is strictly positive in $(0, -\log \xi)$ and $H(1) = H(\xi) = 0$. Therefore, we expect $\widetilde{A}(t)$ to be positive for $t \in (0, -\log \xi)$, and, if so, $\inf_{s \leq t} \widetilde{A}(s) = 0$. The idea is to continue our algorithm in Step 1-Step 3 until the giant component has been found, which implies that $A(t) > 0$ for the time of exploration of the giant component, and $A(t) = 0$ for the first time when we have completed the exploration of the giant component, which is $t = -\log \xi$. Thus, the term $\inf_{s \leq t} \widetilde{A}(s)$ in (10.1.33) ought to be negligible. When Condition 7.5(a)-(b) hold, we further have that $\Delta = o(n)$, so that one can expect $\widetilde{A}(t)$ to be a good approximation of $A(t)$. The remainder of the proof shall make this intuition precise. We start by summarizing the properties of $x \mapsto H(x)$ that we shall rely on:

Lemma 10.6 (Properties of $x \mapsto H(x)$)**.** *Suppose that Condition 7.5(a)-(b) hold and let $H(x)$ be given by* (10.1.11)*.*

(i) If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$ and $p_1 > 0$, then there is a unique $\xi \in (0,1)$ such that $H(\xi) = 0$. Moreover, $H(x) < 0$ for all $x \in (0, \xi)$ and $H(x) > 0$ for all $x \in (\xi, 1)$. If, instead, $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$ and $p_1 = 0$, then $H(x) > 0$ for all $x \in (0,1)$.

(ii) If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \leq 1$, then $H(x) < 0$ for all $x \in (0,1)$.

Proof. As remarked earlier, $H(0) = H(1) = 0$ and $H'(1) = -\mathbb{E}[D(D-2)]$. Furthermore, if we define $\phi(x) := H(x)/x$, then $\phi(x) = \mathbb{E}[D](x - G^*_D(x))$ is a concave function on $(0,1]$, and it is strictly concave unless $p_k = 0$ for all $k \geq 3$, in which case $H'(1) = -\mathbb{E}[D(D-2)] = p_1 > 0$. Indeed, when $p_k = 0$ for all $k \geq 3$, we have that $p_1 + p_2 = 1$. Since we assume that $p_2 < 1$, we thus obtain that $p_1 > 0$ in this case.

In case (ii), we thus have ϕ concave and $\phi'(1) = H'(1) - H(1) \geq 0$, with either the concavity or the inequality strict, and thus $\phi'(x) > 0$ for all $x \in (0,1)$, whence $\phi(x) < \phi(1) = 0$ for $x \in (0,1)$.

In case (i), $H'(1) < 0$, and thus $H(x) > 0$ for x close to 1. Further, when $p_1 > 0$, $H'(0) = -h'(0) = -p_1 < 0$, and thus $H(x) \leq 0$ for x close to 0. When $p_1 = 0$, instead, $H'(0) = 0$, and $H''(0) = 2\mathbb{E}[D](1-\nu) < 0$ since $\nu > 1$. Thus, $H(x) < 0$ for x close to 0. Hence, there is at least one $\xi \in (0,1)$ with $H(\xi) = 0$, and since $H(x)/x$ is strictly concave and also $H(1) = 0$, there is at most one such ξ and the result follows. □

Now we are in the position to complete the proof of Theorem 10.1 in the following section.

10.1.3 Proof of Theorem 10.1

We start with the proof of Theorem 10.1(i). Let ξ be the zero of H given by Lemma 10.6(i) and let $\tau = -\log \xi$, where $\tau = \infty$ if $\xi = 0$. Then, by Lemma 10.6, $H(\mathrm{e}^{-t}) > 0$ for $0 < t < \tau$, and thus $\inf_{t \le \tau} H(\mathrm{e}^{-t}) = 0$. Consequently, (10.1.31) implies

$$n^{-1} \inf_{t \le \tau} \widetilde{A}(t) = \inf_{t \le \tau} n^{-1} \widetilde{A}(t) - \inf_{t \le \tau} H(\mathrm{e}^{-t}) \xrightarrow{\mathbb{P}} 0 \tag{10.1.34}$$

Further, by Condition 7.5(b), $\Delta = o(n)$, and thus $n^{-1}\Delta \to 0$. Consequently, (10.1.32) and (10.1.34) yield

$$\sup_{t \le \tau} n^{-1} |A(t) - \widetilde{A}(t)| = \sup_{t \le \tau} n^{-1} |\widetilde{S}(t) - S(t)| \xrightarrow{\mathbb{P}} 0. \tag{10.1.35}$$

Thus, by (10.1.31),

$$\sup_{t \le \tau} |n^{-1} A(t) - H(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0. \tag{10.1.36}$$

Let $0 < \varepsilon < \tau/2$. To treat both cases $\tau = \infty$ and $\tau < \infty$ at the same time, we abuse notation to write $[\varepsilon, \tau - \varepsilon]$ for the interval $[\varepsilon, \varepsilon^{-1}]$ when $\tau = \infty$. Since $H(\mathrm{e}^{-t}) > 0$ on the compact interval $[\varepsilon, \tau - \varepsilon]$, (10.1.36) implies that $A(t)$ remains **whp** positive on $[\varepsilon, \tau - \varepsilon]$, and thus no new component is started during this interval.

On the other hand, again by Lemma 10.6(i) and when $\tau < \infty$, $H(\mathrm{e}^{-(\tau+\varepsilon)}) < 0$ and (10.1.31) implies $n^{-1}\widetilde{A}(\tau + \varepsilon) \xrightarrow{\mathbb{P}} H(\mathrm{e}^{-(\tau+\varepsilon)})$, while $A(\tau + \varepsilon) \ge 0$. Thus, with $\delta = |H(\mathrm{e}^{-\tau-\varepsilon})|/2 > 0$, **whp**

$$\widetilde{S}(\tau + \varepsilon) - S(\tau + \varepsilon) = A(\tau + \varepsilon) - \widetilde{A}(\tau + \varepsilon) \ge -\widetilde{A}(\tau + \varepsilon) > n\delta, \tag{10.1.37}$$

while (10.1.35) yields $\widetilde{S}(\tau) - S(\tau) < n\delta$ **whp**. Consequently, **whp** $\widetilde{S}(\tau + \varepsilon) - S(\tau + \varepsilon) > \widetilde{S}(\tau) - S(\tau)$, so Step 1 is performed between τ and $\tau + \varepsilon$.

Let T_1 be the last time Step 1 was performed before time $\tau/2$ when $\tau < \infty$, and before time 1 when $\tau = \infty$. Let T_2 be the next time Step 1 is performed (by convention, $T_2 = \infty$ if such a time does not exist). We have shown that for any $\varepsilon > 0$, and **whp** $0 \le T_1 \le \varepsilon$ and $\tau - \varepsilon \le T_2 \le \tau + \varepsilon$, where we recall that $\tau - \varepsilon = 1/\varepsilon$ when $\tau = \infty$. In other words, $T_1 \xrightarrow{\mathbb{P}} 0$ and $T_2 \xrightarrow{\mathbb{P}} \tau$. We conclude that we have found one component that is explored between time $T_1 \xrightarrow{\mathbb{P}} 0$ and time $T_2 \xrightarrow{\mathbb{P}} \tau$. This is our candidate for the giant component, and we continue to study its properties, i.e., its size, its number of edges and its number of vertices of degree k. These properties are stated separately in the next lemma, so that we are able to reuse them later on:

Proposition 10.7 (Cluster properties)**.** *Let T_1^* and T_2^* be two random times when Step 1 is performed, with $T_1^* \le T_2^*$, and assume that $T_1^* \xrightarrow{\mathbb{P}} t_1$ and $T_2^* \xrightarrow{\mathbb{P}} t_2$ where $0 \le t_1 \le t_2 \le \tau < \infty$. If $\mathcal{C}^*$ is the union of all components explored between T_1^* and T_2^*, then*

$$v_k(\mathcal{C}^*)/n \xrightarrow{\mathbb{P}} p_k(\mathrm{e}^{-kt_1} - \mathrm{e}^{-kt_2}), \quad k \ge 0, \tag{10.1.38}$$

$$|\mathcal{C}^*|/n \xrightarrow{\mathbb{P}} G_D(\mathrm{e}^{-t_1}) - G_D(\mathrm{e}^{-t_2}), \tag{10.1.39}$$

$$|E(\mathcal{C}^*)|/n \xrightarrow{\mathbb{P}} \frac{1}{2} h(\mathrm{e}^{-t_1}) - \frac{1}{2} h(\mathrm{e}^{-t_2}). \tag{10.1.40}$$

In particular, if $t_1 = t_2$, then $|\mathcal{C}^*|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}^*)| \xrightarrow{\mathbb{P}} 0$. The same statements hold for $T_2^* = \infty$.

We apply Proposition 10.7 to $T_1 = o_{\mathbb{P}}(1)$ and T_2, where $T_2 = \tau + o_{\mathbb{P}}(1)$ when $\tau < \infty$, and either $T_2 = \infty$ or $T_2 \xrightarrow{\mathbb{P}} \infty$ when $\tau = \infty$. It is here that the case $T_2^* = \infty$ is relevant. We can identify the values of the above constants:

Exercise 10.3 (Limiting constants)**.** *Prove that for $t_1 = 0$ and $t_2 = \tau$, $\mathrm{e}^{-kt_1} = 1, \mathrm{e}^{-kt_2} = \xi$, $G_D(\mathrm{e}^{-t_1}) = 1$, $G_D(\mathrm{e}^{-t_2}) = 1 - \zeta$, $h(\mathrm{e}^{-t_1}) = 2\mathbb{E}[D]$, $h(\mathrm{e}^{-t_2}) = 2\mathbb{E}[D]\xi^2$, where, for $\tau = \infty$, e^{-t_2} should be interpreted as 0.*

By Proposition 10.7 and Exercise 10.3, Theorem 10.1(i) follows when we prove that the cluster found between times T_1 and T_2 is indeed the giant component. We now first prove Proposition 10.7, after which we show that cluster found between times T_1 and T_2 is indeed the giant component.

Proof. The set of vertices $\mathcal{C}^*$ contains all vertices awakened in the interval $[T_1^*, T_2^*)$ and no others, and thus

$$v_k(\mathcal{C}^*) = V_k(T_1^*-) - V_k(T_2^*-), \qquad k \geq 1. \tag{10.1.41}$$

Since $T_2^* \xrightarrow{\mathbb{P}} t_2 \leq \tau$ and H is continuous, we obtain that $\inf_{t \leq T_2^*} H(\mathrm{e}^{-t}) \xrightarrow{\mathbb{P}} \inf_{t \leq t_2} H(\mathrm{e}^{-t}) = 0$, where the latter equality follows since $H(1) = 0$. Now, (10.1.31) and (10.1.32) imply, in analogy with (10.1.34) and (10.1.35), that $n^{-1} \inf_{t \leq T_2^*} \widetilde{A}(t) \xrightarrow{\mathbb{P}} 0$ and

$$\sup_{t \leq T_2^*} n^{-1} |\widetilde{S}(t) - S(t)| \xrightarrow{\mathbb{P}} 0. \tag{10.1.42}$$

Since $\widetilde{V}_j(t) \geq V_j(t)$ for every j and $t \geq 0$,

$$\widetilde{V}_k(t) - V_k(t) \leq k^{-1} \sum_{j=1}^{\infty} j(\widetilde{V}_j(t) - V_j(t)) = k^{-1}(\widetilde{S}(t) - S(t)), \quad k \geq 1. \tag{10.1.43}$$

Hence (10.1.42) implies, for every $k \geq 1$, $\sup_{t \leq T_2^*} |\widetilde{V}_k(t) - V_k(t)| = o_{\mathbb{P}}(n)$. This is further trivially true for $k = 0$ too. Consequently, using Lemma 10.4, for $j = 1, 2$,

$$V_k(T_j^*-) = \widetilde{V}_k(T_j^*-) + o_{\mathbb{P}}(n) = np_k \mathrm{e}^{-kT_j^*} + o_{\mathbb{P}}(n) = np_k \mathrm{e}^{-kt_j} + o_{\mathbb{P}}(n), \tag{10.1.44}$$

and (10.1.38) follows by (10.1.41). Similarly, using $\sum_{k=0}^{\infty} (\widetilde{V}_k(t) - V_k(t)) \leq \widetilde{S}(t) - S(t)$,

$$|\mathcal{C}^*| = \sum_{k=1}^{\infty} (V_k(T_1^*-) - V_k(T_2^*-)) = \sum_{k=1}^{\infty} (\widetilde{V}_k(T_1^*-) - \widetilde{V}_k(T_2^*)) + o_{\mathbb{P}}(n) \tag{10.1.45}$$

$$= nG_D(\mathrm{e}^{-T_1^*}) - nG_D(\mathrm{e}^{-T_2^*}) + o_{\mathbb{P}}(n), \tag{10.1.46}$$

and

$$2|E(\mathcal{C}^*)| = \sum_{k=1}^{\infty} k(V_k(T_1^*-) - V_k(T_2^*)) = \sum_{k=1}^{\infty} k(\widetilde{V}_k(T_1^*-) - \widetilde{V}_k(T_2^*)) + o_{\mathbb{P}}(n) \tag{10.1.47}$$

$$= nh(\mathrm{e}^{-T_1^*}) - nh(\mathrm{e}^{-T_2^*}) + o_{\mathbb{P}}(n), \tag{10.1.48}$$

and (10.1.39) and (10.1.40) follow from the convergence $T_i^* \xrightarrow{\mathbb{P}} t_i$ and the continuity of $t \mapsto G_D(\mathrm{e}^{-t})$ and $t \mapsto h(\mathrm{e}^{-t})$. □

Let $\mathcal{C}'_{\max}$ be the component created at T_1 and explored until T_2, where we recall that T_1 is the last time Step 1 was performed before $\tau/2$ (and before time 1 when $\tau = \infty$), and let

T_2 be the next time it is performed if this occurs and $T_2 = \infty$ otherwise. Then, $T_1 \xrightarrow{\mathbb{P}} 0$ and $T_2 \xrightarrow{\mathbb{P}} \tau$. The cluster $\mathcal{C}'_{\max}$ is our candidate for the giant component $\mathcal{C}_{\max}$, and we next prove that indeed it is, **whp**, the largest connected component.

By Proposition 10.7, with $t_1 = 0$ and $t_2 = \tau$,

$$|v_k(\mathcal{C}'_{\max})|/n \xrightarrow{\mathbb{P}} p_k(1 - \mathrm{e}^{-kt}), \tag{10.1.49}$$

$$|\mathcal{C}'_{\max}|/n \xrightarrow{\mathbb{P}} G_D(1) - G_D(\mathrm{e}^{-\tau}) = 1 - G_D(\xi), \tag{10.1.50}$$

$$|E(\mathcal{C}'_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}(h(1) - h(\mathrm{e}^{-\tau})) = \frac{1}{2}(h(1) - h(\xi)) = \frac{\mathbb{E}[D]}{2}(1 - \xi^2), \tag{10.1.51}$$

using Exercise 10.3.

We have found one large component $\mathcal{C}'_{\max}$ with the claimed numbers of vertices and edges. It remains to show that **whp** there is no other large component.

For this, let $\eta > 0$, and apply Proposition 10.7 to $T_0 = 0$ and T_1. Then, we see that the total number of vertices and edges in *all* components found before $\mathcal{C}'_{\max}$, i.e., before T_1, is $o_{\mathbb{P}}(n)$, because $T_1 \xrightarrow{\mathbb{P}} 0$. Hence, recalling that $\ell_n = \Theta(n)$ by Condition 7.5(b),

$$\mathbb{P}(\text{a component } \mathcal{C} \text{ with } |E(\mathcal{C})| \geq \eta\ell_n \text{ is found before } \mathcal{C}'_{\max}) \to 0. \tag{10.1.52}$$

We conclude that **whp** no component containing at least $\eta\ell_n$ half-edges is found *before* $\mathcal{C}'_{\max}$ is found.

In order to study the probability of finding a component containing at least $\eta\ell_n$ edges *after* $\mathcal{C}'_{\max}$ is found, we start by letting T_3 be the first time after T_2 that Step 1 is performed (if $T_2 < \infty$). The argument below thus only applies when $T_2 < \infty$. Since $\widetilde{S}(t) - S(t)$ increases by at most $\Delta = o(n)$ each time Step 1 is performed, we obtain from (10.1.42) that

$$\sup_{t \leq T_3}(\widetilde{S}(t) - S(t)) \leq \sup_{t \leq T_2}(\widetilde{S}(t) - S(t)) + \Delta = o_{\mathbb{P}}(n). \tag{10.1.53}$$

Comparing this to (10.1.37), for every $\varepsilon > 0$ and **whp**, we have that $\tau + \varepsilon > T_3$. Since also $T_3 > T_2 \xrightarrow{\mathbb{P}} \tau$, it follows that $T_3 \xrightarrow{\mathbb{P}} \tau$. If $\mathcal{C}'$ is the component created between T_2 and T_3, then Proposition 10.7 applied to T_2 and T_3 yields $|\mathcal{C}'|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}')| \xrightarrow{\mathbb{P}} 0$.

On the other hand, if there would exist a component $\mathcal{C} \neq \mathcal{C}'_{\max}$ in $\mathrm{CM}_n(\boldsymbol{d})$ with at least $\eta\ell_n$ edges that has not been found before $\mathcal{C}'_{\max}$, then with probability at least η, the vertex chosen at random by Step 1 at T_2 starting the component $\mathcal{C}'$ would belong to $\mathcal{C}$. When this occurs, we clearly have that $\mathcal{C} = \mathcal{C}'$. Consequently,

$$\mathbb{P}(\text{a component } \mathcal{C} \text{ with } |E(\mathcal{C})| \geq \eta\ell_n \text{ is found after } \mathcal{C}'_{\max}) \leq \eta^{-1}\mathbb{P}(|E(\mathcal{C}')| \geq \eta\ell_n) \to 0, \tag{10.1.54}$$

since $|E(\mathcal{C}')| \xrightarrow{\mathbb{P}} 0$.

Combining (10.1.52) and (10.1.54), we see that **whp** there is no component except $\mathcal{C}'_{\max}$. As a result, we must have that $\mathcal{C}'_{\max} = \mathcal{C}_{\max}$, where $\mathcal{C}_{\max}$ is the largest component. Further, again **whp**, $|E(\mathcal{C}_{(2)})| < \eta\ell_n$. Consequently, the results for $\mathcal{C}_{\max}$ follow from (10.1.49)-(10.1.51). We have further shown $|E(\mathcal{C}_{(2)}|/\ell_n \xrightarrow{\mathbb{P}} 0$, which implies $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$ and $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ because $\ell_n = \Theta(n)$ and $|\mathcal{C}_{(2)}| \leq |E(\mathcal{C}_{(2)})| + 1$. This completes the proof of Theorem 10.1(i). □

The proof of Theorem 10.1(ii) is very similar to the last step in the proof for Theorem 10.1(i). Indeed, let $T_1 = 0$ and let T_2 be the next time Step 1 is performed, or $T_2 = \infty$ when this does not occur. Then,

$$\sup_{t \leq T_2} |A(t) - \widetilde{A}(t)| = \sup_{t \leq T_2} |\widetilde{S}(t) - S(t)| \leq 2\Delta = o(n). \tag{10.1.55}$$

For every $\varepsilon > 0$, we have by (10.1.31) and Lemma 10.6(ii) $n^{-1}\widetilde{A}(\varepsilon) \xrightarrow{\mathbb{P}} H(\mathrm{e}^{-\varepsilon}) < 0$, while $A(\varepsilon) \geq 0$, and it follows from (10.1.55) that **whp** $T_2 < \varepsilon$. Hence, $T_2 \xrightarrow{\mathbb{P}} 0$. We apply Proposition 10.7 (which holds in this case too, with $\tau = 0$) and find that if $\mathcal{C}$ is the first component found, then $|E(\mathcal{C})|/n \xrightarrow{\mathbb{P}} 0$.

Let $\varepsilon > 0$. If $|E(\mathcal{C}_{\max})| \geq \varepsilon \ell_n$, then the probability that the first half-edge chosen by Step 1 belongs to $\mathcal{C}_{\max}$, and thus $\widetilde{\mathcal{C}} = \mathcal{C}_{\max}$, is $2|E(\mathcal{C}_{\max})|/(2\ell_n) \geq \varepsilon$, and hence,

$$\mathbb{P}(|E(\mathcal{C}_{\max})| \geq \varepsilon \ell_n) \leq \varepsilon^{-1}\mathbb{P}(|E(\mathcal{C})| \geq \varepsilon \ell_n) \to 0. \tag{10.1.56}$$

The results follows since $\ell_n = \Theta(n)$ by Condition 7.5(b) and $|\mathcal{C}_{\max}| \leq |E(\mathcal{C}_{\max})| + 1$. This completes the proof of Theorem 10.1(ii), and thus that of Theorem 10.1. □

10.1.4 The giant component of related random graphs

In this section, we extend the results of Theorem 10.1 to some related models, such as uniform simple random graphs with a given degree sequence, as well as generalized random graphs.

In Section 7.4, we have introduced the notation $\mathrm{UG}_n(\boldsymbol{d})$ for a uniform simple random graph with degrees $\boldsymbol{d}$. The results in Theorem 10.1 also hold for $\mathrm{UG}_n(\boldsymbol{d})$:

Theorem 10.8 (Phase transition in $\mathrm{UG}_n(\boldsymbol{d})$)**.** *Let $\boldsymbol{d}$ satisfy Condition 7.5(a)-(c). Then, the results in Theorem 10.1 also hold for a* uniform *simple graph with degree sequence $\boldsymbol{d}$.*

Proof. By Corollary 7.12, and since $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies Condition 7.5(a)-(c), any event $\mathcal{E}_n$ that occurs with high probability for $\mathrm{CM}_n(\boldsymbol{d})$, also occurs with high probability for $\mathrm{UG}_n(\boldsymbol{d})$. By Theorem 10.1, the event $\mathcal{E}_n$ that $\{\big||\mathcal{C}_{\max}|/n - \zeta\big| \leq \varepsilon\}$ occurs with high probability for $\mathrm{CM}_n(\boldsymbol{d})$, so it also holds with high probability for $\mathrm{UG}_n(\boldsymbol{d})$. The proof for the other limits is identical. □

We next prove Theorem 9.31 for $\mathrm{GRG}_n(\boldsymbol{w})$:

Theorem 10.9 (Phase transition in $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *Let $\boldsymbol{w}$ satisfy Condition 6.4(a)-(c). Then, the results in Theorem 10.1 also hold for* $\mathrm{GRG}_n(\boldsymbol{w})$*,* $\mathrm{CL}_n(\boldsymbol{w})$ *and* $\mathrm{NR}_n(\boldsymbol{w})$*.*

Proof. Let D_i be the degree of vertex i in $\mathrm{GRG}_n(\boldsymbol{w})$ defined in (6.2.1). By Theorem 7.13, the law of $\mathrm{GRG}_n(\boldsymbol{w})$ conditionally on $\boldsymbol{D} = \boldsymbol{d}$ and $\mathrm{CM}_n(\boldsymbol{d})$ conditionally on being simple agree. Assume that $(D_i)_{i\in[n]}$ satisfies that Condition 7.5(a)-(c) hold in probability. Then, by Theorems 7.13 and 10.1, the results in Theorem 10.1 also hold for $\mathrm{GRG}_n(\boldsymbol{w})$. By Theorem 6.19, the same result applies to $\mathrm{CL}_n(\boldsymbol{w})$. By Exercise 6.36, the same applies to $\mathrm{NR}_n(\boldsymbol{w})$. Thus, we are left to prove that Condition 6.4(a)-(c) implies Condition 7.5(a)-(c) for $\mathrm{GRG}_n(\boldsymbol{w})$, which is the content of the next proposition:

Proposition 10.10 (Relating the assumptions on weights and degrees)**.** *Let $\boldsymbol{d} = (d_i)_{i\in[n]}$ be the (random) degrees in* $\mathrm{GRG}_n(\boldsymbol{w})$*. Let $\boldsymbol{w}$ satisfy Condition 6.4(a)-(c). Then, Condition 7.5(a)-(c) holds for $\boldsymbol{d}$, where the convergence holds in probability, and the limit law D equals a mixed Poisson random variable with mixing distribution W.*

Proof. Condition 7.5(a) follows from Theorem 6.9. To avoid confusion, we denote by $\mathbb{E}_n$ the conditional expectation given the graph $\mathrm{GRG}_n(\boldsymbol{w})$, so that $\mathbb{E}_n[D_n]$ denotes the degree of a vertex chosen uniformly from $\mathrm{GRG}_n(\boldsymbol{w})$ *given the realization of* $\mathrm{GRG}_n(\boldsymbol{w})$. Thus, for Condition 7.5(b)-(c), we need to show that $\mathbb{E}_n[D_n] \xrightarrow{\mathbb{P}} \mathbb{E}[D] = \mathbb{E}[W]$, $\mathbb{E}_n[D_n^2] \xrightarrow{\mathbb{P}} \mathbb{E}[D^2] = \mathbb{E}[W(W+1)]$.

We perform the proof of Condition 7.5(b)-(c) simultaneously. Let $p \in \{1, 2\}$. Then, we compute

$$\mathbb{E}_n[D_n^p] = \sum_{k=1}^{\infty} k^p P_k^{(n)}, \tag{10.1.57}$$

where we recall that $P_k^{(n)} = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{d_i=k\}}$ denotes the proportion of vertices with degree k. Let $K \geq 1$ be a large constant, and split

$$\mathbb{E}_n[D_n^2] = \sum_{k=1}^{K} k^p P_k^{(n)} + \sum_{k=K+1}^{\infty} k^p P_k^{(n)}. \tag{10.1.58}$$

By Theorem 6.9,

$$\sum_{k=1}^{K} k^p P_k^{(n)} \xrightarrow{\mathbb{P}} \sum_{k=1}^{K} k^p p_k = \mathbb{E}[D^p \mathbb{1}_{\{D\leq K\}}]. \tag{10.1.59}$$

By the Dominated Convergence Theorem (Theorem A.9), $\mathbb{E}[D^p\mathbb{1}_{\{D\leq K\}}] \to \mathbb{E}[D^p]$ when $K\to\infty$. Therefore, we are left to show that, when first $n\to\infty$ followed by $K\to\infty$,

$$\sum_{k=K+1}^{\infty} k^p P_k^{(n)} \xrightarrow{\mathbb{P}} 1. \tag{10.1.60}$$

We rewrite

$$\sum_{k=K+1}^{\infty} k^p P_k^{(n)} = \frac{1}{n} \sum_{i\in[n]:\ d_i\geq K+1} d_i^p. \tag{10.1.61}$$

We split the above sum, depending on whether $w_i \geq d_i/2$ or not, to obtain

$$\sum_{k=K+1}^{\infty} k^p P_k^{(n)} \leq \frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}} + \frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i\geq d_i/2\}}. \tag{10.1.62}$$

We bound each of these contributions separately. We start by bounding

$$\frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i\geq d_i/2\}} \leq \frac{2^p}{n}\sum_{i\in[n]} w_i^p \mathbb{1}_{\{w_i\geq K/2\}} = 2^d \mathbb{E}[W_n^p \mathbb{1}_{\{W_n\geq K/2\}}], \tag{10.1.63}$$

the right-hand side being deterministic. By Condition 7.5(b)-(c), for $p\in\{1,2\}$,

$$\mathbb{E}[W_n^p \mathbb{1}_{\{W_n\geq K/2\}}] \to \mathbb{E}[W^p \mathbb{1}_{\{W\geq K/2\}}] \leq \varepsilon/2, \tag{10.1.64}$$

when we take $K = K(\varepsilon)$ sufficiently large.

We continue to bound the expectation

$$\frac{1}{n}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}}\Big] = \frac{1}{n}\sum_{i\in[n]} \mathbb{E}[d_i^p \mathbb{1}_{\{d_i>(2w_i\vee K)\}}]. \tag{10.1.65}$$

The random degree d_i is stochastically bounded by a $\mathrm{Poi}(w_i)$ random variable (recall Proposition 9.10), so that

$$\mathbb{E}[d_i^p \mathbb{1}_{\{d_i>(2w_i\vee K)\}}] \leq \mathbb{E}[\mathrm{Poi}(w_i)^p \mathbb{1}_{\{\mathrm{Poi}(w_i)>(2w_i\vee K)\}}]. \tag{10.1.66}$$

Let $Y\sim\mathrm{Poi}(\lambda)$. Then, for each $k\geq 2p$,

$$\begin{aligned}
\mathbb{E}[Y^p\mathbb{1}_{\{Y\geq k\}}] &\leq 2^p \mathbb{E}[Y(Y-1)\cdots(Y-p+1)\mathbb{1}_{\{Y\geq k\}}] \\
&= 2^p \sum_{y=k}^{\infty} y(y-1)\cdots(y-p+1)\mathbb{P}(Y=y) \\
&= 2^p \sum_{y=k}^{\infty} y(y-1)\cdots(y-p+1)\frac{\lambda^y \mathrm{e}^{-\lambda}}{y!} = 2^p\lambda^p \mathbb{P}(Y\geq k-p).
\end{aligned} \tag{10.1.67}$$

Therefore,

$$\mathbb{E}[\mathrm{Poi}(w_i)^p \mathbb{1}_{\{\mathrm{Poi}(w_i)>(2w_i\vee K)\}}] \le 2^p w_i^p \mathbb{P}(\mathrm{Poi}(w_i) > (2w_i \vee K)), \tag{10.1.68}$$

and we arrive at

$$\begin{aligned}\frac{1}{n}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\ge K+1, w_i<d_i/2\}}\Big] &\le \frac{2^p}{n}\sum_{i\in[n]} w_i^p \mathbb{P}(\mathrm{Poi}(w_i) > (2w_i\vee K)) \\ &= 2^p \mathbb{E}[W_n^p \mathbb{1}_{\{\mathrm{Poi}(W_n)\ge(2W_n\vee K)\}}].\end{aligned} \tag{10.1.69}$$

Again by Condition 7.5(b)-(c), for $p \in \{1,2\}$,

$$\mathbb{E}[W_n^p \mathbb{1}_{\{\mathrm{Poi}(W_n)\ge(2W_n\vee K)\}}] \to \mathbb{E}[W^p \mathbb{1}_{\{\mathrm{Poi}(W)\ge(2W\vee K)\}}], \tag{10.1.70}$$

which converges to zero when $K\to\infty$. Therefore, for each $\varepsilon>0$, we choose first K and then n sufficiently large, so that

$$\frac{1}{n}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\ge K+1, w_i<d_i/2\}}\Big] \le \varepsilon^2/2. \tag{10.1.71}$$

By the Markov inequality (Theorem 2.14),

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\ge K+1, w_i<d_i/2\}} \ge \varepsilon\Big) \le \frac{1}{n\varepsilon}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\ge K+1, w_i<d_i/2\}}\Big] \le \varepsilon/2. \tag{10.1.72}$$

Combining (10.1.64) and (10.1.72), we see that,

$$\mathbb{P}\Big(\sum_{k\ge K+1} P_k^{(n)} \le \varepsilon\Big) \ge 1-\varepsilon. \tag{10.1.73}$$

As a result, together with (10.1.59),

$$\sum_{k\ge 0} |P_k^{(n)} - p_k| \xrightarrow{\mathbb{P}} 1, \tag{10.1.74}$$

as required. This completes the proof of Proposition 10.10, and thus also that of Theorem 10.9. □

We next extend Theorem 10.9 to the case where $\nu=\infty$:

Theorem 10.11 (Phase transition in $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *Let $\boldsymbol{w}$ satisfy Condition 6.4(a)-(b). Then, the results in Theorem 10.1 also hold for $\mathrm{GRG}_n(\boldsymbol{w})$, $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$.*

Proof. We only prove that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$, the other statements can be proved in a similar fashion. Recall that $\mathcal{N}_m(i)$ denotes the set of vertices at graph distance m from $i\in[n]$, and let

$$|\mathcal{C}'_{\max}| = \#\{i\colon \mathcal{N}_m(i)\neq\varnothing\}. \tag{10.1.75}$$

Then, by Corollary 9.13, $\mathbb{E}[|\mathcal{C}'_{\max}|/n] \to \zeta_m = \mathbb{P}(\mathcal{Z}_m>0)$, and

$$\mathrm{Var}(|\mathcal{C}'_{\max}|/n) \to 0. \tag{10.1.76}$$

Therefore, $|\mathcal{C}'_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_m$. Since $\mathcal{C}_{\max}$ has a diameter that tends to infinity (recall Exercise 9.19), $|\mathcal{C}_{\max}| \le |\mathcal{C}'_{\max}|$, so that $|\mathcal{C}_{\max}|/n \le \zeta_m + o_{\mathbb{P}}(1)$. This proves the required upper bound.

For the lower bound, we bound $\mathrm{NR}_n(\boldsymbol{w})$ from below by a random graph with edge probabilities

$$p_{ij}(K) = 1 - \mathrm{e}^{-(w_i \wedge K)(w_j \wedge K)/\ell_n}. \tag{10.1.77}$$

Therefore, also $|\mathcal{C}_{\max}| \preceq |\mathcal{C}_{\max}(K)|$, where $\mathcal{C}_{\max}(K)$ is the largest connected component in the inhomogeneous random graph with edge probabilities $(p_{ij}(K))_{i,j\in[n]}$. Let

$$w_i(K) = (w_i \wedge K)\frac{1}{\ell_n}\sum_{j\in[n]}(w_j \wedge K), \tag{10.1.78}$$

so that the edge probabilities in (10.1.77) correspond to the Norros-Reittu model with weights $(w_i(K))_{i\in[n]}$. It is not hard to see that when Condition 6.4(a) holds for $(w_i)_{i\in[n]}$, then Condition 6.4(a)-(c) hold for $(w_i(K))_{i\in[n]}$, where the limiting random variable equals $(W \wedge K)$. Therefore, Theorem 10.9 applies, and we deduce that $|\mathcal{C}_{\max}(K)|/n \xrightarrow{\mathbb{P}} \zeta(K)$, which is the survival probability of the two-stage mixed-Poisson branching process with mixing variable $(W \wedge K)$. Since $\zeta(K) \to \zeta$ when $K \to \infty$, we conclude that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$. □

Exercise 10.4 (Number of vertices with degree k)**.** *Prove that also* $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1-\xi^k)$ *by adapting the above proof.*

10.2 Connectivity of $\mathrm{CM}_n(\boldsymbol{d})$

By Theorem 10.1, we see that as soon as $\mathbb{P}(D \geq 2) = 1$ (with an exception when $\mathbb{P}(D = 2) = 1$), $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 1$, as in this case the survival probability equals 1. In this section, we investigate conditions under which $\mathrm{CM}_n(\boldsymbol{d})$ is **whp***connected*, i.e., $\mathcal{C}_{\max} = [n]$ and $|\mathcal{C}_{\max}| = n$.

We first show that $\mathrm{CM}_n(\boldsymbol{d})$ is with positive probability *disconnected* when either $n_1 \gg n^{1/2}$, or when $\mathbb{P}(D \geq 2) > 1$. The main result in this section is Theorem 10.14, which shows that for *all* possible degree sequences with $n_1 = n_2 = 0$, $\mathrm{CM}_n(\boldsymbol{d})$ is **whp** connected. Remarkably, we do not even need Condition 7.5(a) for this result.

Proposition 10.12 (Disconnectivity of $\mathrm{CM}_n(\boldsymbol{d})$ when $n_1 \gg n^{1/2}$)**.** *Let Condition 7.5(a)-(b) hold, and assume that* $n_1 \gg n^{1/2}$*. Then,*

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ } \textit{connected}) = 0. \tag{10.2.1}$$

Proof. We note that $\mathrm{CM}_n(\boldsymbol{d})$ is disconnected when there are two vertices of degree 1 whose half-edges are paired to each other. When the half-edges of two of vertices of degree 1 are paired to each other, we say that a *2-pair* is created. Then,

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ contains no 2-pair}) = \prod_{i=1}^{n_1}\frac{\ell_n - n_1 - 2i + 1}{\ell_n - 2i + 1} = \prod_{i=1}^{n_1}\big(1 - \frac{n_1}{\ell_n - 2i + 1}\big). \tag{10.2.2}$$

Since, for each $i \geq 1$,

$$1 - \frac{n_1}{\ell_n - 2i + 1} \leq 1 - \frac{n_1}{\ell_n} \leq \mathrm{e}^{-n_1/\ell_n}, \tag{10.2.3}$$

we arrive at

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ contains no 2-pair}) \leq \mathrm{e}^{-n_1^2/\ell_n} = o(1), \tag{10.2.4}$$

since $\ell_n = \Theta(n)$ and $n_1 \gg n^{1/2}$. □

Proposition 10.13 (Disconnectivity of $\mathrm{CM}_n(\boldsymbol{d})$ when $p_2 > 0$)**.** *Let Condition 7.5(a)-(b) hold, and assume that $p_2 > 0$. Then,*

$$\limsup_{n\to\infty} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ connected}) < 1. \tag{10.2.5}$$

Proof. We perform a second moment method on the number of pairs $P(2)$ of vertices of degree 2 that are only connected to one another. If this happens, then these two vertices form a connected component of size 2, making the graph disconnected. The expected number of such pairs equals

$$\frac{2n_2(n_2-1)}{(\ell_n-1)(\ell_n-3)} \to 2p_2^2/\mathbb{E}[D]^2 \equiv \lambda_2, \tag{10.2.6}$$

by Condition 7.5(a)-(b), which implies that $n_2/n \to p_2$. We can use Theorem 2.6 to show that $P(2) \xrightarrow{d} \mathrm{Poi}(\lambda_2)$, so that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) \geq \mathbb{P}(P(2) > 0) \to 1 - \mathrm{e}^{-\lambda_2} > 0. \tag{10.2.7}$$

The proof that Theorem 2.6 can be applied is left as an exercise below. □

Exercise 10.5 (Factorial moments of $P(2)$)**.** *Let Condition 7.5(a)-(b) hold, and assume that $p_2 > 0$. Prove that, for every $k \geq 1$ and with $\lambda_2 = 2p_2^2/\mathbb{E}[D]^2$,*

$$\mathbb{E}[(P(2))_k] \to \lambda_2^k. \tag{10.2.8}$$

Conclude that $P(2) \xrightarrow{d} \mathrm{Poi}(\lambda_2)$.

We continue to investigate when the configuration model yields a connected graph:

Theorem 10.14 (Connectivity of $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that $d_i \geq 3$ for every $i \in [n]$. Then $\mathrm{CM}_n(\boldsymbol{d})$ is connected* **whp**. *More precisely, there exists a constant $C > 0$ such that for every $\boldsymbol{d}$,*

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) = O(1/n). \tag{10.2.9}$$

When Condition 7.5(a) holds with $p_1 = p_2 = 0$, then $\nu \geq 2 > 1$ is immediate, so we are always in the supercritical regime. Also, when $p_1 = p_2 = 0$, we have that $\zeta = 1$, since survival of the delayed branching process occurs with probability 1. Therefore, Theorem 10.1 implies that the largest connected component has size $n(1 + o(1))$ when Condition 7.5(a)-(b) hold. Theorem 10.14 extends this to the statement that $\mathrm{CM}_n(\boldsymbol{d})$ is with high probability *connected*. However, we do not assume that Condition 7.5 holds here.

We note that this is an important difference between the generalized random graph and the configuration model. Indeed, for the generalized random graph to be connected with high probability, the degrees must tend to infinity. This can be observed already for $\mathrm{ER}_n(p)$ in Theorem 5.5. For the configuration model, it is possible that the graph is connected while the average degree is bounded. Many real-world networks are connected, which makes the configuration model often more suitable than inhomogeneous random graphs.

Proof. We recall that a *configuration* denotes a pairing of all the half-edges. We note that the probability of a configuration equals $1/(\ell_n - 1)!!$. On the event that $\mathrm{CM}_n(\boldsymbol{d})$ is disconnected, there exists a set of indices $\mathcal{I} \subset [n]$ with $|\mathcal{I}| \leq \lfloor n/2 \rfloor$ such that all half-edges incident to vertices in $\mathcal{I}$ are *only* paired to other half-edges incident to other vertices in $\mathcal{I}$. For $\mathcal{I} \subseteq [n]$, we let

$$\ell_n(\mathcal{I}) = \sum_{i\in\mathcal{I}} d_i. \tag{10.2.10}$$

Clearly, in order for the half-edges incident to vertices in $\mathcal{I}$ to be paired only to other half-edges incident to vertices in $\mathcal{I}$, $\ell_n(\mathcal{I})$ needs to be even. The number of configurations for which this happens is bounded by

$$(\ell_n(\mathcal{I})-1)!!(\ell_n(\mathcal{I}^c)-1)!!. \tag{10.2.11}$$

As a result,

$$\begin{aligned}\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) &\leq \sum_{\mathcal{I}\subset[n]} \frac{(\ell_n(\mathcal{I})-1)!!(\ell_n(\mathcal{I}^c)-1)!!}{(\ell_n-1)!!} \\ &= \sum_{\mathcal{I}\subset[n]} \prod_{j=1}^{\ell_n(\mathcal{I})/2} \frac{\ell_n(\mathcal{I})-2j+1}{\ell_n-2j+1},\end{aligned} \tag{10.2.12}$$

where the sum over $\mathcal{I}\subset[n]$ is restricted to $\mathcal{I}$ for which $|\mathcal{I}|\leq\lfloor n/2\rfloor$.

Exercise 10.6 (Isolated vertex)**.** *Use the above bound to show that, when $d_i\geq 3$ for all $i\in[n]$,*

$$\mathbb{P}(\textit{there exists an isolated vertex}) \leq \frac{3n}{(2\ell_n-1)(2\ell_n-3)}. \tag{10.2.13}$$

Define

$$f(x)=\prod_{j=1}^{x}\frac{2x-2j+1}{\ell_n-2j+1}. \tag{10.2.14}$$

We can rewrite

$$f(x)=\frac{\prod_{j=1}^{x}(2x-2j+1)}{\prod_{j=1}^{x}(\ell_n-2j-1)}=\frac{\prod_{i=0}^{x-1}(2i+1)}{\prod_{k=0}^{x-1}(\ell_n-2k+1)}=\prod_{j=0}^{x-1}\frac{2i+1}{\ell_n-2i-1}, \tag{10.2.15}$$

where we write $i=x-j$ and $k=j-1$ in the second equality. Thus, for $x\leq\ell_n/4$, $x\mapsto f(x)$ is decreasing, since

$$\frac{f(x+1)}{f(x)}=\frac{2x+1}{\ell_n-2x-1}\leq 1. \tag{10.2.16}$$

Now, for every $\mathcal{I}$, since $d_i\geq 3$ for every $i\in[n]$ and since $\ell_n(\mathcal{I})$ is *even*,

$$\ell_n(\mathcal{I})\geq 2\lceil 3|\mathcal{I}|/2\rceil, \tag{10.2.17}$$

which only depends on the number of vertices in $\mathcal{I}$. Since there are precisely $\binom{n}{m}$ ways of choosing m vertices out of $[n]$, we conclude that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) \leq \sum_{\mathcal{I}\subset[n]} f(\lceil 3|\mathcal{I}|/2\rceil) = \sum_{m=1}^{\lfloor n/2\rfloor}\binom{n}{m} f(\lceil 3m/2\rceil), \tag{10.2.18}$$

with $m=|\mathcal{I}|$.

Exercise 10.7 (Isolated vertex)**.** *Use* (10.2.18) *to reprove Exercise 10.7. Hence, the above bounds are quite sharp.*

Exercise 10.8 (A cluster of size two)**.** *Use* (10.2.18) *to prove that*

$$\mathbb{P}(\textit{there exists a cluster of size 2}) \leq \frac{15n(n-1)}{(2\ell_n-1)(2\ell_n-3)(2\ell_n-5)}. \tag{10.2.19}$$

Note that, for m odd,

$$\frac{f(2\lceil 3(m+1)/2\rceil)}{f(2\lceil 3m/2\rceil)} = \frac{f((3m+1)/2+1)}{f((3m+1)/2)} = \frac{3m+3}{\ell_n - 3m - 2}. \tag{10.2.20}$$

while, for m even,

$$\frac{f(\lceil 3(m+1)/2\rceil)}{f(\lceil 3m/2\rceil)} = \frac{f(3m/2+2)}{f(3m/2)} = \frac{3m+5}{\ell_n - 3m - 5}\frac{3m+3}{\ell_n - 3m - 3}, \tag{10.2.21}$$

Define

$$h_n(m) = \binom{n}{m} f(\lceil 3m/2\rceil), \tag{10.2.22}$$

so that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) \leq \sum_{m=1}^{\lfloor n/2\rfloor} h_n(m). \tag{10.2.23}$$

Then,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{n-m}{m+1}\frac{f(\lceil 3(m+1)/2\rceil)}{f(\lceil 3m/2\rceil)}, \tag{10.2.24}$$

so that, for m odd and using $\ell_n \geq 3n$,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{3(n-m)}{\ell_n - 3m - 3} \leq \frac{n-m}{n-m-1}, \tag{10.2.25}$$

while, for m even and using $\ell_n \geq 3n$,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{3(n-m)}{\ell_n - 3m - 5}\frac{3m+5}{\ell_n - 3m - 3} \leq \frac{n-m}{n-m-1}\frac{m+2}{n-m-2}. \tag{10.2.26}$$

Thus, we obtain that, for $m \leq n/2$, there exists a $c > 0$ such that

$$\frac{h_n(m+1)}{h_n(m)} \leq 1 + \frac{c}{n}. \tag{10.2.27}$$

We conclude that, for $m \leq n/2$,

$$\begin{aligned} h_n(m) = h_n(3)\prod_{j=3}^{m}\frac{h_n(j+1)}{h_n(j)} &\leq h_n(3)\prod_{j=3}^{\lfloor n/2\rfloor}(1+c/n) \\ &\leq h_n(3)(1+c/n)^{\lfloor n/2\rfloor} \leq h_n(3)\mathrm{e}^{c/2}, \end{aligned} \tag{10.2.28}$$

so that

$$\begin{aligned} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) &\leq \sum_{m=1}^{\varepsilon n} h_n(m) \leq h_n(1) + h_n(2) + \sum_{m=3}^{\lfloor n/2\rfloor} h_n(m) \\ &\leq h_n(1) + h_n(2) + nh_n(3)\mathrm{e}^{c/2}/2. \end{aligned} \tag{10.2.29}$$

By Exercises 10.6 and 10.8, $h_n(1), h_n(2) = O(1/n)$, so we are left to compute $h_n(3)$. For this, we note that $\lceil 3m/2\rceil = 5$ when $m = 3$, so that

$$h_n(3) = \binom{n}{3} f(5) = \frac{9!!n(n-1)(n-2)}{6(\ell_n - 1)(\ell_n - 3)(\ell_n - 5)(\ell_n - 7)(\ell_n - 9)} = O(1/n^2). \tag{10.2.30}$$

As a result,

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) = O(1/n), \tag{10.2.31}$$

as required. □

10.3 The small-world phenomenon in $\mathrm{CM}_n(\boldsymbol{d})$

In this section, we study distances in the configuration model, both in the case of finite-variance degrees as well as in the case of infinite variance degrees.

Finite-variance degrees. In this section, we give an overview of the results on the typical graph distance in the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ when the (asymptotic) degree sequence has a finite variance. We start by analyzing the typical graph distance.

Theorem 10.15 (Typical distances in $\mathrm{CM}_n(\boldsymbol{d})$ for finite-variance weights)**.** *In the configuration model $\mathrm{CM}_n(\boldsymbol{d})$, where the degrees $\boldsymbol{d}=(d_i)_{i\in[n]}$ satisfy Condition 6.4(a)-(c) and where $\nu>1$, conditionally on $H_n<\infty$,*

$$H_n/\log n \xrightarrow{\mathbb{P}} 1/\log\nu. \tag{10.3.1}$$

Theorem 10.15 shows that the average distances in $\mathrm{CM}_n(\boldsymbol{d})$ are of order $\log_\nu n$. It is similar in spirit as Theorem 9.3. We shall see that also its proof is quite similar.

We continue the discussion of distances in the configuration model by investigating the diameter in the model. Before stating the main result, we introduce some notation. Recall that $G^*_D(x)$ is defined in (10.1.9) as the probability generating function of $g=(g_k)_{k\geq 0}$ defined in (10.1.2). We recall that ξ is the extinction probability of the branching process with offspring distribution g defined in (10.1.6) and further define

$$\mu = G^*_D(\xi) = \sum_{k=1}^{\infty} k\xi^{k-1}g_k. \tag{10.3.2}$$

When $\xi<1$, we also have that $\mu\leq 1$. Then, the main result is as follows:

Theorem 10.16 (Diameter of the configuration model)**.** *Let Condition 7.5(a)-(b) hold. Assume that $n_1=0$ when $p_1=0$, and that $n_2=0$ when $p_2=0$. Then,*

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log\nu} + \frac{(2-\mathbb{1}_{\{p_1=0\}}-\mathbb{1}_{\{p_2=0\}})}{|\log\mu|}. \tag{10.3.3}$$

We note that, by Theorem 10.15 and Theorem 10.16, the diameter of the configuration model is strictly larger than the average distances, except when $p_1=p_2=0$. In the latter case, the degrees are at least three, so that thin lines are not possible, and the configuration model is with high probability *connected* (recall Theorem 10.14). We also remark that Theorem 10.16 applies not only to the finite variance case, but also to the finite mean and infinite variance case. In the latter case, the diameter is of order $\log n$ unless when $p_1=p_2=0$, in which case Theorem 10.16 implies that the diameter is $o(\log n)$. By Corollary 7.12, Theorem 10.16 also applies to uniform random graphs with a given degree sequence. This shall be used in the examples below:

Random regular graphs. Let r be the degree of the random regular graph, where $r\geq 2$. By Corollary 7.12, the diameter of a random regular r-graph has with high probability the same asymptotics as the diameter of $\mathrm{CM}_n(\boldsymbol{d})$, where $d_i=r$ with probability 1. Thus, $p_r=1$ and $p_i=0$ for any $i\neq r$. We assume that nr is even, so that the degree sequence is feasible. It is not hard to see that all assumptions of Theorem 10.16 are satisfied. Moreover, we have that $\nu=r-1$. When $r\geq 3$, we thus obtain that

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log(r-1)}. \tag{10.3.4}$$

When $r = 2$, on the other hand, the graph is *critical*, so that there is no giant component. Since $\nu = 1$, we have that $\mu = \nu = 1$, so that $\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d})) \gg \log n$. This is quite reasonable, since the graph will consist of a collection of *cycles*. The diameter of such a graph is equal to half the longest cycle.

Exercise 10.9 (Diameter of soup of cycles)**.** *Prove that in a graph consisting solely of cycles, the diameter is equal to the longest cycle divided by 2.*

Exercise 10.10 (Longest cycle 2-regular graph)**.** *What is the size of the longest cycle of the 2-regular graph?*

Erdős-Rényi random graph. We next study the diameter of $\mathrm{ER}_n(\lambda/n)$. We let $\lambda > 1$. By Proposition 10.10, Condition 7.5(a)-(b) holds with $p_k = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$. Also, $\mu = \mu_\lambda$, the dual parameter in (3.6.7).

Exercise 10.11 (Parameters for $\mathrm{ER}_n(\lambda/n)$)**.** *Prove that $\nu = \lambda$ and $\mu = \mu_\lambda$.*

We again make essential use of Theorem 7.13, which related the configuration model and the generalized random graph. We note that $\mathrm{ER}_n(\lambda/n)$ is the same as $\mathrm{GRG}_n(\boldsymbol{w})$, where (recall Exercise 6.1)

$$w_i = \frac{n\lambda}{n-\lambda}. \tag{10.3.5}$$

From the above identifications and using Theorem 7.13, we find that

$$\frac{\mathrm{diam}(\mathrm{ER}_n(\lambda/n))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log\lambda} + \frac{2}{|\log\mu_\lambda|}. \tag{10.3.6}$$

This identifies the diameter of the Erdős-Rényi random graph.

Finite mean, infinite variance degrees. In this section, we study the typical distance of the configuration model with degrees having finite mean and infinite variance. We start by formulating the precise condition on the degrees that we shall work with. This condition is identical to the condition on F_n for $\mathrm{NR}_n(\boldsymbol{w})$ formulated in (9.1.14). Recall that $F_n(x)$ denotes the proportion of vertices having degree at most x. Then, we assume that there exists a $\tau \in (2,3)$ such that for all $\delta > 0$, there exists $c_1 = c_1(\delta)$ and $c_2 = c_2(\delta)$ such that, uniformly in n,

$$c_1 x^{-(\tau-1+\delta)} \le [1-F_n](x) \le c_2 x^{-(\tau-1-\delta)}, \tag{10.3.7}$$

where the upper bound is expected to hold for every $x \ge 1$, while the lower bound is only required to hold for $1 \le x \le n^\alpha$ for some $\alpha > 1/2$. The typical distance of $\mathrm{CM}_n(\boldsymbol{d})$ is identified in the following theorem:

Theorem 10.17 (Typical distances in $\mathrm{CM}_n(\boldsymbol{d})$ for $\tau \in (2,3)$)**.** *Let the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ in the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ satisfy Condition 7.5(a)-(b) and* (10.3.7)*. Then, conditionally on $H_n < \infty$,*

$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|}. \tag{10.3.8}$$

Theorem 10.17 is similar in spirit to Theorem 9.4 for $\mathrm{NR}_n(\boldsymbol{w})$. We will give a different proof for Theorem 10.17.

10.4 Proofs of small-world results $\mathrm{CM}_n(\boldsymbol{d})$

10.4.1 Path counting techniques

A path π of length k in $\mathrm{CM}_n(\boldsymbol{d})$ means a sequence

$$\pi = \{(\pi_0, t_0), (\pi_1, s_1, t_1), \ldots, (\pi_{k-1}, s_{k-1}, t_{k-1}), (\pi_k, s_k)\}, \tag{10.4.1}$$

where $\pi_i \in [n]$ denotes the ith vertex along the path, and $s_i \in [d_i]$ denotes the half-edge that is paired to the half-edge $t_{i+1} \in [d_{i+1}]$. In particular, multiple edges between π_i and π_{i+1} give rise to *distinct* paths through the same vertices. We assume throughout that the path is *simple*, i.e., $\pi_0, \ldots, \pi_k$ are distinct vertices. In this section, we perform first and second moment computations on the number of paths present in $\mathrm{CM}_n(\boldsymbol{d})$. We start by proving upper bounds on the expected number of paths.

Upper bounds on the expected number of paths in $\mathrm{CM}_n(\boldsymbol{d})$. For $a, b \in [n]$, $\mathcal{I} \subseteq [n]$ and $k \geq 1$, we let $\mathcal{P}_k(a,b) = \mathcal{P}_k(a,b;\mathcal{I})$ denote the set of k-paths that only use vertices in $\mathcal{I}$, and we let

$$N_k(a,b) = N_k(a,b;\mathcal{I}) = \#\{\pi \in \mathcal{P}_k(a,b) \colon \pi \subset \mathrm{CM}_n(\boldsymbol{d})\} \tag{10.4.2}$$

denote the number of paths of length k between the vertices a and b. Then, we prove the following upper bound on the expected number of paths connecting a and b:

Proposition 10.18 (Expected numbers of paths)**.** *For any $k \geq 1$, $a, b \in [n]$ and $(d_i)_{i\in[n]}$,*

$$\mathbb{E}[N_k(a,b)] \leq \frac{d_a d_b \ell_n}{(\ell_n - 2k + 1)(\ell_n - 2k)} \nu_{\mathcal{I}}^{k-1}, \tag{10.4.3}$$

where

$$\nu_{\mathcal{I}} = \sum_{i \in \mathcal{I} \setminus \{a,b\}} \frac{d_i(d_i - 1)}{\ell_n}. \tag{10.4.4}$$

Proof. The probability that the path π in (10.4.1) is present in $\mathrm{CM}_n(\boldsymbol{d})$ is equal to

$$\prod_{i=1}^{k} \frac{1}{\ell_n - 2i + 1}, \tag{10.4.5}$$

and the number of path with fixed $\pi_0, \ldots, \pi_k$ is equal to

$$d_{\pi_0} \Big(\prod_{i=1}^{k-1} d_{\pi_i}(d_{\pi_i} - 1) \Big) d_{\pi_k}. \tag{10.4.6}$$

Substituting $\pi_0 = a, \pi_k = b$, we arrive at

$$\mathbb{E}[N_k(a,b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1, \ldots, \pi_{k-1}}^{*} \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}, \tag{10.4.7}$$

where the sum is over distinct elements of $\mathcal{I} \setminus \{a,b\}$. Let R denote the subset of vertices of $\mathcal{I} \setminus \{a,b\}$ for which $d_i \geq 2$. Then,

$$\mathbb{E}[N_k(a,b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1, \ldots, \pi_{k-1} \in R}^{*} \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}, \tag{10.4.8}$$

By an inequality of Maclaurin [**?**, Theorem 52], for $r = |R|$, $2 \le k \le r+1$ and any $(a_i)_{i\in R}$ with $a_i \ge 0$,

$$\frac{(r-k+1)!}{r!} \sum_{\pi_1,\dots,\pi_{k-1}\in R}^{*} \prod_{i=1}^{k-1} a_i \le \Big(\frac{1}{r}\sum_{i\in R} a_i\Big)^{k-1}. \tag{10.4.9}$$

Let $a_i = d_i(d_i-1)$, so that

$$\sum_{i\in R} a_i = \ell_n \nu_{\mathcal{I}}. \tag{10.4.10}$$

We arrive at

$$\begin{aligned}\mathbb{E}[N_k(a,b)] &= \frac{d_a d_b}{\ell_n - 2k + 1}(\ell_n\nu_{\mathcal{I}}/r)^{k-1}\prod_{i=1}^{k-1}\frac{(r-i+1)}{(\ell_n - 2i + 1)} \\ &\le \frac{d_a d_b}{\ell_n - 2k + 1}\frac{\ell_n}{\ell_n - 2k}(\nu_{\mathcal{I}})^{k-1}\prod_{i=0}^{k-2}\frac{(1-\frac{i}{r})}{(1-\frac{2i}{\ell_n})}.\end{aligned} \tag{10.4.11}$$

Further, $\ell_n = \sum_{i\in[n]} d_i \ge 2r$, so that $1 - \frac{i}{r} \le 1 - \frac{2i}{\ell_n}$. Substitution yields the required bound. □

Logarithmic lower bound typical distances $\mathrm{CM}_n(\boldsymbol{d})$. With Proposition 10.18 at hand, we can immediately prove the lower bound on the typical graph distance in the case where the degrees have finite second moment (as in Theorem 9.5):

Theorem 10.19 (Logarithmic lower bound typical distances $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that*

$$\limsup_{n\to\infty} \nu_n > 1, \tag{10.4.12}$$

where

$$\nu_n = \mathbb{E}[D_n(D_n-1)]/\mathbb{E}[D_n]. \tag{10.4.13}$$

Then, for any $\varepsilon > 0$,

$$\mathbb{P}(H_n \le (1-\varepsilon)\log_{\nu_n} n) = o(1). \tag{10.4.14}$$

We leave the proof of Theorem 10.19, which is closely related to that of Theorem 9.5, as an exercise:

Exercise 10.12 (Proof Theorem 10.19)**.** *Let V_1, V_2 be two independent vertices chosen uniformly at random from $[n]$. Use Proposition 10.18 with $a = V_1, b = V_2, \mathcal{I} = [n]$ to prove Theorem 10.19.*

Truncated first moment method and loglog lower bound typical distances $\mathrm{CM}_n(\boldsymbol{d})$ for $\tau \in (2,3)$.

Theorem 10.20 (Loglog lower bound on average distances in $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Suppose that the weights $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfy Condition **??**(a) and that there exists a $\tau \in (2,3)$ and c_2 such that, for all $x \ge 1$,*

$$[1 - F_n](x) \le c_2 x^{-(\tau-1)}, \tag{10.4.15}$$

Then, for every $\varepsilon > 0$,

$$\mathbb{P}\Big(H_n \le (1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|}\Big) = o(1). \tag{10.4.16}$$

Second moment method for the number of paths in $\mathrm{CM}_n(\boldsymbol{d})$. Define

$$\bar{n}_k(a,b) = u_a u_b \Big(\sum_{i \in \mathcal{I} \setminus \{a,b\}} d_i(d_1 - 1) \Big)^{k-1}, \qquad \underline{n}_k(a,b) = u_a u_b \Big(\sum_{i \in \mathcal{I}_{a,b,k}} d_i(d_i - 1) \Big)^{k-1}, \tag{10.4.17}$$

where $\mathcal{I}_{a,b,k}$ is the subset of $\mathcal{I}$ in which a and b, as well as the $k+2$ indexes with highest degrees have been removed. Let

$$\nu_{\mathcal{I}} = \frac{1}{\ell_n} \sum_{i \in \mathcal{I}} d_i(d_i - 1), \qquad \gamma_{\mathcal{I}} = \frac{1}{\ell_n^{3/2}} \sum_{i \in \mathcal{I}} d_i(d_i - 1)(d_i - 2). \tag{10.4.18}$$

Proposition 10.21 (Variance of number of paths)**.** *For any $k \geq 1$, $a, b \in \mathcal{I}$ and $(u_i)_{i \in \mathcal{I}}$,*

$$\mathbb{E}[N_k(a,b)] \geq \underline{n}_k(a,b), \tag{10.4.19}$$

while, assuming that $\nu_{\mathcal{I}} > 1$,

$$\mathrm{Var}(N_k(a,b)) \leq n_k(a,b) + \bar{n}_k(a,b)^2 \Big(\frac{\gamma_{\mathcal{I}} \nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}} - 1} \Big(\frac{1}{d_a} + \frac{1}{d_b} \Big) + \frac{\gamma_{\mathcal{I}}^2 \nu_{\mathcal{I}}}{d_a d_b (\nu_{\mathcal{I}} - 1)^2} + e_k \Big), \tag{10.4.20}$$

where

$$e_k = k \Big(1 + \frac{\gamma_{\mathcal{I}}}{d_a \nu_{\mathcal{I}}} \Big) \Big(1 + \frac{\gamma_{\mathcal{I}}}{d_b \nu_{\mathcal{I}}} \Big) \frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}} - 1} \big(\mathrm{e}^{2k^3 \gamma_{\mathcal{I}}^2 / \nu_{\mathcal{I}}^3} - 1 \big). \tag{10.4.21}$$

d proof Prop. 10.21!

Proof. □

ve lower bound H_n for 3 and for $\tau \in (2,3)$ here.

10.4.2 Branching process approximation

In this section, we give a convenient description of the breadth-first exploration in $\mathrm{CM}_n(\boldsymbol{d})$, and relate this to a branching process. We start by describing the algorithm. Suppose we start in a vertex i having degree d_i.

Proposition 10.22 (Weak convergence of neighborhoods of two vertices)**.** *Let the degrees $\boldsymbol{d} = (d_i)_{i \in [n]}$ satisfy Condition **??**(a)-(b). Then, for each $m \geq 1$,*

$$\mathbb{P}(H_n \leq 2m) = o(1), \tag{10.4.22}$$

and

$$(Z_m^{(1)}, Z_m^{(2)}) \xrightarrow{d} (\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)}), \tag{10.4.23}$$

where $(\mathcal{Z}_l^{(1)}, \mathcal{Z}_l^{(2)})_{l \geq 0}$ are two independent branching processes with offspring distribution D in the first generation, and offspring distribution $D^* - 1$ in all further generations.

scribe degrees as size-sed reordering od degree uence!

ve upper bound H_n for 3.

10.4.3 A $\log\log$ upper bound on the diameter core $\tau \in (2,3)$

We take $\sigma > 1/(3-\tau)$ and define the *core* Core_n of the configuration model to be

$$\mathrm{Core}_n = \{i \colon d_i \geq (\log n)^{\sigma}\}, \tag{10.4.24}$$

i.e., the set of vertices with degree at least $(\log n)^{\sigma}$. Then, the diameter of the core is bounded in the following theorem:

Theorem 10.23 (Diameter of the core)**.** *Fix $\tau \in (2,3)$ and assume that (10.3.7) holds. For every $\sigma > \frac{1}{3-\tau}$, the diameter of Core_n is with high probability bounded above by*

$$\frac{2\log\log n}{|\log(\tau-2)|}(1+o(1)). \tag{10.4.25}$$

Proof. Without loss of generality, we assume that the degrees are ordered, i.e., $d_1 \geq d_2 \geq \cdots \geq d_n$. We note that (10.3.7) implies that, for some $\alpha \in (1/2, 1/(\tau-1))$,

$$d_1 \geq u_1, \quad \text{where} \quad u_1 = n^{\alpha}. \tag{10.4.26}$$

Define

$$\Gamma_1 = \{i\colon d_i \geq u_1\}, \tag{10.4.27}$$

so that $\Gamma_1 \neq \varnothing$. For some constant $C > 0$, which will be specified later, and for $k \geq 2$, we define recursively

$$u_k = C\log n\big(u_{k-1}\big)^{\tau-2}. \tag{10.4.28}$$

Then, we define

$$\Gamma_k = \{i\colon d_i \geq u_k\}. \tag{10.4.29}$$

We identify u_k in the following lemma:

Lemma 10.24 (Identification $(u_k)_{k\geq 1}$)**.** *For every $k \geq 1$,*

$$u_k = C^{a_k}(\log n)^{b_k} n^{c_k}, \tag{10.4.30}$$

where

$$c_k = \alpha(\tau-2)^{k-1}, \quad a_k = b_k = \frac{1}{3-\tau}[1-(\tau-2)^{k-1}]. \tag{10.4.31}$$

Proof. We note that c_k, b_k, a_k satisfy the recursions, for $k \geq 2$,

$$c_k = (\tau-2)c_{k-1}, \quad b_k = 1+(\tau-2)b_{k-1}, \quad a_k = 1+(\tau-2)a_{k-1}, \tag{10.4.32}$$

with initial conditions $c_1 = \alpha, a_1 = b_1 = 0$. Solving yields our claim. □

The key step in the proof of Theorem 10.23 is the following lemma:

Lemma 10.25 (Connectivity between Γ_{k-1} and Γ_k)**.** *Fix $\tau \in (2,3)$ and assume that (10.3.7) holds. Fix $k \geq 2$, and take $C > 4\mathbb{E}[D]/c$. Then, the probability that there exists an $i \in \Gamma_k$ that is not directly connected to Γ_{k-1} is $o(n^{-\delta})$, for some $\delta > 0$ independent of k.*

Proof. We note that, by definition,

$$\sum_{i\in\Gamma_{k-1}} d_i \geq u_{k-1}|\Gamma_{k-1}| = u_{k-1}n[1-F_n](u_{k-1}). \tag{10.4.33}$$

By (10.3.7), and since $k \mapsto u_k$ is decreasing with $u_1 = n^{\alpha}$,

$$[1-F_n](u_{k-1}) \geq c(u_{k-1})^{1-\tau}. \tag{10.4.34}$$

As a result, we obtain that for every k, and with high probability

$$\sum_{i\in\Gamma_{k-1}} d_i \geq cn(u_{k-1})^{2-\tau}. \tag{10.4.35}$$

We note that for any two sets of vertices A, B

$$\mathbb{P}(A \text{ not directly connected to } B) \leq \mathrm{e}^{-\frac{d_A d_B}{\ell_n}}, \tag{10.4.36}$$

where, for any $A \subseteq [n]$, we write

$$d_A = \sum_{i \in A} d_i. \tag{10.4.37}$$

By (10.4.36), and Boole's inequality, the probability that there exists an $i \in \Gamma_k$ such that i is not directly connected to Γ_{k-1} is bounded by

$$n\mathrm{e}^{-\frac{u_k n u_{k-1}[1-F(u_{k-1})]}{\ell_n}} \leq n\mathrm{e}^{-\frac{cu_k(u_{k-1})^{2-\tau}}{\mathbb{E}[D_n]}} = n^{1-\frac{cC}{\mathbb{E}[D_n]}}. \tag{10.4.38}$$

By Condition 7.5(a)-(b), $\mathbb{E}[D_n] \to \mathbb{E}[D]$, so that, as $n \to \infty$ and taking $C > \mathbb{E}[D]/c$, we obtain the claim. □

We now complete the proof of Theorem 10.23. Fix

$$k^* = \frac{\log\log n}{|\log(\tau-2)|}. \tag{10.4.39}$$

As a result of Lemma 10.25, with high probability that the diameter of Γ_{k^*} is at most $2k^*+1$, because the distance between any vertex in Γ_{k^*} and Γ_1 is at most k^*. Since $u_1 = n^\alpha \gg \sqrt{n}\log n$, again with high probability, Γ_1 forms a complete graph. Therefore, it suffices to prove that

$$\mathrm{Core}_n \subseteq \Gamma_{k^*}. \tag{10.4.40}$$

For this, we note that

$$\Gamma_{k^*} = \{i\colon d_i \geq u_{k^*}\}, \tag{10.4.41}$$

so that, in turn, it suffices to prove that $u_{k^*} \geq (\log n)^\sigma$, for any $\sigma > \frac{1}{}(3-\tau)$. According to Lemma 10.24,

$$u_{k^*} = C^{a_{k^*}}(\log n)^{b_{k^*}} n^{c_{k^*}}. \tag{10.4.42}$$

We note that $n^{c_{k^*}} = \mathrm{e}^{\log n(\tau-2)^{k^*}}$. Since, for $2 < \tau < 3$,

$$x(\tau-2)^{\frac{\log x}{|\log(\tau-2)|}} = x \cdot x^{-1} = 1, \tag{10.4.43}$$

we find with $x = \log n$ that

$$\log n \cdot (\tau-2)^{\frac{\log\log n}{|\log(\tau-2)|}} = 1, \tag{10.4.44}$$

which implies that $n^{c_{k^*}} = 1$. Further, $b_k \to 1/(\tau-3)$ as $k \to \infty$, so that $(\log n)^{b_{k^*}} = (\log n)^{1/(3-\tau)+o(1)}$, and $a_k = b_k$, so that $C^{a_{k^*}} = C^{1/(\tau-3)+o(1)}$. Thus,

$$u_{k^*} = (\log n)^{1/(3-\tau)+o(1)}, \tag{10.4.45}$$

so that, by picking n sufficiently large, we can make $1/(3-\tau)+o(1) \leq \sigma$. This completes the proof of Theorem 10.23. □

We continue to use Theorem 10.23 to prove a $\log\log n$ upper bound on H_n in the case where $\tau \in (2,3)$. We start by describing the setting. We assume that there exist $\tau \in (2,3)$, $\alpha > 1/2$ and c_1 such that, uniformly in n and $x \leq n^\alpha$,

$$[1-F_n](x) \geq c_1 x^{-(\tau-1)}. \tag{10.4.46}$$

Theorem 10.26 (A log log upper bound on typical distance for $\tau \in (2,3)$)**.** *Suppose that the empirical distribution function F_n of the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies Condition* **??***(a)-(b) and* (10.4.46)*. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(H_n \le \frac{2(1+\varepsilon)\log\log n}{|\log(\tau-2)|} \mid H_n < \infty\Big) = 1. \tag{10.4.47}$$

Proof. In the proof, we make use of the branching process approximation in Section 10.4.2. We let V_1, V_2 denote two vertices chosen uniformly at random from $[n]$, and we recall that $Z_m^{(i)}$ denote the number of unpaired or free half edges incident to vertices in $\mathcal{N}_m(V_i)$. By Proposition 10.22, $(Z_m^{(1)}, Z_m^{(2)}) \xrightarrow{d} (\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)})$, which are two independent two-stage branching processes where the root has offspring distribution D, and individuals in all further generations have offspring distribution $D^* - 1$.

We condition on $\mathcal{N}_{\le m}(V_1), \mathcal{N}_{\le m}(V_2)$ which are such that $Z_m^{(1)} \ge 1, Z_m^{(2)} \ge 1$. Further, conditionally on $Z_m^{(1)} \ge 1, Z_m^{(2)} \ge 1$, in fact, for each $A \ge 1$ large, there exists an $m = m(A) \ge 1$ such that, with probability at least $1-\varepsilon$, $Z_m^{(1)} \ge A, Z_m^{(2)} \ge A$ occurs.

Then, conditionally on $\mathcal{N}_{\le m}(V_1), \mathcal{N}_{\le m}(V_2)$, and collapsing $\mathcal{N}_m(V_1)$ to a single vertex a_1 and $\mathcal{N}_m(V_2)$ to a single vertex a_2, the distribution of the resulting random graph is again a configuration model, with degrees $d_{a_1} = Z_m^{(1)}, d_{a_2} = Z_m^{(2)}$ and vertex set $R = [n] \cup \{a_1, a_2\} \setminus (\mathcal{N}_{\le m}(V_1) \cup \mathcal{N}_{\le m}(V_2))$. We apply Proposition 10.21 with $k = \varepsilon \log\log n$, a_1 and $b = \mathrm{Core}_n$ and with $\mathcal{I} = \{i \in R\colon d_i \le K\}$. Then, Proposition 10.21 gives that, conditionally on $\mathcal{N}_{\le m}(V_1), \mathcal{N}_{\le m}(V_2)$,

$$\mathbb{P}(N_k(a_1,b) = 0) \le \mathrm{Var}(N_k(a_1,b))/\mathbb{E}[N_k(a_1,b)]^2 \le \dots \tag{10.4.48}$$

Theorem 10.26

Prove upper bound H_n for $\tau \in (2,3)$ here.

□

10.5 Diameter of $\mathrm{CM}_n(\boldsymbol{d})$ for $\tau \in (2,3)$

In this section, we use Theorem 10.23 to study the diameter of $\mathrm{CM}_n(\boldsymbol{d})$ when $\tau \in (2,3)$.

We continue by studying the diameter of $\mathrm{CM}_n(\boldsymbol{d})$ in the case where $\tau \in (2,3)$. Note that when $p_1 + p_2 > 0$, then the diameter is equal to a positive constant times $\log n$ by Theorem 10.16. Therefore, we turn to the case where $p_1 = p_2 = 0$. For this result, we assume that there exists a $\tau \in (2,3)$ such that, for some $c > 0$ and all $x \ge 1$,

$$1 - F(x) \ge c x^{1-\tau}. \tag{10.5.1}$$

Observe that (10.5.1) is strictly weaker than (10.8.15). The main result is as follows:

Theorem 10.27 (Diameter of $\mathrm{CM}_n(\boldsymbol{d})$ for $\tau \in (2,3)$)**.** *Suppose that the empirical distribution function F_n of the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies Condition* **??***(a)-(b) and that* (10.3.7) *holds. Assume further that $d_{\min} = \min_{i\in[n]} d_i \ge 3$ and $p_{d_{\min}} = \mathbb{P}(D = d_{\min}) > 0$. Then,*

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|} + \frac{2}{\log(d_{\min}-1)}. \tag{10.5.2}$$

We prove Theorem 10.27 by proving an upper and a lower bound on the diameter. We start with the lower bound, which is the easiest part:

Proof of the lower bound on the diameter. We call a vertex v *minimally-k-connected* when all $i \in \mathcal{N}_{k-1}(v)$ satisfy $d_i = d_{\min}$. Let M_k denote the number of minimally-k-connected vertices. Then, we compute the first moments of M_k in the following lemma:

Lemma 10.28 (Moments of the number of minimally-k-connected vertices)**.** *For all $k \geq 2$,*

$$\mathbb{E}[M_k] = n_{d_{\min}} \prod_{i=1}^{d_{\min}(d_{\min}-1)^{k-2}} \frac{d_{\min}(n_{d_{\min}} - (i-1))}{\ell_n - 2i + 1}, \tag{10.5.3}$$

and

$$\mathbb{E}[M_k^2] \leq \mathbb{E}[M_k]^2 + \mathbb{E}[M_k]\Big[d_{\min}(d_{\min}-1)^{k-1} + \frac{\ell_n}{\ell_n - n_{d_{\min}}}\Big]. \tag{10.5.4}$$

d proof!!

Proof. □

The complete the proof of the lower bound on the diameter, we fix $\varepsilon > 0$ sufficiently small, and take $k = (1-\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)}$. We note that $n_{d_{\min}}/n \to p_2 > 0$ by Condition **??**(a), so that, $d_{\min} n_{d_{\min}}/\ell_n \to \lambda_{d_{\min}}$, where we define $\lambda_{d_{\min}} = d_{\min} p_{d_{\min}}/\mathbb{E}[D]$ by Condition **??**(a)-(b). By Lemma 10.28,

$$\mathbb{E}[M_k] \geq n(p_{d_{\min}} - \delta)(\lambda_{d_{\min}} - \delta)^{d_{\min}(d_{\min}-1)^{k-2}}. \tag{10.5.5}$$

When $k = (1-\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)}$,

$$d_{\min}(d_{\min}-1)^{k-2} \leq (\log n)^{1-\varepsilon}. \tag{10.5.6}$$

As a result, $\mathbb{E}[M_k] \geq n(p_{d_{\min}} - \delta)(\lambda_{d_{\min}} - \delta)^{(\log n)^{1-\varepsilon}}$. Further,

$$\mathrm{Var}(M_k) = o(\mathbb{E}[M_k]^2), \tag{10.5.7}$$

so that

$$M_k/\mathbb{E}[M_k] \xrightarrow{\mathbb{P}} 1. \tag{10.5.8}$$

We conclude that, **whp**, $M_k \geq n^{1-o(1)}$. Since each minimally-k-connected vertex uses up at most

$$1 + \sum_{l=1}^{k-1} d_{\min}(d_{\min}-1)^{l-1} = n^{o(1)} \tag{10.5.9}$$

vertices of degree $d_{\min}$, **whp** there must be at least two minimally-k-connected vertices whose $k-1$ neighborhoods are disjoint. We fix these two vertices and denote them a and b. We note that these vertices have precisely $d_{\min}(d_{\min}-1)^{k-1}$ unpaired half-edges incident to vertices in $\mathcal{N}_{k-1}(a)$ and $\mathcal{N}_{k-1}(b)$. By the truncated first moment method on paths, performed in the proof of Theorem 10.20, it follows that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(\mathcal{N}_{k-1}(a), \mathcal{N}_{k-1}(b)) \leq (1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|}) = o(1). \tag{10.5.10}$$

Therefore, **whp**,

$$\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d})) \geq (1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|} + 2k = (1-\varepsilon)\Big[\frac{2\log\log n}{|\log(\tau-2)|} + \frac{2\log\log n}{\log(d_{\min}-1)}\Big], \tag{10.5.11}$$

as required. □

Proof of the upper bound on the diameter. For the upper bound, it is convenient to explore the neighborhood of a vertex v by only pairing up the first $d_{\min}$ half-edges incident to v and the $d_{\min}-1$ half-edges incident to any other vertex appearing in the tree. We call this exploration graph the k-exploration tree. For k of order $\log\log n$, the probability that more than *two* collisions occur in the k-exploration tree is small, as we prove now:

Lemma 10.29 (At most one collision)**.** *Let $k \le (1+\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)}$. Then, under the conditions in Theorem 10.27, the probability that there exists a vertex whose k-exploration tree has at least two collisions before hitting the core Core_n, is bounded by $(d_{\min})^4(\log n)^d \ell_n^{-2}$, where $d = (1/(\tau-1)+\varepsilon)\log(d_{\min}-1) + 2\sigma$.*

Proof. In this proof, we abbreviate $m = d_{\min}$. For any half-edge in the k-exploration tree, the probability that it create a collision before hitting the core is bounded above by $(m+1)m^{k-1}(\log n)^\sigma \ell_n^{-1}$. The probability that two half-edges will both create a collision is, by similar arguments, bounded above by $\big[(m+1)m^{k-1}(\log n)^\sigma \ell_n^{-1}\big]^2$. The total number of possible pairs of half-edges in the k-exploration tree is bounded by

$$[(m+1)(1+m+\ldots+m^{k-1})]^2 \le [(m+1)m^k]^2,$$

so that, by Boole's inequality, the probability that the k-exploration tree has at least two collisions is bounded by

$$\big[(m+1)m^k\big]^4(\log n)^{2\sigma}\ell_n^{-2}. \tag{10.5.12}$$

When $k \le (1+\varepsilon)\frac{\log\log n}{\log m}$, this is bounded by $\big[(m+1)m^k\big]^4(\log n)^{2\sigma} \le (m+1)^4(\log n)^d$, where $d = (1+\varepsilon)+2\sigma$. □

Finally, we show that for $k = C(m,\varepsilon)\log\log n$, the k-exploration tree, with high probability, connects to Core_n:

Lemma 10.30 (Connecting the exploration tree to the core)**.** *Let $k = (1+\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)}$. Then, under the conditions in Theorem 10.27, the probability that there exists an i such that the distance of i to the core is at least k is $o(n^{-1})$.*

Proof. By Lemma 10.29, the probability that there exists a vertex for which the k-exploration tree has at least 2 collisions before hitting the core is $o(n^{-1})$. When the k-exploration tree from a vertex i has at most one collisions, there are at least $(m-1)m^{k-1}$ half-edges in the kth layer that have not yet been paired. For $k_- = (1+\varepsilon/2)\frac{\log\log n}{\log(d_{\min}-1)}$ this number is at least equal to $N = (\log n)^{1+\varepsilon/2+o(1)}$. We investigate the neighborhoods of these half-edges one by one. Recall that $k = (1+\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)}$, and assume that the k-exploration tree has at most one collision. Then, when we successively pair the half-edges in the k_-th layer, we connect to one of the other half-edges at most once.

Denote the different half-edges incident to the k_-th layer by $x_1,\ldots,x_N$, where $N = (\log n)^{1+\varepsilon/2+o(1)}$. We call a half-edge a *success* when it reaches Core_n in at most $l = A\log\log\log n$ steps. Denote the event that x_i is a success by $\mathcal{A}_{x_i}$. We start by giving a lower bound on $\mathbb{P}(\mathcal{A}_{x_i})$. We start by pairing x_i to y_i, and let $v_{i,0}$ be the vertex incident to y_i. Let $v_{i,l} \in [n]$ denote the vertex (if any) with maximal degree at distance l from $v_{i,0}$. Let $\varepsilon > 0$ be so small that $(1-\varepsilon)/(\tau-2) > 1$. Let $A \ge 1$ be a large constant. When $d_{v_{i,0}} \ge A$, and, $d_{v_{i,j}} \ge (d_{v_{i,j-1}})^{(1-\varepsilon)/(\tau-2)}$ for each $j = 1,\ldots,l$, we have, with $p = (1-\varepsilon)/(\tau-2) > 1$ for

$$d_{v,l} \ge d_{v_{i,0}}^{p^l} \ge A^{p^{A\log\log\log n}} = \mathrm{e}^{\log A(\log\log n)^{A\log p}}, \tag{10.5.13}$$

which is much larger than $(\log n)^\sigma$ when $A > 1/\log p$. In particular, x_i is a success and $\mathcal{A}_{x_i}$ occurs. As a result,

$$\mathbb{P}(\mathcal{A}_{x_i}) \geq \mathbb{P}(d_{v_{i,0}} \geq A, d_{v_{i,j}} \geq (d_{v_{i,j-1}})^{(1-\varepsilon)/(\tau-2)} \ \forall j = 1, \ldots, l). \tag{10.5.14}$$

date from here!!

□

10.6 Branching processes with infinite mean

When $\tau \in (2,3)$, the branching processes $\{\mathcal{Z}_j^{(1)}\}$ and $\{\mathcal{Z}_j^{(2)}\}$ are well-defined, but has infinite mean in generations 2, 3, etc. Under certain conditions on the underlying offspring distribution, Davies [79] proves for this case that $(\tau-2)^j \log(1+\mathcal{Z}_j)$ converges almost surely, as $j \to \infty$, to some random variable Y. Moreover, $\mathbb{P}(Y=0) = 1-\eta$, the extinction probability of $\{\mathcal{Z}_j\}_{j=0}^\infty$. Therefore, also $(\tau-2)^j \log(\mathcal{Z}_j \vee 1)$ converges almost surely to Y.

Since $\tau > 2$, we still have that $\ell_n \approx \mathbb{E}[D]n$, so that the typical value of j for which (**??**) holds satisfies

$$\mathcal{Z}^{(1)}_{\lceil (j+1)/2 \rceil} \mathcal{Z}^{(2)}_{\lfloor (j+1)/2 \rfloor} \approx \mathbb{E}[D]n, \qquad \text{or} \qquad \log(\mathcal{Z}^{(1)}_{\lceil (j+1)/2 \rceil} \vee 1) + \log(\mathcal{Z}^{(2)}_{\lfloor (j+1)/2 \rfloor} \vee 1) \approx \log n.$$

This indicates that the typical value of j is of order

$$m_{\tau,n} = 2\frac{\log\log n}{|\log(\tau-2)|}, \tag{10.6.1}$$

as formulated in Theorem 10.17.

In the analysis of the configuration model, we shall also encounter branching processes with infinite mean. Such branching processes are always supercritical, as the following exercise shows:

Exercise 10.13. *Prove that a branching process with infinite mean is always supercritical, e.g., the probability that it survives forever is strictly positive.*

In this section, we shall give a scaling result for the generation sizes for such branching processes. The main result we shall make use of is the following theorem:

Theorem 10.31 (Branching processes with infinite mean)**.** *Let $\{Z_n\}_{n=0}^\infty$ be a branching process with offspring distribution $Z_1 = X$ having distribution function F_X. Assume that there exist a $\tau \in (1,2)$ and a non-negative, non-increasing function $x \mapsto \gamma(x)$, such that*

$$x^{-(\tau-1)-\gamma(x)} \leq 1 - F_X(x) \leq x^{-(\tau-1)+\gamma(x)}, \qquad \text{for large } x, \tag{10.6.2}$$

where $x \mapsto \gamma(x)$ satisfies

(i) *$x \mapsto x^{\gamma(x)}$ is non-decreasing,*

(ii) *$\int_0^\infty \gamma(\mathrm{e}^{\mathrm{e}^x})\,dx < \infty$, or, equivalently, $\int_{\mathrm{e}}^\infty \frac{\gamma(y)}{y\log y}\,dy < \infty$.*

Then $(\tau-1)^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y$, with $\mathbb{P}(Y=0)$ equal to the extinction probability of $\{Z_n\}$, whereas Y admits a density on $(0,\infty)$.

Theorem 10.31 covers the case where the branching process has an offspring which has very thick tails. Indeed, it is not hard to show that Theorem 10.31 implies that $\mathbb{E}[X^s] = \infty$ for every $s > \alpha \in (0,1)$ (see Exercise 10.15 below).

We shall not prove Theorem 10.31 in full generality. Rather, we shall prove it in a simpler, yet still quite general case, in which $\gamma(x) = (\log x)^{\gamma-1}$ for some $\gamma \in [0,1)$.

Exercise 10.14. *Prove that* $\gamma(x) = (\log x)^{\gamma-1}$ *for some* $\gamma \in [0,1)$ *satisfies the assumptions in Theorem 10.31.*

Proof of Theorem 10.31 for $\gamma(x) = (\log x)^{\gamma-1}$. The proof shall be divided into four main steps. Let $\alpha = \tau - 1$. Define

$$M_n = \alpha^n \log(Z_n \vee 1). \tag{10.6.3}$$

We shall first assume that $\mathbb{P}(Z_1 \geq 1) = 1$, so that $\eta = 1$. We start by splitting M_n in a suitable way.

The split. For $i \geq 1$, we define

$$Y_i = \alpha^i \log\Big(\frac{Z_i \vee 1}{(Z_{i-1} \vee 1)^{1/\alpha}}\Big). \tag{10.6.4}$$

We can write

$$M_n = Y_1 + Y_2 + \cdots + Y_n. \tag{10.6.5}$$

From this split, it is clear that almost sure convergence of M_n follows when the sum $\sum_{i=0}^{\infty} Y_i$ converges, which, in turn, is the case when

$$\sum_{i=1}^{\infty} \mathbb{E}\big[|Y_i|\big] < \infty. \tag{10.6.6}$$

This is what we shall prove in the following steps.

Inserting normalization sequences. We shall next investigate $\mathbb{E}\big[|Y_i|\big]$. We shall prove by induction on i that there exist constants $\kappa < 1$ and $C > 0$ such that

$$\mathbb{E}\big[|Y_i|\big] \leq K\kappa^i. \tag{10.6.7}$$

For $i = 0$, this follows from the fact that, when (10.6.2) holds, the random variable $Y_1 = \alpha \log(Z_1 \vee 1)$ has a bounded absolute expectation. This initializes the induction hypothesis. We next turn to the advancement of the induction hypothesis. For this, we recall the definition of u_n in (2.6.7) and define

$$U_i = \alpha^i \log\Big(\frac{u_{Z_{i-1}\vee 1}}{(Z_{i-1} \vee 1)^{1/\alpha}}\Big), \qquad V_i = \alpha^i \log\Big(\frac{Z_i \vee 1}{u_{Z_{i-1}\vee 1}}\Big). \tag{10.6.8}$$

Then we have that $Y_i = U_i + V_i$, so that

$$\mathbb{E}\big[|Y_i|\big] \leq \mathbb{E}\big[|U_i|\big] + \mathbb{E}\big[|V_i|\big]. \tag{10.6.9}$$

We shall bound each of these terms separately.

Bounding the normalizing constants. In this step, we analyse the normalizing constants $n \mapsto u_n$, assuming (10.6.2), and use this, as well as the induction hypothesis, to bound $\mathbb{E}\big[|U_i|\big]$.

When (10.6.2) holds, then it follows from [39] that there exists a slowly varying function $n \mapsto \ell(n)$ such that

$$u_n = n^{1/\alpha}\ell(n). \tag{10.6.10}$$

Since $n \mapsto \ell(n)$ is slowly varying, by Potter's Theorem, we have that, for every $\varepsilon > 0$, there exists a constant $C_\varepsilon \geq 1$ such that, for all $n \geq 1$,

$$C_\varepsilon^{-1} n^{-\varepsilon} \leq \ell(n) \leq C_\varepsilon n^{\varepsilon}, \tag{10.6.11}$$

so that, in turn, $u_n \leq C_\varepsilon n^{\frac{1}{\alpha}+\varepsilon}$. This gives a first bound on $n \mapsto u_n$. We next substitute this bound into (10.6.2) and use that $x \mapsto x^{\gamma(x)}$ is non-decreasing together with $\gamma(x) = (\log x)^{\gamma-1}$, to obtain that

$$1 + o(1) = n[1 - F_X(u_n)] \geq n\big[u_n^{-(\tau-1)-\gamma(u_n)}\big] \geq n\Big[u_n^{-(\tau-1)} \mathrm{e}^{\log\left(C_\varepsilon n^{\frac{1}{\alpha}+\varepsilon}\right)^\gamma}\Big], \quad (10.6.12)$$

which, in turn, implies that there exists a constant $c > 0$ such that

$$u_n \leq n^{1/\alpha} \mathrm{e}^{c(\log n)^\gamma}. \quad (10.6.13)$$

In a similar way, we can show the matching lower bound $u_n \geq n^{1/\alpha} \mathrm{e}^{-c(\log n)^\gamma}$. As a result, we obtain that

$$\mathbb{E}\big[|U_i|\big] \leq c\alpha^i \mathbb{E}\big[(\log(Z_{i-1} \vee 1))^\gamma\big]. \quad (10.6.14)$$

Using the concavity of $x \mapsto x^\gamma$ for $\gamma \in [0, 1)$, as well as Jensen's Inequality, we arrive at

$$\mathbb{E}\big[|U_i|\big] \leq c\alpha^i \Big(\mathbb{E}\big[(\log(Z_{i-1} \vee 1))\big]\Big)^\gamma = \alpha^{i(1-\gamma)} \mathbb{E}[M_{i-1}]^\gamma. \quad (10.6.15)$$

By (10.6.5) and (10.6.7), which implies that $\mathbb{E}[M_{i-1}] \leq K\kappa/(1-\kappa)$, we arrive at

$$\mathbb{E}\big[|U_i|\big] \leq \alpha^{i(1-\gamma)} c\Big(\frac{K\kappa}{1-\kappa}\Big)^\gamma. \quad (10.6.16)$$

Bounding the logarithmic moment of an asymptotic stable random variable. In this step, we bound $\mathbb{E}\big[|V_i|\big]$. We note that by Theorem 2.28 and for Z_i quite large, the random variable $(Z_i \vee 1)/(u_{Z_{i-1}\vee 1})$ should be close to a stable random variable. We make use of this fact by bounding

$$\mathbb{E}\big[|V_i|\big] \leq \alpha^i \sup_{m\geq 1} \mathbb{E}\big[\big|\log\big(S_m/u_m\big)\big|\big], \quad (10.6.17)$$

where $S_m = X_1 + \cdots + X_m$, and $\{X_i\}_{i=1}^m$ are i.i.d. copies of the offspring distribution X. We shall prove that there exists a constant $C > 0$ such that, for all $m \geq 1$,

$$\mathbb{E}\big[\big|\log\big(S_m/u_m\big)\big|\big] \leq C. \quad (10.6.18)$$

In order to prove (10.6.18), we note that it suffices to bound

$$\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_+\big] \leq C_+, \qquad \mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_-\big] \leq C_-, \quad (10.6.19)$$

where, for $x \in \mathbb{R}$, $x_+ = \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$. Since $|x| = x_+ + x_-$, we then obtain (10.6.18) with $C = C_+ + C_-$. In order to prove (10.6.18), we start by investigating $\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_-\big]$. We note that $(\log x)_- = \log(x^{-1} \vee 1)$, so that

$$\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_-\big] = \mathbb{E}\big[\log\big(u_m/(S_m \wedge u_m)\big)\big], \quad (10.6.20)$$

where $x \wedge y = \min\{x, y\}$. The function $x \mapsto \log\big((u_m/(x \wedge u_m)\big)$ is non-increasing, and, since $S_m \geq X_{(m)}$, where $X_{(m)} = \max_{1\leq i\leq m} X_i$, we arrive at

$$\mathbb{E}\big[\log\big(u_m/(S_m \wedge u_m)\big)\big] \leq \mathbb{E}\big[\log\big(u_m/(X_{(m)} \wedge u_m)\big)\big]. \quad (10.6.21)$$

We next use that, for $x \geq 1$, $x \mapsto \log(x)$ is concave, so that, for every s,

$$\begin{aligned}\mathbb{E}\big[\log\big(u_m/(X_{(m)} \wedge u_m)\big)\big] &= \frac{1}{s}\mathbb{E}\big[\log\big((u_m/(X_{(m)} \wedge u_m))^s\big)\big] \leq \frac{1}{s}\log\Big(\mathbb{E}\big[\big(u_m/(X_{(m)} \wedge u_m)\big)^s\big]\Big) \\ &\leq \frac{1}{s} + \frac{1}{s}\log\Big(u_m^p \mathbb{E}\big[X_{(m)}^{-s}\big]\Big), \end{aligned} \quad (10.6.22)$$

where, in the last step, we made us of the fact that $u_m/(x \wedge u_m) \leq 1 + u_m/x$. Now rewrite $X_{(m)}^{-s} = (-Y_{(m)})^s$, where $Y_j = -X_j^{-1}$ and $Y_{(m)} = \max_{1 \leq j \leq m} Y_j$. Clearly, since $X_i \geq 1$, we have that $Y_j \in [-1, 0]$, so that $\mathbb{E}[(-Y_1)^s] < \infty$. Also, $u_m Y_{(m)} = -u_m/X_{(m)}$ converges in distribution to $-E^{-1/(\tau-1)}$, where E is exponential with mean 1, so it follows from [189, Theorem 2.1] that, as $m \to \infty$,

$$\mathbb{E}\left[(u_m Y_{(m)})^p\right] \to \mathbb{E}[E^{-1/(\tau-1)}] < \infty. \tag{10.6.23}$$

We proceed with $\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_+\big]$, for which the proof is a slight adaptation of the above argument. Now we make use of the fact that $(\log x)_+ = \log(x \vee 1) \leq 1 + x$ for $x \geq 0$, so that we must bound

$$\mathbb{E}\big[\log\big(S_m \vee u_m/u_m\big)\big] = \frac{1}{s}\mathbb{E}\big[\log\big((S_m \vee u_m/u_m)\big)^s\big)\big] \leq \frac{1}{s} + \log\Big(\mathbb{E}\big[\big(\frac{S_m}{u_m}\big)^s\big]\Big). \tag{10.6.24}$$

The discussion on page 565 and Corollary 1 of [113] yields, for $s < \tau - 1$, $\mathbb{E}[S_m^s = \mathbb{E}[|S_m|^s] \leq 2^{s/2}\lambda_s(t)$, for some function $\lambda_s(m)$ depending on s, m and F_X. Using the discussion on [113, Page 564], we have that $\lambda_s(m) \leq C_s m^{s/(\tau-1)} l(m^{1/(\tau-1)})^s$, where $l(\,\cdot\,)$ is a slowly varying function. With some more effort, it can be shown that we can replace $l(m^{1/(\tau-1)})$ by $\ell(m)$, which gives

$$\mathbb{E}\big[\log\big(S_m \vee u_m/u_m\big)\big] \leq \frac{1}{s} + \log\Big(\mathbb{E}\big[\big(\frac{S_m}{u_m}\big)^s\big]\Big) \leq \frac{1}{s} + \frac{C_s}{s} m^{s/(\tau-1)} \ell(m)^s u_m^{-s} = \frac{1}{s} + 2^{s/2}\frac{C_s}{s}, \tag{10.6.25}$$

which proves the first bound in (10.6.19) with $C_+ = \frac{1}{s} + 2^{s/2}\frac{C_s}{s}$.

Completion of the proof of Theorem 10.31 when $X \geq 1$. Combining (10.6.9) with (10.6.16) and (10.6.17)–(10.6.18), we arrive at

$$\mathbb{E}\big[|Y_i|\big] \leq \alpha^{i(1-\gamma)} c\Big(\frac{K\kappa}{1-\kappa}\Big)^\gamma + C\alpha^i \leq K\kappa^i, \tag{10.6.26}$$

when we take $\kappa = \alpha^{1-\gamma}$ and we take K to be sufficiently large, for example $K \geq 2C$ and $K \geq 2c\Big(\frac{K\kappa}{1-\kappa}\Big)^\gamma$. We shall not prove that Y admits a density. This completes the proof when the offspring distribution X satisfies $X \geq 1$.

Completion of the proof of Theorem 10.31. We finally extend the result to the setting where $X = 0$ with positive probability. Since $\mathbb{E}[X] = \infty$, we have that the survival probability $\xi > 0$ (recall Exercise 10.13). Conditionally on extinction, we have that $Z_n \xrightarrow{a.s.} 0$, so that, on the survival event, $(\tau - 1)^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y$, where, conditionally on extinction, $Y = 0$.

It remains to prove that $(\tau - 1)^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y$ on the survival event. By Theorem 3.12, we have that, conditionally on survival,

$$\frac{Z_n^{(\infty)}}{Z_n} \xrightarrow{a.s.} \xi > 0, \tag{10.6.27}$$

where we recall that $Z_n^{(\infty)}$ are the individuals in the n^{th} generation which have an infinite line of descent. By Theorem 3.11 and conditionally on survival, $\{Z_n^{(\infty)}\}_{n=0}^\infty$ is again a branching process, now with offspring distribution $p^{(\infty)}$ given in (3.4.2). Note that, in particular, $\mathbb{P}(Z_1^{(\infty)} \geq 1) = 1$, and we wish to apply Theorem 10.31 to $Z_n^{(\infty)}$ instead of Z_n. We shall assume that also $p^{(\infty)}$ in (3.4.2) satisfies the conditions in Theorem 10.31 with

the function $x \mapsto \gamma^*(x)$, given by $\gamma^*(x) = \gamma(x) + c$ for some constant $c \geq 0$. Thus, we have that, conditionally on survival,

$$(\tau - 1)^n \log(Z_n^{(\infty)} \vee 1) \xrightarrow{a.s.} Y^{(\infty)}, \tag{10.6.28}$$

and combining (10.6.27) and (10.6.28), it immediately follows that, conditionally on survival, we also obtain

$$(\tau - 1)^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y^{(\infty)}. \tag{10.6.29}$$

We conclude that Theorem 3.12 holds, where $Y = 0$ with probability $\eta = 1 - \zeta$ and $Y = Y^{(\infty)}$ with probability ζ. □

Exercise 10.15. *Prove that $\mathbb{E}[X] = \infty$ when the conditions in Theorem 10.31 are satisfied. Extend this to show that $\mathbb{E}[X^s] = \infty$ for every $s > \alpha \in (0, 1)$.*

Exercise 10.16. *Show that, under the conditions of Theorem 10.31, also $\alpha^n \log(Z_n + 1)$ converges to Y almost surely.*

We finally state some properties of the a.s. limit Y of $\{(\tau - 1)^n \log(Z_n \vee 1)\}_{n=0}^{\infty}$ taken from [79], of which we omit a proof:

Theorem 10.32 (The limiting variable for infinite mean branching processes)**.** *Under the conditions of Theorem 10.31,*

$$\lim_{x \to \infty} \frac{\log \mathbb{P}(Y > x)}{x} = -1, \tag{10.6.30}$$

where is the a.s. limit of $(\tau - 1)^n \log(Z_n \wedge 1)$.

Theorem 10.32 can be understood from the fact that by (10.6.3)–(10.6.4), we have

$$Y = \sum_{n=1}^{\infty} Y_i, \tag{10.6.31}$$

where

$$Y_1 = (\tau - 1) \log \big(Z_1 \vee 1\big). \tag{10.6.32}$$

By (10.6.2),

$$\mathbb{P}(Y_1 > x) = \mathbb{P}(Z_1 > \mathrm{e}^{x^{1/(\tau-1)}}) = \mathrm{e}^{-x(1+o(1))}, \tag{10.6.33}$$

which shows that Y_1 satisfies (10.6.30). The equality in (10.6.31) together with (10.6.4) suggests that the tails of Y_1 are equal to those of Y, which heuristically explains (10.6.30).

10.7 Related random graph models

In this section, we describe a few related models that have been investigated in the literature.

The directed configuration model. Many real-world networks are *directed*, in the sense that edges are oriented. For example, in the World-Wide Web, the vertices are web pages, and the edges are the hyperlinks between the pages, which are clearly oriented. One could naturally forget about these directions, but that would discard a wealth of information. For example, in citation networks, it makes a substantial difference whether my paper links to a paper, or that paper links to mine.

One way to obtain a directed version of $\mathrm{CM}_n(\boldsymbol{d})$ is to give each edge a direction, chosen with probability 1/2, independently of all other edges. In this model, however, the correlation coefficient between the in- and out-degree of vertices is close to one, particularly when the degrees are large. In real-world applications, correlations between in- and out-degrees can be positive or negative, depending on the precise application. Therefore, we formulate a general model of directed graphs, where we can prescribe both the in- and out-degrees of vertices.

Fix $\boldsymbol{d}^{(\mathrm{in})} = (d_i^{(\mathrm{in})})_{i\in[n]}$ to be a sequence of in-degrees, where $d_i^{(\mathrm{in})}$ denotes the in-degree of vertex i. Similarly, we let $\boldsymbol{d}^{(\mathrm{out})} = (d_i^{(\mathrm{out})})_{i\in[n]}$ be a sequence of out-degrees. Naturally, we need that

$$\sum_{i\in[n]} d_i^{(\mathrm{in})} = \sum_{i\in[n]} d_i^{(\mathrm{out})} \tag{10.7.1}$$

in order for a graph with in- and out-degree sequence $\boldsymbol{d} = (\boldsymbol{d}^{(\mathrm{in})}, \boldsymbol{d}^{(\mathrm{out})})$ to exist. We think of $d_i^{(\mathrm{in})}$ as the number of in-half-edges incident to vertex i and $d_i^{(\mathrm{out})}$ as the number of out-half-edges incident to vertex i. The directed configuration model $\mathrm{DCM}_n(\boldsymbol{d})$ is obtained by pairing each in-half-edge to a uniformly chosen out-half-edge. The resulting graph is a random multigraph, where each vertex i has in-degree $d_i^{(\mathrm{in})}$ and out-degree $d_i^{(\mathrm{out})}$. Similarly to $\mathrm{CM}_n(\boldsymbol{d})$, $\mathrm{DCM}_n(\boldsymbol{d})$ can have self-loops as well as multiple edges. A self-loop arises at vertex i when one of its in-half-edges pairs to one of its out-half-edges.

The household model.

Random intersection graph.

10.8 Related results for the configuration model

In this section, we discuss related results for the configuration model. We start by investigating the setting of

Infinite mean degrees. In this section, we investigate the distances in $\mathrm{CM}_n(\boldsymbol{d})$ when the degrees $\boldsymbol{d}$ are i.i.d. and have *infinite* mean. More specifically, we assume that there exist $\tau \in (1,2)$ and $c > 0$ such that

$$\lim_{x\to\infty} x^{\tau-1}[1-F](x) = c. \tag{10.8.1}$$

In the statement, we make heavy use of the notation used in Theorem 7.17, which we first recall. Recall that the random probability distribution $P = (P_i)_{i=1}^{\infty}$ is given by

$$P_i = Z_i/Z, \tag{10.8.2}$$

where $Z_i = \Gamma_i^{-1/(\tau-1)}$ and $\Gamma_i = \sum_{j=1}^{i} E_i$ with $(E_i)_{i\geq 1}$ an i.i.d. sequence of exponential random variables with parameter 1, and where $Z = \sum_{i\geq 1} \Gamma_i$. Recall further that $M_{P,k}$ is a multinomial distribution with parameters k and probabilities $P = \{P_i\}_{i=1}^{\infty}$. Thus, $M_{P,k} = (B_1, B_2, \ldots)$, where, conditionally on $P = (P_i)_{i\geq 1}$, B_i is the number of outcomes i in k independent trials such that each outcome is equal to i with probability P_i. In Theorem

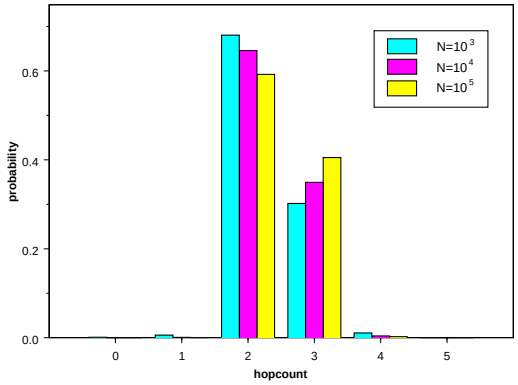

Figure 10.1: Empirical probability mass function of the hopcount for $\tau = 1.8$ and $N = 10^3, 10^4, 10^5$.

7.17, the random variable M_{P,D_1} appears, where D_1 is independent of $P = (P_i)_{i\geq 1}$. We let $M^{(1)}_{P,D_1}$ and $M^{(2)}_{P,D_2}$ be two random variables which are conditionally independent given $P = (P_i)_{i\geq 1}$.

In terms of this notation, the main result on distances in $\mathrm{CM}_n(\boldsymbol{d})$ when the degrees have infinite mean is the following:

Theorem 10.33 (Distances in $\mathrm{CM}_n(\boldsymbol{d})$ with i.i.d. infinite mean degrees)**.** *Fix $\tau \in (1,2)$ in* (10.8.1) *and let $d_1, d_2, \ldots, d_n$ be a sequence of i.i.d. copies of D. Then,*

$$\lim_{n\to\infty} \mathbb{P}(H_n = 2) = 1 - \lim_{n\to\infty} \mathbb{P}(H_n = 3) = p_F \in (0,1). \tag{10.8.3}$$

The probability p_F can be identifies as the probability that $M^{(1)}_{P,D_1}$ and $M^{(2)}_{P,D_2}$ have an identical outcome, i.e., there is an outcome that occurs both in $M^{(1)}_{P,D_1}$ and $M^{(2)}_{P,D_2}$.

Proof. We sketch the proof of Theorem 10.33. First, with high probability, both $d_1 \leq \log n$ and $d_2 \leq \log n$. The event that $H_n = 1$ occurs precisely when one of the d_1 half-edges of vertex 1 is attached to one of the d_2 half-edges of vertex 2. Also, with high probability, $\ell_n \geq n^{1/(\tau-1)-\varepsilon}$. Therefore, on the event that $\ell_n \geq n^{1/(\tau-1)-\varepsilon}$ and $d_1 \leq \log n$ and $d_2 \leq \log n$, the probability that $H_n = 1$ is bounded above by

$$\frac{(\log n)^2}{n^{1/(\tau-1)-\varepsilon}} = o(1). \tag{10.8.4}$$

Exercise 10.17 (Typical distance is at least 2 **whp**)**.** *Complete the argument that* $\mathbb{P}(H_n = 1) = o(1)$.

We note that the proof of Theorem 7.17 implies that $M^{(1)}_{P,D_1}$ denotes the number of edges between vertex 1 and the largest order statistics. Indeed, $M^{(1)}_{P,D_1} = (B^{(1)}_1, B^{(1)}_2, \ldots)$, where $B^{(1)}_i$ is the number of edges between vertex i and the vertex with degree $D_{(n+1-i)}$. The same applies to vertex 2. As a result, when $M^{(1)}_{P,D_1}$ and $M^{(2)}_{P,D_2}$ have an identical outcome,

then the typical graph distance equals 2. We are left to prove that the typical graph distance is bounded by 3 with high probability. By (2.6.17), we have that $\xi_k k^{1/(\tau-1)} \xrightarrow{\mathbb{P}} 1$ as $k \to \infty$. Thus, when K is large, the probability that vertex 1 is not connected to any of the vertices corresponding to $\{D_{(n+1-i)}\}_{i=1}^{K}$ converges to 0 when K tends to infinity.

Recall that $\mathbb{P}_n$ denotes the conditional probability given the degrees $(d_i)_{i\in[n]}$. For $i = 1, 2, \ldots,$, we let v_i be the vertex corresponding to the i^{th} order statistic $D_{(n+1-i)}$. By (11.4.11), we have that

$$\mathbb{P}_n(v_i \text{ not directly connected to } v_j) \le \mathrm{e}^{-\frac{D_{(n+1-i)}D_{(n+1-j)}}{\ell_n}}. \tag{10.8.5}$$

Moreover, $D_{(n+1-i)}, D_{(n+1-j)} \ge n^{1/(\tau-1)-\varepsilon}$ with high probability for n sufficiently large and any $\varepsilon > 0$, while $\ell_n \le n^{1/(\tau-1)+\varepsilon}$ with high probability. As a result, with high probability,

$$\mathbb{P}_n(v_i \text{ not directly connected to } v_j) \le \mathrm{e}^{-n^{1/(\tau-1)-3\varepsilon}}, \tag{10.8.6}$$

which converges to 0 exponentially fast. Therefore, for every $i, j \in \{1, \ldots, K\}$, the vertices v_i and v_j are connected with high probability. This implies that the vertices corresponding to the high order statistics form a *complete graph*. We have already concluded that 1 is connected to v_i for some $i \le K$. In the same way, we conclude that vertex 2 is connected to v_j for some $j \le K$. Since v_i is with high probability connected to v_j, we conclude that

$$\mathbb{P}_n\big(H_n \le 3\big) = 1 - o(1). \tag{10.8.7}$$

This completes the proof. □

Fluctuation of distances for finite-variance degrees. We continue to study the *fluctuations* of the distances in the configuration model. We need a limit result from branching process theory before we can identify the limiting random variables $(R_a)_{a\in(-1,0]}$. We introduce a *delayed* branching process $\{\mathcal{Z}_k\}$, where in the first generation, the offspring distribution is chosen according to F and in the second and further generations, the offspring is chosen in accordance to g given by

$$g_j = \frac{(j+1)p_{j+1}}{\mathbb{E}[D]}, \quad j = 0, 1, \ldots. \tag{10.8.8}$$

The process $(\mathcal{Z}_k/\mathbb{E}[D]\nu^{k-1})_{k\ge1}$ is a martingale with uniformly bounded expectation and consequently converges almost surely to a limit (see e.g., Theorem 2.21 and Exercise 2.23):

$$\lim_{n\to\infty} \frac{\mathcal{Z}_n}{\mathbb{E}[D]\nu^{n-1}} = \mathcal{W} \qquad a.s. \tag{10.8.9}$$

In the theorem below we need two independent copies $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ of $\mathcal{W}$.

Theorem 10.34 (Limit law for the average distance)**.** *Assume that* (**??**) *is satisfied, and let $\nu > 1$. For $k \ge 1$, let $a_k = \lfloor \log_\nu k \rfloor - \log_\nu k \in (-1, 0]$. Then, there exist random variables $(R_a)_{a\in(-1,0]}$ such that as $n \to \infty$,*

$$\mathbb{P}\big(H_n - \lfloor \log_\nu n \rfloor = k \mid H_n < \infty\big) = \mathbb{P}(R_{a_n} = k) + o(1), \qquad k \in \mathbb{Z}, \tag{10.8.10}$$

where, for $a \in (-1, 0]$,

$$\mathbb{P}(R_a > k) = \mathbb{E}\big[\exp\{-\kappa\nu^{a+k}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big|\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\big], \tag{10.8.11}$$

where $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ are independent limit copies of $\mathcal{W}$ in (10.8.9) and where $\kappa = \mathbb{E}[D](\nu - 1)^{-1}$.

In words, Theorem 10.34 states that for $\tau > 3$, the graph distance H_n between two randomly chosen connected vertices grows like the $\log_\nu n$, where n is the size of the graph, and that the fluctuations around this mean remain uniformly bounded in n. Theorem 10.34 proves the conjecture in Newman, Strogatz and Watts [179, Section II.F, (54)], where a heuristic is given that the number of edges between arbitrary vertices grows like $\log_\nu n$. In addition, Theorem 10.34 improves upon that conjecture by specifying the fluctuations around the value $\log_\nu n$.

The law of R_a is involved, and can in most cases not be computed exactly. The reason for this is the fact that the random variables $\mathcal{W}$ that appear in its statement are hard to compute explicitly. For example, for the configuration model with $\tau > 3$, we do not know the law of $\mathcal{W}$. See also Chapter 3. There are two examples where the law of $\mathcal{W}$ is known. The first is when all degrees in the graph are equal to some $r > 2$, and we obtain the r-regular graph (see also [46], where the diameter of this graph is studied). In this case, we have that $\mathbb{E}[D] = r, \nu = r - 1$, and $\mathcal{W} = 1$ a.s. In particular, $\mathbb{P}(H_n < \infty) = 1 + o(1)$. Therefore, we obtain that

$$\mathbb{P}(R_a > k) = \exp\{-\frac{r}{r-2}(r-1)^{a+k}\}, \tag{10.8.12}$$

and H_n is asymptotically equal to $\log_{r-1} n$. The second example is when g is the probability mass function of a geometric random variable, in which case the branching process with offspring q conditioned to be positive converges to an exponential random variable with parameter 1. This example corresponds to

$$q_j = p(1-p)^{j-1}, \quad \text{so that} \quad p_j = \frac{1}{jc_p}p(1-p)^{j-2}, \quad \forall j \geq 1, \tag{10.8.13}$$

and c_p is a normalization constant. For $p > \frac{1}{2}$, the law of $\mathcal{W}$ has the same law as the sum of D_1 copies of a random variable $\mathcal{Y}$, where $\mathcal{Y} = 0$ with probability $\frac{1-p}{p}$ and equal to an exponential random variable with parameter 1 with probability $\frac{2p-1}{p}$. Even in this simple case, the computation of the exact law of R_a is non-trivial.

We finally give two exercises discussing consequences of Theorem 10.34:

Exercise 10.18. *([117]) Fix an integer n_1. Prove that, under the assumptions in Theorem 10.34, and conditionally on $H_n < \infty$, along the subsequence $n_k = \lfloor n_1 \nu^{k-1} \rfloor$, the sequence of random variables $H_{n_k} - \lfloor \log_\nu n_k \rfloor$ converges in distribution to $R_{a_{n_1}}$ as $k \to \infty$.*

Exercise 10.19. *([117]) Prove that, under the assumptions in Theorem 10.34,*

(i) *with probability $1 - o(1)$ and conditionally on $H_n < \infty$, the random variable H_n is in between $(1 \pm \varepsilon) \log_\nu n$ for any $\varepsilon > 0$;*

(ii) *conditionally on $H_n < \infty$, the random variables $H_n - \log_\nu n$ form a tight sequence, i.e.,*

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(|H_n - \log_\nu n| \leq K \big| H_n < \infty\big) = 1. \tag{10.8.14}$$

As a consequence, prove that the same result applies to a uniform random graph with degrees $(d_i)_{i\in[n]}$. Hint: Make use of Theorem 7.15.

Fluctuation of distances for infinite-variance degrees. We next study the fluctuations of the average distance in $\mathrm{CM}_n(\boldsymbol{d})$, for which we need to make a stronger assumption than in (**??**). For $\tau \in (2, 3)$, we assume that there exists $\gamma \in [0, 1)$ and $C < \infty$ such that

$$x^{-\tau+1-C(\log x)^{\gamma-1}} \leq 1 - F(x) \leq x^{-\tau+1+C(\log x)^{\gamma-1}}, \qquad \text{for large } x. \tag{10.8.15}$$

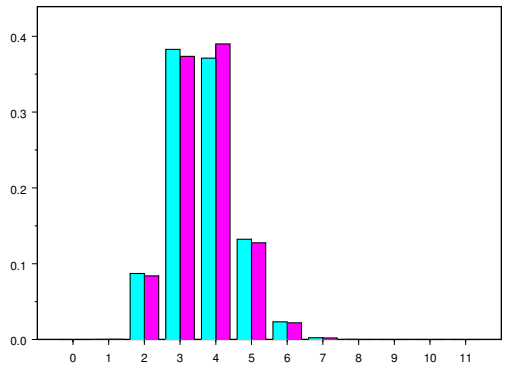

Figure 10.2: Number of AS traversed in hopcount data (blue) compared to the model (purple) with $\tau = 2.25, n = 10,940$.

The condition in (10.8.15) is slightly stronger than (**??**), and is such that the results in Theorem 10.31 apply. Then, we can identify the fluctuations of the typical graph distance in $\mathrm{CM}_n(\boldsymbol{d})$ as follows:

Theorem 10.35 (Fluctuations graph distance $\mathrm{CM}_n(\boldsymbol{d})$ for infinite variance degrees)**.** *Fix $\tau \in (2,3)$ and assume that* (10.8.15) *holds. Then there exist random variables $(R_a)_{a\in(-1,0]}$ such that, as $n \to \infty$,*

$$\mathbb{P}\Big(H_n = 2\Big\lfloor \frac{\log\log n}{|\log(\tau-2)|}\Big\rfloor + l \;\Big|H_n < \infty\Big) = \mathbb{P}(R_{a_n} = l) + o(1), \qquad l \in \mathbb{Z}, \tag{10.8.16}$$

where $a_n = \lfloor \frac{\log\log n}{|\log(\tau-2)|}\rfloor - \frac{\log\log n}{|\log(\tau-2)|} \in (-1,0]$, and where, for $a \in (-1,0]$,

$$\mathbb{P}(R_a > l) = \mathbb{P}\Big(\min_{s\in\mathbb{Z}}\big[(\tau-2)^{-s}Y^{(1)} + (\tau-2)^{s-c_l}Y^{(2)}\big] \le (\tau-2)^{\lceil l/2\rceil + a}\big|Y^{(1)}Y^{(2)} > 0\Big),$$

where $c_l = 1$ if l is even, and zero otherwise, and $Y^{(1)}, Y^{(2)}$ are two independent copies of the limit random variable in Theorem 10.31.

In words, Theorem 10.17 states that for $\tau \in (2,3)$, the graph distance H_n between two randomly chosen connected vertices grows proportional to $\log\log$ of the size of the graph, and that the fluctuations around this mean remain uniformly bounded in n.

Simulations indicating the properties of the typical graph distance for $\tau \in (2,3)$ can bee seen in Figure 10.2. In it, the distances of the AS-graph in Figure 1.2, and these distances are compared to the ones in $\mathrm{CM}_n(\boldsymbol{d})$ where n is equal to the number of AS and the best approximation to the exponent of the power-law for the degree sequence of the AS-graph, which is $\tau = 2.25$.

The near-critical behavior in the configuration model. In [133], also partial results appear on the near-critical behavior of $\mathrm{CM}_n(\boldsymbol{d})$:

Theorem 10.36. *Suppose that Condition 7.5(a)-(c) hold and that $\nu = 1$. Assume further that $\alpha_n = \mathbb{E}[D_n(D_n-2)] = \sum_{i\in[n]} d_i(d_i-2)/n > 0$ and, moreover, $n^{1/3}\alpha_n \to \infty$, and that*

$$\sum_{i\in[n]} d_i^{4+\varepsilon} = O(n) \tag{10.8.17}$$

for some $\varepsilon > 0$. Let $\beta := \mathbb{E}[D(D-1)(D-2)]$. Then, $\beta > 0$ and

$$\begin{aligned} |\mathcal{C}_{\max}| &= \frac{2\mathbb{E}[D]}{\beta} n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \\ |v_k(\mathcal{C}_{\max})| &= \frac{2}{\beta} k p_k n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \text{ for every } k \geq 0, \\ |E(\mathcal{C}_{\max})| &= \frac{2\mathbb{E}[D]}{\beta} n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \end{aligned}$$

while $|\mathcal{C}_{(2)}| = o_{\mathbb{P}}(n\alpha_n)$ and $|E(\mathcal{C}_{(2)}) = o_{\mathbb{P}}(n\alpha_n)$.

d results critical behavior
!

The critical behavior in the configuration model.

10.9 Notes and discussion

Notes on Section 10.1 This section is based on [133].

Notes on Section 10.2. These results are folklore, even though we were not able to find a proof.

Notes on Section 10.3. Distances in the configuration model were first obtained in [179, 180]. Theorem 10.16 is [107, Theorem 4.1]. A $\log n$ lower bound on the diameter is also proved in [119]. Theorem 10.34 is proved in [117]. Theorem 10.35 is proved in [118]. The identification of $\mathbb{P}(H_n = 2)$ is new. The proof of Theorem 10.23 is close in spirit to the analysis in [195], the only difference being that we have simplified the argument slightly.

One might argue that including degrees larger than $n-1$ is artificial in a network with n vertices. In fact, in many real networks, the degree is bounded by a physical constant. Therefore, in [103], also the case where the degrees are conditioned to be smaller than n^α is considered, where α is an arbitrary positive number. Of course, we cannot condition on the degrees to be at most M, where M is fixed and independent on n, since in this case, the degrees are uniformly bounded, and this case is treated in [117] as well. Therefore, [103] considers cases where the degrees are conditioned to be at most a given power of n. In this setting, it turns out that the average distance is equal to $k+3$ with high probability, whenever $\alpha \in (1/(\tau+k), 1/(\tau+k-1))$.

d notes on Section 10.4.

Notes on Section 10.4. Proposition 10.18 is adapted from [**?**, Lemma 5.1]

Notes on Section 10.5. Theorem 10.27 is novel, as far as we are aware.

Notes on Section 10.6. Theorem 10.31 is proved in [79]. A related result, under stronger conditions, appeared in [78]. Branching processes with infinite mean have attracted considerable attention, see e.g., [198, 200] and the references therein. There is a balance between the generality of the results and the conditions on the offspring distribution, and, in our opinion, Theorem 10.31 strikes a nice balance in that the result is relatively simple and the conditions fairly general.

Notes on Section 10.8. Theorem 10.33 is proved in [103].

Chapter 11
Preferential attachment revisited

In this chapter, we state some further results on preferential attachment models. We start by discussing an important tool in this chapter: exchangeable random variables and their distribution described in De Finetti's Theorem. These results shall be applied to Polya urn schemes, which can be used to describe the distribution of the degrees in preferential attachment models. In Section 11.2, we investigate the connectivity of $\mathrm{PA}_{m,\delta}(t)$. In Section 11.2.1, we investigate graph distances in $\mathrm{PA}_{m,\delta}(t)$.

Throughout this chapter, we shall work with the preferential attachment model defined in Section 8.1 and denoted by $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$, unless stated otherwise. We recall that this model starts with a single vertex with m self-loops at time $t = 1$, and at each time, a vertex is added with m edges which are attached to the vertices in the graph with probabilities given in (8.1.1) for $m = 1$, and as described on page 164 for $m \geq 2$. This model can also be obtained by identifying blocks of m vertices in $\{\mathrm{PA}_{1,\delta}(t)\}_{t=1}^{\infty}$. We shall sometimes also discuss other variants of the model, such as $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$, in which the $m = 1$ model does not have any self-loops.

11.1 Exchangeable random variables and Polya urn schemes

In this section, we discuss the distribution of infinite sequences of exchangeable random variables and their applications to Polya urn schemes. We start by discussing De Finetti's Theorem.

De Finetti's Theorem for infinite sequences of exchangeable random variables. A sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is called *exchangeable* when, for every $n \geq 1$, the distribution of $\{X_i\}_{i=1}^{n}$ is the same as the one of $\{X_{\sigma(i)}\}_{i=1}^{n}$ for any permutation $\sigma\colon \{1,\ldots,n\} \to \{1,\ldots,n\}$ of $\{1,\ldots,n\}$. Clearly, when a sequence of random variables is independent and identically distributed, then it is also exchangeable. Remarkably, however, the distribution of an infinite sequence of random variables is always a *mixture* of independent and identically distributed random variables. This is the content of De Finetti's Theorem, which we shall state and prove here in the case where $\{X_i\}_{i=1}^{\infty}$ are indicator variables:

Theorem 11.1 (De Finetti's Theorem)**.** *Let $\{X_i\}_{i=1}^{\infty}$ be an infinite sequence of exchangeable random variables, and assume that $X_i \in \{0,1\}$. Then there exists a random variable U with $\mathbb{P}(U \in [0,1]) = 1$ such that, for all $1 \leq k \leq n$,*

$$\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \mathbb{E}[U^k(1-U)^{n-k}]. \tag{11.1.1}$$

De Finetti's Theorem (Theorem 11.1) states that an infinite exchangeable sequence of indicators has the same distribution as an independent Bernoulli sequence with a *random* success probability U. Thus, the different elements of the sequence are *not* independent, but their dependence enters only through the random success probability.

From the proof, we shall see that Theorem 11.1 in fact holds more generally, for example, when X_i takes on at most a *finite* number of values. Since we shall use Theorem 11.1 only for indicator variables, we refrain from stating this version.

When writing

$$S_n = \sum_{k=1}^{n} X_k, \tag{11.1.2}$$

Theorem 11.1 implies that

$$\mathbb{P}(S_n = k) = \mathbb{E}\Big[\mathbb{P}\big(\mathrm{Bin}(n, U) = k\big)\Big]. \tag{11.1.3}$$

Equation (11.1.3) also allows us to compute the distribution of U. Indeed, when we would have that

$$\lim_{n\to\infty} n\mathbb{P}(S_n = \lceil un \rceil) = f(u), \tag{11.1.4}$$

where f is a density, then (11.1.3) implies that f is in fact the density of the random variable U. This shall prove to be useful in applications of De Finetti's Theorem (Theorem 11.1). Furthermore, by the strong law of large numbers applied to the conditional law given U, we have that $S_n/n \xrightarrow{a.s.} U$.

Exercise 11.1. *Prove* (11.1.3).

Proof of Theorem 11.1. The proof of this deep result is in fact quite simple. It makes use of Helly's Theorem, which states that any sequence of *bounded* random variables has a weakly converging subsequence.

We fix $m \geq n$ and condition on S_m to write

$$\begin{aligned}\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0)\\ = \sum_{j=k}^{m} \mathbb{P}\big(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 \big| S_m = j\big)\mathbb{P}(S_m = j).\end{aligned} \tag{11.1.5}$$

Now, when conditioning on $S_m = j$ and by exchangeability, each sequence $\{X_i\}_{i=1}^m$ containing precisely j ones is equally likely. Since there are precisely $\binom{m}{j}$ such sequences, and precisely $\binom{n}{k}\binom{m-n}{j-k}$ of them start with k ones and $n-k$ zeros, we obtain

$$\mathbb{P}\big(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 \big| S_m = j\big) = \frac{\binom{n}{k}\binom{m-n}{j-k}}{\binom{m}{j}}. \tag{11.1.6}$$

We therefore arrive at

$$\begin{aligned}\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0)\\ = \sum_{j=k}^{m} \frac{j\cdots(j-k+1)\cdot(m-j)\cdots(m-j-(n-k)+1)}{m\cdots(m-n+1)}\mathbb{P}(S_m = j).\end{aligned} \tag{11.1.7}$$

When $m \to \infty$, we see that, for $k \leq n$ fixed,

$$\frac{j\cdots(j-k+1)\cdot(m-j)\cdots(m-j-(n-k)+1)}{m\cdots(m-n+1)} = \Big(\frac{j}{m}\Big)^k\Big(1-\frac{j}{m}\Big)^{n-k} + o(1). \tag{11.1.8}$$

Recall that $S_m = j$, so that

$$\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \lim_{m\to\infty} \mathbb{E}\big[Y_m^k(1-Y_m)^{n-k}\big], \tag{11.1.9}$$

where we write $Y_m = S_m/m$. Note that it is here that we make use of the fact that $\{X_i\}_{i=1}^\infty$ is an *infinite* exchangeable sequence of random variables.

Since $0 \leq S_m \leq m$, we have that $0 \leq Y_m \leq 1$, so that the sequence of random variables $\{Y_m\}_{m=1}^\infty$ is a bounded sequence. By Helly's Theorem, it contains a weakly converging subsequence, i.e., there exists a $\{Y_{m_l}\}_{l=1}^\infty$ with $\lim_{l\to\infty} m_l = \infty$ and a random variable U

such that $Y_{m_l} \xrightarrow{d} U$. Since, for each k, n, the random variable $Y_m^k(1-Y_m)^{n-k}$ is uniformly bounded, we thus obtain by bounded convergence that

$$\lim_{m\to\infty} \mathbb{E}\big[Y_m^k(1-Y_m)^{n-k}\big] = \lim_{l\to\infty} \mathbb{E}\big[Y_{m_l}^k(1-Y_{m_l})^{n-k}\big] = \mathbb{E}\big[U^k(1-U)^{n-k}\big] \tag{11.1.10}$$

This completes the proof. □

We close this paragraph by discussing some consequences of De Finetti's Theorem. De Finetti's Theorem implies that when X_k and X_n are coordinates of an infinite exchangeable sequence of indicators, then they are *positively correlated*:

Exercise 11.2. *Let $\{X_i\}_{i=1}^\infty$ be an infinite sequence of exchangeable random variables. Prove that*

$$\mathbb{P}(X_k = X_n = 1) \geq \mathbb{P}(X_k = 1)\mathbb{P}(X_n = 1). \tag{11.1.11}$$

Prove that equality holds if and only if there exists a p such that $\mathbb{P}(U = p) = 1$.

By Exercise 11.2, an infinite sequence of exchangeable random variables is positively correlated. This also allows us to discuss examples of *finite* exchangeable sequences of random variables and their relation to De Finetti's Theorem (Theorem 11.1). Indeed, in the proof of De Finetti's Theorem, it is imperative that the sequence $\{X_i\}_{i=1}^\infty$ is infinite. This is not mere a technicality of the proof. Rather, there are finite exchangeable sequences of random variables for which the equality (11.1.1) does *not* hold. Indeed, take an urn filled with b blue and r red balls and draw balls successively without replacement. Let X_i denote the indicator that the i^{th} ball drawn is blue. Then, clearly, the sequence $\{X_i\}_{i=1}^{r+b}$ is exchangeable. However,

$$\mathbb{P}(X_1 = X_2 = 1) = \frac{b(b-1)}{(b+r)(b+r-1)} < \big(\frac{b}{b+r}\big)^2 = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1), \tag{11.1.12}$$

so that X_1 and X_2 are *negatively* correlated. As a result, (11.1.1) fails.

Polya urn schemes. An important application of De Finetti's Theorem (Theorem 11.1) arises in Polya urn schemes. An urn consists of a number of balls, and we successively draw balls and replace them in the urn. We start with $B_0 = b_0$ blue balls and $R_0 = r_0$ red balls at time $n = 0$. Let $W_b, W_r \colon \mathbb{N} \to (0,\infty)$ be two weight functions. Then, at time $n+1$, the probability of drawing a blue ball, conditionally on the number B_n of blue balls at time n, is proportional to the weight of the blue balls at time n, i.e., the probability of drawing a blue ball is equal to

$$\frac{W_b(B_n)}{W_b(B_n) + W_r(R_n)}. \tag{11.1.13}$$

After drawing a ball, it is replaced together with a *second* ball of the same color. We shall denote this Polya urn scheme by $\{(B_n, R_n)\}_{n=1}^\infty$. Naturally, we have that $B_n + R_n = b_0 + r_0 + n$.

In this section, we shall restrict to the case where there are $a_r, a_b \geq 0$ such that

$$W_b(k) = a_b + k, \qquad W_r(k) = a_r + k, \tag{11.1.14}$$

i.e., both weight functions are linear with the same slope, but possibly a different intercept. The main result concerning Polya urn schemes is the following theorem:

Theorem 11.2 (Limit theorem for linear Polya urn schemes)**.** *Let $\{(B_n, R_n)\}_{n=1}^{\infty}$ be a Polya urn scheme with linear weight functions W_b and W_r as in* (11.1.14)*. Then, as $n \to \infty$,*

$$\frac{B_n}{B_n + R_n} \xrightarrow{a.s.} U, \tag{11.1.15}$$

where U has a Beta-distribution with parameters $a = b_0 + a_b$ and $b = r_0 + a_r$, and

$$\mathbb{P}(B_n = B_0 + k) = \mathbb{E}\Big[\mathbb{P}\big(\mathrm{Bin}(n, U) = k\big)\Big]. \tag{11.1.16}$$

Proof of Theorem 11.2. Let X_n denote the indicator that the n^{th} ball drawn is blue. We first show that $\{X_n\}_{n=1}^{\infty}$ is an infinite exchangeable sequence. For this, we note that

$$B_n = b_0 + \sum_{j=1}^{n} X_j, \qquad R_n = r_0 + \sum_{j=1}^{n}(1 - X_j) = r_0 - b_0 + n - B_n. \tag{11.1.17}$$

Now, for any sequence $\{x_t\}_{t=1}^{n}$, we have that

$$\mathbb{P}\big(\{X_t\}_{t=1}^{n} = \{x_t\}_{t=1}^{n}\big) = \prod_{t=1}^{n} \frac{W_b(b_t)^{x_t} W_r(r_t)^{1-x_t}}{W_b(b_t) + W_r(r_t)}, \tag{11.1.18}$$

where $b_t = b_0 + \sum_{j=1}^{t} x_j$ and $r_t = R_0 - B_0 + t - b_t$. Denote $k = \sum_{t=1}^{n} x_t$. Then, by (11.1.14) and (11.1.17), we have that

$$\prod_{t=1}^{n}(W_b(b_t) + W_r(r_t)) = \prod_{t=0}^{n-1}(b_0 + r_0 + a_b + a_r + t), \tag{11.1.19}$$

while

$$\prod_{t=1}^{n} W_b(b_t)^{x_t} = \prod_{m=0}^{k-1}(b_0 + a_b + m), \qquad \prod_{t=1}^{n} W_r(r_t)^{1-x_t} = \prod_{j=0}^{n-k-1}(r_0 + a_r + j). \tag{11.1.20}$$

Thus, we arrive at

$$\mathbb{P}\big(\{X_t\}_{t=1}^{n} = \{x_t\}_{t=1}^{n}\big) = \frac{\prod_{m=0}^{k-1}(b+m)\prod_{j=0}^{n-k-1}(r+j)}{\prod_{t=0}^{n-1}(b+r+t)}, \tag{11.1.21}$$

where $b = b_0 + a_b$ and $r = r_0 + a_r$. In particular, (11.1.21) does not depend on the *order* in which the elements of $\{x_t\}_{t=1}^{n}$ appear, so that the sequence $\{X_n\}_{n=1}^{\infty}$ is an infinite exchangeable sequence. Thus, by De Finetti's Theorem (Theorem 11.1), the sequence $\{X_n\}_{n=1}^{\infty}$ is a mixture of Bernoulli random variables with a random success probability U, and we are left to compute the distribution of U. We also observe that the distribution of depends only on b_0, r_0, a_b, a_r through $b = b_0 + a_b$ and $r = r_0 + a_r$.

We next verify (11.1.4). For fixed $0 \le k \le n$, there are $\binom{n}{k}$ sequences of k ones and $n - k$ zeros. Each sequence has the same probability given by (11.1.21). Thus,

$$\begin{aligned}
\mathbb{P}(S_n = k) &= \binom{n}{k} \frac{\prod_{m=0}^{k-1}(b+m)\prod_{j=0}^{n-k-1}(r+j)}{\prod_{t=0}^{n-1}(b+r+t)} \\
&= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \times \frac{\Gamma(k+b)}{\Gamma(b)} \times \frac{\Gamma(n-k+r)}{\Gamma(r)} \times \frac{\Gamma(b+r)}{\Gamma(n+b+r)} \\
&= \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)} \times \frac{\Gamma(k+b)}{\Gamma(k+1)} \times \frac{\Gamma(n-k+r)}{\Gamma(n-k+1)} \times \frac{\Gamma(n+1)}{\Gamma(n+b+r)}.
\end{aligned} \tag{11.1.22}$$

For k and $n-k$ large, by (8.2.8),

$$\mathbb{P}(S_n = k) = \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)} \frac{k^{b-1}(n-k)^{r-1}}{n^{b+r-1}}(1+o(1)). \tag{11.1.23}$$

Taking $k = \lceil un \rceil$ (recall (11.1.4))

$$\lim_{n\to\infty} n\mathbb{P}(S_n = \lceil un \rceil) = \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)} u^{b-1}(1-u)^{r-1}, \tag{11.1.24}$$

which is the density of a Beta-distribution with parameters b and r. □

Applications to scale-free trees. We close this section by discussing applications of Polya urn schemes to scale-free trees. We start at time $t = 2$ with two vertices of which vertex 1 has degree d_1 and vertex 2 has degree d_2. After this, we successively attach vertices to older vertices with probability proportional to the degree plus $\delta > -1$. We do not allow for self-loops, so that indeed we obtain a tree. This is a generalization of $\{\mathrm{PA}_{1,\delta}(t)\}_{t=2}^{\infty}$, in which we are are more flexible in choosing the initial graph.

We now decompose the growing tree in two trees. For $i = 1, 2$, we let $T_i(t)$ be the tree of vertices which are closer to i than to $2-i$. Thus, the tree $T_2(t)$ consists of those vertices for which the path in the tree from the vertex to the root passes through vertex 2, and $T_1(t)$ consists of the remainder of the scale-free tree. Let $S_i(t) = |T_i(t)|$ denote the number of vertices in $T_i(t)$. Clearly, $S_1(t) + S_2(t) = t$, which is the total number of vertices of the tree at time t.

Theorem 11.3 (Tree decomposition for scale-free trees)**.** *As $t \to \infty$,*

$$\frac{S_1(t)}{S_1(t)+S_2(t)} \xrightarrow{a.s.} U, \tag{11.1.25}$$

where U has a Beta-distribution with parameters $a = \frac{d_1+\delta}{2+\delta}$ and $b = \frac{d_2+\delta}{2+\delta}$, and

$$\mathbb{P}(S_1(t) = k) = \mathbb{E}\Big[\mathbb{P}\big(\mathrm{Bin}(t-1, U) = k-1\big)\Big]. \tag{11.1.26}$$

Theorem 11.3 implies that we can decompose a scale-free tree into two disjoint scale-free trees, each of which contains a positive proportion of the vertices. This proportion converges almost surely to a Beta-distribution with parameters $a = \frac{d_1+\delta}{2+\delta}$ and $b = \frac{d_2+\delta}{2+\delta}$, which is strictly bounded away from 0 and 1.

Proof of Theorem 11.3. The evolution of $\{S_1(t)\}_{t=2}^{\infty}$ can be viewed as a Polya urn scheme. Indeed, when $S_1(t) = s_1(t)$, then the probability of attaching the $(t+1)^{\text{st}}$ vertex to $T_1(t)$ is equal to

$$\frac{(2s_1(t) + d_1 - 2) + \delta s_1(t)}{(2s_1(t) + d_1 - 2) + \delta s_1(t) + 2(s_2(t) + d_2) + \delta s_2(t)}, \tag{11.1.27}$$

since the number of vertices in $T_i(t)$ equals $S_i(t)$, while the total degree of $T_i(t)$ equals $(2S_i(t) + d_i - 2)$. We can rewrite this as

$$\frac{s_1(t) + \frac{d_1-2}{2+\delta}}{s_1(t) + s_2(t) + \frac{d_1+d_2-4}{2+\delta}}, \tag{11.1.28}$$

which is equal to (11.1.13) in the case (11.1.14) when $r_0 = b_0 = 1$ and $a_b = \frac{d_1-2}{2+\delta}, a_r = \frac{d_2-2}{2+\delta}$. Therefore, the proof of Theorem 11.3 follows directly from Theorem 11.2. □

Exercise 11.3. *A uniform recursive tree is obtained by starting with a single vertex, and successively attaching the* $(n+1)^{\text{st}}$ *vertex to a uniformly chosen vertex in* $\{1,\dots,n\}$*. Prove that for uniform recursive trees, the tree decomposition described above is such that*

$$\frac{S_1(n)}{S_1(n)+S_2(n)} \xrightarrow{a.s.} U, \tag{11.1.29}$$

where U *is uniform on* $[0,1]$*. Use this to prove that, for each* $k \in \{1,\dots,n\}$*,*

$$\mathbb{P}(S_1(n)=k) = \frac{1}{n}. \tag{11.1.30}$$

We continue by discussing an application of Polya urn schemes to the relative sizes of the initial degrees. For this, we fix an integer $k \geq 2$, and only regard times $t \geq k$ at which an edge is attached to one of the k initial vertices. We work with $\{\mathrm{PA}_{m,\delta}(t)\}_{t\geq 1}$, so that we start at time $t=1$ with one vertices with one self-loop, after which we successively attach vertices to older vertices with probability proportional to the degree plus $\delta > -1$, allowing for self-loops. The main result is as follows:

Theorem 11.4 (Relative degrees in scale-free trees)**.** *As* $t \to \infty$*,*

$$\frac{D_k(t)}{D_1(t)+\cdots+D_k(t)} \xrightarrow{a.s.} B_k, \tag{11.1.31}$$

where B_k *has a Beta-distribution with parameters* $a = 1+\delta$ *and* $b = (k-1)(2+\delta)$*.*

By Theorem 8.8, $B_k = \xi_k/(\xi_1+\cdots+\xi_k)$ (where we also use that these random variables are positive almost surely). Theorem 11.4 allows to identify properties of the law of the limiting degrees.

Proof of Theorem 11.4. Denote the sequence of stopping times $\{\tau_k(n)\}_{n=2k-1}^{\infty}$, by $\tau_k(2k-1) = k-1$, and

$$\tau_k(n) = \inf\{t : D_1(t)+\cdots+D_k(t) = n\}, \tag{11.1.32}$$

i.e., $\tau_k(n)$ is the time where the total degree of vertices $1,\dots,k$ equals n. Since $\mathbb{P}(\xi_j = 0) = 0$ for every j, $\tau_k(n) < \infty$ for every n. Moreover, since $\lim_{n\to\infty}\tau_k(n) = \infty$, we have that

$$\lim_{t\to\infty}\frac{D_k(t)}{D_1(t)+\cdots+D_k(t)} = \lim_{n\to\infty}\frac{D_k(\tau_k(n))}{D_1(\tau_k(n))+\cdots+D_k(\tau_k(n))} = \lim_{n\to\infty}\frac{D_k(\tau_k(n))}{n}. \tag{11.1.33}$$

Now, $\{(D_k(\tau_k(n)), D_1(\tau_k(n))+\cdots+D_{k-1}(\tau_k(n)))\}_{n=2k-1}^{\infty}$ is a Polya urn scheme, with $D_k(\tau_k(2k-1)) = 1$, and $D_1(\tau_k(2k-1))+\cdots+D_{k-1}(\tau_k(2k-1)) = 2k-2$. The edge at time $\tau_k(n)$ is attached to vertex k with probability

$$\frac{D_k(t)+\delta}{n+k\delta}, \tag{11.1.34}$$

which are the probabilities of a Polya urn scheme in the linear weight case in (11.1.14) when $a_b = \delta, a_r = (k-1)\delta$, $b_0 = 1$, $r_0 = 2(k-1)$. Thus, the statement follows from Theorem 11.2. □

11.2 Connectivity of preferential attachment models

In this section, we investigate the connectivity of $\{\mathrm{PA}_{m,\delta}(t)\}_{t\geq 1}$. We start by describing the connectivity when $m = 1$, which is special. For $m = 1$, we note that N_t, the number of connected components of $\mathrm{PA}_{1,\delta}(t)$, has distribution given by

$$N_t = I_1 + I_2 + \cdots + I_t, \tag{11.2.1}$$

where I_i is the indicator that the i^{th} edge connects to itself, so that $\{I_i\}_{i=1}^t$ are independent indicator variables with

$$\mathbb{P}(I_i = 1) = \frac{1+\delta}{(2+\delta)(i-1)+1+\delta}. \tag{11.2.2}$$

As a result, it is not hard to see that $N_t/\log t$ converges in probability to $(1+\delta)/(2+\delta) < 1$, so that **whp** there exists a largest connected component of size at least $t/\log t$. As a result, we see that $\mathrm{PA}_{1,\delta}(t)$ is not connected, but has few connected components which are almost all quite large. We shall not elaborate more on the connectivity properties when $m = 1$, but leave the asymptotics of the number of connected components as an exercise:

Exercise 11.4. *Show that the number of connected components N_t in $\mathrm{PA}_{1,\delta}(t)$ satisfies a central limit theorem with equal asymptotic mean and variance given by*

$$\mathbb{E}[N_t] = \frac{1+\delta}{2+\delta}\log t(1+o(1)), \qquad \mathrm{Var}(N_t) = \frac{1+\delta}{2+\delta}\log t(1+o(1)). \tag{11.2.3}$$

When $m \geq 2$, then the situation is entirely different. Indeed, in this case, $\mathrm{PA}_{m,\delta}(t)$ turns out to be *connected* with high probability:

Theorem 11.5 (Connectivity of $\mathrm{PA}_{m,\delta}(t)$ for $m \geq 2$)**.** *Fix $m \geq 2$. Then, with high probability, $\mathrm{PA}_{m,\delta}(t)$ is connected.*

Proof of Theorem 11.5. Again we let N_t denote the number of connected components of $\mathrm{PA}_{m,\delta}(t)$. We note that, $I_t = N_t - N_{t-1} = 1$ precisely when all m edges of vertex t are attached to vertex t. Thus,

$$\mathbb{P}(I_t = 1) = \prod_{e=1}^{m} \frac{2e-1+\delta}{(2m+\delta)t + (2e-1+\delta)}. \tag{11.2.4}$$

When $m \geq 2$, then

$$\sum_{t=2}^{\infty} \mathbb{P}(I_t = 1) < \infty, \tag{11.2.5}$$

so that, almost surely, $I_t = 1$ only occurs *finitely* often. As a result, since also $N_t \leq 1 + \sum_{t=2}^{\infty} I_t$, we have that $N_t < \infty$ almost surely. This implies that, for $m \geq 2$, $\mathrm{PA}_{m,\delta}(t)$ almost surely contains only finitely many connected components. However, $\mathrm{PA}_{m,\delta}(t)$ has a *positive* probability of being disconnected at a certain time $t \geq 2$ (see Exercise 11.5 below). We next extend the proof to show that, with probability converging to 1, $N_t = 1$. For this, we first fix $K \geq 1$ large. Then, with probability converging to 1 as $K \to \infty$, $\sum_{t=K}^{\infty} I_t = 0$. We condition on $\sum_{t=K}^{\infty} I_t = 0$, so that no new connected components are formed after time K, and we are left to prove that the vertices $1, \ldots, K$ are all connected in $\mathrm{PA}_{m,\delta}(t)$ for any $m \geq 2$. This proof proceeds in two steps. In the first step, we prove that with high probability, the number of vertices that are connected to $i \in \{1, \ldots, K\}$ is large when

$t \to \infty$. In the next step, we show that this implies that in fact *all* vertices in $\{1, \ldots, K\}$ are connected to i in $\mathrm{PA}_{m,\delta}(t)$.

To prove a lower bound on the number of vertices attached to vertex $i \in \{1, \ldots, K\}$, we lower bound this number by a Polya urn scheme. We denote $b_0 = 1$ and $r_0 = i$. We recursively define the Polya urn scheme $\{(B_n, R_n)\}_{n=1}^{\infty}$ as follows. We draw a ball, where the color of the ball is blue with probability proportional to the number of the blue balls. If we draw a blue ball, we replace it with an extra blue ball. If we draw a red ball, then we replace it with an extra red ball. Then, we claim that $C_i(t)$, the number of vertices connected to vertex i at time t, is bounded below by B_{t-i}. We shall prove this by induction on $t \geq i$. At time i, the number of vertices connected to vertex i is at least i (namely, vertex i itself), and this proves the claim since $B_0 = b_0 = 1$. Suppose that $C_i(t) \geq B_{t-i}$. Then, the total weight of the vertices connected to vertex i at time t is *at least* $B_{t-i}(2m+\delta)$, while the total weight is equal to $t(2m+\delta) + (1+\delta) \leq (t+1)(2m+\delta) = (2m+\delta)(B_{t-i} + R_{t-i})$, since $B_{t-i} + R_{t-i} = (t-i) + b_0 + r_0 = t+1$. Now, we say that at time $t+1$, we draw a red ball when the first edge of vertex $t+1$ is attached to a vertex which is connected to vertex i at time t. If this happens, then indeed vertex $t+1$ *will* be connected to vertex i, and $C_i(t)$ is increased by (at least) one. Thus, this advances the induction hypothesis, and proves the claim.

Now, by Theorem 11.2, we have that $C_i(t) \geq B_{t-i}$, and $B_t/t \xrightarrow{a.s.} U$, where U has a Beta-distribution with parameters $a = 1$ and $b = i$. Since $\mathbb{P}(U = 0) = 0$, we have that, with high probability, $C_i(t) \geq \varepsilon(t-i)$ for all t sufficiently large when $\varepsilon > 0$ is small. We conclude that, as $t \to \infty$,

$$\mathbb{P}(\liminf_{t\to\infty} C_i(t)/t \geq \varepsilon \ \ \forall i = 1, \ldots, K) = 1 - o(1) \tag{11.2.6}$$

as $\varepsilon \downarrow 0$. We claim that when $C_i(t)/t \geq \varepsilon$ and $C_j(t)/t \geq \varepsilon$ for $i \neq j$, then, with high probability, i is connected to j in $\mathrm{PA}_{m,\delta}(2t)$. Indeed, let $l \in \{t+1, \ldots, 2t\}$. Then, when the first edge of vertex l is attached to a vertex connected to vertex i, and the second to a vertex connected to vertex j, then i and j will be connected in $\mathrm{PA}_{m,\delta}(2t)$. In this case, we say that l is a *t-connector* for vertices i and j. Independently of the attachment of the edges of vertices $t+1, \ldots, l-1$, and conditionally on $\mathrm{PA}_{m,\delta}(2t)$, the probability that l is *t-connector* for vertices i and j is at least

$$\frac{C_i(t)(2m+\delta)}{l(2m+\delta)+1+\delta} \frac{C_j(t)(2m+\delta)}{l(2m+\delta)+2+\delta} \geq \frac{C_i(t)C_j(t)}{(2t+1)^2}. \tag{11.2.7}$$

When $C_i(t)/t \geq \varepsilon$ and $C_j(t)/t \geq \varepsilon$, this is at least $\varepsilon^2/16$. Thus, the probability that there is no t-connector for vertices i and j is at most

$$\left(1 - \frac{\varepsilon^2}{16}\right)^t \leq e^{-\frac{\varepsilon^2}{16}t}. \tag{11.2.8}$$

This tends to 0 exponentially when $t \to \infty$, so that we obtain that the probability that the probability that there exists $i, j \in \{1, \ldots, K\}$ that are not connected in $\mathrm{PA}_{m,\delta}(2t)$ tends to 0 as $t \to \infty$. This proves that $\mathrm{PA}_{m,\delta}(2t)$ is connected with high probability for t large, which implies Theorem 11.5. □

Exercise 11.5. *Compute the probability that $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ is connected for all times $t \geq 1$, and show that this probability is in $(0, 1)$.*

11.2.1 Small-world effect in preferential attachment models

In the next sections, we shall investigate distances in preferential attachment models. While the results are far from being as complete as the results for inhomogeneous random,

graphs or the configuration model discussed in Chapter 9 and 10, respectively, the available results are all consistent with the fact that distances in preferential attachment models have the same asymptotics as distances in the configuration model with the same degree sequence. This suggest a strong form of *universality*, which is interesting in its own right. For distances in preferential attachment models, we shall focus on the *diameter*, since results on the distances between a uniform pair of vertices have not yet been obtained. By Theorem 11.5, $\mathrm{PA}_{m,\delta}(t)$ is connected with high probability for $m \geq 2$. When $m = 1$, this is not true for $\mathrm{PA}_{m,\delta}(t)$, but $\mathrm{PA}^{(b)}_{1,\delta}(t)$ is, by definition, connected. Thus, we shall define the diameter of $\mathrm{PA}_{m,\delta}(t)$ to be the maximum graph distance between any pair of vertices that lie in the same connected component.

The results on distances are organized as follows. In Section 11.3, we state and prove logarithmic bounds on the diameter of $\mathrm{PA}_{m,\delta}(t)$ when $\delta \geq 0$ and when $m = 1$. In Section 11.4, we prove doubly logarithmic bounds on the diameter of $\mathrm{PA}_{m,\delta}(t)$ when $\delta < 0$ and $m \geq 2$.

11.3 Small-world effect in PA models: logarithmic bounds

In this section, we prove logarithmic bounds on distances in $\mathrm{PA}_{m,\delta}(t)$. In Section 11.3.1, we investigate the case where $m = 1$, in which case $\mathrm{PA}_{1,\delta}(t)$ consists of a collection of trees. The main results in this case are proved in [192]. In Section 11.3, we prove logarithmic bound on the diameter in $\mathrm{PA}^{(b)}_{m,\delta}(t)$ when $m \geq 2$.

11.3.1 Logarithmic distances in preferential attachment trees

When $m = 1$, one of the main results in [192] is as follows:

Theorem 11.6 (The diameter of scale-free trees)**.** *Fix $m = 1$ and $\delta > -1$. Let γ be the non-negative solution of*

$$\gamma + (1 + \delta)(1 + \log \gamma) = 0. \tag{11.3.1}$$

Then, for every $\varepsilon > 0$ and, with high probability, the diameter of $\mathrm{PA}_{1,\delta}(t)$ is in between $(\frac{2(1+\delta)}{(2+\delta)\gamma} - \varepsilon) \log t$ and $(\frac{2(1+\delta)}{(2+\delta)\gamma} + \varepsilon) \log t$.

In order to prove Theorem 11.6, it will be convenient to first consider a variation of the $\mathrm{PA}_{1,\delta}(t)$ model where there are no self-loops. Recall the growth rule in (8.1.2) and the discussion below it. The main result for $\{\mathrm{PA}^{(b)}_{1,\delta}(t)\}_{t=1}^{\infty}$ is as follows:

Theorem 11.7 (The diameter of scale-free trees $\mathrm{PA}^{(b)}_{1,\delta}(t)$)**.** *Fix $m = 1$ and $\delta > -1$, and let γ be the solution of* (11.3.1)*. Then, for every $\varepsilon > 0$ and with high probability, the diameter of $\mathrm{PA}^{(b)}_{1,\delta}(t)$ is in between $(\frac{2(1+\delta)}{(2+\delta)\gamma} - \varepsilon) \log t$ and $(\frac{2(1+\delta)}{(2+\delta)\gamma} + \varepsilon) \log t$.*

Theorem 11.6 implies that for $m = 1$, the diameter grows proportionally to the logarithm of the graph size. Scale-free trees have received substantial attention in the literature, we refer to [49, 192] and the references therein.

The proof of Theorem 11.6 relies on the fact that the case when $m = 1$ is special. In this case, $\mathrm{PA}_{1,\delta}(t)$ consists of a collection of trees each with precisely one self-loop. There is a close analogy between these trees and so-called *uniform recursive trees*. In uniform recursive trees, we grow a tree such that at time 1, we have a unique vertex called the root, with label 1. At time t, we add a vertex and connect it to a uniformly chosen vertex in the tree. See [204] for a survey of recursive trees.

A variant of a uniform recursive tree is the case where the probability that a newly added vertex is attached to a vertex is proportional to the degree of the vertices (and, for the root, the degree of the root plus one). This process is called a *random plane-oriented recursive tree*. For a uniform recursive tree of size t, it is proved by Pittel in [192] that the maximal distance between the root and any other vertex is with high probability equal to $\frac{1}{2\gamma}\log t(1+o(1))$, where γ satisfies (11.3.1) with $\delta = 0$. It is not hard to see that this implies that the maximal graph distance between any two vertices in the uniform recursive tree is equal to $\frac{1}{\gamma}\log t(1+o(1))$. We shall make use of similar arguments below to prove Theorem 11.6.

In order to prove Theorem 11.7, we make use of a result on the *height* of scale-free trees, which is proved by Pittel in [192]. The height of a tree is the maximal distance between any of the vertices of the tree to the root. For a tree T, we denote the height of T by $H(T)$. Then the asymptotics of the height of scale-free trees is as follows:

Theorem 11.8 (The height of scale-free trees)**.** *Fix $m = 1$ and $\delta > -1$, and let γ be the solution of* (11.3.1)*. Then,*

$$\frac{H\big(\mathrm{PA}_{1,\delta}^{(b)}(t)\big)}{\log t} \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}. \tag{11.3.2}$$

We shall only prove the upper bound in Theorem 11.8. The lower bound, which is somewhat more delicate, is proved in [192] using a result by Kingman in [145]. We remark that the almost sure limit of the height in Theorem 11.8 does not depend on the precise starting configuration of the graph $\mathrm{PA}_{1,\delta}^{(b)}(2)$. We shall crucially rely on this fact.

In the proof of the upper bound, we shall make use of the following result which computes the probability mass function of the distance between vertex v_s and the root v_1. Before stating the result, we need some more notation. We write $t \longrightarrow s$ when in $\{\mathrm{PA}_{1,\delta}^{(b)}(t)\}_{t\geq 1}$ one of the edges of vertex v_t is connected to vertex v_s. Note that for this to happen, we need that $t > s$. For $s_1 = t > s_2 > \cdots > s_k = 1$, and denoting $\vec{s}_k = (s_1, s_2, \ldots, s_k)$, we write

$$E_{\vec{s}_k} = \bigcap_{i=1}^{k-1} \{s_i \longrightarrow s_{i+1}\}. \tag{11.3.3}$$

For a configuration of $\mathrm{PA}_{1,\delta}^{(b)}(t)$, we let $\mathrm{dist}(v_t, v_1)$ denote the unique value of k such that $t = s_1 \longrightarrow s_2 \longrightarrow \cdots \longrightarrow s_{k-1} \longrightarrow s_k = 1$. Then the probability mass function of $\mathrm{dist}(v_t, v_1)$ can be identified as follows:

Proposition 11.9 (The distribution of $\mathrm{dist}(v_t, v_1)$ in $\mathrm{PA}_{1,\delta}^{(b)}(t)$)**.** *Fix $m = 1$ and $\delta > -1$. Then,*

$$\mathbb{P}(\mathit{dist}(v_t, v_1) = k) = \Big(\frac{1+\delta}{2+\delta}\Big)^{k-1} \frac{\Gamma(s+\frac{1}{2+\delta})}{\Gamma(1+\frac{1}{2+\delta})\Gamma(s+1)} \sum_{\vec{s}_k} \prod_{i=2}^{k-1} \frac{1}{s_i}, \tag{11.3.4}$$

where the sum is over ordered vectors $\vec{s}_k$ of length k, for which $s_1 = t$ and $s_k = 1$.

Proof of Proposition 11.7. We note that since the path between vertex v_t and v_1 is *unique*, we have

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) = \sum_{\vec{s}_k} \mathbb{P}\Big(\bigcap_{i=1}^{k-1} \{s_i \longrightarrow s_{i+1}\}\Big), \tag{11.3.5}$$

where the sum is over all *ordered* vectors $\vec{s}_k$ of length k, for which $s_1 = t$ and $s_k = 1$.

We claim that the events $\{s_i \longrightarrow s_{i+1}\}$ are *independent*, i.e.,

$$\mathbb{P}\Big(\bigcap_{i=1}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big) = \prod_{i=1}^{k-1}\mathbb{P}(s_i \longrightarrow s_{i+1}). \tag{11.3.6}$$

We prove this independence by induction. First note that

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{i=1}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big) &= \mathbb{E}\Big[\mathbb{P}\Big(\bigcap_{i=1}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big|\mathrm{PA}_{1,\delta}^{(b)}(s_1-1)\Big)\Big]\\ &= \mathbb{E}\Big[\mathbb{1}_{\{\bigcap_{i=2}^{k-1}\{s_i\longrightarrow s_{i+1}\}\}}\mathbb{P}\big(s_1 \longrightarrow s_2|\mathrm{PA}_{1,\delta}^{(b)}(s_1-1)\big)\Big],\end{aligned} \tag{11.3.7}$$

since the event $\bigcap_{i=2}^{l-1}\{s_i \longrightarrow s_{i+1}\}$ is measurable with respect to $\mathrm{PA}_{1,\delta}^{(b)}(s_1-1)$. Furthermore, from (8.1.2),

$$\mathbb{P}\big(s_1 \longrightarrow s_2|\mathrm{PA}_{1,\delta}^{(b)}(s_1-1)\big) = \frac{D_{s_2}(s_1-1)+\delta}{(2+\delta)(s_1-1)}. \tag{11.3.8}$$

In particular, we have that

$$\mathbb{P}\big(s_1 \longrightarrow s_2\big) = \mathbb{E}\Big[\frac{D_{s_2}(s_1-1)+\delta}{(2+\delta)(s_1-1)}\Big]. \tag{11.3.9}$$

Therefore,

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{i=1}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big) &= \mathbb{E}\Big[\mathbb{1}_{\{\bigcap_{i=2}^{k-1}\{s_i\longrightarrow s_{i+1}\}\}}\frac{D_{s_2}(s_1-1)+\delta}{(2+\delta)(s_1-1)}\Big]\\ &= \mathbb{P}\Big(\bigcap_{i=2}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big)\mathbb{E}\Big[\frac{D_{s_2}(s_1-1)+\delta}{(2+\delta)(s_1-1)}\Big],\end{aligned} \tag{11.3.10}$$

since the random variable $D_{s_2}(s_1-1)$ only depends on how many edges are connected to s_2 after time s_2, which is independent of the event $\bigcap_{i=2}^{l-1}\{s_i \longrightarrow s_{i+1}\}$, which depends on the attachment of the edges *up to and including* time s_2 only. We conclude that

$$\mathbb{P}\Big(\bigcap_{i=1}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big) = \mathbb{P}\big(\{s_1 \longrightarrow s_2\}\big)\mathbb{P}\Big(\bigcap_{i=2}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big). \tag{11.3.11}$$

Iteration leads to the independence claim in (11.3.6).

Combining (11.3.5) with (11.3.6) and (11.3.9), we obtain that

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) = \sum_{\vec{s}_k}\prod_{i=1}^{k-1}\mathbb{E}\Big[\frac{D_{s_{i+1}}(s_i-1)+\delta}{(2+\delta)(s_i-1)}\Big]. \tag{11.3.12}$$

By (8.2.12),

$$\mathbb{E}\Big[\frac{D_i(t)+\delta}{(2+\delta)t}\Big] = (1+\delta)\frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(i)}{(2+\delta)t\Gamma(t)\Gamma(i+\frac{1}{2+\delta})} = \frac{1+\delta}{2+\delta}\frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(i)}{\Gamma(t+1)\Gamma(i+\frac{1}{2+\delta})}, \tag{11.3.13}$$

so that

$$\begin{aligned}\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) &= \sum_{\vec{s}_k} \Big(\frac{1+\delta}{2+\delta}\Big)^{k-1} \prod_{i=1}^{k-1} \frac{\Gamma(s_i + \frac{1}{2+\delta})\Gamma(s_{i+1})}{\Gamma(s_i+1)\Gamma(s_{i+1}+\frac{1}{2+\delta})} \\ &= \Big(\frac{1+\delta}{2+\delta}\Big)^{k-1} \sum_{\vec{s}_k} \frac{\Gamma(s_1 + \frac{1}{2+\delta})\Gamma(s_k)}{\Gamma(s_k+\frac{1}{2+\delta})\Gamma(s_1)} \prod_{i=1}^{k-1} \frac{1}{s_i} \\ &= \Big(\frac{1+\delta}{2+\delta}\Big)^{k-1} \frac{\Gamma(t+\frac{1}{2+\delta})}{\Gamma(1+\frac{1}{2+\delta})\Gamma(t+1)} \sum_{\vec{s}_k} \prod_{i=2}^{k-1} \frac{1}{s_i} \end{aligned} \tag{11.3.14}$$

where we recall that the sum is over ordered vectors $\vec{s}_k$ of length k, for which $s_1 = t$ and $s_k = 1$. This completes the proof of Proposition 11.7.

Proof of the upper bound in Theorem 11.8. We first use Proposition 11.7 to prove that, almost surely,

$$\limsup_{t\to\infty} \frac{\mathrm{dist}(v_t, v_1)}{\log t} \le \frac{(1+\delta)}{(2+\delta)\gamma}. \tag{11.3.15}$$

To see (11.3.15), by (11.3.4) and by symmetry,

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) = \Big(\frac{1+\delta}{2+\delta}\Big)^{k-1} \frac{\Gamma(t+\frac{1}{2+\delta})}{\Gamma(1+\frac{1}{2+\delta})\Gamma(t+1)} \sum_{\vec{t}_{k-2}}^{*} \frac{1}{(k-2)!} \prod_{i=1}^{k-2} \frac{1}{t_i}, \tag{11.3.16}$$

where the sum now is over all vectors $\vec{t}_{k-2} = (t_1, \ldots, t_{k-2})$ with $1 < t_i < t$ with *distinct* coordinates. We can upper bound this sum by leaving out the restriction that the coordinates of $\vec{t}_{k-2}$ are distinct, so that

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) \le \Big(\frac{1+\delta}{2+\delta}\Big)^{k-1} \frac{\Gamma(t+\frac{1}{2+\delta})}{\Gamma(1+\frac{1}{2+\delta})\Gamma(t+1)} \frac{1}{(k-2)!} \Big(\sum_{s=2}^{t-1} \frac{1}{s}\Big)^{k-2}. \tag{11.3.17}$$

Since $x \mapsto 1/x$ is monotonically decreasing, we have that

$$\sum_{s=2}^{t-1} \frac{1}{s} \le \int_1^t \frac{1}{x} dx = \log t. \tag{11.3.18}$$

Also, we use (8.2.8) to bound, for some constant $C_\delta > 0$,

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) \le C_\delta t^{-\frac{1+\delta}{2+\delta}} \frac{\Big(\frac{1+\delta}{2+\delta}\log t\Big)^{k-2}}{(k-2)!}. \tag{11.3.19}$$

The solution γ to (11.3.1) is strictly in between $(0, e^{-1})$, so that, for any $k \ge \frac{(1+\delta)}{(2+\delta)\gamma}\log t + 2$, we have that

$$\frac{\frac{1+\delta}{2+\delta}\log t}{k} \le \gamma, \tag{11.3.20}$$

and we conclude that, for $k \geq \frac{(1+\delta)}{(2+\delta)\gamma}\log t + 1$,

$$\begin{aligned}\mathbb{P}(\mathrm{dist}(v_t, v_1) > k) &\leq C_\delta t^{-\frac{1+\delta}{2+\delta}} \frac{\left(\frac{1+\delta}{2+\delta}\log t\right)^{k-1}}{(k-1)!} \sum_{l=0}^{\infty} \frac{\left(\frac{1+\delta}{2+\delta}\log t\right)^{l}(k-1)!}{(k+l-1)!} \\ &\leq C_\delta t^{-\frac{1+\delta}{2+\delta}} \frac{\left(\frac{1+\delta}{2+\delta}\log t\right)^{k-1}}{(k-1)!} \sum_{l=0}^{\infty} \gamma^l = \frac{C_\delta}{1-\gamma} t^{-\frac{1+\delta}{2+\delta}} \frac{\left(\frac{1+\delta}{2+\delta}\log t\right)^{k-1}}{(k-1)!}.\end{aligned} \tag{11.3.21}$$

Now we fix $\varepsilon > 0$ and we take $k_t = \frac{(1+\delta)}{(2+\delta)}(\frac{1}{\gamma} + \varepsilon)\log t + 1$. We use

$$k! \geq k^k e^{-k}, \tag{11.3.22}$$

to arrive at

$$\begin{aligned}\mathbb{P}(\mathrm{dist}(v_t, v_1) > k_t) &\leq \frac{C_\delta}{1-\gamma} t^{-\frac{1+\delta}{2+\delta}} \Big(\frac{e(1+\delta)\log t}{(k_t-1)}\Big)^{k_t-1} = \frac{C_\delta}{1-\gamma} t^{-\frac{1+\delta}{2+\delta}} \big(\frac{e\gamma}{1+\gamma\varepsilon}\big)^{\frac{(1+\delta)}{(2+\delta)}(\frac{1}{\gamma}+\varepsilon)\log t} \\ &\leq \frac{C_\delta}{1-\gamma} t^{-\frac{1+\delta}{2+\delta}} \big(e\gamma\big)^{\frac{(1+\delta)}{(2+\delta)}(\frac{1}{\gamma}+\varepsilon)\log t} \\ &= \frac{C_\delta}{1-\gamma} t^{-\frac{1+\delta}{2+\delta}\left(1-(\frac{1}{\gamma}+\varepsilon)[1+\log\gamma]\right)} = \frac{C_\delta}{1-\gamma} t^{-p},\end{aligned} \tag{11.3.23}$$

where

$$p = \frac{1+\delta}{2+\delta}\Big(1 - (\frac{1}{\gamma}+\varepsilon)[1+\log\gamma]\Big). \tag{11.3.24}$$

Since $\gamma < e^{-1}$, $1 + \log\gamma < 0$, so that

$$p > \frac{1+\delta}{2+\delta}\big(1 - \frac{1}{\gamma}[1+\log\gamma]\big) = \frac{1+\delta}{2+\delta}\big(1 + \frac{1}{1+\delta}\big) = 1, \tag{11.3.25}$$

where the equality follows from (11.3.1). Thus, we conclude that, for $k_t = \frac{(1+\delta)}{(2+\delta)}(\frac{1}{\gamma} + \varepsilon)\log t + 1$ with $\varepsilon >$, we have

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) > k_t) = O(t^{-p}), \tag{11.3.26}$$

for some $p > 1$. As a result, by the Borel-Cantelli Lemma, the event $\{\mathrm{dist}(v_t, v_1) > k_t\}$ occurs only finitely often, and we conclude that (11.3.15) holds. We finally complete the proof of Theorem 11.8. For this, we note that

$$H\big(\mathrm{PA}^{(b)}_{1,\delta}(t)\big) = \max_{s=1}^{t} \mathrm{dist}(v_s, v_1), \tag{11.3.27}$$

so that when $\{\mathrm{dist}(v_t, v_1) > k_t\}$ occurs only finitely often, we also have that $\{H\big(\mathrm{PA}^{(b)}_{1,\delta}(t)\big) > k_t\}$ occurs only finitely often. Since $\varepsilon > 0$ was arbitrary, we obtain Theorem 11.8. □

We next prove Theorems 11.6–11.7:

Proof of Theorems 11.6 and 11.7. We first prove the upper bound on the diameter of $\mathrm{PA}^{(b)}_{1,\delta}(t)$ in Theorem 11.7, for which we use that

$$\mathrm{diam}(\mathrm{PA}^{(b)}_{1,\delta}(t)) \leq 2H\big(\mathrm{PA}^{(b)}_{1,\delta}(t)\big). \tag{11.3.28}$$

Equation (11.3.28) together with the upper bound in Theorem 11.8 imply that

$$\limsup_{t\to\infty} \frac{\mathrm{diam}(\mathrm{PA}^{(b)}_{1,\delta}(t))}{\log t} \leq \frac{2(1+\delta)}{\gamma(2+\delta)}. \tag{11.3.29}$$

For the lower bound, we make use of the lower bound on the diameter of $\mathrm{PA}^{(b)}_{1,\delta}(t)$ in Theorem 11.7, we make use of the decomposition of scale-free trees in Theorem 11.3, which states that the scale-free tree $\mathrm{PA}^{(b)}_{1,\delta}(t)$ can be decomposed into two scale-free trees, having a similar distribution as copies $\mathrm{PA}^{(b1)}_{1,\delta}(S_1(t))$ and $\mathrm{PA}^{(b2)}_{1,\delta}(t-S_1(t))$, where $\{\mathrm{PA}^{(b1)}_{1,\delta}(t)\}_{t=1}^{\infty}$ and $\{\mathrm{PA}^{(b2)}_{1,\delta}(t)\}_{t=1}^{\infty}$ are independent scale-free trees. By this tree decomposition, we have

$$\mathrm{diam}\big(\mathrm{PA}^{(b)}_{1,\delta}(t)\big) \geq H\big(\mathrm{PA}^{(b1)}_{1,\delta}(S_1(t))\big) + H\big(\mathrm{PA}^{(b2)}_{1,\delta}(t-S_1(t))\big). \tag{11.3.30}$$

The two trees $\{\mathrm{PA}^{(b1)}_{1,\delta}(t)\}_{t=1}^{\infty}$ and $\{\mathrm{PA}^{(b2)}_{1,\delta}(t)\}_{t=1}^{\infty}$ are not *exactly* equal in distribution to $\mathrm{PA}^{(b)}_{1,\delta}(t)$, because the initial degree of the starting vertices at time $t=2$ is different. However, the precise almost sure scaling in Theorem 11.3 does not depend sensitively on d_1 and d_2, and also the height of the scale-free tree in Theorem 11.8 does not depend on the starting graphs $\mathrm{PA}^{(b1)}_{1,\delta}(2)$ and $\mathrm{PA}^{(b1)}_{1,\delta}(2)$ (see the remark below Theorem 11.8). Since $S_1(t)/t \xrightarrow{a.s.} U$, with U having a Beta-distribution, $H\big(\mathrm{PA}^{(b1)}_{1,\delta}(S_1(t))\big)/\log t \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}$ and $H\big(\mathrm{PA}^{(b2)}_{1,\delta}(t-S_1(t))\big)/\log t \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}$, we conclude that

$$\limsup_{t\to\infty} \frac{\mathrm{diam}\big(\mathrm{PA}^{(b)}_{1,\delta}(t)\big)}{\log t} \geq \frac{2(1+\delta)}{(2+\delta)\gamma}. \tag{11.3.31}$$

Combining (11.3.29) and (11.3.31) proves Theorem 11.7.

To prove Theorem 11.6, we note that the connected components of $\mathrm{PA}_{1,\delta}(t)$ are similar in distribution to single scale-free tree $\mathrm{PA}^{(b1)}_{1,\delta}(t_1), \ldots, \mathrm{PA}^{(bN_t)}_{1,\delta}(t_{N_t})$, apart from the initial degree of the root. Since $N_t/\log t \xrightarrow{d} (1+\delta)/(2+\delta)$ (recall Exercise 11.4), we have that, with high probability, the largest connected component has size at least $\varepsilon t/\log t$. Since

$$\log\Big(\varepsilon t/\log t\Big) = \log t(1+o(1)), \tag{11.3.32}$$

the result follows along the same lines as in the proof of Theorem 11.7. □

We complete this section by proving that the logarithmic bounds on the diameter for $\{\mathrm{PA}^{(b)}_{1,\delta}(t)\}_{t=1}^{\infty}$ imply similar upper bounds for $\{\mathrm{PA}^{(b)}_{m,\delta}(t)\}_{t=1}^{\infty}$:

Theorem 11.10 (A $\log t$ upper bound for the diameter in $\{\mathrm{PA}^{(b)}_{m,\delta}(t)\}_{t=1}^{\infty}$)**.** *Fix $m \geq 1$ and $\delta > -m$. Then, with γ the solution of* (11.3.1)*,*

$$\mathbb{P}\Big(\mathrm{diam}(\mathrm{PA}^{(b)}_{m,\delta}(t)) > \big(\frac{2(m+\delta)}{(2m+\delta)\gamma} + \varepsilon\big)\log t\Big) = o(1), \quad t\to\infty. \tag{11.3.33}$$

Proof of Theorem 11.10. Since $\{\mathrm{PA}^{(b)}_{m,\delta}(t)\}_{t=1}^{\infty}$ is obtained from $\{\mathrm{PA}^{(b)}_{1,\delta/m}(mt)\}_{t=1}^{\infty}$ by collapsing m successive vertices, we have that $\mathrm{diam}(\mathrm{PA}^{(b)}_{m,\delta}(t)) \preceq \mathrm{diam}(\mathrm{PA}^{(b)}_{1,\delta/m}(mt))$, and the result follows from Theorem 11.7. □

11.3.2 Path counting in preferential attachment models

Denote by

$$\{g(t,j)=s\},\quad 1\le j\le m, \tag{11.3.34}$$

the event that at time t the j^{th} edge of vertex t is attached to the earlier vertex s. For the models (a) and (b), this event means that in $\{\mathrm{PA}_{1,\delta}(mt)\}$ the edge from vertex $m(t-1)+j$ is attached to one of the vertices $m(s-1)+1,\dots,ms$. It is a direct consequence of the definition of PA-models that the event (11.3.34) increases the preference for vertex s, and hence decreases (in a relative way) the preference for the vertices u, $1\le u\le t$, $u\neq s$. It should be intuitively clear that another way of expressing this effect is to say that, for different $s_1\neq s_2$, the events $\{g(t_1,j_1)=s_1\}$ and $\{g(t_2,j_2)=s_2\}$ are negatively correlated. In order to state such a result, we introduce some notation. For integer $n_s\ge 1$, we denote by

$$E_s=\bigcap_{i=1}^{n_s}\big\{g(t_i^{(s)},j_i^{(s)})=s\big\}, \tag{11.3.35}$$

the event that at time t_i the j_i^{th} edge of vertex t_i is attached to the earlier vertex s. We will start by proving that for each $k\ge 1$ and all possible choices of $t_i^{(s)},j_i^{(s)}$, the events E_s, for different s, are negatively correlated:

Lemma 11.11. *For distinct $s_1,s_2,\dots,s_k$,*

$$\mathbb{P}\Big(\bigcap_{i=1}^{k}E_{s_i}\Big)\le\prod_{i=1}^{k}\mathbb{P}(E_{s_i}). \tag{11.3.36}$$

Proof. We will use induction on the largest edge number present in the events E_s. Here, for an event $\{g(t,j)=s\}$, we let the edge number be $m(t-1)+j$, which is the order of the edge when we consider the edges as being attached in sequence. The induction hypothesis is that (11.3.36) holds for all k and all choices of $t_i^{(s)},j_i^{(s)}$ such that $\max_{i,s}m(t_i^{(s)}-1)+j_i^{(s)}\le e$, where induction is performed with respect to e. We now complete the induction argument. To initialize the induction, we note that for $e=1$, the induction hypothesis holds trivially, since $\bigcap_{i=1}^{k}E_{s_i}$ can be empty or consist of exactly one event, and in the latter case there is nothing to prove. This initializes the induction.

To advance the induction, we assume that (11.3.36) holds for all k and all choices of $t_i^{(s)},j_i^{(s)}$ such that $\max_{i,s}m(t_i^{(s)}-1)+j_i^{(s)}\le e-1$, and we extend it to all k and all choices of $t_i^{(s)},j_i^{(s)}$ such that $\max_{i,s}m(t_i^{(s)}-1)+j_i^{(s)}\le e$. Clearly, for k and $t_i^{(s)},j_i^{(s)}$ such that $\max_{i,s}m(t_i^{(s)}-1)+j_i^{(s)}\le e-1$, the bound follows from the induction hypothesis, so we may restrict attention to the case that $\max_{i,s}m(t_i^{(s)}-1)+j_i^{(s)}=e$. We note that there is a unique choice of t,j such that $m(t-1)+j=e$. In this case, there are again two possibilities. Either there is exactly one choice of s and $t_i^{(s)},j_i^{(s)}$ such that $t_i^{(s)}=t,j_i^{(s)}=j$, or there are at least *two* of such choices. In the latter case, we immediately have that $\bigcap_{s=1}^{k}E_s=\varnothing$, since the e^{th} edge can only be connected to a *unique* vertex. Hence, there is nothing to prove. Thus, we are left to investigate the case where there exists unique s and $t_i^{(s)},j_i^{(s)}$ such that $t_i^{(s)}=t,j_i^{(s)}=j$. Denote by

$$E_s'=\bigcap_{i=1:(t_i^{(s)},j_i^{(s)})\neq(t,j)}^{n_s}\big\{g(t_i^{(s)},j_i^{(s)})=s\big\}, \tag{11.3.37}$$

the restriction of E_s to the *other* edges. Then we can write

$$\bigcap_{i=1}^{k}E_{s_i}=\big\{g(t,j)=s\big\}\cap E_s'\cap\bigcap_{i=1:s_i\neq s}^{k}E_{s_i}. \tag{11.3.38}$$

By construction, all the edge numbers of the events in $E'_s \cap \bigcap_{i=1:s_i\neq s}^{k} E_{s_i}$ are at most $e-1$. Thus, we obtain

$$\mathbb{P}\Big(\bigcap_{i=1}^{k} E_{s_i}\Big) \leq \mathbb{E}\Big[\mathbb{1}_{E'_s\cap\bigcap_{i=1:s_i\neq s}^{k} E_{s_i}} \mathbb{P}_{e-1}(g(t,j)=s)\Big], \tag{11.3.39}$$

where $\mathbb{P}_{e-1}$ denotes the conditional probability given the edge attachments up to the $(e-1)^{\text{st}}$ edge connection.

We can compute

$$\mathbb{P}_{e-1}(g(t,j)=s) = \frac{D_s(t-1,e-1)+\delta}{(2m+\delta)(t-1)+(e-1)(2+\delta/m)+1+\delta}. \tag{11.3.40}$$

We wish to use the induction hypothesis. For this, we note that

$$D_s(t-1,e-1) = m + \sum_{(t',j'):t'\leq t-1} \mathbb{1}_{\{g(t',j')=s\}}. \tag{11.3.41}$$

We note that each of the terms in (11.3.41) has edge number strictly smaller than e and occurs with a non-negative multiplicative constant. As a result, we may use the induction hypothesis for each of these terms. Thus, we obtain, using also $m+\delta\geq 0$, that,

$$\begin{aligned}(2m+\delta)(t-1)\mathbb{P}\Big(\bigcap_{i=1}^{k} E_{s_i}\Big) \leq (m+\delta)\mathbb{P}(E'_s)\prod_{i=1:s_i\neq s}^{k}\mathbb{P}(E_{s_i}) \\ + \sum_{(t',j'):t'\leq t-1}\mathbb{P}(E'_s\cap\{g(t',j')=s\})\prod_{i=1:s_i\neq s}^{k}\mathbb{P}(E_{s_i}).\end{aligned} \tag{11.3.42}$$

We can recombine to obtain

$$\mathbb{P}\Big(\bigcap_{i=1}^{k} E_{s_i}\Big) \leq \mathbb{E}\Big[\mathbb{1}_{E'_s}\frac{D_s(t-1,e-1)+\delta}{(2m+\delta)(t-1)+(e-1)(2+\delta/m)+1+\delta}\Big]\prod_{i=1:s_i\neq s}^{k}\mathbb{P}(E_{s_i}), \tag{11.3.43}$$

and the advancement is completed when we note that

$$\mathbb{E}\Big[\mathbb{1}_{E'_s}\frac{D_s(t-1,e-1)+\delta}{(2m+\delta)(t-1)+(e-1)(2+\delta/m)+1+\delta}\Big] = \mathbb{P}(E_s). \tag{11.3.44}$$

□

We next study the probabilities of E_s when $n_s \leq 2$:

Lemma 11.12. *There exist absolute constants M_1, M_2, such that*
(i) for each $1\leq j\leq m$, and $t>s$,

$$\mathbb{P}\big(g(t,j)=s\big) \leq \frac{M_1}{t^{1-\gamma}s^{\gamma}}, \tag{11.3.45}$$

and
(ii) for $t_2>t_1>s$, and any $1\leq j_1,j_2\leq m$,

$$\mathbb{P}\Big(g(t_1,j_1)=s, g(t_2,j_2)=s\Big) \leq \frac{M_2}{(t_1t_2)^{1-\gamma}s^{2\gamma}}, \tag{11.3.46}$$

where we abbreviate $\gamma=\frac{m}{2m+\delta}$.

Proof. We prove (11.3.45) and (11.3.46) only for $m = 1$, for which we abbreviate $g(t) = g(t, 1)$. By the definition of $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ in terms of $\{\mathrm{PA}_{1,\delta/m}(t)\}_{t=1}^{\infty}$, this implies the result for general $m \geq 1$. Note, in particular, that $g(t, j) = s$ for $m \geq 1$ in $\{\mathrm{PA}_{m,\delta}(t)\}_{t=1}^{\infty}$ is equivalent to $g(m(t-1)+j) \in \{m(s-1)+1, \ldots, ms\}$ in $\{\mathrm{PA}_{1,\delta/m}(t)\}_{t=1}^{\infty}$.

For (11.3.45), we use Theorem 8.1 to compute

$$\begin{aligned}\mathbb{P}\big(g(t)=s\big) &= \mathbb{E}\big[\mathbb{E}[\mathbb{1}_{\{g(t)=s\}} \mid \mathrm{PA}_{1,\delta}(t-1)]\big] = \mathbb{E}\big[\frac{D_s(t-1)+\delta}{(2+\delta)(t-1)+1+\delta}\big] \\ &= (1+\delta)\frac{\Gamma(t)\Gamma(i+\frac{1+\delta}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})\Gamma(i+1)}, \end{aligned} \tag{11.3.47}$$

so that the result follows from (8.2.8).

We proceed with the proof of (11.3.46) for $m = 1$ by computing

$$\begin{aligned}\mathbb{P}\Big(g(t_1)=s, g(t_2)=s\Big) &= \mathbb{E}\Big[\mathbb{P}\big(g(t_1)=s, g(t_2)=s\big|\mathrm{PA}_{m,\delta}(t_2-1)\big)\Big] \\ &= \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(\frac{D_s(t_2-1)+\delta}{(t_2-1)(2+\delta)+1+\delta}\right)\right] \\ &= \frac{1}{(t_2-1)(2+\delta)+1+\delta}\frac{\Gamma(t_2)\Gamma(t_1+\frac{1+\delta}{2+\delta})}{\Gamma(t_2-1+\frac{1+\delta}{2+\delta})\Gamma(t_1+1)}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right] \\ &= \frac{\Gamma(t_2)\Gamma(t_1+\frac{1+\delta}{2+\delta})}{\Gamma(t_2+\frac{1+\delta}{2+\delta})\Gamma(t_1+1)}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right] \\ &\leq \frac{M_3}{t_1^{\gamma}t_2^{1-\gamma}}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right], \end{aligned} \tag{11.3.48}$$

where we use, for $t_1 < u \leq t_2 - 1$, the iteration

$$\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(u)+\delta\right)\right] = \Big(1+\frac{1}{(2+\delta)(u-1)+1+\delta}\Big)\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(u-1)+\delta\right)\right]. \tag{11.3.49}$$

We are lead to compute $\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(d_s(t_1)+\delta\right)\right]$. We do so by recursion as follows

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\big|\mathrm{PA}_{m,\delta}(t_1-1)\right] &= \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)-D_s(t_1-1)\right)\big|\mathrm{PA}_{m,\delta}(t_1-1)\right] \\ &\quad + \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1-1)+\delta\right)\big|\mathrm{PA}_{m,\delta}(t_1-1)\right] \\ &= \frac{(D_s(t_1-1)+\delta)(D_s(t_1-1)+1+\delta)}{(t_1-1)(2+\delta)+1+\delta}. \end{aligned} \tag{11.3.50}$$

By Proposition 8.9, there exists a uniform constant M_4 such that

$$\mathbb{E}[(D_s(t)+\delta)(D_s(t)+1+\delta)] = \frac{2}{c_2(t)}\mathbb{E}[Z_{s,2}(t)] \leq M_4(t/s)^{2a}. \tag{11.3.51}$$

Combining (11.3.48), (11.3.50) and (11.3.51), we find

$$\mathbb{P}\big(g(t_1)=s, g(t_2)=s\big) \leq \left(\frac{M_3}{t_1^{\gamma}t_2^{1-\gamma}}\right)\left[\frac{M_4}{(2+\delta)(t_1-1)+1+\delta}\Big(\frac{t_1-1}{s}\Big)^{2\gamma}\right] \leq \frac{M_2}{(t_1t_2)^{1-\gamma}s^{2\gamma}}, \tag{11.3.52}$$

for some M_2, as required. □

We combine the results of Lemmas 11.11 and 11.12 into the following corollary, yielding an upper bound for the probability of the existence of a path. In its statement, we call a path $\Gamma = (s_0, s_1, \dots, s_l)$ self-avoiding when $s_i \neq s_j$ for all $1 \leq i < j \leq l$.

Corollary 11.13. *Let $\Gamma = (s_0, s_1, \dots, s_l)$ be a self-avoiding path of length l consisting of the $l+1$ unordered vertices $s_0, s_1, \dots, s_l$, then*

$$\mathbb{P}(\Gamma \in \mathrm{PA}_{m,\delta}(t)) \leq (m^2 C)^l \prod_{i=0}^{l-1} \frac{1}{(s_i \wedge s_{i+1})^{\gamma} (s_i \vee s_{i+1})^{1-\gamma}}. \tag{11.3.53}$$

Proof. Since Γ is self-avoiding, we can write $\{\Gamma \in G\} = \cap_{i=1}^{k} E_{s_i}$, where either

$$E_s = \{g(t, j) = s\} \tag{11.3.54}$$

for some $t > s$ and some $1 \leq j \leq m$, or

$$E_s = \{g(t_1, j_1) = g(t_2, j_2) = s\}, \tag{11.3.55}$$

for some $t_1, t_2 > s$ and some $1 \leq j_1, j_2 \leq m$. In the first case we have according to (11.3.45),

$$\mathbb{P}(E_s) = \mathbb{P}\big(g(t, j) = s\big) \leq \frac{M_1}{t^{1-\gamma} s^{\gamma}}, \tag{11.3.56}$$

whereas in the second case, according to (11.3.46),

$$\mathbb{P}(E_s) = \mathbb{P}(g(t_1, j_1) = s, g(t_2, j_2) = s) \leq \frac{M_2}{(t_1 t_2)^{1-\gamma} s^{2\gamma}} = \frac{M_2}{t_1^{1-\gamma} s^{\gamma} t_2^{1-\gamma} s^{\gamma}}. \tag{11.3.57}$$

In both cases M_i, $i = 1, 2$, is an absolute constant. Lemma 11.11 then yields (11.3.53), where the factor m^{2l} originates from the choices of $j \in \{1, 2, \dots, m\}$. □

Observe that for integers $j > i$ we have

$$\left(\frac{j}{i}\right)^{\gamma - \frac{1}{2}} \leq 1, \tag{11.3.58}$$

if and only if $\gamma \leq \frac{1}{2}$, which happens if and only if $\delta \geq 0$, because $\gamma = \frac{m}{2m+\delta}$. Hence, $j^{\gamma-1} i^{-\gamma} \leq (ij)^{-1/2}$ precisely when $\delta \geq 0$. It now follows from (11.3.53) and the above inequality that for $\delta \geq 0$,

$$\mathbb{P}(\Gamma \in \mathrm{PA}_{m,\delta}(t)) \leq (m^2 C)^l \prod_{i=0}^{l-1} \frac{1}{\sqrt{s_i s_{i+1}}}. \tag{11.3.59}$$

11.3.3 Logarithmic lower bounds on distances for $\delta > 0$

We next improve the bound in the previous section in the case when $\delta > 0$, in which case $\gamma = m/(2m + \delta) < 1/2$. From the above discussion, we conclude that

$$\mathbb{P}\Big(\mathrm{dist}_{G_{m,\delta}(t)}(1, t) = k\Big) \leq c^k \sum_{\vec{s}} \prod_{j=0}^{k-1} \frac{1}{(s_j \wedge s_{j+1})^{\gamma} (s_j \vee s_{j+1})^{1-\gamma}}, \tag{11.3.60}$$

where $c = m^2 C$, and where the sum is over $\vec{s} = (s_0, \dots, s_k)$ with $s_k = t$, $s_0 = 1$, $s_l \geq 1$ for all $l = 1, \dots, k-1$ and $s_l \neq s_n$ for all $l \neq n$, since we may assume that our path $(s_0, \dots, s_k)$ is self-avoiding. Define

$$f_k(i, t) = \sum_{\vec{s}} \prod_{j=0}^{k-1} \frac{1}{(s_j \wedge s_{j+1})^{\gamma} (s_j \vee s_{j+1})^{1-\gamma}}, \tag{11.3.61}$$

where now the sum is over $\vec{s} = (s_0, \dots, s_k)$ with $s_k = t, s_0 = i$, $s_l \geq 1$ for all $l = 1, \dots, k-1$ and $s_l \neq s_n$ for all $l \neq n$, so that

$$\mathbb{P}\Big(\mathrm{dist}_{G_{m,\delta(t)}}(i,t) = k\Big) \leq c^k f_k(i,t). \tag{11.3.62}$$

We study the function $f_k(i,t)$ in the following lemma:

Lemma 11.14 (A bound on f_k)**.** *Fix $\gamma < 1/2$. Then, for every $b > \gamma$ such that $\gamma + b < 1$, there exists a $C_{\gamma,b} > 0$ such that, for every $1 \leq i < t$ and all $k \geq 1$,*

$$f_k(i,t) \leq \frac{C_{\gamma,b}^k}{i^b t^{1-b}}. \tag{11.3.63}$$

Proof. We prove the lemma using induction on $k \geq 1$. To initialize the induction hypothesis, we note that, for $1 \leq i < t$ and every $b \geq a$,

$$f_1(i,t) = \frac{1}{(i \wedge t)^\gamma (i \vee t)^{1-\gamma}} = \frac{1}{i^\gamma t^{1-\gamma}} = \frac{1}{t}\Big(\frac{t}{i}\Big)^\gamma \leq \frac{1}{t}\Big(\frac{t}{i}\Big)^b = \frac{1}{i^b t^{1-b}}. \tag{11.3.64}$$

This initializes the induction hypothesis as long as $C_{\gamma,b} \geq 1$.

To advance the induction hypothesis, note that we have the recursion relation

$$f_k(i,t) \leq \sum_{s=1}^{i-1} \frac{1}{s^\gamma i^{1-\gamma}} f_{k-1}(s,t) + \sum_{s=i+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} f_{k-1}(s,t). \tag{11.3.65}$$

We now bound each of these two contributions, making use of the induction hypothesis. For the first sum, we bound

$$\sum_{s=1}^{i-1} \frac{1}{s^\gamma i^{1-\gamma}} f_{k-1}(s,t) \leq C_{\gamma,b}^{k-1} \sum_{s=1}^{i-1} \frac{1}{s^\gamma i^{1-\gamma}} \frac{1}{s^b t^{1-b}} = \frac{C_{\gamma,b}^{k-1}}{i^{1-\gamma} t^{1-b}} \sum_{s=1}^{i-1} \frac{1}{s^{\gamma+b}} \leq \frac{1}{1-\gamma-b} \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}}, \tag{11.3.66}$$

since $\gamma + b < 1$. For the second sum, we bound

$$\begin{aligned} \sum_{s=i+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} f_{k-1}(s,t) &\leq C_{\gamma,b}^{k-1} \sum_{s=i+1}^{t-1} \frac{1}{i^\gamma s^{1-\gamma}} \frac{1}{s^b t^{1-b}} + C_{\gamma,b}^{k-1} \sum_{s=t+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} \frac{1}{t^b s^{1-b}} \qquad (11.3.67) \\ &= \frac{C_{\gamma,b}^{k-1}}{i^\gamma t^{1-b}} \sum_{s=i+1}^{t-1} \frac{1}{s^{1-\gamma+b}} + \frac{C_{\gamma,b}^{k-1}}{i^\gamma t^b} \sum_{s=t+1}^{\infty} \frac{1}{s^{2-\gamma-b}} \leq \frac{1}{b-\gamma} \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}} + \frac{1}{1-\gamma-b} \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}}, \end{aligned}$$

since $1 + b - a > 1$, $2 - \gamma - b > 1, b > \gamma$ and $(t/i)^\gamma \leq (t/i)^b$. We conclude that

$$f_k(i,t) \leq \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}} \Big(\frac{1}{b-\gamma} + \frac{2}{1-\gamma-b}\Big) \leq \frac{C_{\gamma,b}^k}{i^b t^{1-b}}, \tag{11.3.68}$$

when

$$C_{\gamma,b} = \frac{1}{b-\gamma} + \frac{2}{1-\gamma-b} \geq 1. \tag{11.3.69}$$

This advances the induction hypothesis, and completes the proof. □

Using Lemma 11.14 and (11.3.62), we obtain that

$$\mathbb{P}\Big(\mathrm{dist}_{G_{m,\delta(t)}}(1,t) = k\Big) \leq \frac{(cC_{\gamma,b})^k}{t^{1-b}}. \tag{11.3.70}$$

As a result, we obtain that

$$\mathbb{P}\Big(\mathrm{diam}(G_{m,\delta}(t)) \le k\Big) \le \mathbb{P}\Big(\mathrm{dist}_{G_{m,\delta}(t)}(1,t) \le k\Big) \le \frac{(cC_{\gamma,b})^{k+1}}{t^{1-b}(cC_{\gamma,b}-1)} = o(1), \quad (11.3.71)$$

whenever $k \le \frac{1-b}{\log(cC_{\gamma,b})}\log t$. We conclude that there exists $c_2 = c_2(m,\delta)$ such that, with high probability $\mathrm{diam}\Big(G_{m,\delta}(t)\Big) \ge c_2 \log t$. □

We next extend the above discussion to *typical* distances.

Lemma 11.15 (Typical distances for $\delta > 0$)**.** *Fix $m \ge 1$ and $\delta > 0$. Let $H_t = \mathrm{dist}_t(A_1, A_2)$ be the distance between two uniformly chosen vertices. Then, for $c_2 = c_2(m,\delta) > 0$ sufficiently small,* **whp**, *$H_t \ge c_2 \log t$.*

Proof. For $c_2 = c_2(m,\delta) > 0$, define

$$B_t \equiv \#\big\{i,j \in [t] : i < j : \mathrm{dist}_{G_{m,\delta}(t)}(i,j) \le c_2 \log t\big\}, \quad (11.3.72)$$

where $\#\{A\}$ denotes the cardinality of A.

By Lemma 11.14, with $K = \log(cC_{a,b} \vee 2)$ and $\gamma < b < 1-\gamma$, and for all $1 \le i < j \le t$,

$$\mathbb{P}\Big(\mathrm{dist}_{G_{m,\delta}(t)}(i,j) = k\Big) \le c^k f_k(i,j) \le \frac{e^{Kk}}{i^b j^{1-b}}. \quad (11.3.73)$$

As a result, we obtain that

$$\mathbb{P}\Big(\mathrm{dist}_{G_{m,\delta}(t)}(i,j) \le c_2 \log t\Big) \le \frac{t^{Kc_2}}{i^b j^{1-b}} \frac{e^K}{e^K - 1}, \quad (11.3.74)$$

and thus, using also $\sum_{i=1}^{j-1} i^{-b} \le j^{1-b}/(1-b)$,

$$\mathbb{E}[B_t] \le O(1) \sum_{1 \le i < j \le t} \frac{t^{Kc_2}}{i^b j^{1-b}} = O(t^{Kc_2+1}). \quad (11.3.75)$$

It now suffices to note that

$$\mathbb{P}(H_t \le c_2 \log t) = \mathbb{E}\big[I[\mathrm{dist}_{G_{m,\delta}(t)}(A_1,A_2) \le c_2 \log t]\big] = \frac{2\mathbb{E}[B_t] + t}{t^2} = o(1), \quad (11.3.76)$$

by (11.3.75), for every $c_2 > 0$ such that $Kc_2 + 1 < 2$. □

d referenced equation?

Note that (**??**) is also a lower bound on typical distances in case $\delta = 0$, which can be proven as above.

11.3.4 Logarithmic bounds on distances when $m \ge 2$

In this section, we investigate logarithmic bounds on the distances in $\mathrm{PA}_{m,\delta}(t)$ when $m \ge 2$.

Theorem 11.16 (Diameter of $\mathrm{PA}_{m,\delta}(t)$)**.** *Fix $m \ge 2$, $\delta \ge 0$ and let $\varepsilon > 0$ be fixed. Then, with high probability, the largest distance between two vertices in the same component of $\mathrm{PA}_{m,\delta}(t)$ is bounded below by $(1-\varepsilon)\frac{\log t}{\log\log t}$, i.e.,*

$$\mathrm{diam}(\mathrm{PA}_{m,\delta}(t)) \ge (1-\varepsilon)\frac{\log t}{\log\log t}. \quad (11.3.77)$$

The proof of Theorem 11.16 proceeds in two steps. The first step is to use the collapsing of vertices to obtain

$$\mathrm{PA}_{m,\delta}(t)$$

from $\mathrm{PA}_{1,\delta/m}(mt)$, so that we fix $m = 1$ for the time being. For the lower bound, it is proved that the probability that vertex j sends it (unique) edge to vertex i is $O\Big((ij)^{-1/2}\Big)$. This suggests that the model is related to the Chung-Lu model with $w_i = ci^{-1/2}n^{-1/4}$ and an appropriately chosen constant c. However, the edge statuses are *not* independent. A second result says that the probability that j and k both send their edges to i is $O\Big((ij)^{-1/2}(ik)^{-1/2}\Big)$, so that a weak form of independence is valid.

We finally state the complementary result from [48], which states that, when $\delta = 0$, the diameter of $\mathrm{PA}_{m,\delta}(t)$ is asymptotically equal to $\frac{\log t}{\log\log t}(1 + o(1))$:

Theorem 11.17 (Diameter of $\mathrm{PA}_{m,0}(t)$)**.** *Fix $m \geq 2$, $\delta = 0$ and let $\varepsilon > 0$ be fixed. Then, with high probability, the largest distance between two vertices in the same component in $\mathrm{PA}_{m,\delta}(t)$ is bounded above by*

$$\mathrm{diam}(\mathrm{PA}_{m,0}(t)) \geq (1-\varepsilon)\frac{\log t}{\log\log t}. \tag{11.3.78}$$

The proof of Theorem 11.17 is given in [48] and is considerably more involved than the one of Theorem 11.16, and will be omitted here. We now give the proof of Theorem 11.16.

Proof of Theorem 11.16. We prove that (11.3.59) implies that for $\delta \geq 0$,

$$L = \frac{\log(t-1)}{\log(3Cm^2\log t)} \tag{11.3.79}$$

is a lower bound for the diameter of $\mathrm{PA}_{m,\delta}(t)$. Consider a path of length l consisting of the vertices $v_0, v_1 \ldots, v_l$, then (11.3.59) implies that

$$\mathbb{P}(\Gamma \in \mathrm{PA}_{m,\delta}(t)) \leq (Cm^2)^l \prod_{k=0}^{l-1} \frac{1}{\sqrt{v_k v_{k+1}}} = \frac{(Cm^2)^l}{\sqrt{v_0 v_l}} \prod_{k=1}^{l-1} \frac{1}{v_k}. \tag{11.3.80}$$

Thus, the expected number of paths of length l between $v_0 = t$ and $v_l = t-1$ is bounded by

$$\begin{aligned}\frac{(Cm^2)^l}{\sqrt{t(t-1)}} \sum_{1\leq v_1,\ldots,v_{l-1}\leq t-2} \prod_{k=1}^{l-1} \frac{1}{v_k} &= \frac{(Cm^2)^l}{\sqrt{t(t-1)}} \left(\sum_{1}^{t-2} \frac{1}{v_k}\right)^{l-1} \\ &\leq \frac{(Cm^2)^l}{t-1}(\log t)^{l-1} \leq (1/2)^l (\log t)^{-1} \to 0\end{aligned}$$

precisely when $(2Cm^2\log t)^l \leq t-1$, or, equivalently,

$$l \leq \frac{\log(t-1)}{\log((2Cm^2)\log t)}. \tag{11.3.81}$$

Equality in (11.3.81) holds for L in (11.3.79). This implies that the diameter is at least L in (11.3.79), and completes the proof of Theorem 11.16. □

11.4 Small-world effect in PA models: doubly logarithmic bounds

In this section, we investigate $\mathrm{PA}_{m,\delta}(t)$ with $m \geq 2$ and $\delta \in (-m, 0)$. Since the degree power-law exponent equals $\tau = 3 + \delta/m$, such graphs are such that $\tau \in (2, 3)$. We know, for $\mathrm{CM}_n(\boldsymbol{D})$, that when $\tau \in (2, 3)$, the graph distances grow doubly logarithmically in the size of the graph. We now state and prove a result from [**?**] that proves a similar result for $\mathrm{PA}_{m,\delta}(t)$ when $m \geq 2$:

Theorem 11.18 (A $\log\log t$ upper bound on the diameter for $\delta < 0$)**.** *Fix $m \geq 2$ and assume that $\delta \in (-m, 0)$. Then, for every $\sigma > \frac{1}{3-\tau}$ and with*

$$C_G = \frac{4}{|\log(\tau - 2)|} + \frac{4\sigma}{\log m}. \tag{11.4.1}$$

the diameter of $\mathrm{PA}_{m,\delta}(t)$ is, **whp**, *bounded above by $C_G \log\log t$, as $t \to \infty$.*

The proof of Theorem 11.18 is divided into two key steps. In the first, in Theorem 11.19, we give a bound on the diameter of the *core* which consists of the vertices with degree at least a certain power of $\log t$. This argument is close in spirit to the argument in [195] used to prove bounds on the average distance for the configuration model, but substantial adaptations are necessary to deal with preferential attachment. After this, in Theorem 11.23, we derive a bound on the distance between vertices with a small degree and the core. We start by defining and investigating the core of the preferential attachment model. In the sequel, it will be convenient to prove Theorem 11.18 for $2t$ rather than for t. Clearly, this does not make any difference for the results.

11.4.1 The diameter of the core

We adapt the proof of Theorem **??** to $\mathrm{PA}_{m,\delta}(t)$. We recall that

$$\tau = 3 + \frac{\delta}{m}, \tag{11.4.2}$$

so that $-m < \delta < 0$ corresponds to $\tau \in (2, 3)$. Throughout this section, we fix $m \geq 2$.

We take $\sigma > \frac{1}{3-\tau} = -\frac{m}{\delta} > 1$ and define the *core* Core_t of the PA-model $\mathrm{PA}_{m,\delta}(2t)$ to be

$$\mathrm{Core}_t = \big\{ i \in \{1, 2, \ldots, t\} : D_i(t) \geq (\log t)^\sigma \big\}, \tag{11.4.3}$$

i.e., all the vertices which at time t have degree at least $(\log t)^\sigma$.

For a graph G with vertex set $\{1, 2, \ldots, t\}$ and a given edge set, we write $d_G(i, j)$ for the shortest-path distance between i and j in the graph G. Also, for $A \subseteq \{1, 2, \ldots, t\}$, we write

$$\mathrm{diam}_t(A) = \max_{i,j \in A} d_{\mathrm{PA}_{m,\delta(t)}}(i, j). \tag{11.4.4}$$

Then, the diameter of the core in the graph $\mathrm{PA}_{m,\delta}(2t)$, which we denote by $\mathrm{diam}_{2t}(\mathrm{Core}_t)$, is bounded in the following theorem:

Theorem 11.19 (The diameter of the core)**.** *Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$,* **whp**,

$$\mathrm{diam}_{2t}(\mathrm{Core}_t) \leq (1 + o(1)) \frac{4 \log\log t}{|\log(\tau - 2)|}. \tag{11.4.5}$$

The proof of Theorem 11.19 is divided into several smaller steps. We start by proving that the diameter of the *inner core* Inner_t, which is defined by

$$\mathrm{Inner}_t = \big\{ i \in \{1,2,\ldots,t\} : D_i(t) \geq t^{\frac{1}{2(\tau-1)}} (\log t)^{-\frac{1}{2}} \big\}, \tag{11.4.6}$$

is, **whp** bounded by $C_\delta < \infty$. After this, we will show that the distance from the *outer core*, which is defined to be equal to $\mathrm{Outer}_t = \mathrm{Core}_n \backslash \mathrm{Inner}_t$, to the inner core can be bounded by a fixed constant times $\log\log t$. This also shows that the diameter of the outer core is bounded by a different constant times $\log\log t$. We now give the details.

Proposition 11.20 (The diameter of the inner core)**.** *Fix $m \geq 2$ and $\delta \in (-m, 0)$. Then,* **whp**,

$$\mathrm{diam}_{2t}(\mathrm{Inner}_t) < C_\delta. \tag{11.4.7}$$

Proof. We first introduce the important notion of a t-connector between a vertex $i \in \{1,2,\ldots,t\}$ and a set of vertices $A \subseteq \{1,2,\ldots,t\}$, which plays a crucial role throughout the proof. Fix a set of vertices A and a vertex i. We say that the vertex $j \in \{t+1, t+2, \ldots, 2t\}$ is a *t-connector* between i and A if one of the edges incident to j connects to i and another edge incident to j connects to a vertex in A. Thus, when there exists a t-connector between i and A, the distance between i and A in $\mathrm{PA}_{m,\delta}(2t)$ is at most 2.

We note that for a set of vertices A and a vertex i with degree at time t equal to $D_i(t)$, we have that, conditionally on $\mathrm{PA}_{m,\delta}(t)$, the probability that $j \in \{t+1, t+2, \ldots, 2t\}$ is a t-connector for i and A is at least

$$\frac{(D_A(t) + \delta|A|)(D_i(t)+\delta)}{[2t(2m+\delta)]^2}, \tag{11.4.8}$$

independently of the fact whether the other vertices are t-connectors or not, and where, for any $A \subseteq \{1,2,\ldots,t\}$, we write

$$D_A(t) = \sum_{i \in A} d_i(t). \tag{11.4.9}$$

Since $d_i(t) + \delta \geq m + \delta > 0$ for every $i \leq t$, and $\delta < 0$, we have that

$$D_i(t) + \delta = D_i(t)\Big(1 + \frac{\delta}{D_i(t)}\Big) \geq D_i(t)(1 + \frac{\delta}{m}) = D_i(t)\frac{m+\delta}{m}, \tag{11.4.10}$$

and, thus, also $D_A(t) + \delta|A| \geq D_A(t)\frac{m+\delta}{m}$. As a result, for $\eta = (m+\delta)^2/(2m(2m+\delta))^2 > 0$, the probability that $j \in \{t+1, t+2, \ldots, 2t\}$ is a t-connector for i and A is at least $\frac{\eta D_A(t) D_i(t)}{t^2}$, independently of the fact whether the other vertices are t-connectors or not. Therefore, the probability that there is no t-connector for i and A is, conditionally on $\mathrm{PA}_{m,\delta}(t)$, bounded above by

$$\Big(1 - \frac{\eta D_A(t) D_i(t)}{t^2}\Big)^t \leq \exp\Big\{ -\frac{\eta D_A(t) D_i(t)}{t} \Big\}. \tag{11.4.11}$$

We shall make use of (11.4.11) in several places throughout the proof.

From Theorem 8.2 **whp**, Inner_t contains at least $\sqrt{t}$ vertices and denote the first $\sqrt{t}$ vertices of Inner_t by I. Observe that for $\tau > 2$ we have $t^{(\tau-1)^{-1}-1} \downarrow 0$ so that, for any $i, j \in I$, the probability that there exists a t-connector for i and j is bounded below by

$$1 - \exp\{-\eta t^{\frac{1}{\tau-1}-1} (\log t)^{-1}\} \geq p_t \equiv t^{\frac{1}{\tau-1}-1} (\log t)^{-2}, \tag{11.4.12}$$

for t sufficiently large.

We wish to couple Inner_t with an Erdős-Rényi random graph with $n_t = \sqrt{t}$ vertices and edge probability p_t, which we denote by $\text{ER}(n_t, p_t)$. For this, for $i, j \in \{1, 2, \ldots, n_t\}$, we say that an edge between i and j is present when there exists a t-connector connecting the i^{th} and j^{th} vertex in I. We now prove that this graph is bounded below by $\text{ER}(n_t, p_t)$. Note that (11.4.12) does not guarantee this coupling, instead we should prove that the lower bound holds uniformly, when i and j belong to I.

For this, we order the $n_t(n_t-1)/2$ edges in an arbitrary way, and bound the conditional probability that the l^{th} edge is present conditionally on the previous edges from below by p_t, for every l. This would prove the claimed stochastic domination by $\text{ER}(n_t, p_t)$.

Indeed, the l^{th} edge is present precisely when there exists a t-connector connecting the corresponding vertices which we call i and j in I. Moreover, we shall not make use of the first vertices which were used to t-connect the previous edges. This removes at most $n_t(n_t-1)/2 \leq t/2$ possible t-connectors, after which at least another $t/2$ remain. The probability that one of them is a t-connector for the i^{th} and j^{th} vertex in I is bounded below by, for t sufficiently large,

$$1 - \exp\{-\eta t^{\frac{1}{\tau-1}-2}(\log t)^{-1} t/2\} = 1 - \exp\{-\eta t^{\frac{1}{\tau-1}-1}(\log t)^{-1}/2\} \geq p_t \equiv t^{\frac{1}{\tau-1}-1}(\log t)^{-2}, \tag{11.4.13}$$

using $1 - e^{-x} \geq x/2$ for $x \in [0,1]$ and $\eta/2 \geq \log t^{-1}$ for t sufficiently large.

This proves the claimed stochastic domination of the random graph on the vertices I and $\text{ER}(n_t, p_t)$. Next, we show that $\text{diam}(\text{ER}(n_t, p_t))$ is, **whp**, bounded by a uniform constant.

For this we use the result in [44, Corollary 10.12], which gives sharp bounds on the diameter of an Erdős-Rényi random graph. Indeed, this result implies that if $p^d n^{d-1} - 2\log n \to \infty$, while $p^{d-1}n^{d-2} - 2\log n \to -\infty$, then $\text{diam}(\text{ER}(n,p)) = d$, **whp**. In our case, $n = n_t = t^{1/2}$ and $p = p_t = t^{\frac{1}{\tau-1}-1}(\log t)^{-2}$, which implies that, **whp**, $\frac{\tau-1}{3-\tau} < d \leq \frac{\tau-1}{3-\tau} + 1$. Thus, we obtain that the diameter of I in $\text{PA}_{m,\delta}(2t)$ is **whp** bounded by $2(\frac{\tau-1}{3-\tau}+1)$ in this case.

We finally show that for any $i \in \text{Inner}_t \setminus I$, the probability that there does not exist a t-connector connecting i and I is small. Indeed, this probability is, since $D_I(t) \geq \sqrt{t}t^{\frac{1}{\tau-1}}(\log t)^{-1/2}$, and $D_i(t) \geq t^{\frac{1}{2(\tau-1)}}(\log t)^{-1/2}$, the probability of there not existing a t-connector is bounded above by $e^{-\eta t^{1/(\tau-1)-1/2}(\log t)^{-1}}$, which is tiny since $\tau < 3$. This proves that **whp** the distance between any vertex $i \in \text{Inner}_t \setminus I$ and I is bounded by 2, and, together with the fact that $\text{diam}_{2t}(I) \leq 2(\frac{\tau-1}{3-\tau}+1)$ thus implies that $\text{diam}_{2t}(\text{Inner}_t) \leq 2(\frac{\tau-1}{3-\tau}+2)$. □

Proposition 11.21 (Distance between outer and inner core)**.** *Fix $m \geq 2$. With high probability, the inner core Inner_t can be reached from any vertex in the outer core Outer_t using no more than $\frac{2\log\log t}{|\log(\tau-2)|}$ edges in $\text{PA}_{m,\delta}(2t)$. More precisely, with high probability,*

$$\max_{i \in \text{Outer}_t} \min_{j \in \text{Inner}_t} d_{\text{PA}_{m,\delta}(2t)}(i,j) \leq \frac{2\log\log t}{|\log(\tau-2)|}. \tag{11.4.14}$$

Proof. Recall that

$$\text{Outer}_t = \text{Core}_t \setminus \text{Inner}_t. \tag{11.4.15}$$

and define

$$\mathcal{N}1 = \text{Inner}_t = \{i : D_i(t) \geq \mathcal{W}1\}, \tag{11.4.16}$$

where

$$\mathcal{W}1 = l_t = t^{\frac{1}{2(\tau-1)}}(\log t)^{-\frac{1}{2}}. \tag{11.4.17}$$

We now recursively define a sequence u_k, for $k \geq 2$, so that for any vertex $i \in \{1, 2, \ldots, t\}$ with degree at least u_k, the probability that there is no t-connector for the vertex i and the set

$$\mathcal{N}k - 1 = \{j : D_j(t) \geq \mathcal{W}k - 1\}, \tag{11.4.18}$$

conditionally on $\mathrm{PA}_{m,\delta}(t)$ is tiny. According to (11.4.11) and Exercise 8.17, this probability is at most

$$\exp\Big\{ - \frac{\eta B t [u_{k-1}]^{2-\tau} u_k}{t} \Big\} = o(t^{-1}), \tag{11.4.19}$$

where we define

$$\mathcal{W}k = D \log t \big(\mathcal{W}k - 1\big)^{\tau - 2}, \tag{11.4.20}$$

with D exceeding $(\eta B)^{-1}$. By Lemma **??** we have

$$\mathcal{W}k = D^{a_k} (\log t)^{b_k} t^{c_k}, \tag{11.4.21}$$

where

$$c_k = \frac{(\tau - 2)^{k-1}}{2(\tau - 1)}, \qquad b_k = \frac{1 - (\tau - 2)^{k-1}}{3 - \tau} - \frac{1}{2}(\tau - 2)^{k-1}, \qquad a_k = \frac{1 - (\tau - 2)^{k-1}}{3 - \tau}. \tag{11.4.22}$$

Then, the key step in the proof of Proposition 11.21 is the following lemma:

Lemma 11.22 (Connectivity between $\mathcal{N}k - 1$ and $\mathcal{N}k$)**.** *Fix $m, k \geq 2$. Then the probability that there exists an $i \in \mathcal{N}k$ that is not at distance two from $\mathcal{N}k - 1$ in $\mathrm{PA}_{m,\delta}(2t)$ is $o(t^{-1})$.*

Proof. We note that, by Exercise 8.17, with probability exceeding $1 - o(t^{-1})$, for all k,

$$\sum_{i \in \mathcal{N}k - 1} D_i(t) \geq Bt[\mathcal{W}k - 1]^{2-\tau}. \tag{11.4.23}$$

On the event that the bounds in (11.4.23) hold, we obtain by (11.4.11) that the conditional probability, given $\mathrm{PA}m, \delta(t)$, that there exists an $i \in \mathcal{N}k$ such that there is no t-connector between i and $\mathcal{N}k - 1$ is bounded, using Boole's inequality, by

$$te^{-\eta B [\mathcal{W}k - 1]^{2-\tau} \mathcal{W}k} = te^{-\eta B D \log t} = o(t^{-1}), \tag{11.4.24}$$

where we have used (11.4.20) and we have taken $D > 2(\eta B)^{-1}$. □

We now complete the proof of Proposition 11.21. Fix

$$k^* = \Big\lfloor \frac{\log \log t}{|\log (\tau - 2)|} \Big\rfloor. \tag{11.4.25}$$

As a result of Lemma 11.22, we have that the distance between $\mathcal{N}k^*$ and Inner_t is at most $2k^*$. Therefore, we are done when we can show that

$$\mathrm{Outer}_t \subseteq \{i : D_i(t) \geq (\log t)^\sigma\} \subseteq \mathcal{N}k^* = \{i : D_i(t) \geq \mathcal{W}k^*\}, \tag{11.4.26}$$

so that it suffices to prove that $(\log t)^\sigma \geq \mathcal{W}k^*$, for any $\sigma > \frac{1}{3-\tau}$. For this, we note that, by Lemma **??**, we have that

$$\mathcal{W}k^* = D^{a_{k^*}} (\log t)^{b_{k^*}} t^{c_{k^*}}. \tag{11.4.27}$$

We have that $t^{c_{k^*}} = O(1) = (\log t)^{o(1)}$, $(\log t)^{b_{k^*}} = (\log t)^{\frac{1}{3-\tau}+o(1)}$, and $D^{a_{k^*}} = (\log t)^{o(1)}$. Thus,

$$\mathcal{W}k^* = (\log t)^{\frac{1}{3-\tau}+o(1)}, \tag{11.4.28}$$

so that, by picking t sufficiently large, we can make $\sigma \geq \frac{1}{3-\tau} + o(1)$. This completes the proof of Proposition 11.21. □

Proof of Theorem 11.19. We note that **whp** $\mathrm{diam}_{2t}(\mathrm{Core}_t) \leq C_\delta + 2k^*$, where k^* is the upper bound on $\max_{i\in \mathrm{Outer}_t} \min_{j\in \mathrm{Inner}_t} d_{\mathrm{PA}_{m,\delta}(2t)}(i,j)$ in Proposition 11.21, and we have made use of Proposition 11.20. This proves Theorem 11.19. □

11.4.2 Connecting the periphery to the core

In this section, we extend the results of the previous section and, in particular, study the distance between the vertices not in the core Core_n and the core. The main result in this section is the following theorem:

Theorem 11.23 (Connecting the periphery to the core)**.** *Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$,* **whp***, the maximal distance between any vertex and Core_t in $G_m(2t)$ is bounded from above by $2\sigma \log\log t/\log m$.*

Together with Theorem 11.19, Theorem 11.23 proves the main result in Theorem 11.18.

The proof of Theorem 11.23 again consists of two key steps. The first in Proposition 11.24 states that, for any two vertices $i, j \in \{1, 2, \ldots, t\}$, the distance $d_{\mathrm{PA}_{m,\delta}(2t)}(i, j)$ is bounded by a constant times $\log\log t$, i.e., $\mathrm{diam}_{2t}(\mathrm{PA}_{m,\delta}(t))$ is bounded by some constant times $\log\log t$. The second in Proposition 11.27 shows that the distance between any vertex in $\{t+1, t+2, \ldots, 2t\}$ and $\{1, 2, \ldots, t\}$ is bounded by another constant times $\log\log t$.

Proposition 11.24 (Connecting half of the periphery to the core)**.** *Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$,* **whp***, the maximal distance between any vertex in $\{1, 2, \ldots, t\}$ and the core Core_t in $\mathrm{PA}_{m,\delta}(2t)$ is bounded from above by $\sigma \log\log t/\log m$.*

Proof. We start from a vertex $i \in \{1, 2, \ldots, t\}$ and will show that the probability that the distance between i and Core_t is at least $C \log\log t$ is $o(t^{-1})$ where $C = \sigma/\log m$. This proves the claim. For this, we explore the neighborhood of i as follows. From i, we connect its $m \geq 2$ edges. Then, successively, we connect the m edges from each of the at most m vertices that i has connected to and have not yet been explored. We continue in the same fashion. We call the arising process when we have explored up to distance k from the initial vertex i the *k-exploration tree*.

When we never connect two edges to the same vertex, then the number of vertices we can reach within k steps is *precisely* equal to m^k. We call an event where an edge connects to a vertex which already was in the exploration tree a *collision*. When k increases, the probability of a collision increases. However, the probability that there exists a vertex for which *many* collisions occur in the k-exploration tree before it hits the core is small, as we prove now:

Lemma 11.25 (A bound on the probability of multiple collisions)**.** *Fix $m \geq 2$ and $\delta \in (-m, 0)$. Fix $C = \sigma/\log m$, $l \geq 1$, $b \in (0, 1]$ and take $k \leq C \log\log t$. Then, for every vertex $i \in \{1, 2, \ldots, t\}$, the probability that its k-exploration tree has at least l collisions before it hits $\mathrm{Core}_t \cup \{j : j \leq t^b\}$ is bounded above by*

$$\left((\log t)^d t^{-b}\right)^l = (\log t)^{dl} t^{-bl},$$

for some $d > 0$.

Proof. Take $i \in \{\lceil t^b \rceil + 1, \lceil t^b \rceil + 2, \ldots, t\}$ and consider its k-exploration tree $\mathcal{T}ki$. Since we add edges after time t^b the denominator in (8.1.1) is at least t^b. Moreover, before hitting the core, any vertex in the k-exploration tree has degree at most $(\log t)^\sigma$. Hence, for $l = 1$, the probability mentioned in the statement of the lemma is at most

$$\sum_{v \in \mathcal{T}ki} \frac{D_v(t) + \delta}{t^b} \leq \sum_{v \in \mathcal{T}ki} \frac{(\log t)^\sigma}{t^b} \leq \frac{m^k (\log t)^\sigma}{t^b} \tag{11.4.29}$$

where the bound follows from $\delta < 0$ and $|\mathcal{T}ki| \leq m^k$. For general l this upper bound becomes

$$\left(\frac{m^k (\log t)^\sigma}{t^b} \right)^l$$

When $k = C \log\log t$ with $C = \sigma / \log m$, we have that $m^{lk} = (\log t)^{l\sigma}$. Therefore, the claim in Lemma 11.25 holds with $d = 2\sigma$. □

Lemma 11.25 shall prove to be extremely useful, as it will imply that the shortest path graph from any vertex is, with high probability, close to a tree up to the moment when $\mathrm{Core}_t \cup \{j : j \leq t^b\}$ is hit. We recall that Core_t consists of the vertices with high degree, and it is reasonable to suspect that the early vertices are precisely the ones which have high degree. This suggests that $\{j : j \leq t^b\} \subseteq \mathrm{Core}_t$. We next prove that indeed, **whp**, $\{j : j \leq t^b\}$ is a subset of the core Core_t:

Lemma 11.26 (Early vertices have large degrees **whp**)**.** *Fix $m \geq 1$. There exists a $b > 0$ such that,* **whp**, $\min_{j \leq t^b} d_j(t) \geq (\log t)^\sigma$, *for some $\sigma > \frac{1}{3-\tau}$. As a result,* **whp**, $\{j : j \leq t^b\} \subseteq \mathrm{Core}_t$.

Proof. Note that, by Exercise 8.19, for all $m \geq 1$ and for $\delta < 0$,

$$\mathbb{P}(D_i(t) \leq j) \leq j \frac{\Gamma(mt)\Gamma\big(mi + \frac{m+\delta}{2m+\delta}\big)}{\Gamma\big(mt + \frac{m+\delta}{2m+\delta}\big)\Gamma(mi)}. \tag{11.4.30}$$

Indeed, note that by (8.6.4), and when $\delta \leq 0$, we have that $C_j \leq 1$. Thus, for $m = 1$, the bound in (11.4.30) follows in this case. Furthermore, to conclude (11.4.30) from Exercise 8.19 for $m \geq 2$, we note that, by the relation between $\mathrm{PA}_{m,\delta}(t)$ and $\mathrm{PA}_{1,\delta/m}(mt)$, the degree of vertex i in $\mathrm{PA}_{m,\delta}(t)$ is bounded from below by the degree of vertex im in $\mathrm{PA}_{1,\delta}(mt)$. As a result,

$$\mathbb{P}(D_i^{(m)}(t) \leq j) \leq \mathbb{P}(D_{im}^{(1)}(mt) \leq j), \tag{11.4.31}$$

after which the result follows from (11.4.30) for $m = 1$.

Below, we shall rely on the obvious consequence of (11.4.30) that

$$\mathbb{P}(D_i(t) \leq j) \leq j \frac{\Gamma(mt)\Gamma\big(mi + \frac{m+\delta}{2m+\delta}\big)}{\Gamma\big(mt + \frac{m+\delta}{2m+\delta}\big)\Gamma(mi)}. \tag{11.4.32}$$

Obviously, for t and i large, we have that

$$\mathbb{P}(D_i(t) \leq j) \leq j t^{-\frac{m+\delta}{2m+\delta}} i^{\frac{m+\delta}{2m+\delta}} (1 + o(1)). \tag{11.4.33}$$

We finally use (11.4.32) to complete the proof of Lemma 11.26. Take $0 < b < \frac{\frac{m+\delta}{2m+\delta}}{\frac{m+\delta}{2m+\delta}+1} = \frac{m+\delta}{3m+2\delta}$. Then, by Boole's inequality,

$$\begin{aligned}\mathbb{P}(\exists i \le t^b : D_i(t) \le (\log t)^\sigma) &\le \sum_{i=1}^{t^b} \mathbb{P}(d_i(t) \le (\log t)^\sigma)\\ &\le (\log t)^\sigma \frac{\Gamma(mt)}{\Gamma\big(mt+\frac{m+\delta}{2m+\delta}\big)} \sum_{i=1}^{t^b} \frac{\Gamma\big(mi+\frac{m+\delta}{2m+\delta}\big)}{\Gamma(mi)}\\ &\le (\log t)^\sigma \big(\frac{m+\delta}{2m+\delta}+1\big)^{-1} \frac{\Gamma(mt)}{\Gamma\big(mt+\frac{m+\delta}{2m+\delta}\big)} \frac{\Gamma\big(mt^b+\frac{m+\delta}{2m+\delta}+1\big)}{\Gamma(mt^b)}\\ &= o(1), \qquad (11.4.34)\end{aligned}$$

by a similar equality as in (11.4.33). This completes the proof of Lemma 11.26. □

Now we are ready to complete the proof of Proposition 11.24:

Proof of Proposition 11.24. By combining Lemmas 11.25 and 11.26, the probability that there exists an $i \in \{1, 2, \dots, t\}$ for which the exploration tree $\mathcal{T}ki$ has at least l collisions before hitting the core is $o(1)$, whenever $l > 1/b$, since, by Boole's inequality, it is bounded by

$$\sum_{i=1}^{t} (\log t)^{dl} t^{-bl} = (\log t)^{2\sigma l} t^{-bl+1} = o(1), \qquad (11.4.35)$$

precisely when $l > \frac{1}{b}$. When the k-exploration tree hits the core, then we are done by Theorem 11.19. When the k-exploration tree from a vertex i does not hit the core, but has less than l collisions, then there are at least m^{k-l} vertices in k-exploration tree. Indeed, when there are at most l collisions, the minimal size of the tree is obtained by identifying at most l vertices and their complete offspring, and the size of the pruned tree has size at least m^{k-l}.

When $k = C \log\log t$ with $C = \sigma/\log m$, this number is at least equal to $(\log t)^{\sigma+o(1)}$. The total weight of the core is, by Exercise 8.17, at least

$$\sum_{i \in \mathrm{Core}_t} (D_i(t) + \delta) \ge Bt(\log t)^{-(\tau-2)\sigma}. \qquad (11.4.36)$$

The probability that there does not exist a t-connector between the k-exploration tree and the core is, by (11.4.11) bounded above by

$$\exp\left\{-\frac{\eta Bt(\log t)^{-(\tau-2)\sigma}(\log t)^{\sigma+o(1)}}{t}\right\} = o(t^{-1}), \qquad (11.4.37)$$

by picking B sufficiently large, since $\sigma > 1/(3-\tau)$. This completes the proof. □

Proposition 11.27. *Fix $m \ge 2$ and $\delta \in (-m, 0)$. For every $\sigma > \frac{1}{3-\tau}$,* **whp**, *the maximal distance between any vertex and $\mathrm{Core}_n \cup \{1, 2, \dots, t\}$ in $\mathrm{PA}_{m,\delta}(2t)$ is bounded from above by $\frac{\sigma \log\log t}{\log m}$.*

Proof. Denote $k = \lfloor \frac{\sigma \log\log t}{\log m} \rfloor - 1$. We again grow the k-exploration trees from the vertices $i \in \{t+1, t+2, \dots, 2t\}$.

By Lemma 11.25 for $b = 1$, the probability that there exists a vertex whose k-exploration tree contains at least two collisions before hitting the vertex set $\mathrm{Core}_t \cup \{1, 2, \ldots, t\}$ is bounded above by $t^{-2}(\log t)^{d_1}$ for some d_1 sufficiently large. When the k-exploration tree contains a vertex in $\mathrm{Core}_t \cup \{1, 2, \ldots, t\}$, then we are done by Proposition 11.24 and Theorem 11.19. If not, and there are at most 2 collisions, then there are at least $m_k = (m-1)m^{k-1}$ vertices in $\{t+1, t+2, \ldots, 2t\}$ at distance precisely equal to k from the original vertex. Denote these vertices by $i_1, \ldots, i_{m_k}$, and denote the k-exploration tree of vertex $i \in \{t+1, t+2, \ldots, 2t\}$ by $\mathcal{T}ki$. We write

$$\begin{aligned}\mathbb{P}(\nexists j \in \{1, 2, \ldots, m_k\} \text{ such that } i_j \longrightarrow \{1, 2, \ldots, t\} \big| \mathcal{T}ki) \\ = \prod_{j=1}^{m_k} \mathbb{P}\big(i_j \not\longleftrightarrow \{1, 2, \ldots, t\} \big| i_s \not\longleftrightarrow \{1, 2, \ldots, t\} \forall s < j, \mathcal{T}ki\big).\end{aligned} \tag{11.4.38}$$

Now we note that, uniformly in the way all edges in $\mathrm{PA}_{m,\delta}(2t)$ are formed, we have that for every $s \in \{t+1, t+2, \ldots, 2t\}$,

$$\frac{\sum_{i=1}^{t}(D_i(s) + \delta)}{(2m+\delta)s} \geq \frac{1}{2}. \tag{11.4.39}$$

Thus, for any vertex i_j in the boundary of $\mathcal{T}ki$, the probability that it will be directly connected to $\{1, 2, \ldots, t\}$ is at least $1/2$. As a result, we have that, uniformly in t, i and j,

$$\mathbb{P}\big(i_j \not\longleftrightarrow \{1, 2, \ldots, t\} \big| i_s \not\longleftrightarrow \{1, 2, \ldots, t\} \forall s < j, \mathcal{T}ki\big) \leq \frac{(2m+\delta)t}{(2m+\delta)(2t)} = \frac{1}{2}. \tag{11.4.40}$$

Therefore, we obtain that

$$\mathbb{P}(\nexists j = 1, \ldots, m_k \text{ such that } i_j \longleftrightarrow \{1, 2, \ldots, t\} | \mathcal{T}ki) \leq 2^{-m_k}. \tag{11.4.41}$$

Since $m_k = \frac{m-1}{m}(\log t)^{\sigma}$, with $\sigma = \frac{1}{3-\tau} > 1$, we have that $2^{-m_k} = o(t^{-1})$. Therefore, any vertex $i \in \{t+1, t+2, \ldots, 2t\}$ is, **whp**, within distance $k+1$ from $\{1, 2, \ldots, t\}$. □

Proof of Theorem 11.23. Proposition 11.27 states that **whp** every vertex in $\mathrm{PA}_{m,\delta}(2t)$ is within distance $k+1 = \lfloor \frac{\sigma \log\log t}{\log m} \rfloor$ of $\mathrm{Core}_t \cup \{1, 2, \ldots, t\}$. Proposition 11.24 states that **whp** every vertex in $\{1, 2, \ldots, t\}$ is at most distance $k+1$ from the core Core_t. This shows that every vertex in $\mathrm{PA}_{m,\delta}(2t)$ is **whp** within distance $2(k+1)$ from the core. □

Proof of Theorem 11.18. Theorem 11.23 states that every vertex in $\mathrm{PA}_{m,\delta}(2t)$ is within distance $\frac{2\sigma \log\log t}{\log m}$ of the core Core_t. Theorem 11.19 states that the diameter of the core is at most $\frac{4 \log\log t}{|\log(\tau-2)|}(1 + o(1))$, so that the diameter of $\mathrm{PA}_{m,\delta}(2t)$ is at most $C_G \log\log t$, where C_G is given in (11.4.1). This completes the proof of Theorem 11.18. □

11.5 Notes and discussion

Notes on Section 11.1. The proof of Theorem 11.1 is adapted from [196]. More recent discussions on exchangeable random variables and their properties can be found in [8] and [188], the latter focussing on random walks with self-interaction, where exchangeability is a crucial tool. There is a lot of work on urn schemes, also in cases where the weight functions are not linear with equal slope, in which case the limits can be seen to obey rather different characteristics. See e.g., [16, Chapter 9].

Notes on Section 11.3. There is a beautiful result on the height of trees using branching processes due to Kingman [145], which Pittel [192] makes crucial use of. This approach is based on exponential martingales, and allows for a relatively short proof of the lower bound on the height of the tree.

Chapter 12
Shortest weight routing on random graphs

Let a random weight X_{ij} be assigned to every edge ij of the complete graph K_n. (Thus $X_{ji} = X_{ij}$. We do not define X_{ij} for $i = j$.) We assume that th $\binom{n}{2}$ weights X_{ij}, $1 \le i < j \le n$, are independent and identically distributed; moreover, we assume that they are non-negative and that their distribution function $\mathbb{P}(X_{ij} \le t) = t + o(t)$ as $t \searrow 0$, the main examples being the uniform $U(0,1)$ and the exponential Exp(1) distributions.

For two vertices i and j, let W_{ij} be the minimal total weight of a path between i and j. our main theorem is a set of three different asymptotic results for W_{ij} (log denotes the natural logarithm).

Theorem 12.1. *Under the assumptions above, as $n \to \infty$*

(i) for any fixed i and j,

$$\frac{W_{ij}}{\log n/n} \xrightarrow{\mathbb{P}} 1; \tag{12.0.1}$$

(ii) for any fixed i,

$$\frac{\max_{j \le n} W_{ij}}{\log n/n} \xrightarrow{\mathbb{P}} 2; \tag{12.0.2}$$

(iii)

$$\frac{\max_{i,j \le n} W_{ij}}{\log n/n} \xrightarrow{\mathbb{P}} 3. \tag{12.0.3}$$

Hence, with high probability, W_{ij} is about $\log n/n$ for any fixed (or random) pair of vertices, but there are pairs of vertices for which it is larger: up to $2\log n/n$ if i is fixed and up to $3\log n/n$ globally.

Similarly, defining $Y_i = \max_{j \le n} W_{ij}$, we see from (ii) and (iii) that Y_i typically is about $2\log n/n$, but that it is larger for a few vertices with $\max_i Y_i$ being about $3\log n/n$. A companion result shows that, in contrast, Y_i is not significantly smaller than $2\log n/n$ for any vertex i.

Theorem 12.2. *As $n \to \infty$,*

$$\frac{\min_{i \le n} \max_{j \le n} W_{ij}}{\log n/n} \xrightarrow{\mathbb{P}} 2. \tag{12.0.4}$$

In other words, interpreting the weights as distances, most pairs of vertices are at a distance of about $\log n/n$, the radius of the graph is about $2\log n/n$ and the diameter is about $3\log n/n$.

Remark 1. Theorem 12.1(i),(ii) may alternatively be stated in terms of first-passage percolation on the complete graph (the time to reach a given vertex is about $\log n/n$ and the time to reach all is $2\log n/n$).

For completeness and comparison, we also state the corresponding simple (and well-known) results for the *minimal* distance from a vertex. In this case there is less concentration and we obtain convergence (in distribution) to a nondegenerate random variable instead of to a constant.

Theorem 12.3. *Let $Z_i = \min_{j\neq i} W_{ij} = \min_{j\neq i} X_{ij}$. As $n \to \infty$:*

(i) for any fixed i,

$$nZ_i \xrightarrow{d} \mathrm{Exp}(1); \tag{12.0.5}$$

(ii)

$$n^2 \min_{i\leq n} Z_i = n^2 \min_{i,j\leq n} X_{ij} \xrightarrow{d} \mathrm{Exp}(2); \tag{12.0.6}$$

(iii)

$$\frac{\max_{i\leq n} Z_i}{\log n/n} \xrightarrow{\mathbb{P}} 1. \tag{12.0.7}$$

The proofs of (i) and (ii) are simple exercises, while (iii) is, in disguise, the well-known threshold for existence of isolated vertices in a random graph [44, Exercise III.2]; consider the graph with edges $\{ij : X_{ij} < t\}$. We leave the details to the reader. (Note that if $X_{ij} \in \mathrm{Exp}(1)$, then $(n-1)Y_i \in \mathrm{Exp}(1)$ and $n(n-1)\min_i Y_i \in \mathrm{Exp}(2)$ exactly.) Using Theorem 12.3(iii), we can give a simple informal explanation of the discrepancy between the three parts of Theorem 12.1 as follows, interpreting the weights as travel times. Most vertices are connected by efficient highways, which take you to almost any other vertex within about $\log n/n$ (but rarely much quicker). Some vertices, however, are remote villages (like Oberwolfach), from which it takes up to $\log n/n$ to get to any other vertex at all. Hence, starting at a typical vertex, most travel times are about $\log n/n$, but it takes an extra $\log n/n$ (just for the final step in the path) to reach a few remote vertices. Similarly, if we start at one of the very remote vertices, it takes about $\log n/n$ to get to any other vertex at all, $2\log n/n$ to get to most other vertices and $3\log n/n$ to get to the other very remote vertices.

Some further results on asymptotic distributions and moments are given in Section 12.2. The lengths of the minimum weight paths are studied in Section 12.3.

12.1 Proofs

We first observe that the distribution of X_{ij} does not affect the results, as long as it satisfies the condition above. This is seen by the following standard coupling argument, which we include for completeness.

Let $F^{-1} : [0,1) \to [0,\infty)$ be the inverse function of the distribution function $F(t) = \mathbb{P}(X_{ij} \leq t)$ of X_{ij}. If $U_{ij} \in U(0,1)$ are independent uniform random variables, then $F^{-1}(U_{ij})$ has the same distribution as X_{ij}, so we may without loss of generality assume that $X_{ij} = F^{-1}(U_{ij})$. By assumption, $F(t)/t \to 1$ as $t \searrow 0$, and thus also $F^{-1}(t)/t \to 1$. Let $\varepsilon > 0$. If $W_{ij} < 10\log n/n$, for some i and j, then $X_{kl} = F^{-1}(U_{kl} < 10\log n/n$ for each edge kl in the minimum weight path from i to j, and thus, provided n is large enough, $1-\varepsilon < X_{kl}/U_{kl} < 1+\varepsilon$. Consequently, the sum of the U_{kl} for the same path is at most $(1-\varepsilon)^{-1}W_{ij}$, and thus, using W'_{ij} to denote the minimal path weight defined by $\{U_{ij}\}$, $W'_{ij} \leq (1-\varepsilon)^{-1}W_{ij}$. Conversely, by the same argument, if $W'_{ij} < 10\log n/n$ then $W_{ij} < (1+\varepsilon)W'_{ij}$. It follows that, if either $W_{ij} < 9\log n/n$ or $W'_{ij} < 9\log n/n$, and n is large enough, then both $W_{ij} < 10\log n/n$ and $W'_{ij} < 10\log n/n$ hold, and moreover $(1-\varepsilon)W'_{ij} < W_{ij} < (1+\varepsilon)W'_{ij}$. It now follows immediately that, if any part of Theorem 12.1 or 12.2 holds either for X_{ij} or for the uniform U_{ij}, then it holds for both. In particular, a proof of these results for any distribution with $F(t)/t \to 1$ as $t \searrow 0$ implies the same result for $U(0,1)$, and then for any other such distribution.

We may thus choose a convenient distribution of X_{ij}; we use the exponential distribution because of its excellent Markov properties. Hence, in the sequel we assume that $X_{ij} \in \mathrm{Exp}(1)$.

Proof of Theorem 12.1. for parts (i) and (ii), we may assume that $i = 1$. We adopt the first-passage percolation viewpoint (see Remark 1), so we regard vertex 1 as initially infected, and assume that the infection spreads along each edge with an Exp(1)-distributed waiting time. We first study when the other vertices get infected, considering them in order of infection and ignoring their labels.

Since there are $n-1$ neighbours of the initially infected vertex, the time V_1 until the second vertex is infected is exponentially distributed with expectation $1/(n-1)$. More generally, when $k < n$ vertices have been infected, there are $k(n-k)$ edges connecting the infected and non-infected vertices, and thus the time V_k until the next vertex is infected is $\mathrm{Exp}\,(1/(k(n-k)))$; moreover, this time is independent of $V_1, \ldots, V_{k-1}$. In other words, te time S_m until m vertices have become infected can be written

$$S_m = \sum_{k=1}^{m-1} V_k \tag{12.1.1}$$

where $V_1, \ldots, V_{m-1}$ are independent with $V_k \in \mathrm{Exp}\,(1/(k(n-k)))$.

The times $(S_m)_{m=2}^n$ are just the minimal path weights $(W_{1j})_{j=2}^n$, arranged in increasing order. In particular,

$$Y_1 = \max_{j \geq 2} X_{1j} = S_n = \sum_{k=1}^{n-1} V_k. \tag{12.1.2}$$

Hence

$$\begin{aligned}\mathbb{E}[Y_1] = \sum_{k=1}^{n-1} \mathbb{E}[V_k] = \sum_{k=1}^{n-1} \frac{1}{k(n-k)} = \frac{1}{n} \sum_{k=1}^{n-1} \left(\frac{1}{k} + \frac{1}{n-k} \right) = \frac{2}{n} \sum_{k=1}^{n-1} \frac{1}{k} \\ = 2\frac{\log n}{n} + O\left(\frac{1}{n}\right),\end{aligned} \tag{12.1.3}$$

and similarly

$$\begin{aligned}\mathrm{Var}(Y_i) = \sum_{k=1}^{n-1} \mathrm{Var}(V_k) = \sum_{k=1}^{n-1} \left(\frac{1}{k(n-k)} \right)^2 \leq 2 \sum_{k=1}^{n/2} \frac{1}{b} b k^2 (n-k)^2 \\ \leq \frac{8}{n^2} \sum_{k=1}^{n/2} \frac{1}{k^2} = O(n^{-2}).\end{aligned} \tag{12.1.4}$$

Part (ii) now follows by Chebyshev's inequality.

For part (i), fix $j = 2$. Observe that, if N is the number of vertices before vertex 2, then

$$W_{12} = S_{N+1} = \sum_{k=1}^{N} V_k, \tag{12.1.5}$$

where, by symmetry, N is uniformly distributed over $1, \ldots, n-1$ and independent of $V_1, \ldots, V_{n-1}$. We rewrite this equation as $W_{12} = \sum_{k=1}^{n-1} \mathbb{1}_{\{N \geq k\}} V_k$, using indicator functions to eliminate the random summation limit. Hence,

$$\begin{aligned}\mathbb{E}[W_{12}] = \sum_{k=1}^{n-1} \mathbb{E}[\mathbb{1}_{\{N \geq k\}} V_k] = \sum_{k=1}^{n-1} \mathbb{P}(N \geq k) \mathbb{E}[V_k] \\ = \sum_{k=1}^{n-1} \frac{n-k}{n-1} \frac{1}{k(n-k)} = \sum_{k=1}^{n-1} \frac{1}{k(n-1)} \\ = \frac{\log n}{n} + O\left(\frac{1}{n}\right).\end{aligned} \tag{12.1.6}$$

In order to estimate the variance, we further rewrite the sum as

$$\begin{aligned} W_{12} &= \sum_{k=1}^{N}(V_k - \mathbb{E}[V_k]) + \sum_{k=1}^{N}\frac{1}{n}\left(\frac{1}{k}+\frac{1}{n-k}\right) \\ &= \sum_{k=1}^{N}(V_k - \mathbb{E}[V_k]) + \frac{1}{n}\left(\log N + \log n - \log(n-N)\right) + O\left(\frac{1}{n}\right). \end{aligned} \tag{12.1.7}$$

We consider the three terms on the right-hand side separately. Since $N, V_1, \ldots, V_{n-1}$ are independent,

$$\begin{aligned} \mathrm{Var}\left(\sum_{k=1}^{N}(V_k - \mathbb{E}[V_k])\right) &= \mathbb{E}\left[\left(\sum_{k=1}^{N}(V_k - \mathbb{E}[V_k])\right)^2\right] = \mathbb{E}\left[\sum_{k=1}^{N}\mathrm{Var}(V_k)\right] \\ &\le \sum_{k=1}^{n-1}\mathrm{Var}V_k = \sum_{k=1}^{n-1}\frac{1}{k(n-k)^2} \\ &\le \sum_{k=1}^{n/2}\frac{4}{k^2n^2} + \sum_{k=n/2}^{n-1}\frac{4}{n^2(n-k)^2} = O\left(\frac{1}{n^2}\right). \end{aligned} \tag{12.1.8}$$

For the second term, we observe that

$$\mathbb{E}[(\log N - \log(n-1))^2] = \mathbb{E}\left[\left(\log\frac{N}{n-1}\right)^2\right] \to \int_0 61(\log x)^2 dx < \infty \tag{12.1.9}$$

as $n \to \infty$. Hence $\mathrm{Var}(\log N) = \mathrm{Var}(\log(n-N)) = O(1)$, and it follows that the variance of the second term in (12.1.7) is also $O(n^{-2})$. The same is trivially true for the third term.

Consequently, $\mathrm{Var}(W_{ij}) = O(n^{-2})$, which together with (12.1.6) yields part (i).

The proof of (iii) is divided into two parts, considering upper and lower bounds separately. First, by (12.1.2), for $-\infty \le t < 1 - 1/n$.

$$\mathbb{E}[\mathrm{e}^{tnY_1}] = \prod_{k=1}^{n-1}\mathbb{E}[\mathrm{e}^{ntV_k}] = \prod_{k=1}^{n-1}\left(1-\frac{nt}{k(n-k)}\right)^{-1}. \tag{12.1.10}$$

Hence, for every $a > 0$, choosing $t = 1 - 1/\log n$ $(n \ge 3)$,

$$\begin{aligned} \mathbb{P}(Y_1 > a\log n/n) &\le \mathbb{E}[\mathrm{e}^{tnY_1 - ta\log n}] = \mathrm{e}^{ta\log n}\prod_{k=1}^{n-1}\left(1-\frac{nt}{k(n-k)}\right)^{-1} \\ &= \left(1-\frac{nt}{n-1}\right)^{-2}\exp\left(-ta\log n + \sum_{k=2}^{n-2} -\log\left(1\!\!1 - \frac{nt}{k(n-k)}\right)\right) \\ &\le \left(1-\frac{nt}{n-1}\right)^{-2}\exp\left(-ta\log n + \sum_{k=2}^{n-2}\left(\frac{nt}{k(n-k)} + \left(\frac{nt}{k(n-k)}\right)^2\right)\right) \\ &= (1-t+O(n^{-1}))^{-2}\exp(-ta\log n + 2t\log n + O(1)) = O(n^{2-a}\log^2 n). \end{aligned} \tag{12.1.11}$$

This evidently implies

$$\mathbb{P}(\max_i Y_i > a\log n/n) \le n\mathbb{P}(Y_1 > a\log n/n) = O(n^{3-a}\log^2 n), \tag{12.1.12}$$

which tends to 0 as $n \to \infty$ for every fixed $a > 3$.

For the lower bound, let $\varepsilon > 0$ be small. Partition the vertex set $[n]$ of K_n into the sets $A = \{1, \dots, b_A\}$ and $B = \{n_A + 1, \dots, n\}$, where $n_A = \lceil n^{1-\varepsilon} \rceil$. Let $n_B = |B| = n - n_A$.

For $i \in A$, let $U_i = \min_{j \in B} X_{ij}$. Then the random variables U_i, $i \in A$, are independent with $U_i \in \mathrm{Exp}(1/n_B)$. In particular,

$$\begin{aligned} \mathbb{P}(U_i > (1-2\varepsilon)\log n/n) &= \exp\left(-(1-2\varepsilon)\frac{n_B}{n}\log n\right) \\ &\geq \exp(-(1-2\varepsilon)\log n) = n^{2\varepsilon-1} \end{aligned} \tag{12.1.13}$$

and thus

$$\mathbb{P}(U_i \leq (1-2\varepsilon)\log n/n \text{ for every } i \in A) \leq (1 - n^{2\varepsilon-1})^{n^{1-\varepsilon}} < \mathrm{e}^{-n^{\varepsilon}}. \tag{12.1.14}$$

For $k \in A$, let $\mathcal{E}_k$ be the event that $U_k > (1-2\varepsilon)\log n/n$ but $U_i \leq (1-2\varepsilon)\log n/n$ for $i < k$. Then the events $\mathcal{E}_k$ are disjoint and, by (12.1.14),

$$\sum_{k \in A} \mathbb{P}(\mathcal{E}_k) = \mathbb{P}\left(\bigcup_{k \in A} \mathcal{E}_k\right) > 1 - \mathrm{e}^{-n^{\varepsilon}}. \tag{12.1.15}$$

The idea of the proof is to show that, conditioned on $\mathcal{E}_k$, Y_k is with high probability close to $3\log n/n$; in fact, as is shown in detail below, conditioning on $U_k > (1-2\varepsilon)\log n/n$ typically increases Y_k (which is usually about $2\log n/n$) by $(1-2\varepsilon)\log n/n$, while conditioning on $U_i \leq (1-2\varepsilon)\log n/n$ for $i < k$ hardly affects the result.

We will use the following lemma.

Lemma 12.4. *Suppose that $\mu, b > 0$ and $X \in \mathrm{Exp}(\mu)$, and define*

$$f(x) = -\mu \log(\mathrm{e}^{-b/\mu} + (1 - \mathrm{e}^{b/\mu})\mathrm{e}^{-x/\mu}). \tag{12.1.16}$$

(i) The distribution of $f(X)$ equals the conditional distribution of X given $X \leq b$.

(ii) If, further, $0 \leq \alpha < 1$ and $b/\mu \geq \alpha(1 - \log\alpha)/(1-\alpha)$, then $f(x) \geq \alpha x$ when $0 \leq x \leq \alpha^{-1}b - \mu$. Consequently,

$$\mathbb{P}(f(X) < \alpha X) \leq \mathbb{P}(X > \alpha^{-1}b - \mu) = \mathrm{e}^{1-\alpha^{-1}b/\mu}. \tag{12.1.17}$$

Proof. We may for simplicity, by homogeneity, assume that $\mu = 1$. Then e^X is uniformly distributed on $[0,1]$, and thus for $0 \leq t \leq b$,

$$\begin{aligned} \mathbb{P}(f(X) \leq t) &= \mathbb{P}(\mathrm{e}^{-b} + (1-\mathrm{e}^{-b})\mathrm{e}^{-X} \geq \mathrm{e}^{-t}) = \mathbb{P}\left(\mathrm{e}^{-X} \geq \frac{\mathrm{e}^{-t} - \mathrm{e}^{-b}}{1-\mathrm{e}^{-b}}\right) \\ &= \frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-b}} = \mathbb{P}(X \leq t | X \leq B), \end{aligned} \tag{12.1.18}$$

which proves part (i).

For part (ii) we observe that (when $\mu = 1$) $f(x) \geq \alpha x$ if and only if

$$\mathrm{e}^{-b} + (1-\mathrm{e}^{-b})\mathrm{e}^{-x} \leq \mathrm{e}^{-\alpha x}. \tag{12.1.19}$$

Letting $y = \mathrm{e}^{-x}$, the left-hand side of (12.1.19) is a linear function of y, while the right-hand side y^{α} is concave; hence, in order to verify (12.1.19) for the interval $0 \leq x \leq \alpha^{-1}b - 1$, it suffices to verify it for the endpoints.

For $x = 0$, (12.1.19) is a trivial identity, while for $x = \alpha^{-1}b - 1$ it is

$$\mathrm{e}^{-b} + (1 - \mathrm{e}^{-b})\mathrm{e}^{-\alpha^{-1}b+1} \le \mathrm{e}^{-b+\alpha}. \tag{12.1.20}$$

Now, by assumption, $\alpha^{-1}b = b + b(1-\alpha)\alpha^{-1} \ge b + 1 - \log\alpha$, and thus

$$\mathrm{e}^{-b} + \mathrm{e}^{-\alpha^{-1}b+1} \le \mathrm{e}^{-b} + \mathrm{e}^{-b+\log\alpha} = (1+\alpha)\mathrm{e}^{-b} \le \mathrm{e}^{\alpha}\mathrm{e}^{-b}; \tag{12.1.21}$$

this implies (12.1.20), which completes the proof of the lemma. □

Continuing with the proof of Theorem 12.1(iii), let $k \in A$ be fixed, let f be as in Lemma 12.4 with $\mu = 1/n_B$ and $b = (1-2\varepsilon)\log n/n$, and define

$$U_i' = \begin{cases} f(u_i), & i < k, \\ U_i + b, & i = k, \\ U_i, & i > k. \end{cases} \tag{12.1.22}$$

Then, by Lemma 12.4(i) for $i < k$ and the standard lack-of-memory property of exponential distributions for $i = k$, the distribution of U_i' equals the conditional distribution of U_i' equals the conditional distribution of U_i given $\mathcal{E}_k$ for every $i \in A$; moreover, by our independence assumptions, this extends to the joint distribution. Furthermore, by the same lack-of-memory-property, the family of random variables $\{X_{ij} - U_i\}_{j\in B}$ is independent of U_i, for each $i \in A$ separately and thus for all $i \in A$ jointly too; hence the joint distribution of $\{X_{ij} - U_i\}_{j\in B}$ is not affected by conditioning on $\mathcal{E}_k$. It follows that if we define X_{ij}' for $1 \le i < j \le n$ by

$$X_{ij}' = \begin{cases} X_{ij} - U_i + U_i', & i \in A \text{ and } j \in B, \\ X_{ij}, & \text{otherwise}, \end{cases} \tag{12.1.23}$$

and let $X_{ji}' = X_{ij}'$ for $j > i$, then the family $\{X_i'j\}$ has the same distribution as the conditional distribution of $\{X_{ij}\}$ given $\mathcal{E}_k$. Note in particular that $X_{kj}' = X_{kj} + b$ when $j \in B$.

Suppose that $\{X_{ij}\}$ are such that

$$U_i' \ge (1-2\varepsilon)U_i, \quad \text{for every } i \in A, \tag{12.1.24}$$

$$X_{ik} \ge 3\frac{\log n}{n}, \quad \text{for every } i \in A, \tag{12.1.25}$$

and

$$Y_k \ge (2-\varepsilon)\frac{\log n}{n}. \tag{12.1.26}$$

We observe first that, by (12.1.23) and (12.1.24), then

$$X_{ij}' \ge (1-2\varepsilon)X_{ij}, \quad \text{for every } i \text{ and } j \ne i. \tag{12.1.27}$$

Now consider the minimal path weights W_{ij}' defined by the edge weights X_{ij}' and the corresponding $Y_i' = \max_j W_{ij}'$. By (12.1.26), there exists a vertex l such that every path $i_0 = k, i_1, \dots, i_m = l$ from k to l has weight $W = \sum_{s=1}^m X_{i_{s-1}i_s} \ge (2-\varepsilon)\log n/n$. Consider such a path and the corresponding weight $W' = \sum_{s=1}^m X_{i_{s-1}i_s}'$. Either $i_1 \in A$, and then, by (12.1.23) and (12.1.25), $W' \ge X_{ki_1}' = X_{ki_1} \ge 3\log n/n$, or $i_1 \in B$, and then $X_{ki_1}' = X_{ki_1} + b$, which together with (12.1.27) yields

$$W' \ge b + (1-2\varepsilon)W \ge (1-2\varepsilon)\frac{\log n}{n} + (1-2\varepsilon)(2-\varepsilon)\frac{\log n}{n} \ge (3-7\varepsilon)\frac{\log n}{n}. \tag{12.1.28}$$

Hence $W' \ge (3-7\varepsilon)\log n/n$ for every path from k to l, and thus $W_{kl}' \ge (3-7\varepsilon)\log n/n$ and finally $Y_k' \ge W_{kl}' \ge (3-7\varepsilon)\log n/n$.

We have shown that, if (12.1.24)-(12.1.26) hold, then $Y_k' \geq (3-7\varepsilon)\log n/n$. Consequently,

$$\begin{aligned}\mathbb{P}(Y_k \geq (3-7\varepsilon)\log n/n|\mathcal{E}_k) &= \mathbb{P}(Y_k' \geq (3-7\varepsilon)\log n/n)\\ &\geq \mathbb{P}((12.1.24)-(12.1.26)\text{ hold}).\end{aligned} \tag{12.1.29}$$

Let q denote the probability that (12.1.24)-(12.1.26) hold. We have so far kept k fixed, but q is independent of k, and summing over k we obtain

$$\begin{aligned}\mathbb{P}(\max_i Y_i \geq (3-7\varepsilon)\log n/n) &\geq \sum_{k\in A}\mathbb{P}(Y_k \geq (3-7\varepsilon)\log n/n|\mathcal{E}_k)\mathbb{P}(\mathcal{E}_k)\\ &\geqq \sum_{k\in A}\mathbb{P}(\mathcal{E}_k).\end{aligned} \tag{12.1.30}$$

Now, by Lemma 12.4(ii) with $\alpha = 1-2\varepsilon$, if n is large enough,

$$\begin{aligned}\mathbb{P}((12.1.24)\text{ fails}) &\leq \sum_{i\in A}\mathbb{P}(U_i' < (1-2\varepsilon)U_i) \leq n_A \mathrm{e}^{1-n_B\log n/n}\\ &= O(n^{1-\varepsilon}n^{-1}) = o(1).\end{aligned} \tag{12.1.31}$$

Similarly,

$$\mathbb{P}((12.1.25)\text{ fails}) \leq \sum_{i\in A}\mathbb{P}\left(X_{ik} < 3\frac{\log n}{n}\right) \leq 3n_A\frac{\log n}{n} = o(1), \tag{12.1.32}$$

while $\mathbb{P}((12.1.26)\text{ fails}) = o(1)$ by the already proven part (ii) of the theorem.

Consequently, $q = 1-o(1)$, which by (12.1.30) and (12.1.15) yields $\mathbb{P}(\max_i Y_i \geq (3-7\varepsilon)\log n/n) \to 1$ as $n\to\infty$. This completes the proof of (iii). □

Proof of Theorem 12.2. We use (12.1.10), replacing t by $-t$, and, for every a and $t>0$, obtain

$$\begin{aligned}\mathbb{P}(Y_1 < a\log n/n) &\leq \mathbb{E}[\mathrm{e}^{ta\log n - tnY_1}] \leq \mathrm{e}^{ta\log n}\prod_{k=1}^{n-1}\left(1+\frac{nt}{k(n-k)}\right)^{-1}\\ &= \exp\left(ta\log n + \sum_{k=1}^{n-1} -\log\left(1+\frac{nt}{k(n-k)}\right)\right)\\ &\leq \exp\left(ta\log n + \sum_{k=1}^{n-1}\left(-\frac{nt}{k(n-k)} + \frac{1}{2}\left(\frac{nt}{k(n-k)}\right)^2\right)\right)\\ &= \exp\left(at\log n - 2t\log n + O(t) + O(t^2)\right).\end{aligned} \tag{12.1.33}$$

If $0<a<2$, we thus obtain for any constant t

$$\mathbb{P}(\min_i Y_i < a\log n/n) \leq n\mathbb{P}(Y_1 < a\log n/n) = O(n^{1+(a-2)t}), \tag{12.1.34}$$

which is $o(1)$ provided $t > 1/(2-a)$. On the other hand, Theorem 12.1(ii) implies

$$\mathbb{P}(\min_i Y_i > (2+\varepsilon)\log n/n) \leq \mathbb{P}(Y_1 > (2+\varepsilon)\log n/n) \to 0 \tag{12.1.35}$$

for every $\varepsilon > 0$, and the proof is complete. □

12.2 Asymptotic distribution and moments

The method above also yields the asymptotic distributions of W_{ij} and Y_i: these are not normal. More precisely, we have the following result. (We have to impose a slightly stronger condition on the distribution of the X_{ij}; the condition is satisfied for the exponential and uniform distributions.)

Theorem 12.5. *Suppose that the distribution function* $\mathbb{P}(X_{ij} \le t) = t + o(t/|\log t|)$ *as* $t \searrow 0$. *Then, as* $n \to \infty$,

$$nW_{ij} - \log n - \gamma \xrightarrow{d} \sum_1^\infty \frac{1}{k}(\xi_k - 1) + \zeta \tag{12.2.1}$$

and

$$nY_i - 2\log n - 2\gamma \xrightarrow{d} \sum_1^\infty \frac{1}{k}(\xi_k - 1) + \sum_1^\infty \frac{1}{k}(\xi'_k - 1), \tag{12.2.2}$$

where γ *is Euler's constant, and the random variables* ξ_k, ξ'_k, $k \ge 1$, *and* ζ *are independent with* $\xi_k, \xi'_k \in \mathrm{Exp}(1)$ *while* ζ *has the logistic distribution* $\mathbb{P}(\zeta \le x) = \mathrm{e}^x/(1+\mathrm{e}^x)$.

Proof. By a slight modification of the coupling argument in the proof of Theorem 12.1, it suffices to consider the case $X_{ij} \in \mathrm{Exp}(1)$; we omit the details.

We write $A_n \approx B_n$ to mean that $\mathbb{E}[(A_n - B_n)^2] = o(1)$ as $n \to \infty$. In the exponential case (12.1.5) and (12.1.2) imply that

$$\begin{aligned} nW_{12} &\stackrel{d}{=} \sum_1^N \frac{n}{k(n-k)}\xi_k = \sum_1^N \frac{n}{k(n-k)}(\xi_k - 1) + \sum_1^N \left(\frac{1}{k} + \frac{1}{n-k}\right) \\ &\approx \sum_1^N \frac{1}{k}(\xi_k - 1) + \log N + \gamma + \log n - \log(n-N) \\ &\approx \sum_1^\infty \frac{1}{k}(\xi_k - 1) + \log\frac{N/n}{1-N/n} + \log n + \gamma, \end{aligned} \tag{12.2.3}$$

and

$$\begin{aligned} nY_1 &\stackrel{d}{=} \sum_1^{n-1} \frac{n}{k(n-k)}\xi_k = \sum_1^{n-1} \frac{n}{k(n-k)}(\xi_k - 1) + 2\sum_1^{n-1}\frac{1}{k} \\ &\approx \sum_1^{\lfloor n/2\rfloor} \frac{1}{k}(\xi_k - 1) \sum_{\lfloor n/2\rfloor + 1}^{n-1} \frac{1}{n-k}(\xi_k - 1) + 2\log n + 2\gamma \\ &\stackrel{d}{=} \sum_1^{\lfloor n/2\rfloor} \frac{1}{k}(\xi_k - 1) + \sum_1^{\lceil n/2\rceil - 1} \frac{1}{k}(\xi'_k - 1) + 2\log n + 2\gamma. \end{aligned} \tag{12.2.4}$$

The result follows, since $N/n \xrightarrow{d} \eta$, where $\eta \in U(0,1)$, and $\zeta = \log(\eta/(1-\eta))$ has the logistic distribution. □

Since the moment generating function $\mathbb{E}[\mathrm{e}^{t\xi_k}]$ of ξ_k equals $(1-t)^{-1}$, $\mathrm{R}et < 1$, it follows that the moment generating function of $\sum \frac{1}{k}(\xi_k - 1)$ is

$$\begin{aligned}\prod_{k=1}^{\infty}(1-t/k)^{-1}\mathrm{e}^{-t/k} &= \lim_{n\to\infty}\left(\prod_{k=1}^{n}\frac{k}{k-t}\right)\mathrm{e}^{-t\sum_{l=1}^{n}\frac{1}{l}} \\ &= \lim_{n\to\infty}\frac{\Gamma(n+1)\Gamma(1-t)}{\Gamma(n+1-t)}\mathrm{e}^{-t\log n - t\gamma + o(1)} \\ &= \Gamma(1-t)\mathrm{e}^{-t\gamma}, \quad \mathrm{R}et < 1; \end{aligned} \tag{12.2.5}$$

hence the moment generating function of $\Lambda = \sum \frac{1}{k}(\xi_k - 1) + \gamma$ equals $\Gamma(1-t)$, $\mathrm{R}et < 1$. Now, if $X \in \mathrm{Exp}(1)$, then $-\log X$ has the moment generating function $\mathbb{E}[\mathrm{e}^{-t\log X}] = \mathbb{E}[X^{-t}] = \int_0^\infty x^{-t}\mathrm{e}^{-x}dx = \Gamma(1-t)$ too. Thus, $\Lambda \stackrel{d}{=} -\log X$. (Recall that the restriction of the moment generating function to the imaginary axis yields the characteristic function, which determines the distribution.) Hence,

$$\mathbb{P}(\Lambda \le x) = \mathbb{P}(\log X \ge -x) = \mathbb{P}(X \ge \mathrm{e}^{-x}) = \mathrm{e}^{-\mathrm{e}^{-x}}, \quad -\infty < x < \infty, \tag{12.2.6}$$

which is one of the standard extreme value distributions [156]. Consequently, the right-hand side of (12.2.2) can be written $\Lambda + \Lambda' - 2\gamma$, where Λ and Λ' are independent random variables with the distributions (12.2.6).

Moreover, the logistic distribution has the moment generating function, for $|\mathrm{R}et| < 1$, with $\eta \in U(0,1)$ as above,

$$\mathbb{E}[\mathrm{e}^{t\log(\eta/(1-\eta))}] = \int_0^1 x^t(1-x)^{-t}dx = B(1+t, 1-t) = \Gamma(1+t)\Gamma(1-t), \tag{12.2.7}$$

which equals the moment generating function of the symmetrization $\Lambda - \Lambda'$. Thus $\zeta \stackrel{d}{=} \Lambda - \Lambda'$. We can now restate Theorem 12.5 as follows.

Theorem 12.6. *Suppose that the distribution function* $\mathbb{P}(X_{ij} \le t) = t + o(t/|\log t|)$ *as* $t \searrow 0$. *Then, as* $n \to \infty$,

$$nW_{ij} - \log n \xrightarrow{d} \Lambda_1 + \Lambda_2 - \Lambda_3 \tag{12.2.8}$$

and

$$nY_i - 2\log n \xrightarrow{d} \Lambda_1 + \Lambda_2, \tag{12.2.9}$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ *are independent random variables with the same extreme value distribution* $P(\Lambda_i \le x) = \mathrm{e}^{-\mathrm{e}^{-x}}$. □

The variables on the right-hand sides of (12.2.8) and (12.2.9) have the moment generating functions $\Gamma(1-t)^2\Gamma(1+t)$, $|\mathrm{R}et| < 1$, and $\Gamma(1-t)^2$, $\mathrm{R}et < 1$, respectively, and thus the characteristic functions $\Gamma(1-it)^2\Gamma(1+it)$ and $Gamma(1-it)^2$. The limit $\Lambda_1 + \Lambda_2$ in (12.2.9) has a density function that can be expressed using modified Bessel functions as $2\mathrm{e}^{-x}K_0(2\mathrm{e}^{-x/2})$ (cf., for instance, [**?**, (5.10.23)]). We do not know any simple expression for the density function of $\Lambda_1 + \Lambda_2 - \Lambda_3$.

Using the fact that the variance of the logistic distribution is $\pi^2/3$ (which follows from its moment generating function $\Gamma(1+t)\Gamma(1-t) = \pi t/\sin \pi t$, $|\mathrm{R}et < 1$, or from the representation $\Lambda - \Lambda'$ above), it is easily seen that the limiting variables in (12.2.1) and (12.2.2) have expectations 0 and variances $\sum_1^\infty k^{-2} + \pi^2/3 = \pi^2/2$ and $2\sum_1^\infty k^{-2} = \pi^2/3$, respectively. Since all approximations and limits in the proof hold in L^2 sense, we obtain that these are the limits of the expectations and variances of the left-hand sides in (12.2.1) and (12.2.2) too, provided $X_{ij} \in \mathrm{Exp}(1)$. This carries over to other distributions as well, in particular to the uniform distribution; we have the following theorem.

Theorem 12.7. *Suppose that the distribution function* $\mathbb{P}(X_{ij} \le t) = t + o(t/|\log t|)$ *as* $t \searrow 0$ *and that* $\mathbb{E}[X_{ij}^p] < \infty$ *for some* $p > 0$. *Then all moments converge in* (12.2.1), (12.2.2), (12.2.8) *and* (12.2.9)*; in particular*

$$\mathbb{E}[W_{ij}] = \frac{\log n}{n} + \frac{\gamma}{n} + o\left(\frac{1}{n}\right), \tag{12.2.10}$$

$$\mathbb{E}[Y_i] = 2\frac{\log n}{n} + \frac{2\gamma}{n} + o\left(\frac{1}{n}\right), \tag{12.2.11}$$

$$\mathrm{Var}(W_{ij}) \sim \frac{pi^2}{2n^2}, \tag{12.2.12}$$

$$\mathrm{Var}(Y_i) \sim \frac{pi^2}{3n^2}. \tag{12.2.13}$$

Proof. It suffices to prove that $\mathbb{E}[(nW_{ij} - \log n)^m] = O(1)$ and $\mathbb{E}[(nY_i - 2\log n)^m] = O(1)$ for every even integer m and n large enough, since this implies convergence of all moments of order $< m$ by a standard result on uniform integrability.

When T_{ij} is exponentially distributed, this can be done as for the case $m = 2$ in the proof of Theorem 12.1; we omit the details.

In general, we let a and b be two constants, to be chosen later, and split the expectation into three parts. (We treat only W_{ij}; the same argument applies to Y_i.)

First, $\mathbb{E}[(nW_{ij} - \log n)^m \mathbb{1}_{\{W_{ij} \le a \log n/n\}}] = O(1)$ by comparison with the exponential case, using the coupling argument as in earlier proofs.

Secondly, by (12.1.11), $\mathbb{E}[(nW_{ij})^m \mathbb{1}_{\{a \log n/n < W_{ij} < b\}}] \le b^m n^m \mathbb{P}(W_{ij} > a \log n/n) = O(n^{m+2-a} \log^2 n)$; choosing $a = m + 3$ this becomes bounded.

Finally, considering only the $n - 2$ paths of length 2 between i and j, we see that

$$\begin{aligned}\mathbb{P}(W_{ij} > x) &\le \mathbb{P}(X_{ik} > x/2 \text{ or } X_{jk} > x/2 \text{ for every } k \ne i, j)\\ &\le (2\mathbb{P}(X_{ij} > x/2))^{n-2}.\end{aligned} \tag{12.2.14}$$

Now, if $\mathbb{E}[X_{ij}^p] < \infty$, then $x6p\mathbb{P}(X_{ij} > x) \to 0$ as $x \to \infty$; it follows that if b is large enough, then $2\mathbb{P}(X_{ij} > x/2) < x^{-p}$ when $x \ge b$, and thus

$$\mathbb{P}(W_{ij} > x) \le x^{-(n-2)p}, \quad x \ge b. \tag{12.2.15}$$

Consequently,

$$\begin{aligned}\mathbb{E}[(nW_{ij})^m \mathbb{1}_{\{W_{ij} > b\}}] &= n^m b^m \mathbb{P}(W_{ij} > m) + n^m \int_b^\infty m x^{m-1} \mathbb{P}(W_{ij} > x) dx\\ &= O(n^m b^{-np}) = O(1),\end{aligned} \tag{12.2.16}$$

provided $n > 2 + m/p$.

Combining these estimates we find $\mathbb{E}[(nW_{ij} - \log n)^m] = O(1)$ as required. □

Remark 2. The asymptotic variances can also be obtained by refining the estimates used in the proof of Theorem 12.1.

Remark 3. The condition that $\mathbb{E}[X_{ij}^p] < \infty$ for some $p > 0$ is necessary tool if it fails then W_{ij} has no finite moment for any n. In fact, suppose that, for instance, $\mathbb{E}[W_{ij}] < \infty$ for some n; then $\mathbb{P}(W_{ij} > t) < 1/t$ for large t. Since $\mathbb{P}(W_{ij} > t \ge \mathbb{P}(X_{ik} > t \text{ for every } k \ne i) = \mathbb{P}(X_{ij} > t)^{n-1}$, this yields $\mathbb{P}(X_{ij} > t) < t^{-1/(n-1)}$ (t large), and thus for example $\mathbb{E}[X_{ij}^{1/n}] < \infty$.

We do not know any similar results for $\max_{i,j} W_{ij}$.

What is the asymptotic distribution of $\max_{i,j} W_{ij}$ (presuming that some exists)?

What is the order of $\mathrm{Var}(\max_{i,j} W_{ij})$? Is it $\sim c/n^2$? If so, what is the constant c?

12.3 Lengths of minimal paths

We have so far studied the weights of the minimal paths, but one might also ask how long they are, disregarding their weights, that is, how many edges they contain. Let H_{ij} be the length of the path between i and j that has minimal weight.

For the case of exponentially distributed T_{ij}, these lengths can be studied by observing that the proof of Theorem 12.1 shows that the collection of minimal weight paths from a given vertex, 1 say, form a tree (rooted at 1) which can be constructed as follows. Begin with a single root and add $n-1$ vertices one by one, each time joining the new vertex to a (uniformly) randomly chosen old vertex. This type of random tree is known as a random recursive tree, and it is known that, if D_k is the depths of the kth vertex, then $D_n/\log n \xrightarrow{\mathbb{P}} 1$ [171] and $\max_{k\le n} D_k/\log n \xrightarrow{\mathbb{P}} \mathrm{e}$ [192] as $n\to\infty$; see also the survey [204].

This leads to the following result; our condition on the distribution of X_{ij} is presumably much stronger than really required.

Theorem 12.8. *Suppose that X_{ij} has a density function $f(t) = 1 + O(t)$ for $t > 0$. Then, as $n\to\infty$:*

(i) for any fixed i and j,

$$\frac{H_{ij}}{\log n} \xrightarrow{\mathbb{P}} 1; \tag{12.3.1}$$

(ii) for any fixed i,

$$\frac{\max_{j\le n} H_{ij}}{\log n} \xrightarrow{\mathbb{P}} \mathrm{e}. \tag{12.3.2}$$

Proof. The case when $X_{ij} \in \mathrm{Exp}(1)$ follows from the discussion before the theorem: we have $H_{ij} = D_N$, where N is uniformly distributed over $2,\dots,n$ and $\max_{j\le n} H_{ij} = \max_{k\le n} D_k$.

In general, we first observe that we may, for a given n, modify the distribution of X_{ij} on the interval $t \ge 5\log n/n$ without affecting the result, since, by Theorem 12.1, edges with such large weights are hardly ever used. Hence we may assume that its density function is $1 + O(\log n/n)$ times the density function e^{-t} of the exponential distribution, uniformly for all $t > 0$. It is now easy to see that the minimum weight paths from $i = 1$ form a random tree, obtained by adding vertices one by one as above, with the modification that the probability that the kth vertex (in order of insertion) is joined to the lth, for $l < k$, may depend on the previous history of the tree but is always $(1 + O(\log n/n))/(k-1)$. We may couple this random tree growing process with the one with equal probabilities $1/(k-1)$ in such a way that the probability that a vertex k is joined to a different preceding vertices in the two trees is $O(\log n/n)$, even if we condition on the previous history. It follows that, if we fix the end vertex j, the path from $i = 1$ to j is the same in both trees with probability $1 - O(\log^2 n/n)$, which, by the result for the exponential case, implies (i) for a general distribution.

For (ii) we observe that, if D_k is the depth of the kth vertex (in order of insertion) in the tree, and $\bar{D}_k$ is the depth in the random recursive tree with uniformly chosen ancestors, then, by the above, $D_k = \bar{D}_k$, for every $k \le n_1 = n/\log^2 n$ with probability $1 - O(n_1 \log n/n) = 1 - O(1/\log n)$. Since $\max_{k\le n_1} \bar{D}_k/\log n_1 \xrightarrow{\mathbb{P}} \mathrm{e}$ by the result quoted above [192], it follows that for every $\varepsilon > 0$, with probability $1 - o(1)$,

$$\max_{k\le n} D_k \ge \max_{k\le n_1} D_k = \max_{k\le n_1} \bar{D}_k \ge (\mathrm{e}-\varepsilon)\log n_1 = (\mathrm{e} - \varepsilon - o(1))\log n, \tag{12.3.3}$$

which by $\max_{j\le n} H_{ij} = \max_{k\le n} D_k$ proves one half of the result.

For the other half, define the generating functions

$$F_m(t) = \mathbb{E}\left[\sum_{k=1}^{m} t^{D_k}\right] \tag{12.3.4}$$

and

$$\bar{F}_m(t) = \mathbb{E}\left[\sum_{k=1}^{m} t^{\bar{D}_k}\right]. \tag{12.3.5}$$

The recursive definition of the tree yields $\mathbb{E}[t^{\bar{D}_{m+1}}] = \frac{1}{m}\bar{F}_m(t)$ and thus

$$\bar{F}_{m+1}(t) = \left(1 + \frac{t}{m}\right)\bar{F}_m(t), \tag{12.3.6}$$

which together with $\bar{D}_1 = 0$ yields

$$\bar{F}_m(t) = \frac{\Gamma(m+t)}{\Gamma(m)\Gamma(1+t)}. \tag{12.3.7}$$

Choosing $t = \mathrm{e}$ we obtain, for every $a > \mathrm{e}$,

$$\mathbb{P}(\max_{k\le n} \bar{D}_k \ge a\log n) \le \mathbb{P}\left(\sum_{k=1}^{n} \mathrm{e}^{\bar{D}_k} \ge n^a\right) \le n^{-a}\bar{F}_n(\mathrm{e}) \sim n^{-a+\mathrm{e}}/\Gamma(\mathrm{e}+1), \tag{12.3.8}$$

which tends to 0 as $n \to \infty$.

For D_k we similarly obtain the inequalities, for some $C < \infty$ and all $t > 0$,

$$\mathbb{E}[t^{D_{m+1}}] \le \frac{t}{m}\left(1 + C\frac{\log n}{n}\right)F_m(t), \tag{12.3.9}$$

$$F_{m+1}(t) \le \left(1 + \frac{t}{m}\left(1 + C\frac{\log n}{n}\right)\right)F_m(t), \tag{12.3.10}$$

and thus

$$F_m(t) \le \bar{F}_m\left(t\left(1 + C\frac{\log n}{n}\right)\right), \tag{12.3.11}$$

which yields, as above,

$$\mathbb{P}(\max_{k\le n} D_k \ge a\log n) \le n^{-a}F_n(\mathrm{e}) \le n^{-a}\bar{F}_n(\mathrm{e} + C\mathrm{e}\log n/n) \sim n^{-a+\mathrm{e}}/\Gamma(\mathrm{e}+1), \tag{12.3.12}$$

which tends to 0 as $n \to \infty$ for $a > \mathrm{e}$.

How large is $\max_{i,j} H_{ij}$?

We can show that, if $\alpha \approx 3.591$ is defined by $\alpha\log\alpha - \alpha = 1$, then, for every $\varepsilon > 0$, $\mathbb{P}(\mathrm{e} - \varepsilon < \max_{i,j} H_{ij}/\log n < \alpha + \varepsilon) \to 1$. Hence it is natural to conjecture that $\max_{i,j} H_{ij}/\log n$ converges in probability to a constant in $[\mathrm{e}, \alpha]$. Which?

Appendix A
Some measure and integration results

In this section, we give some classical results from the theory of measure and integration, which will be used in the course of the proofs. For details and proofs of these results, we refer to the books [38, 106, 87, 114]. For the statements of the results below, we refer to [106, Pages 110-111].

Theorem A.9 (Lebesque's dominated convergence theorem)**.** *Let X_n and Y satisfy $\mathbb{E}[Y] < \infty$, $X_n \xrightarrow{a.s.} X$, and $|X_n| \leq Y$ almost surely. Then*

$$\mathbb{E}[X_n] \to \mathbb{E}[X], \tag{A.13}$$

and $\mathbb{E}[|X|] < \infty$.

We shall also make use of a slight extension, where almost sure convergence is replaced with convergence in distribution:

Theorem A.10 (Lebesque's dominated convergence theorem)**.** *Let X_n and Y satisfy $\mathbb{E}[|X_n|] < \infty$, $\mathbb{E}[Y] < \infty$, $X_n \xrightarrow{d} X$, and $|X_n| \leq Y$. Then*

$$\mathbb{E}[X_n] \to \mathbb{E}[X], \tag{A.14}$$

and $\mathbb{E}[|X|] < \infty$.

Theorem A.11 (Monotone convergence theorem)**.** *Let X_n be a monotonically increasing sequence, i.e., $X_n \leq X_{n+1}$ such that $\mathbb{E}[|X_n|] < \infty$. Then $X_n(\omega) \uparrow X(\omega)$ for all ω and some limiting random variable X, and*

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X]. \tag{A.15}$$

In particular, when $\mathbb{E}[X] = \infty$, then $\mathbb{E}[X_n] \uparrow \infty$.

Theorem A.12 (Fatou's lemma)**.** *If $X_n \geq 0$ and $\mathbb{E}[|X_n|] < \infty$, then*

$$\mathbb{E}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]. \tag{A.16}$$

In particular, if $X_n(\omega) \to X(\omega)$ for every ω, then

$$\mathbb{E}[X] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]. \tag{A.17}$$

Appendix B
Solutions to selected exercises

Solutions to the exercises of Chapter 1.

Solution to Exercise 1.1. When (1.1.6) holds with equality, then

$$1 - F_X(x) = \sum_{k=x+1}^{\infty} f_k = \sum_{k=x+1}^{\infty} k^{-\tau}.$$

Therefore, by monotonicity of $x \mapsto x^{-\tau}$,

$$1 - F_X(x) \le \int_x^{\infty} y^{-\tau} dy = \frac{x^{1-\tau}}{\tau - 1},$$

while

$$1 - F_X(x) \ge \int_{x+1}^{\infty} y^{-\tau} dy = \frac{(x+1)^{1-\tau}}{\tau - 1}.$$

As a result, we obtain that

$$1 - F_X(x) = \frac{x^{1-\tau}}{\tau - 1}(1 + O(\frac{1}{x})).$$

For an example where (**??**) holds, but (1.1.6) fails, we can take $f_{2k+1} = 0$ for $k \ge 0$ and, for $k \ge 1$,

$$f_{2k} = \frac{1}{k^{\tau-1}} - \frac{1}{(k+1)^{\tau-1}}.$$

Then (1.1.6) fails, while

$$1 - F_X(x) = \sum_{k>x} f_k \sim \frac{1}{\lfloor x/2 \rfloor^{\tau-1}} \sim \frac{1}{x^{\tau-1}}.$$

□

Solution to Exercise 1.2. Recall that a function $x \mapsto L(x)$ is slowly varying when, for every $c > 0$,

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1.$$

For $L(x) = \log x$, we can compute

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = \lim_{x\to\infty} \frac{\log(cx)}{\log x} = \lim_{x\to\infty} \frac{\log x + \log c}{\log x} = 1.$$

For $L(x) = e^{(\log x)^\gamma}$, we compute similarly

$$\begin{aligned} \lim_{x\to\infty} \frac{L(cx)}{L(x)} &= \lim_{x\to\infty} e^{(\log(cx))^\gamma - (\log x)^\gamma} \\ &= \lim_{x\to\infty} e^{\log(x)^\gamma \left((1+\frac{\log c}{\log x})^\gamma - 1\right)} \\ &= \lim_{x\to\infty} e^{\log(x)^{\gamma-1}\gamma \log c} = 1. \end{aligned}$$

When $\gamma = 1$, however, we have that $L(x) = e^{\log x} = x$, which is regularly varying with exponent 1. □

Solutions to the exercises of Chapter 2.

Solution to Exercise 2.1. Take

$$X_n = \begin{cases} Y_1 & \text{for } n \text{ even}, \\ Y_2 & \text{for } n \text{ odd}, \end{cases}$$

where Y_1 and Y_2 are two independent copies of a random variable which is such that $\mathbb{P}(Y_i = \mathbb{E}[Y_i]) < 1$. Then, since Y_1 and Y_2 are identical in distribution, the sequence $\{X_n\}_{n=1}^{\infty}$ converges in distribution. In fact, $\{X_n\}_{n=1}^{\infty}$ is constant in distribution.

Moreover, $X_{2n} \equiv Y_1$ and $X_{2n+1} \equiv Y_2$. Since subsequences of converging sequences are again converging, if $\{X_n\}_{n=1}^{\infty}$ converges in probability, the limit of $\{X_n\}_{n=1}^{\infty}$ should be equal to Y_1 and to Y_2. Since $\mathbb{P}(Y_1 \neq Y_2) > 0$, we obtain a contradiction. □

Solution to Exercise 2.2. Note that for any $\varepsilon > 0$, we have

$$\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n = n) = \frac{1}{n} \to 0. \tag{B.1}$$

Therefore, $X_n \xrightarrow{\mathbb{P}} 0$, which in turn implies that $X_n \xrightarrow{d} 0$. □

Solution to Exercise 2.3. The random variable X with density

$$f_X(x) = \frac{1}{\pi(1+x^2)},$$

which is a *Cauchy random variable*, does the job. □

Solution to Exercise 2.4. Note that, by a Taylor expansion of the moment generating function, if $M_X(t) < \infty$ for all t, then

$$M_X(t) = \sum_{r=0}^{\infty} \mathbb{E}[X^r] \frac{t^r}{r!}.$$

As a result, when $M_X(t) < \infty$ for all t, we must have that

$$\lim_{r\to\infty} \mathbb{E}[X^r] \frac{t^r}{r!} = 0.$$

Thus, when $t > 1$, (2.1.8) follows. Thus, it is sufficient to show that the moment generating function $M_X(t)$ of the Poisson distribution is finite for all t. For this, we compute

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = \exp\{-\lambda(1-e^t)\} < \infty,$$

for all t. □

Solution to Exercise 2.5. We write out

$$\begin{aligned}\mathbb{E}[(X)_r] &= \mathbb{E}[X(X-1)\cdots(X-r+1)] = \sum_{x=0}^{\infty} x(x-1)\cdots(x-r+1)\mathbb{P}(X=x)\\ &= \sum_{x=r}^{\infty} x(x-1)\cdots(x-r+1)e^{-\lambda}\frac{\lambda^x}{x!}\\ &= \lambda^r \sum_{x=r}^{\infty} e^{-\lambda}\frac{\lambda^{x-r}}{(x-r)!} = \lambda^r. \end{aligned} \tag{B.2}$$

□

Solution to Exercise 2.6. Compute that

$$\mathbb{E}[X^m] = e^{-\lambda}\sum_{k=1}^{\infty} k^m \frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty} k^{m-1}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{l=0}^{\infty}(l+1)^{m-1}\frac{\lambda^l}{l!} = \lambda\mathbb{E}[(X+1)^{m-1}].$$

□

Solution to Exercise 2.7. By the discussion around (2.1.16), we have that the sum $\sum_{r=k}^{n}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!}$ is alternatingly larger and smaller than $\mathbb{P}(X=k)$. Thus, it suffices to prove that, when (2.1.18) holds, then also

$$\lim_{n\to\infty}\sum_{r=k}^{n}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!} = \sum_{r=k}^{\infty}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!}. \tag{B.3}$$

This is equivalent to the statement that

$$\lim_{n\to\infty}\sum_{r=n}^{\infty}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!} = 0. \tag{B.4}$$

To prove (B.4), we bound

$$\Big|\sum_{r=n}^{\infty}(-1)^{k+r}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!}\Big| \le \sum_{r=n}^{\infty}\frac{\mathbb{E}[(X)_r]}{(r-k)!k!} \to 0, \tag{B.5}$$

by (2.1.18). □

Solution to Exercise 2.8. For $r=2$, we note that

$$\mathbb{E}[(X)_r] = \mathbb{E}[X^2] - \mathbb{E}[X], \tag{B.6}$$

and, for $X = \sum_{i\in\mathcal{I}} I_i$ a sum of indicators,

$$\mathbb{E}[X^2] = \sum_{i,j}\mathbb{E}[I_iI_j] = \sum_{i\neq j}\mathbb{P}(I_i=I_j=1) + \sum_i \mathbb{P}(I_i=1). \tag{B.7}$$

Using that $\mathbb{E}[X] = \sum_i \mathbb{P}(I_i=1)$, we thus arrive at

$$\mathbb{E}[(X)_r] = \sum_{i\neq j}\mathbb{P}(I_i=I_j=1), \tag{B.8}$$

which is (2.1.21) for $r=2$. □

Solution to Exercise 2.9. For the Poisson distribution factorial moments are given by

$$\mathbb{E}[(X)_k] = \lambda^k$$

(recall Exercise 2.5.) We make use of Theorems 2.4 and 2.5. If X_n is binomial with parameters n and $p_n = \lambda/n$, then

$$\mathbb{E}[(X_n)_k] = \mathbb{E}[X_n(X_n-1)\cdots(X_n-k+1)] = n(n-1)\dots(n-k+1)p^k \to \lambda^k,$$

when $p = \lambda/n$ and $n \to \infty$. □

Solution to Exercise 2.10. We prove Theorem 2.7 by induction on $d \geq 1$. The induction hypothesis is that (2.1.21) holds for all measures $\mathbb{P}$ with corresponding expectations $\mathbb{E}$ and all $r_1, \dots, r_d$.

Theorem 2.7 for $d = 1$ is Theorem 2.5, which initializes the induction hypothesis. We next advance the induction hypothesis by proving (2.1.21) for $d+1$. For this, we first note that we may assume that $\mathbb{E}[(X_{d+1,n})_{r_{d+1}}] > 0$, since $(X_{d+1,n})_{r_{d+1}} \geq 0$ and when $\mathbb{E}[(X_{d+1,n})_{r_{d+1}}] = 0$, then $(X_{d+1,n})_{r_{d+1}} \equiv 0$, so that (2.1.21) follows. Then, we define the measure $\mathbb{P}_{X,d}$ by

$$\mathbb{P}_{X,d}(\mathcal{E}) = \frac{\mathbb{E}\Big[(X_{d+1,n})_{r_{d+1}} \mathbb{1}_{\mathcal{E}}\Big]}{\mathbb{E}[(X_{d+1,n})_{r_{d+1}}]}, \tag{B.9}$$

for all possible measurable events $\mathcal{E}$. Then,

$$\mathbb{E}[(X_{1,n})_{r_1}\cdots(X_{d,n})_{r_d}(X_{d+1,n})_{r_{d+1}}] = \mathbb{E}[(X_{d+1,n})_{r_{d+1}}]\mathbb{E}_{X,d}\Big[(X_{1,n})_{r_1}\cdots(X_{d,n})_{r_d}\Big]. \tag{B.10}$$

By the induction hypothesis applied to the measure $\mathbb{P}_{X,d}$, we have that

$$\mathbb{E}_{X,d}\Big[(X_{1,n})_{r_1}\cdots(X_{d,n})_{r_d}\Big] = \sum_{i_1^{(1)},\dots,i_{r_1}^{(1)} \in \mathcal{I}_1}^{*} \cdots \sum_{i_1^{(d)},\dots,i_{r_d}^{(d)} \in \mathcal{I}_d}^{*} \mathbb{P}_{X,d}\big(I_{i_s}^{(l)} = 1 \forall l = 1,\dots,d \& s = 1,\dots,r_l\big). \tag{B.11}$$

Next, we define the measure $\mathbb{P}_{\vec{i}_d}$ by

$$\mathbb{P}_{\vec{i}_d}(\mathcal{E}) = \frac{\mathbb{E}\Big[\prod_{l=1}^{d} I_{i_s}^{(l)} \mathbb{1}_{\mathcal{E}}\Big]}{\mathbb{P}\big(I_{i_s}^{(l)} = 1 \ \forall l = 1,\dots,d,\ s = 1,\dots,r_l\big)}, \tag{B.12}$$

so that

$$\begin{aligned}\mathbb{E}[(X_{d+1,n})_{r_{d+1}}]&\mathbb{P}_{X,d}\big(I_{i_s}^{(l)} = 1 \ \forall l = 1,\dots,d,\ s = 1,\dots,r_l\big)\\ &= \mathbb{E}_{\vec{i}_d}[(X_{d+1,n})_{r_{d+1}}]\mathbb{P}\big(I_{i_s}^{(l)} = 1 \ \forall l = 1,\dots,d,\ s = 1,\dots,r_l\big).\end{aligned} \tag{B.13}$$

Again by Theorem 2.5,

$$\mathbb{E}_{\vec{i}_d}[(X_{d+1,n})_{r_{d+1}}] = \sum_{i_1^{(d+1)},\dots,i_{r_1}^{(d+1)} \in \mathcal{I}_{d+1}}^{*} \mathbb{P}_{\vec{i}_d}(I_{i_1}^{(d+1)} = \cdots = I_{i_{r_{d+1}}}^{(d+1)} = 1). \tag{B.14}$$

Then, the claim for $d+1$ follows by noting that

$$\begin{aligned}\mathbb{P}\big(I_{i_s}^{(l)} = 1 \ \forall l = 1,\dots,d,\ s = 1,\dots,r_l\big)&\mathbb{P}_{\vec{i}_d}(I_{i_1}^{(d+1)} = \cdots = I_{i_{r_{d+1}}}^{(d+1)} = 1)\\ &= \mathbb{P}\big(I_{i_s}^{(l)} = 1 \ \forall l = 1,\dots,d+1,\ s = 1,\dots,r_l\big).\end{aligned} \tag{B.15}$$

□

Solution to Exercise 2.11. Observe that

$$\sum_x |p_x - q_x| = \sum_x (p_x - q_x)\mathbb{1}_{\{p_x > q_x\}} + \sum_x (q_x - p_x)\mathbb{1}_{\{q_x > p_x\}} \tag{B.16}$$

$$0 = 1 - 1 = \sum_x (p_x - q_x) = \sum_x (p_x - q_x)\mathbb{1}_{\{p_x > q_x\}} + \sum_x (p_x - q_x)\mathbb{1}_{\{q_x > p_x\}}. \tag{B.17}$$

We add the two equalities to obtain

$$\sum_x |p_x - q_x| = 2\sum_x (p_x - q_x)\mathbb{1}_{\{p_x > q_x\}}.$$

Complete the solution by observing that

$$\sum_x (p_x - \min(p_x, q_x)) = \sum_x (p_x - q_x)\mathbb{1}_{\{p_x > q_x\}}.$$

□

Solution to Exercise 2.12. The proof of (2.2.11) is the continuous equivalent of the proof of (2.2.9). Therefore, we will only prove (2.2.9).

Let Ω be the set of possible outcomes of the probability mass functions $\{p_x\}$ and $\{q_x\}$. The set Ω can be partitioned into two subsets

$$\Omega_1 = \{x \in \Omega : p_x \geq q_x\} \qquad \text{and} \qquad \Omega_2 = \{x \in \Omega : p_x < q_x\}.$$

Since $\{p_x\}$ and $\{q_x\}$ are probability distribution functions, the sum $\sum_{x\in\Omega}(p_x - q_x)$ equals zero. Therefore,

$$\begin{aligned} \sum_{x\in\Omega} |p_x - q_x| &= \sum_{x\in\Omega_1} (p_x - q_x) - \sum_{x\in\Omega_2} (p_x - q_x) \\ 0 = \sum_{x\in\Omega} (p_x - q_x) &= \sum_{x\in\Omega_1} (p_x - q_x) + \sum_{x\in\Omega_2} (p_x - q_x) \end{aligned}$$

Adding and subtracting the above equations yields

$$\sum_{x\in\Omega} |p_x - q_x| = 2\sum_{x\in\Omega_1} (p_x - q_x) = -2\sum_{x\in\Omega_2} (p_x - q_x).$$

Hence, there exists a set $A \subseteq \Omega$ such that $|F(A) - G(A)| \geq \frac{1}{2}\sum_{x\in\Omega} |p_x - q_x|$. It remains to show that $|F(A) - G(A)| \leq \frac{1}{2}\sum_{x\in\Omega} |p_x - q_x|$ for all $A \subseteq \Omega$.

Let A be any subset of Ω. Just as the set Ω, the set A can be partitioned into two subsets

$$A_1 = A \cap \Omega_1 \qquad \text{and} \qquad A_2 = A \cap \Omega_2,$$

so that

$$|F(A) - G(A)| = |\sum_{x\in A_1} (p_x - q_x) + \sum_{x\in A_2} (p_x - q_x)| = |\alpha_A + \beta_A|.$$

Since α_A is non-negative and β_A non-positive, it holds that

$$|\alpha_A + \beta_A| \leq \max_A \big(\alpha_A, -\beta_A\big).$$

The quantity α_A satisfies

$$\alpha_A \leq \sum_{x\in\Omega_1} (p_x - q_x) = \frac{1}{2}\sum_{x\in\Omega} |p_x - q_x|,$$

while β_A satisfies

$$\beta_A \geq \sum_{x\in\Omega_2}(p_x - q_x) = -\frac{1}{2}\sum_{x\in\Omega}|p_x - q_x|.$$

Therefore,

$$|F(A) - G(A)| \leq \frac{1}{2}\sum_{x\in\Omega}|p_x - q_x| \qquad \forall A \subseteq \Omega,$$

which completes the proof. □

Solution to Exercise 2.13. By (2.2.13) and (2.2.18)

$$d_{\mathrm{TV}}(f,g) \leq \mathbb{P}(\hat{X} \neq \hat{Y}). \tag{B.18}$$

Therefore, the first claim follows directly from Theorem 2.9. The second claim follows by (2.2.9).

Solution to Exercise 2.15. Without any loss of generality we can take $\sigma^2 = 1$. Then for each t, and with Z a standard normal variate

$$\mathbb{P}(X \geq t) = \mathbb{P}(Z \geq t - \mu_X) \leq \mathbb{P}(Z \geq t - \mu_Y) = \mathbb{P}(Y \geq t),$$

whence $X \preceq Y$. □

Solution to Exercise 2.16. The answer is negative. Take X standard normal and $Y \sim N(0,2)$, then $X \preceq Y$ implies

$$\mathbb{P}(Y \geq t) \geq \mathbb{P}(X \geq t) = \mathbb{P}(Y \geq t\sqrt{2}),$$

for each t. However, this is false for $t < 0$. □

Solution to Exercise 2.17. Let X be Poisson distributed with parameter λ, then

$$\mathbb{E}[e^{tX}] = \sum_{n=0}^{\infty} e^{tn} e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda e^t)^n}{n!} = e^{\lambda(e^t-1)}.$$

Put

$$g(t) = at - \log \mathbb{E}[e^{tX}] = at + \lambda - \lambda e^t$$

then $g'(t) = a - \lambda e^t = 0 \Leftrightarrow t = \log(a/\lambda)$. Hence, $I(a)$ in (2.4.12) is equal to $I(a) = I_\lambda(a) = a(\log(a/\lambda) - 1) + \lambda$ and with $a > \lambda$ we obtain from (2.4.9),

$$\mathbb{P}(\sum_{i=1}^{n} X_i \geq an) \leq e^{-nI_\lambda(a)}.$$

This proves (2.4.17). For $a < \lambda$, we get $g'(t) = a - \lambda e^t = 0$ for $t = \log(a/\lambda) < 0$ and we get again

$$I_\lambda(a) = a(\log a/\lambda - 1) + \lambda.$$

By (2.4.9), with $a < \lambda$, we obtain (2.4.18).

$I_\lambda(\lambda) = 0$ and $\frac{d}{da}I_\lambda(a) = \log a - \log\lambda$, so that for $a < \lambda$ the function $a \mapsto I_\lambda(a)$ decreases, whereas for $a > \lambda$ the function $a \mapsto I_\lambda(a)$ increases. Because $I_\lambda(\lambda) = 0$, this shows that for all $a \neq \lambda$, we have $I_\lambda(a) > 0$. □

Solution to Exercise 2.19. By taking expectations on both sides of (2.5.2),

$$\mathbb{E}[M_n] = \mathbb{E}[\mathbb{E}[M_{n+1}|M_1, M_2, \ldots, M_n]] = \mathbb{E}[M_{n+1}],$$

since according to the theorem of total probability:

$$\mathbb{E}[\mathbb{E}[X|Y_1, \ldots, Y_n]] = \mathbb{E}[X].$$

□

Solution to Exercise 2.20. First we show that $\mathbb{E}[|M_n|] < \infty$. Indeed, since $\mathbb{E}[|X_i|] < \infty$, $\forall i$, and since the fact that X_i is an independent sequence implies that the sequence $|X_i|$ is independent we get

$$\mathbb{E}[|M_n|] = \prod_{i=0}^{n} \mathbb{E}[|X_i|] < \infty.$$

To verify the martingale condition, we write

$$\begin{aligned}
\mathbb{E}[M_{n+1}|X_1, X_2, \ldots, X_n] &= \mathbb{E}\Big[\prod_{i=1}^{n+1} X_i \Big| X_1, X_2, \ldots, X_n\Big] \\
&= \Big(\prod_{i=1}^{n} X_i\Big) \cdot \mathbb{E}[X_{n+1}|X_1, X_2, \ldots, X_n] = M_n \mathbb{E}[X_{n+1}] = M_n \qquad \text{a.s.}
\end{aligned}$$

□

Solution to Exercise 2.21. First we show that $\mathbb{E}[|M_n|] < \infty$. Indeed, since $\mathbb{E}[|X_i|] < \infty \forall i$,

$$\mathbb{E}[|M_n|] = \mathbb{E}\Big|\sum_{i=1}^{n} X_i\Big| \leq \sum_{i=1}^{n} \mathbb{E}|X_i| < \infty.$$

To verify the martingale condition, we write

$$\begin{aligned}
\mathbb{E}[M_{n+1}|M_1, M_2, \ldots, M_n] &= \mathbb{E}[\sum_{i=1}^{n+1} X_i|X_0, X_1, \ldots, X_n] \\
&= \sum_{i=1}^{n} X_i + \mathbb{E}[X_{n+1}|X_0, X_1, \ldots, X_n] = M_n + \mathbb{E}[X_{n+1}] = M_n \qquad \text{a.s.}
\end{aligned}$$

□

Solution to Exercise 2.22. Again we first that $\mathbb{E}[|M_n|] < \infty$. Indeed, since $\mathbb{E}[|X_i|] < \infty \forall i$,

$$\mathbb{E}[|M_n|] = \mathbb{E}\Big|\mathbb{E}[Y|X_0, \ldots, X_n]\Big| \leq \mathbb{E}\Big[\mathbb{E}\big[|Y|\big|X_0, \ldots, X_n\big]\Big] = \mathbb{E}[|Y|] < \infty.$$

To verify the martingale condition, we write

$$\begin{aligned}
\mathbb{E}[M_{n+1}|X_0, \ldots, X_n] &= \mathbb{E}\Big[\mathbb{E}[Y|X_0, \ldots, X_{n+1}]\Big|X_0, \ldots, X_n\Big] \\
&= \mathbb{E}[Y|X_0, \ldots, X_n] = M_n + \mathbb{E}[X_{n+1}] = M_n \qquad \text{a.s.}
\end{aligned}$$

□

Solution to Exercise 2.23. Since M_n is non-negative we have $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = \mu \leq M$, by Exercise 2.19. Hence, according to Theorem 2.21 we have convergence to some limiting random variable M_∞. □

Solution to Exercise 2.24. Since $X_i \geq 0$, we have $M_n = \prod_{i=0}^n X_i \geq 0$, hence the claim is immediate from Exercise 2.23. □

Solution to Exercise 2.25. First,

$$\mathbb{E}[|M_n|] \leq \sum_{i=1}^m \mathbb{E}[|M_n^{(i)}|] < \infty. \tag{B.19}$$

Secondly, since $\mathbb{E}[\max\{X, Y\}] \geq \max\{\mathbb{E}[X], \mathbb{E}[Y]\}$, we obtain

$$\mathbb{E}[M_{n+1}|X_0, \ldots, X_n] = \mathbb{E}\Big[\max_{i=0}^m M_{n+1}^{(i)}|X_0, \ldots, X_n\Big] \geq \max_{i=0}^m \mathbb{E}[M_{n+1}^{(i)}|X_0, \ldots, X_n] \tag{B.20}$$

$$= \max_{i=0}^m M_n^{(i)} = M_n, \tag{B.21}$$

where we use that $\{M_n^{(i)}\}_{n=0}^\infty$ is a sequence of martingales with respect to $\{X_n\}_{n=0}^\infty$.

Solution to Exercise 2.26. We can write

$$M_n = \sum_{i=1}^n I_i - p, \tag{B.22}$$

where $\{I_i\}_{i=1}^\infty$ are i.i.d. indicator variables with $\mathbb{P}(I_i = 1) = 1 - \mathbb{P}(I_i = 0) = p$. Then, $M - n$ has the same distribution as $X - np$, while, by Exercise 2.21, the sequence $\{M_n\}_{n=0}^\infty$ is a martingale with

$$|M_n - M_{n-1}| = |I_n - p| \leq \max\{p, 1-p\} \leq 1 - p, \tag{B.23}$$

since $p \leq 1/2$. Thus, the claim follows from the Azuma-Hoeffding inequality (Theorem 2.23).

Solution to Exercise 2.27. Since $\mathbb{E}[X_i] = 0$, we have, by Exercise 2.21, that $M_n = \sum_{i=1}^n X_i$ is a martingale, with by hypothesis,

$$-1 \leq M_n - M_{n-1} = X_n \leq 1,$$

so that the condition of Theorem 2.23 is satisfied with $\alpha_i = \beta_i = 1$. Since $\mathbb{E}[M_n] = 0$, we have $\mu = 0$ and $\sum_{i=0}^n (\alpha_i + \beta_i)^2 = 4(n+1)$, hence from (2.5.18) we get (2.5.31).

We now compare the Azuma-Hoeffding bound (2.5.31) with the central limit approximation. With $a = x\sqrt{n+1}$, and $\sigma^2 = \mathrm{Var}(X_i)$,

$$\mathbb{P}(|M_n| \geq a) = \mathbb{P}(|M_n| \geq x\sqrt{n+1}) = \mathbb{P}(|M_n|/\sigma\sqrt{n+1} \geq x/\sigma) \to 2(1 - \Phi(x/\sigma)),$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2}\, du$. A well-known approximation tells us that

$$2(1 - \Phi(t)) \sim 2\phi(t)/t = \frac{\sqrt{2}}{t\sqrt{\pi}} e^{-t^2/2},$$

so that by the central limit theorem and this approximation

$$\mathbb{P}(|M_n| \geq a) \sim \frac{\sigma\sqrt{2}}{x\sqrt{\sigma\pi}} e^{-x^2/2\sigma^2} = \frac{\sigma\sqrt{2(n+1)}}{a\sqrt{\pi}} e^{-a^2/2(n+1)\sigma^2}$$

Finally $\sigma^2 \leq 1$, so that the leading order term and with $a = x\sqrt{n+1}$, the inequality of Azuma-Hoefding is quite sharp! □

Solutions to the exercises of Chapter 3.

Solution to Exercise 3.1. When $\eta = 0$, then, since η is a solution of $\eta = G_X(\eta)$, we must have that

$$p_0 = G_X(0) = 0. \tag{B.24}$$

□

Solution to Exercise 3.2. We note that for $p = \{p_x\}_{x=0}^{\infty}$ given in (3.1.15), and writing $q = 1 - p$, we have that $\mathbb{E}[X] = 2p$, so that $\eta = 1$ when $p \leq 1/2$, and

$$G_X(s) = q + ps^2. \tag{B.25}$$

Since η satisfies $\eta = G(\eta)$, we obtain that

$$\eta = q + p\eta^2, \tag{B.26}$$

of which the solutions are

$$\eta = \frac{1 \pm \sqrt{1 - 4pq}}{2p}. \tag{B.27}$$

Noting further that $1 - 4pq = 1 - 4p(1-p) = 4p^2 - 4p + 1 = (2p-1)^2$, and $p > 1/2$, we arrive at

$$\eta = \frac{1 \pm (2p-1)}{2p}. \tag{B.28}$$

Since $\eta \in [0, 1)$ for $p > 1/2$, we must have that

$$\eta = \frac{1 - (2p-1)}{2p} = \frac{1-p}{p}. \tag{B.29}$$

□

Solution to Exercise 3.3. We compute that

$$G_X(s) = 1 - b/p + \sum_{k=1}^{\infty} b(1-p)^{k-1}s^k = 1 - \frac{b}{p} + \frac{bs}{1-qs}, \tag{B.30}$$

so that

$$\mu = G_X'(1) = \frac{b}{p^2}. \tag{B.31}$$

As a result, $\eta = 1$ if $\mu = b/p^2 \leq 1$ follows from Theorem 3.1. Now, when $\mu = b/p^2 > 1$, then $\eta < 1$ is the solution of $G_X(\eta) = \eta$, which becomes

$$1 - \frac{b}{p} + \frac{b\eta}{1-q\eta} = \eta, \tag{B.32}$$

which has the solution given by (3.1.18). □

Solution to Exercise 3.4. We note that $s \mapsto G_X(s)$ in (B.30) has the property that for any points s, u, v

$$\frac{G_X(s) - G_X(u)}{G_X(s) - G_X(v)} = \frac{s-u}{s-v}\frac{1-qv}{1-qu}. \tag{B.33}$$

Taking $u = \eta, v = 1$ and using that $G_X(\eta) = \eta$ by Theorem 3.1, we obtain that, if $\eta < 1$,

$$\frac{G_X(s) - \eta}{G_X(s) - 1} = \frac{s-\eta}{s-1}\frac{p}{1-q\eta}. \tag{B.34}$$

By (3.1.18), we further obtain that

$$\frac{p}{1-q\eta} = \mu^{-1} = p^2/b, \tag{B.35}$$

so that we arrive at

$$\frac{G_X(s) - \eta}{G_X(s) - 1} = \frac{1}{\mu}\frac{s-\eta}{s-1}. \tag{B.36}$$

Since $G_n(s)$ is the n-fold iteration of $s \mapsto G_X(s)$, we thus arrive at

$$\frac{G_n(s) - \eta}{G_n(s) - 1} = \frac{1}{\mu^n}\frac{s-\eta}{s-1}, \tag{B.37}$$

of which the solution is given by the first line of (3.1.19).

When $\mu = 1$, then we have that $b = p^2$, so that

$$G_X(s) = \frac{q - (q-p)s}{1-qs}. \tag{B.38}$$

We now prove by induction that $G_n(s)$ is equal to the second line of (3.1.19). For $n = 1$, we have that $G_1(s) = G_X(s)$, so that the induction is initialized by (B.38).

To advance the induction, we assume it for n and advance it to $n+1$. For this, we note that, since $G_n(s)$ is the n-fold iteration of $s \mapsto G_X(s)$, we have

$$G_{n+1}(s) = G_n(G_X(s)). \tag{B.39}$$

By the induction hypothesis, we have that $G_n(s)$ is equal to the second line of (3.1.19), so that

$$G_{n+1}(s) = \frac{nq - (nq-p)G(s)}{p + nq - nqG_X(s)} = \frac{nq(1-qs) - (nq-p)(q-(q-p)s)}{(p+nq)(1-qs) - nq(q-(q-p)s)}. \tag{B.40}$$

Note that, using $p = 1 - q$,

$$\begin{aligned} nq(1-qs) - (nq-p)(q-(q-p)s) &= \big[nq - (nq-p)q\big] + s\big[(q-p)(nq-p) - nq^2\big] \\ &= (n+1)qp - s[qp(n+1) - p^2], \end{aligned} \tag{B.41}$$

while

$$\begin{aligned} (p+nq)(1-qs) - nq(q-(q-p)s) &= \big[(p+nq) - nq^2\big] + s\big[(q-p)nq - (p+nq)q\big] \\ &= [p + nqp] - s(n+1)pq = p[p + (n+1)q] - s(n+1)pq, \end{aligned} \tag{B.42}$$

and dividing (B.41) by (B.42) advances the induction hypothesis. □

Solution to Exercise 3.5. We first note that

$$\mathbb{P}(Z_n > 0,\ \exists m > n \text{ such that } Z_m = 0) = \mathbb{P}(\exists m > n \text{ such that } Z_m = 0) - \mathbb{P}(Z_n = 0) = \eta - \mathbb{P}(Z_n = 0). \tag{B.43}$$

We next compute, using (3.1.19),

$$\mathbb{P}(Z_n = 0) = G_n(0) = \begin{cases} 1 - \mu^n \frac{1-\eta}{\mu^n - \eta} & \text{when } b \neq p^2; \\ \frac{nq}{p+nq} & \text{when } b = p^2. \end{cases} \tag{B.44}$$

Using that $\eta = 1$ when $b \leq p^2$ gives the first two lines of (3.1.20). When $\eta < 1$, so that $\mu > 1$, we thus obtain

$$\mathbb{P}(Z_n > 0,\ \exists m > n \text{ such that } Z_m = 0) = (1-\eta)\Big[\frac{\mu^n}{\mu^n - \eta} - 1\Big] = \frac{(1-\eta)\eta}{\mu^n - \eta}. \tag{B.45}$$

This proves the third line of (3.1.20). □

Solution to Exercise 3.6. By (B.25), we have that $G(s) = q + ps^2$. Thus, by (3.1.23), we obtain

$$G_T(s) = s\big(q + pG_T(s)^2\big), \tag{B.46}$$

of which the solutions are given by

$$G_T(s) = \frac{1 \pm \sqrt{1 - 4s^2pq}}{2sp}. \tag{B.47}$$

Since $G_T(0) = 0$, we must that that

$$G_T(s) = \frac{1 - \sqrt{1 - 4s^2pq}}{2sp}. \tag{B.48}$$

□

Solution to Exercise 3.7. By (B.30), we have $G_X(s) = 1 - \frac{b}{p} + \frac{bs}{1-qs}$. Thus, by (3.1.23), we obtain

$$G_T(s) = s\Big[1 - \frac{b}{p} + \frac{bG_T(s)}{1 - qG_T(s)}\Big]. \tag{B.49}$$

Multiplying by $p(1 - qG_T(s))$, and using that $p + q = 1$, leads to

$$pG_T(s)(1-qG_T(s)) = s\big[(p-b)(1-qG_T(s)) + bpG_T(s)\big] = s\big[(p-b) + (b-pq)G_T(s)\big]. \tag{B.50}$$

We can simplify the above to

$$pqG_T(s)^2 + (p + s(b-pq))G_T(s) + s(p-b) = 0, \tag{B.51}$$

of which the two solutions are given by

$$G_T(s) = \frac{-(p + sbq) \pm \sqrt{(p + s(b-pq))^2 - 4pqs(p-b)}}{2pq}. \tag{B.52}$$

Since $G_T(s) \geq 0$ for all $s \geq 0$, we thus arrive at

$$G_T(s) = \frac{\sqrt{(p + s(b-pq))^2 - 4pqs(p-b)} - (p + sbq)}{2pq}. \tag{B.53}$$

□

Solution to Exercise 3.8. Compute

$$\begin{aligned}\mathbb{E}[Z_n|Z_{n-1}=m] &= \mathbb{E}[\textstyle\sum_{i=1}^{Z_{n-1}} X_{n,i}|Z_{n-1}=m] = \mathbb{E}[\textstyle\sum_{i=1}^{m} X_{n,i}|Z_{n-1}=m] \\ &= \textstyle\sum_{i=1}^{m} \mathbb{E}[X_{n,i}] = m\mu,\end{aligned}$$

so that, by taking double expectations,

$$\mathbb{E}[Z_n] = \mathbb{E}[\mathbb{E}[Z_n|Z_{n-1}]] = \mathbb{E}[\mu Z_{n-1}] = \mu\mathbb{E}[Z_{n-1}].$$

□

Solution to Exercise 3.9. Using induction we conclude from the previous exercise that

$$\mathbb{E}[Z_n] = \mu\mathbb{E}[Z_{n-1}] = \mu^2\mathbb{E}[Z_{n-2}] = \ldots = \mu^n\mathbb{E}[Z_0] = \mu^n.$$

Hence,

$$\mathbb{E}[\mu^{-n}Z_n] = \mu^{-n}\mathbb{E}[Z_n] = 1.$$

Therefore, we have that, for all $n \geq 0$, $\mathbb{E}[|\mu^{-n}Z_n|] = \mathbb{E}[\mu^{-n}Z_n] < \infty$

By the Markov property and the calculations in the previous exercise

$$\mathbb{E}[Z_n|Z_1, \ldots, Z_{n-1}] = \mathbb{E}[Z_n|Z_{n-1}] = \mu Z_{n-1},$$

so that, with $M_n = Z_n/\mu^n$,

$$\mathbb{E}[M_n|Z_1, \ldots, Z_{n-1}] = \mathbb{E}[M_n|Z_{n-1}] = \frac{1}{\mu^n}\mu Z_{n-1} = M_{n-1},$$

almost surely. Therefore, $M_n = \mu^{-n}Z_n$ is a martingale with respect to $\{Z_n\}_{n=1}^{\infty}$. □

Solution to Exercise 3.10. For a critical BP we have $\mu = 1$, and so Z_n is a martingale. Therefore, for all n,

$$\mathbb{E}[Z_n] = \mathbb{E}[Z_0] = 1.$$

On the other hand, if $\mathbb{P}(X = 1) < 1$, then, $\eta = 1$ by Theorem 3.1, and by monotonicity,

$$\lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \mathbb{P}(\lim_{n\to\infty} Z_n = 0) = \eta = 1.$$

□

Solution to Exercise 3.11.

$$\mathbb{P}(Z_n > 0) = \mathbb{P}(Z_n \geq 1) \leq \mathbb{E}[Z_n] = \mu^n,$$

by Theorem 3.3. □

Solution to Exercise 3.12. Since $T = 1 + \sum_{n=1}^{\infty} Z_n$, we obtain by (3.2.1) that

$$\mathbb{E}[T] = 1 + \sum_{n=1}^{\infty} \mathbb{E}[Z_n] = 1 + \sum_{n=1}^{\infty} \mu^n = 1/(1-\mu). \tag{B.54}$$

□

Solution to Exercise 3.13. For $k = 1$, we note that, in (3.3.2), $\{T = 1\} = \{X_1 = 0\}$, so that

$$\mathbb{P}(T = 1) = p_0. \tag{B.55}$$

On the other hand, in (3.1.21), $T = 1$ precisely when $Z_1 = X_{1,1} = 0$, which occurs with probability p_0 as well.

For $k = 2$, since $X_i \geq 0$, we have that $\{T = 2\} = \{X_1 = 1, X_2 = 0\}$, so that

$$\mathbb{P}(T = 2) = p_0 p_1. \tag{B.56}$$

On the other hand, in (3.1.21), $T = 2$ precisely when $Z_1 = X_{1,1} = 1$ and $Z_2 = X_{2,1} = 0$, which occurs with probability $p_0 p_1$ as well, as required.

For $k = 3$, since $X_i \geq 0$, we have that $\{T = 3\} = \{X_1 = 2, X_2 = X_3 = 0\} \cup \{X_1 = X_2 = 1, X_3 = 0\}$, so that

$$\mathbb{P}(T = 3) = p_0^2 p_2 + p_0 p_1^2. \tag{B.57}$$

On the other hand, in (3.1.21),

$$\{T = 3\} = \{Z_1 = Z_2 = 1, Z_3 = 0\} \cup \{Z_1 = 2, Z_2 = 0\}, \tag{B.58}$$

so that $\{T = 3\} = \{X_{1,1} = X_{2,1} = 1, X_{3,1} = 0\} \cup \{X_{1,1} = 2, X_{2,1} = X_{2,2} = 0\}$, which occurs with probability $p_0^2 p_2 + p_0 p_1^2$ as well, as required. This proves the equality of $\mathbb{P}(T = k)$ for T in (3.3.2) and (3.1.21) and $k = 1, 2$ and 3. □

Solution to Exercise 3.14. We note that

$$\mathbb{P}\big(S_0 = S_{k+1} = 0, S_i > 0 \ \forall 1 \leq i \leq k\big) = p\mathbb{P}\big(S_1 = 1, S_i > 0 \ \forall 1 \leq i \leq k, S_{k+1} = 0\big), \tag{B.59}$$

since the first step must be upwards. By (3.3.2),

$$\mathbb{P}\big(S_1 = 1, S_i > 0 \ \forall 1 \leq i \leq k, S_{k+1} = 0\big) = \mathbb{P}(T = k), \tag{B.60}$$

which completes the proof. □

Solution to Exercise 3.15. We note that $p'_x \geq 0$ for all $x \in \mathbb{N}$. Furthermore,

$$\sum_{x=0}^{\infty} p'_x = \sum_{x=0}^{\infty} \eta^{x-1} p_x = \eta^{-1} \sum_{x=0}^{\infty} \eta^x p_x = \eta^{-1} G(\eta). \tag{B.61}$$

Since η satisfies $\eta = G(\eta)$, it follows also that $p' = \{p'_x\}_{x=0}^{\infty}$ sums up to 1, so that p' is a probability distribution. □

Solution to Exercise 3.16. We compute

$$G_d(s) = \sum_{x=0}^{\infty} s^x p'_x = \sum_{x=0}^{\infty} s^x \eta^{x-1} p_x = \eta^{-1} \sum_{x=0}^{\infty} (\eta s)^x p_x = \frac{1}{\eta} G_X(\eta s). \tag{B.62}$$

□

Solution to Exercise 3.17. We note that

$$\mathbb{E}[X'] = \sum_{x=0}^{\infty} x p'_x = \sum_{x=0}^{\infty} x \eta^{x-1} p_x = G'_X(\eta). \tag{B.63}$$

Now, η is the smallest solution of $\eta = G_X(\eta)$, and, when $\eta > 0$, $G_X(0) = p_0 > 0$ by Exercise 3.1. Therefore, since $s \mapsto G'_X(s)$ is increasing, we must have that $G'_X(\eta) < 1$. □

Solution to Exercise 3.18. Since $M_n = \mu^{-n} Z_n \xrightarrow{a.s.} W_\infty$ by Theorem 3.9, by Lebesques dominated convergence theorem and the fact that, for $y \geq 0$ and $s \in [0,1]$, we have that $s^y \leq 1$, it follows that

$$\mathbb{E}[s^{M_n}] \to \mathbb{E}[s^{W_\infty}]. \tag{B.64}$$

However,

$$\mathbb{E}[s^{M_n}] = \mathbb{E}[s^{Z_n/\mu_n}] = G_n(s^{\mu^{-n}}). \tag{B.65}$$

Since $G_n(s) = G_X(G_{n-1}(s))$, we thus obtain

$$\mathbb{E}[s^{M_n}] = G_X\big(G_{n-1}(s^{\mu^{-n}})\big) = G_X\Big(G_{n-1}\big((s^{\mu^{-1}})^{\mu^{-n-1}}\big)\Big) \to G_X\big(G_W(s^{1/\mu})\big), \tag{B.66}$$

again by (B.64).

Solution to Exercise 3.19. If $M_n = 0$, then $M_m = 0$ for all $m \geq n$, so that

$$\{M_\infty = 0\} = \lim_{n\to\infty} \{M_n = 0\} = \cap_{n=0}^\infty \{M_n = 0\}.$$

On the other hand, $\{\text{extinction}\} = \{\exists n : M_n = 0\}$ or $\{\text{survival}\} = \{\forall n, M_n > 0\}$. We hence conclude that $\{\text{survival}\} \subset \{M_\infty > 0\} = \cup_{n=0}^\infty \{M_n > 0\}$, and so

$$\mathbb{P}(M_\infty > 0 | \text{survival}) = \frac{\mathbb{P}(M_\infty > 0 \cap \{\text{survival}\})}{\mathbb{P}(\text{survival})} = \frac{\mathbb{P}(M_\infty > 0)}{1-\eta} = 1,$$

because it is given that $\mathbb{P}(W_\infty > 0) = 1 - \eta$. □

Solution to Exercise 3.20. By Theorem 3.9, we have that $M_n = \mu^{-n} Z_n \xrightarrow{a.s.} W_\infty$. By Fubini's theorem, we thus obtain that

$$\mathbb{E}[W_\infty] \leq \lim_{n\to\infty} \mathbb{E}[M_n] = 1, \tag{B.67}$$

where the equality follows from Theorem 3.3. □

Solution to Exercise 3.30. The total offspring equals $T = 1 + \sum_{n=1}^\infty Z_n$, see (3.1.21). Since we search for $T \leq 3$, we must have $\sum_{n=1}^\infty Z_n \leq 2$ or $\sum_{n=1}^2 Z_n \leq 2$, because $Z_k > 0$ for some $k \geq 3$ implies $Z_3 \geq 1, Z_2 \geq 1, Z_1 \geq 1$, so that $\sum_{n=1}^\infty Z_n \geq \sum_{n=1}^3 Z_n \geq 3$. Then, we can write out

$$\begin{aligned}
\mathbb{P}(T=1) &= \mathbb{P}(\sum_{n=1}^2 Z_n = 0) = \mathbb{P}(Z_1 = 0) = e^{-\lambda}, \\
\mathbb{P}(T=2) &= \mathbb{P}(\sum_{n=1}^2 Z_n = 1) = \mathbb{P}(Z_1 = 1, Z_2 = 0) = \mathbb{P}(X_{1,1} = 1)\mathbb{P}(X_{2,1} = 0) = \lambda e^{-2\lambda} \\
\mathbb{P}(T=3) &= \mathbb{P}(\sum_{n=1}^2 Z_n = 2) = \mathbb{P}(Z_1 = 1, Z_2 = 1, Z_3 = 0) + \mathbb{P}(Z_1 = 2, Z_2 = 0) \\
&= \mathbb{P}(X_{1,1} = 1, X_{2,1} = 1, X_{3,1} = 0) + \mathbb{P}(X_{1,1} = 2, X_{2,1} = 0, X_{2,2} = 0) \\
&= (\lambda e^{-\lambda})^2 \cdot e^{-\lambda} + e^{-\lambda}(\lambda^2/2) \cdot e^{-\lambda} \cdot e^{-\lambda} = e^{-3\lambda} \frac{3\lambda^2}{2}.
\end{aligned}$$

These answers do coincide with $\mathbb{P}(T = n) = e^{-n\lambda} \frac{(n\lambda)^{n-1}}{n!}$, for $n \leq 3$.

□

Solutions to the exercises of Chapter 4.

Solution to Exercise 4.3. We start by computing $\mathbb{P}(T=m)$ for $m=1,2,3$. For $m=1$, we get

$$\mathbb{P}(T=1)=\mathbb{P}(S_1=0)=\mathbb{P}(X_1=0)=\mathbb{P}(\mathrm{Bin}(n-1,p)=0)=(1-p)^{n-1}.$$

For $m=2$, we get

$$\begin{aligned}\mathbb{P}(T=2) &= \mathbb{P}(S_1>0,S_2=0)=\mathbb{P}(X_1>0,X_1+X_2=1)=\mathbb{P}(X_1=1,X_2=0)\\ &= \mathbb{P}(X_1=1)\mathbb{P}(X_2=0|X_1=1)=\mathbb{P}(\mathrm{Bin}(n-1,p)=1)\mathbb{P}(\mathrm{Bin}(n-2,p)=0)\\ &= (n-1)p(1-p)^{n-2}\cdot(1-p)^{n-2}=(n-1)p(1-p)^{2n-4}.\end{aligned}$$

For $m=3$, we get

$$\begin{aligned}\mathbb{P}(T=3) &= \mathbb{P}(S_1>0,S_2>0,S_3=0)=\mathbb{P}(X_1>0,X_1+X_2>1,X_1+X_2+X_3=2)\\ &= \mathbb{P}(X_1=1,X_2=1,X_3=0)+\mathbb{P}(X_1=2,X_2=0,X_3=0)\\ &= \mathbb{P}(X_3=0|X_2=1,X_1=1)\mathbb{P}(X_2=1|X_1=1)\mathbb{P}(X_1=1)\\ &\qquad +\mathbb{P}(X_3=0|X_2=0,X_1=2)\mathbb{P}(X_2=0|X_1=2)\mathbb{P}(X_1=2)\\ &= \mathbb{P}(X_3=0|S_2=1)\mathbb{P}(X_2=1|S_1=1)\mathbb{P}(X_1=1)\\ &\quad +\mathbb{P}(X_3=0|S_2=1)\mathbb{P}(X_2=0|S_1=2)\mathbb{P}(X_1=2)\\ &= \mathbb{P}(\mathrm{Bin}(n-3,p)=0)\mathbb{P}(\mathrm{Bin}(n-2,p)=1)\mathbb{P}(\mathrm{Bin}(n-1,p)=1)\\ &\quad +\mathbb{P}(\mathrm{Bin}(n-3,p)=0)\mathbb{P}(\mathrm{Bin}(n-3,p)=0)\mathbb{P}(\mathrm{Bin}(n-1,p)=2)\\ &= (1-p)^{n-3}(n-2)p(1-p)^{n-3}(n-1)p(1-p)^{n-2}\\ &\quad +(1-p)^{n-3}(1-p)^{n-3}(n-1)(n-2)p^2(1-p)^{n-3}/2\\ &= (n-1)(n-2)p^2(1-p)^{3n-8}+(n-1)(n-2)p^2(1-p)^{3n-9}/2\\ &= (n-1)(n-2)p^2(1-p)^{3n-9}(\frac{3}{2}-p).\end{aligned}$$

We now give the combinatoric proof. For $m=1$,

$$\mathbb{P}(|\mathcal{C}(v)|=1)=(1-p)^{n-1},$$

because all connections from vertex 1 have to be closed. For $m=2$,

$$\mathbb{P}(|\mathcal{C}(v)|=2)=(n-1)p(1-p)^{2n-4}$$

because you must connect one of $n-1$ vertices to vertex v and then isolate these two vertices which means that $2n-4$ connections should not be present.

For $m=3$, the first possibility is to attach one vertex a to 1 and then a second vertex b to a, with the edge vb being closed. This gives

$$(n-1)p(1-p)^{n-2}(n-2)p(1-p)^{n-3}(1-p)^{n-3}=(n-1)(n-2)p^2(1-p)^{3n-8}.$$

The second possibility is to attach one vertex a to v and then a second vertex b to a, with the edge vb being occupied. This gives

$$\binom{n-1}{2}p(1-p)^{n-3}p(1-p)^{n-3}(1-p)^{n-3}p=\binom{n-1}{2}p^3(1-p)^{3n-9}.$$

The final possibility is that you pick two vertices attached to vertex v, and then leave both vertices without any further attachments to the other $n-3$ and being unconnected (the connected case is part of the second possibility)

$$\binom{n-1}{2}p^2(1-p)^{n-3}\cdot(1-p)^{2n-5} = \binom{n-1}{2}p^2(1-p)^{3n-8}.$$

In total, this gives

$$\begin{aligned}(n-1)(n-2)p^2(1-p)^{3n-8} &+ \binom{n-1}{2}p^3(1-p)^{3n-9} + \binom{n-1}{2}p^2(1-p)^{3n-9} \qquad \text{(B.68)}\\ &= (n-1)(n-2)p^2(1-p)^{3n-9}(1-p+\frac{p}{2}+\frac{(1-p)}{2})\\ &= (n-1)(n-2)p^2(1-p)^{3n-9}(\frac{3}{2}-p).\end{aligned}$$

□

Solution to Exercise 4.5. We first pick 3 different elements i,j,k from $\{1,2,\ldots,n\}$ without order. This can be done in

$$\binom{n}{3}$$

different ways. Then all three edges ij, ik, jk have to be present, which has probability p^3. The number of triangles is the sum of indicators running over all unordered triples. These indicators are dependent, but that is of no importance for the expectation, because the expectation of a sum of dependent random variables equals the sum of the expected values. Hence the expected number of occupied triangles equals:

$$\binom{n}{3}p^3.$$

□

Solution to Exercise 4.6. We pick 4 elements i,j,k,l from $\{1,2,\ldots,n\}$ This kan be done in

$$\binom{n}{4}$$

different ways. This quadruple may form an occupied square in 3 different orders, that is (i,j,k,l), (i,k,j,l) and (i,j,l,k). Hence there are

$$3\cdot\binom{n}{4}$$

squares in which all four sides should be occupied. Hence the expected number of occupied squares equals

$$3\binom{n}{4}p^4.$$

□

Solution to Exercise 4.7. We define the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ where X_n is the number of occupied triangles in an Erdős-Rényi random graph with edge probability $p = \lambda/n$. Next we introduce the indicator function

$$I_{a,n} := \begin{cases} 0 & \text{triangle } a \text{ not connected;} \\ 1 & \text{triangle } i \text{ connected.} \end{cases}$$

Now, according to (2.1.21) we have

$$\lim_{n\to\infty} \mathbb{E}[(X_n)_r] = \lim_{n\to\infty} \sum_{a_1,a_2,\ldots,a_r\in\mathcal{I}}^{*} \mathbb{P}(I_{a_1,n} = 1, I_{a_2,n} = 1, \ldots, I_{a_r,n} = 1). \tag{B.69}$$

Now, there are two types of collections of triangles, namely, sets of triangles in which all edges are distinct, or the set of triangles for which at least one edge occurs in two different triangles. In the first case, we see that the indicators $I_{a_1,n}, I_{a_2,n}, \ldots, I_{a_r,n}$ are *independent*, in the second case, they are not. We first claim that the collection of $(a_1, a_2, \ldots, a_r)$ for which all triangles contain different edges has size

$$(1 + o(1))\binom{n}{3}^r. \tag{B.70}$$

To see this, we note that the upper bound is obvious (since $\left(\binom{n}{3}\right)^r$ is the number of collections of r triangles without any restriction). For the lower bound, we note that $a_i = (k_i, l_i, m_i)$ for $k_i, l_i, m_i \in [n]$ such that $k_i < l_i < m_i$. We obtain a lower bound on the number of triangles containing different edges when we assume that all vertices k_i, l_i, m_i for $i = 1, \ldots, r$ are distinct. There are precisely

$$\prod_{i=0}^{r-1} \binom{n-i}{3} \tag{B.71}$$

of such combinations. When r is fixed, we have that

$$\prod_{i=0}^{r-1} \binom{n-i}{3} = (1 + o(1))\binom{n}{3}^r. \tag{B.72}$$

Thus, the contribution to the right-hand side of (B.69) of collections $(a_1, a_2, \ldots, a_r)$ for which all triangles contain different edges is, by independence and (B.70), equal to

$$(1 + o(1))\binom{n}{3}^r \Big(\frac{\lambda^3}{n^3}\Big)^r = (1 + o(1))\Big(\frac{\lambda^3}{6}\Big)^r. \tag{B.73}$$

We next prove that the contribution to the right-hand side of (B.69) of collections $(a_1, a_2, \ldots, a_r)$ for which at least one edge occurs in two different triangles. We give a crude upper bound for this. We note that each edge which occurs more that once reduces the number of possible vertices involved. More precisely, when the collection of triangles $(a_1, a_2, \ldots, a_r)$ contains precisely $3r-l$ edges for some $l \geq 1$, then the collection of triangles $(a_1, a_2, \ldots, a_r)$ contains at most $3r-2l$ vertices, as can easily be seen by induction. As a result, the contribution to the right-hand side of (B.69) of collections $(a_1, a_2, \ldots, a_r)$ $(a_1, a_2, \ldots, a_r)$ contains precisely $3r - l$ edges is bounded by

$$n^{3r-2l}(\lambda/n)^{3r-l} = \lambda^{3r-l}n^{-l} = o(1). \tag{B.74}$$

Since this is negligible, we obtain that

$$\lim_{n\to\infty} \mathbb{E}[(X_n)_r] = \left(\frac{\lambda^3}{6}\right)^r. \tag{B.75}$$

Hence, due to Theorem 2.4 we have that the number of occupied triangles in an Erdős-Rényi random graph with edge probability $p = \lambda/n$ has an asymptotic Poisson distribution with parameter $\lambda^3/6$. □

Solution to Exercise 4.8. We have

$$\begin{aligned}\mathbb{E}[\Delta_G] &= \mathbb{E}\left[\sum_{i,j,k\in G} \mathbb{1}_{\{ij,ik,jk \text{ occupied}\}}\right] = \sum_{i,j,k\in G} \mathbb{E}\left[\mathbb{1}_{\{ij,ik,jk \text{ occupied}\}}\right] \\ &= n(n-1)(n-2)\left(\frac{\lambda}{n}\right)^3,\end{aligned} \tag{B.76}$$

and

$$\begin{aligned}\mathbb{E}[W_G] &= \mathbb{E}\left[\sum_{i,j,k\in G} I[ij,jk \text{ occupied}]\right] = \sum_{i,j,k\in G} \mathbb{E}\left[\mathbb{1}_{\{ij,jk \text{ occupied}\}}\right] \\ &= n(n-1)(n-2)\left(\frac{\lambda}{n}\right)^2.\end{aligned} \tag{B.77}$$

This yields for the clustering coefficient

$$CC_G = \lambda/n.$$

□

Solution to Exercise 4.9. We have $\mathbb{E}[W_G] = n(n-1)(n-2)p^2(1-p)$. According to the Chebychev inequality we obtain:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}[|W_G - \mathbb{E}[W]| > \epsilon] &\leq \lim_{n\to\infty} \frac{\sigma^2_{W_G}}{\epsilon^2}, \\ \lim_{n\to\infty} \mathbb{P}[|W_G - (n)(n-1)(n-2)(\frac{\lambda}{n})^2(\frac{n-\lambda}{n})| > \epsilon] &\leq \lim_{n\to\infty} \frac{\sigma^2_{W_G}}{\epsilon^2}, \\ \lim_{n\to\infty} \mathbb{P}[|W_G - n\lambda^2| > \epsilon] &\leq 0.\end{aligned}$$

Hence, $W_G/n \xrightarrow{\mathbb{P}} \lambda^2$ and, therefore, $n/W_G \xrightarrow{\mathbb{P}} 1/\lambda^2$. We have already shown in previous exercise that the number of occupied triangles has an asymptotic Poisson distribution with parameter $\frac{\lambda^3}{6}$. Δ_G is three times the number of triangles and thus $\Delta_G \xrightarrow{d} 3 \cdot \text{Poi}(\frac{\lambda^3}{6})$. Slutsky's Theorem states that

$$X_n \xrightarrow{\mathbb{P}} c \text{ and } Y_n \xrightarrow{d} Y \Rightarrow X_n Y_n \xrightarrow{d} cY$$

Hence $\frac{n\Delta_G}{W_G} \xrightarrow{d} \frac{3}{\lambda^2} Y$ where $Y \sim \text{Poi}(\lambda^3/6)$. □

Solution to Exercise 4.10. We have to show that for each x, the event $\{|\mathcal{C}(v)| \geq x\}$ remains true if the the number of edges increases.

Obviously by increasing the number of edges the number $|\mathcal{C}(v)|$ increases or stays the same depending on whether or not some of the added edges connect new vertices to the cluster. In both cases $\{|\mathcal{C}(v)| \geq x\}$ remains true.

□

Solution to Exercise 4.11. This is not true. Take two disjoint clusters which differ by one in size, and suppose that the larger component equals $\mathcal{C}_{\max}$, before adding the edges. Take any $v \in \mathcal{C}_{\max}$. Now add edges between the second largest component and isolated vertices. If you add two of such edges, then the new $\mathcal{C}_{\max}$ equals the union of the second largest component and the two isolated vertices. Since originally v did not belong to the second largest component and v was not isolated, because it was a member of the previous largest component, we now have $v \notin \mathcal{C}_{\max}$.

□

Solution to Exercise 4.12. As a result of (4.2.1) we have

$$\mathbb{E}_\lambda[|\mathcal{C}(v)|] = \sum_{k=1}^{\infty} \mathbb{P}(|\mathcal{C}(v)| \geq k) \leq \sum_{k=1}^{\infty} \mathbb{P}_{n,p}(T \geq k) = \mathbb{E}[T] = \frac{1}{1-\mu}, \tag{B.78}$$

where

$$\mu = \mathbb{E}[\text{Offspring}] = np = \lambda.$$

Hence,

$$\mathbb{E}_\lambda[|\mathcal{C}(v)|] \leq 1/(1-\lambda).$$

□

Solution to Exercise 4.14. We recall that $Z_{\geq k} = \sum_{i=1}^{n} \mathbb{1}_{\{|\mathcal{C}(i)| \geq k\}}$.

$$|\mathcal{C}_{\max}| < k \Rightarrow |\mathcal{C}(i)| < k \forall i, \text{ which implies that } Z_{\geq k} = 0$$
$$|\mathcal{C}_{\max}| \geq k \Rightarrow |\mathcal{C}(i)| \geq k \text{ for at least } k \text{ vertices } \Rightarrow Z_{\geq k} \geq k.$$

□

Solution to Exercise 4.15. Intuitively the statement is logical, for we can see M as doing n trails with succes probability p and for each trial we throw an other coin with succes probability q. The eventual amount of successes are the successes where both trails ended in succes and is thus equal to throwing n coins with succes probability pq.
There are several ways to prove this, we give two of them.

Suppose we have two binomial trials N and Y both of length n and with succes rates p, q respectively. We thus create two vectors filled with ones and zeros. For each index $i = 1, 2, \ldots, n$ we compare the vectors and in case both entries are 1, we will see this as a succes. The now counted amount of successes is of course $\text{Bin}(n, pq)$ distributed.
Now we produce the first vector similarly by denoting ones and zeros for the successes and losses in trail N. For each 'one', we produce an other outcome by a $\text{Be}(q)$ experiment. We count the total number of successes of these experiments and those are of course $\text{Bin}(N, q)$ distributed. But now, this is the same as the experiment described above, since all Bernoulli outcomes are independent. Hence if $N \sim \text{Bin}(n, p)$ and $M \sim \text{Bin}(N, q)$, then $M \sim \text{Bin}(n, pq)$.

We will also give an analytical proof, which is somewhat more enhanced. We wish to show that $\mathbb{P}(M=m) = \binom{n}{m}(pq)^m(1-pq)^{n-m}$. Off course we have

$$\begin{aligned}\mathbb{P}(M=m) &= \sum_{i=m}^{n} \mathbb{P}(N=i)\cdot\binom{i}{m}\cdot q^m\cdot(1-q)^i - m,\\ &= \sum_{i=m}^{n}\binom{n}{i}\cdot(p)^i\cdot(1-p)^{n-i}\cdot\binom{i}{m}\cdot q^m\cdot(1-q)^i - m.\end{aligned}$$

Rearranging terms yields

$$\mathbb{P}(M=m) = \frac{(1-p)^n q^m}{(1-q)^m}\sum_{i=m}^{n}\binom{n}{i}\binom{i}{m}\frac{p^i}{(1-p)^i}(1-q)^i.$$

Further analysis yields

$$\begin{aligned}\mathbb{P}(M=m) &= (1-p)^n\Big(\frac{q}{1-q}\Big)^m\sum_{i=m}^{n}\frac{n!}{i!(n-i)!}\,\frac{i!}{m!(i-m)!}\Big(\frac{p(1-q)}{1-p}\Big)^i\\ &= (1-p)^n\Big(\frac{q}{1-q}\Big)^m\frac{n!}{m!}\sum_{i=m}^{n}\frac{1}{(n-i)!(i-m)!}\Big(\frac{p(1-q)}{1-p}\Big)^i\\ &= (1-p)^n(\frac{q}{1-q})^m\frac{n!m!}{\sum}^{n-m}_{k=0}\frac{1}{(n-k-m)!(m+k-m)!}\Big(\frac{p(1-q)}{1-p}\Big)^{k+m}\\ &= (1-p)^n(\frac{q}{1-q})^m\frac{n!}{m!(n-m)!}\sum_{k=0}^{n-m}\frac{(n-m)!}{(n-k-m)!k!}\Big(\frac{p(1-q)}{1-p}\Big)^{k+m}\\ &= \binom{n}{m}\sum_{k=0}^{n-m}\binom{n-m}{k}p^{k+m}(1-p)^{n-m-k}q^m(1-q)^{k+m-m}\\ &= \binom{n}{m}p^m q^m\sum_{k=0}^{n-m}\binom{n-m}{k}p^k(1-p)^{n-m-k}(1-q)^k\end{aligned}$$

It is now sufficient to show that $\sum_{k=0}^{n-m}\binom{n-m}{k}p^k(1-p)^{n-m-k}(1-q)^k = (1-pq)^{n-m}$.

$$\begin{aligned}\sum_{k=0}^{n-m}\binom{n-m}{k}p^k(1-p)^{n-m-k}(1-q)^k &= (1-p)^{n-m}\sum_{k=0}^{n-m}\binom{n-m}{k}\Big(\frac{p-pq}{1-p}\Big)^k\\ &= (1-p)^{n-m}\Big(1+\frac{p-pq}{1-p}\Big)^{n-m}\\ &= (1-p)^{n-m}\Big(\frac{1-p+p-pq}{1-p}\Big)^n - m\\ &= (1-pq)^{n-m}.\end{aligned}$$

Now we can use this result to proof that $N_t \sim \text{Bin}(n,(1-p)^t)$ by using induction. The initial value $N_0 = n-1$ is given, hence

$$\begin{aligned}N_0 &= n-1;\\ N_1 &= \text{Bin}(n-1,1-p);\\ N_2 &= \text{Bin}(N_1,1-p) = \text{Bin}(n-1,(1-p)^2);\\ &\vdots\\ N_t &= \text{Bin}(n-1,(1-p)^t).\end{aligned}$$

□

Solution to Exercise 4.17. The extinction probability η satisfies

$$\eta_\lambda = G_X(\eta_\lambda) = \mathbb{E}[\eta_\lambda^X] = e^{-\lambda+\lambda\eta_\lambda}$$

Hence,

$$\zeta_\lambda = 1 - \eta_\lambda = 1 - e^{-\lambda+\lambda\eta} = 1 - e^{-\lambda\zeta_\lambda}.$$

This equation has only two solutions, one of which is $\zeta_\lambda = 0$, the other must be the survival probability. □

Solution to Exercise 4.18. We compute that

$$\begin{aligned}\chi(\lambda) = \mathbb{E}_\lambda[|\mathcal{C}(1)|] = \mathbb{E}_\lambda\left[\sum_{j=1}^n \mathbb{1}_{\{j\in\mathcal{C}(1)\}}\right] = 1 + \sum_{j=2}^n \mathbb{E}_\lambda[\mathbb{1}_{\{j\in\mathcal{C}(1)\}}]\\ = 1 + \sum_{j=2}^n \mathbb{E}_\lambda[\mathbb{1}_{\{1\leftrightarrow j\}}] = 1 + \sum_{j=2}^n \mathbb{P}_\lambda(1\leftrightarrow j) = 1 + (n-1)\mathbb{P}_\lambda(1\leftrightarrow 2). \qquad \text{(B.79)}\end{aligned}$$

□

Solution to Exercise 4.19. In this exercise we denote by $|\mathcal{C}_{(1)}| \geq |\mathcal{C}_{(2)}| \geq \ldots$, the components ordered by their size. Relation (4.4.1) reads that for $\nu \in (\frac{1}{2}, 1)$:

$$\mathbb{P}\Big(\big||\mathcal{C}_{\max}| - n\zeta_\lambda\big| \geq n^\nu\Big) = O(n^{-\delta}).$$

Observe that

$$\begin{aligned}\mathbb{P}_\lambda(1\leftrightarrow 2) &= \mathbb{P}_\lambda(\exists\mathcal{C}(k) : 1\in\mathcal{C}(k), 2\in\mathcal{C}(k))\\ &= \sum_{l\geq 1}\mathbb{P}_\lambda(1,2\in\mathcal{C}_{(l)}) = \mathbb{P}_\lambda(1,2\in\mathcal{C}_{(1)}) + \sum_{l\geq 2}\mathbb{P}_\lambda(1,2\in\mathcal{C}_{(l)})\\ &= \frac{(n\zeta_\lambda \pm n^\nu)^2}{n^2} + O(n^{-\delta}) + \sum_{l\geq 2}\mathbb{P}_\lambda(1,2\in\mathcal{C}_{(l)}).\end{aligned}$$

For $l \geq 2$, we have $|\mathcal{C}_{(l)}| \leq K\log n$ with high probability, hence

$$\mathbb{P}_\lambda(1,2\in\mathcal{C}_{(l)}) \leq \frac{K^2\log^2 n}{n^2} + O(n^{-2}),$$

so that

$$\sum_{l\geq 2}\mathbb{P}_\lambda(1,2\in\mathcal{C}_{(l)}) \leq \frac{K^2\log^2 n}{n} + O(n^{-1}) \to 0.$$

Together, this shows that

$$\mathbb{P}_\lambda(1\leftrightarrow 2) = \zeta_\lambda^2 + O(n^{-\delta}),$$

for some $\delta > 0$. □

Solution to Exercise 4.20. Combining Exercise 4.18 and Exercise 4.19, yields

$$\chi(\lambda) = 1 + (n-1)\zeta_\lambda^2(1+o(1)) = n\zeta_\lambda^2(1+o(1)).$$

□

Solution to Exercise 4.16. We have that the cluster of i has size l. Furthermore, we have $\mathbb{P}_\lambda\big(i \longleftrightarrow j \big| |\mathcal{C}(i)| = l\big) + \mathbb{P}_\lambda\big(i \longleftrightarrow\!\!\!\!\!\!/\;\; j \big| |\mathcal{C}(i)| = l\big) = 1$ Of course $i, j \in [n]$ and $j \neq i$. So, having i fixed, gives us $n-1$ choices for j in $\mathrm{ER}_n(p)$ and $l-1$ choices for j in $\mathcal{C}(i)$. Hence,

$$\mathbb{P}_\lambda\big(i \longleftrightarrow j \big| |\mathcal{C}(i)| = l\big) = \frac{l-1}{n-1},$$

and thus

$$\mathbb{P}_\lambda\big(i \longleftrightarrow\!\!\!\!\!\!/\;\; j \big| |\mathcal{C}(i)| = l\big) = 1 - \frac{l-1}{n-1}.$$

□

Solution to Exercise 4.21. According to the duality principle we have that the random graph obtained by removing the largest component of a supercritical Erdős-Rényi random graph is again an Erdős-Rényi random graph of size $m \sim n\eta_\lambda = \frac{\mu_\lambda n}{\lambda}$ where $\mu_\lambda < 1 < \lambda$ are conjugates as in (3.6.7) and the remaining graph is thus in the subcritical regime. Hence, studying the second largest component in a supercritical graph is close to studying the largest component in the remaining graph.
Now, as a result of Theorems 4.4 and 4.5 we have that for some $\epsilon > 0$

$$\lim_{n\to\infty} \Big(\mathbb{P}\big(\frac{|\mathcal{C}_{\max}|}{\log m} > I^{-1}_{\mu_\lambda} + \epsilon\big) + \mathbb{P}\big(\frac{|\mathcal{C}_{\max}|}{\log m} < I^{-1}_{\mu_\lambda} - \epsilon\big)\Big) = 0.$$

Hence, $\frac{|\mathcal{C}_{\max}|}{\log m} \xrightarrow{\mathbb{P}} I^{-1}_{\mu_\lambda}$. But since we have that $n - m = \zeta_\lambda n(1 + o(1))$ and thus $m = n(1-\zeta_\lambda)$, we have that $\frac{\log m}{\log n} \to 1$ as $n \to \infty$. Hence $\frac{|\mathcal{C}_{\max}|}{\log n} \xrightarrow{\mathbb{P}} I^{-1}_{\mu_\lambda}$. □

Solution to Exercise 4.22. Denote

$$Z_n = \frac{X_n - a_n p_n}{\sqrt{a_n p_n (1-p_n)}}, \tag{B.80}$$

so that we need to prove that Z_n converges is distribution to a standard normal random variable Z. For this, it suffices to prove that the moment generating function $M_{Z_n}(t) = \mathbb{E}[e^{tZ_n}]$ of Z_n converges to that of Z.

Since the variance of X_n goes to infinity, the same holds for a_n. Now we write X_n as to be a sum of a_n Bernoulli variables $X_n = \sum_{i=1}^{a_n} Y_i$, where $\{Y_i\}_{1\le i\le a_n}$ are independent random variables with $Y_i \sim \mathrm{Be}(p_n)$. Thus, we note that the moment generating function of X_n equals

$$M_{X_n}(t) = \mathbb{E}[e^{tX_n}] = \mathbb{E}[e^{tY_1}]^{a_n}. \tag{B.81}$$

We further prove, using a simple Taylor expansion,

$$\log \mathbb{E}[e^{tY_1}] = \log\big(p_n e^t + (1-p_n)\big) = p_n t + \frac{t^2}{2} p_n(1-p_n) + O(|t|^3 p_n). \tag{B.82}$$

Thus, with $t_n = t/\sqrt{a_n p_n(1-p_n)}$, we have that

$$M_{Z_n}(t) = M_{X_n}(t_t) e^{a_n p_n t_n} = e^{a_n \log \mathbb{E}[e^{tY_1}]} = e^{\frac{t_n^2}{2} p_n(1-p_n) + O(|t_n|^3 a_n p_n)} = e^{t^2/2 + o(1)}. \tag{B.83}$$

We conclude that $\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2}$, which is the moment generating function of a standard normal distribution. Theorem 2.3(b) implies that $Z_n \xrightarrow{d} Z$, as required. Hence, the CLT follows and (4.5.14) implies (4.5.15). □

Solution to Exercise 4.25. We have that $n\lambda/2$ edges are added in a total system of $n(n-1)/2$ edges. This intuitively yields for p in the classical notation for the ER graphs to be $p = \frac{n\lambda/2}{n(n-1)/2}$ and $\lambda' = n \cdot p$, so that one would expect subcritical behavior $|\mathcal{C}_{\max}|/\log n \xrightarrow{\mathbb{P}} I_\lambda^{-1}$. We now provide the details of this argument.

We make use of the crucial relation (4.6.1), and further note that when we increase M, then we make the event $|\mathcal{C}_{\max}| \geq k$ more likely. This is a related version of monotonicity as in Section 4.1.1. In particular, from (4.6.1), it follows that for any increasing event E, and with $p = \lambda/n$,

$$\begin{aligned}\mathbb{P}_\lambda(E) &= \sum_{m=1}^{n(n-1)/2} \mathbb{P}_m(E)\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) = m\big) \\ &\geq \sum_{m=M}^{\infty} \mathbb{P}_m(E)\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) = m\big) \\ &\geq \mathbb{P}_M(E)\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) \geq M\big).\end{aligned} \tag{B.84}$$

In particular, when p is chosen such that $\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) \geq M\big) = 1 - o(1)$, then $\mathbb{P}_M(E) = o(1)$ follows when $\mathbb{P}_\lambda(E) = o(1)$.

Take $a > I_\lambda^{-1}$ and let $k_n = a \log n$. Then we shall first show that $\mathbb{P}_{n,M}(|\mathcal{C}_{\max}| \geq k_n) = o(1)$. For this, we use the above monotonicity to note that, for every λ',

$$\mathbb{P}_{n,M}(|\mathcal{C}_{\max}| \geq k_n) \leq \mathbb{P}_{\lambda'}(|\mathcal{C}_{\max}| \geq k_n)/\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,\lambda'/n) \geq M\big). \tag{B.85}$$

For any $\lambda' > \lambda$, we have $\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,\lambda'/n) \geq M\big) = 1 + o(1)$. Now, since $\lambda \mapsto I_\lambda^{-1}$ is continuous, we can take $\lambda' > \lambda$ such that $I_{\lambda'}^{-1} < a$, we further obtain by Theorem 4.4 that $\mathbb{P}_{\lambda'}(|\mathcal{C}_{\max}| \geq k_n) = o(1)$, so that $\mathbb{P}_{n,M}(|\mathcal{C}_{\max}| \geq k_n) = o(1)$ follows.

Next, take $a < I_\lambda^{-1}$, take $k_n = a \log n$, and we next wish to prove that $\mathbb{P}_{n,M}(|\mathcal{C}_{\max}| \leq k_n) = o(1)$. For this, we make use of a related bound as in (B.84), namely, for a *decreasing event* F, we obtain

$$\begin{aligned}\mathbb{P}_\lambda(F) &= \sum_{m=1}^{n(n-1)/2} \mathbb{P}_m(F)\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) = m\big) \\ &\geq \sum_{m=1}^{M} \mathbb{P}_m(F)\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) = M\big) \\ &\geq \mathbb{P}_M(F)\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) \leq M\big).\end{aligned} \tag{B.86}$$

Now, we take $p = \lambda'/n$ where $\lambda' < \lambda$, so that $\mathbb{P}\big(\mathrm{Bin}(n(n-1)/2,p) \leq M\big) = 1 - o(1)$. Then, we pick $\lambda' < \lambda$ such that $I_{\lambda'}^{-1} > a$ and use Theorem 4.5. We conclude that, with high probability, $|\mathcal{C}_{\max}|/\log n \leq I_\lambda^{-1} + \varepsilon)$ for any $\varepsilon > 0$, and, again with high probability, $|\mathcal{C}_{\max}|/\log n \geq I_\lambda^{-1} - \varepsilon)$ for any $\varepsilon > 0$. This yields directly that $|\mathcal{C}_{\max}|/\log n \xrightarrow{\mathbb{P}} I_\lambda^{-1}$. □

Solutions to the exercises of Chapter 5.

Solution to Exercise ??. Using (3.6.20) we see that

$$\mathbb{P}_\lambda^*(T^* \geq k) = (2\pi)^{-1/2} \sum_{n=k}^{\infty} n^{-3/2}[1 + O(n^{-1})]. \tag{B.87}$$

The sum can be bounded from above and below by an integral as follows

$$\int_k^\infty x^{-3/2}dx \leq \sum_{n=k}^\infty n^{-3/2} \leq \int_{k-1}^\infty x^{-3/2}dx$$

Computing these integrals gives

$$2k^{-1/2} \leq \sum_{n=k}^\infty n^{-3/2} \leq 2(k-1)^{-1/2}$$

Similar bounds can be derived such that

$$\sum_{n=k}^\infty n^{-3/2}O(n^{-1}) = O(k^{-3/2}).$$

Combining both bounds, it follows that

$$\mathbb{P}_\lambda^*(T^* \geq k) = \left(\frac{2}{\pi}\right)^{1/2} k^{-1/2}[1 + O(k^{-1})].$$

□

Solution to Exercise 5.1. Fix some $r > 0$, then

$$\chi(1) \geq \sum_{k=1}^{rn^{2/3}} \mathbb{P}(|\mathcal{C}(1)| \geq k) = \sum_{k=1}^{rn^{2/3}} P_{\geq k}(1). \tag{B.88}$$

By Proposition 5.2, we have the bounds

$$P_{\geq k}(1) \geq \frac{c_1}{\sqrt{k}}.$$

Substituting this bounds into (B.88) yields

$$\chi(1) \geq \sum_{k=1}^{rn^{2/3}} \frac{c_1}{\sqrt{k}} \geq c_1' r n^{1/3},$$

where $c_1' > 0$ and $r > 0$. □

Solution to Exercise ??. By Theorem 3.16, we have that

$$\frac{1}{\lambda}\, e^{-I_\lambda t}\mathbb{P}_1^*(T^* = t) = \frac{1}{\lambda}\, e^{-(\lambda - 1 - \log\lambda)t} \frac{t^{t-1}}{t!} e^{-t}.$$

Rearranging the terms in this equation we get

$$\frac{1}{\lambda}\, e^{-I_\lambda t}\mathbb{P}_1^*(T^* = t) = \frac{1}{\lambda}\, \left(e^{\log\lambda}\right)^t \frac{t^{t-1}}{t!}\, e^{-\lambda t} = \frac{(\lambda t)^{t-1}}{t!}\, e^{-\lambda t}.$$

□

Solution to Exercise ??. Let $\mathcal{G}(n)$ be the collection of all possible simple graphs on n points. The set $\mathcal{G}(n,m)$ is the subset of $\mathcal{G}$ which contains all possible simple graphs on n points which have m edges. Then,

$$\begin{aligned}\mathbb{P}(1 \longleftrightarrow 2) &= |\mathcal{G}(n)|^{-1} \sum_{m=1}^{\binom{n}{2}} \sum_{G \in \mathcal{G}(n,m)} \mathbb{P}(G) \mathbb{1}_{\{1 \longleftrightarrow 2 \text{ in } G\}} \\ &= 2^{-\binom{n}{2}} \sum_{m=1}^{n} \sum_{G \in \mathcal{G}(n,m)} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{\binom{n}{2}-m} \mathbb{1}_{\{1 \longleftrightarrow 2 \text{ in } G\}},\end{aligned}$$

which is polynomial in λ. Furthermore, the maximal degree of the polynoom is $\binom{n}{2}$.

Solution to Exercise ??. Take some $l \in \mathbb{N}$ such that $l < n$, then $\chi_{n-l}(\lambda \frac{n-l}{n})$ is the expected component size in the graph $\mathrm{ER}(n-l,p)$. We have to prove that the expected component size in the graph $\mathrm{ER}(n-l,p)$ is smaller than the expected component size in the graph $\mathrm{ER}(n-l+1,p)$ for all $0 < p \leq 1$. Consider the graph $\mathrm{ER}(n-l+1,p)$. This graph can be created from $\mathrm{ER}(n-l,p)$ by adding the vertex $n-l+1$ and independently connecting this vertex to each of the vertices $1, 2, \ldots, n-l$.

Let $\mathcal{C}'(1)$ denote the component of $\mathrm{ER}(n-l,p)$ which contains vertex 1 and $\mathcal{C}(1)$ represents the component of $\mathrm{ER}(n-l+1,p)$ which contains vertex 1. By the construction of $\mathrm{ER}(n-l+1,p)$, it follows that

$$\mathbb{P}(|\mathcal{C}(1)| = k) = \begin{cases} (1-p)^{n-l+1} & \text{if } k = 1, \\ \mathbb{P}(|\mathcal{C}'(1)| = k)(1-p)^k + \mathbb{P}(|\mathcal{C}'(1)| = k-1)(1-(1-p)^{k-1}) & \text{if } 2 \leq k \leq n, \\ \mathbb{P}(|\mathcal{C}'(1)| = n)(1-(1-p)^n) & \text{if } k = n+1. \end{cases}$$

Hence, the expected size of $\mathcal{C}(1)$ is

$$\begin{aligned}\mathbb{E}[|\mathcal{C}(1)|] &= \sum_{k=1}^{n+1} \mathbb{P}(|\mathcal{C}(1)| = k) k \\ &= (1-p)^{n-l+1} + \sum_{k=2}^{n} \Big[\mathbb{P}(|\mathcal{C}'(1)| = k)(1-p)^k + \mathbb{P}(|\mathcal{C}'(1)| = k-1)(1-(1-p)^{k-1})\Big] k \\ &\quad + \mathbb{P}(|\mathcal{C}'(1)| = n)(1-(1-p)^n)(n+1).\end{aligned}$$

Rewriting this expression for the expected size of $\mathcal{C}(1)$ yields

$$\begin{aligned}\mathbb{E}[|\mathcal{C}(1)|] &= (1-p)^{n-l+1} + \mathbb{P}(|\mathcal{C}'(1)| = 1)2p + \sum_{k=2}^{n-1} \mathbb{P}(|\mathcal{C}'(1)| = k) k \\ &\quad + \sum_{k=2}^{n-1} \mathbb{P}(\mathcal{C}'(1) = k)(1-(1-p)^{k-1}) + \mathbb{P}(|\mathcal{C}'(1)| = n)(n + (1-(1-p)^n)) \\ &\geq (1+p)\mathbb{P}(|\mathcal{C}'(1)| = 1) + \sum_{k=2}^{n-1} k \mathbb{P}(\mathcal{C}'(1) = k) \geq \mathbb{E}[|\mathcal{C}(1)'|].\end{aligned}$$

□

Solution to Exercise ??. By (**??**), we have that

$$\frac{\partial}{\partial\lambda}\chi_n(\lambda) = (n-1)\frac{\partial}{\partial\lambda}\tau_n(\lambda).$$

For the derivative of $\tau_n(\lambda)$ we use (**??**) to obtain

$$\frac{\partial}{\partial\lambda}\chi_n(\lambda) \leq \sum_{l=1}^{n} l\mathbb{P}_\lambda(|\mathcal{C}(1)| = l)\chi_{n-l}(\lambda\frac{n-l}{n}).$$

The function $l \mapsto \chi_{n-l}(\lambda\frac{n-l}{n})$ is decreasing (see Exercise **??**), hence

$$\frac{\partial}{\partial\lambda}\chi_n(\lambda) \leq \chi_n(\lambda)\sum_{l=1}^{n} l\mathbb{P}_\lambda(|\mathcal{C}(1)| = l) = \chi_n(\lambda)^2,$$

or

$$\frac{\frac{\partial}{\partial\lambda}\chi_n(\lambda)}{\chi_n(\lambda)^2} \leq 1. \tag{B.89}$$

The second part of the exercise relies on integration. Integrate both the left-hand and the right-hand side of (B.89) between λ and 1.

$$\frac{1}{\chi_n(\lambda)} - \frac{1}{\chi_n(1)} \leq 1 - \lambda$$

Bring a term to the other side to obtain

$$\frac{1}{\chi_n(\lambda)} \leq \frac{1}{\chi_n(1)} + 1 - \lambda,$$

which is equivalent to

$$\chi_n(\lambda) \geq \frac{1}{\chi_n(1)^{-1} + (1-\lambda)}.$$

□

Solution to Exercise 5.2. Using (5.2.8) and (5.2.10) we see that

$$\begin{aligned}
\mathbb{E}_\lambda[Y^2] &= n\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1) + n(n-1)\left(\frac{\lambda}{n(1-\frac{\lambda}{n})} + 1\right)\mathbb{P}_\lambda(|\mathcal{C}(1)| = 1)^2 \\
&= n\left(1-\frac{\lambda}{n}\right)^{n-1} + n(n-1)\left(1-\frac{\lambda}{n}\right)^{2n-3} \\
&= n\left(1-\frac{\lambda}{n}\right)^{n-1}\left(1 + (n-1)\left(1-\frac{\lambda}{n}\right)^{n-2}\right).
\end{aligned}$$

Consider the first power, taking the logarithm yields

$$\log n + (n-1)\log(1-\frac{\lambda}{n}) = \log n + (n-1)\log(1-\frac{\log n + t}{n}).$$

Taylor expanding the logarithm gives

$$\log n + (n-1)\log(1-\frac{\log n + t}{n}) = \log n - (n-1)\Big[\frac{\log n + t}{n} + O\Big((\frac{\log n + t}{n})^2\Big)\Big].$$

The latter expression can be simplified to

$$\log n - (n-1)\Big[\frac{\log n + t}{n} + O\Big(\big(\frac{\log n + t}{n}\big)^2\Big)\Big] = \log n - \frac{n-1}{n}\log n - \frac{n-1}{n}\ t + O\Big(\frac{(\log n + t)^2}{n}\Big)$$
$$= -t + \frac{\log n}{n} + \frac{t}{n}\ + O\Big(\frac{(\log n + t)^2}{n}\Big),$$

and, as n tends to infinity,

$$-t + \frac{\log n}{n} + \frac{t}{n}\ + O\Big(\frac{(\log n + t)^2}{n}\Big) \to -t.$$

Hence,

$$\lim_{n\to\infty} n\left(1 - \frac{\lambda}{n}\right)^{n-1} = e^{-t}.$$

A similar argument gives that as $n \to \infty$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-2} = e^{-t}.$$

Therefore, we conclude

$$\lim_{n\to\infty} \mathbb{E}_\lambda[Y^2] = e^{-t}(1 - e^{-t}),$$

which is the second moment of a Poisson random variable with mean e^{-t}.

□

Solutions to the exercises of Chapter 6.

Solution to Exercise 6.1. By the definition of p_{ij} (6.1.1), the numerator of p_{ij} is $(n\lambda)^2(n-\lambda)^{-2}$. The denominator of p_{ij} is

$$\sum_{i=1}^{n} \frac{n\lambda}{n-\lambda} + \left(\frac{n\lambda}{n-\lambda}\right)^2 = \frac{n^2\lambda}{n-\lambda} + \left(\frac{n\lambda}{n-\lambda}\right)^2 = \frac{n^2\lambda(n-\lambda) + (n\lambda)^2}{(n-\lambda)^2} = \frac{n^3\lambda}{(n-\lambda)^2}.$$

Dividing the numerator of p_{ij} by its denominator gives

$$p_{ij} = \frac{(n\lambda)^2}{n^3\lambda} = \frac{\lambda}{n}.$$

□

Solution to Exercise 6.2. Consider the distribution function $F_n(x) = \mathbb{P}(w_V \le x)$ of a uniformly chosen vertex V and let $x \ge 0$. The law of total probability gives that

$$\mathbb{P}(w_V \le x) = \sum_{i=1}^{n} \mathbb{P}(w_V \le x | V = i)\mathbb{P}(V = i)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{w_i \le x\}}, \qquad x \ge 0, \tag{B.90}$$

as desired. □

Solution to Exercise ??. By (6.1.13), $F_n(x) = \frac{1}{n}(\lfloor nF(x)\rfloor + 1) \wedge 1$. To prove pointwise convergence of this function to $F(x)$, we shall first examine its behavior when $F(x)$ gets close to 1. Consider the case where $\frac{1}{n}(\lfloor nF(x)\rfloor + 1) > 1$, or equivalently, $\lfloor nF(x)\rfloor > n-1$, which is in turn equivalent to $F(x) > \frac{n-1}{n}$. Now fixing x gives us two possibilities: either $F(x) = 1$ or there is an n such that $F(x) \leq \frac{n-1}{n}$. In the first case, we have that

$$\begin{aligned}\left|\left[\frac{1}{n}(\lfloor nF(x)\rfloor + 1) \wedge 1\right] - F(x)\right| &= \left|\left[\frac{1}{n}(\lfloor n\rfloor + 1) \wedge 1\right] - 1\right| \\ &= |1-1| = 0. \end{aligned} \tag{B.91}$$

In the second case, we have that for large enough n

$$\begin{aligned}\left|\left[\frac{1}{n}(\lfloor nF(x)\rfloor + 1) \wedge 1\right] - F(x)\right| &= \left|\frac{1}{n}(\lfloor nF(x)\rfloor + 1) - \frac{nF(x)}{n}\right| \\ &= \left|\frac{\lfloor nF(x)\rfloor - nF(x) + 1}{n}\right| \leq \left|\frac{1}{n}\right| \to 0, \end{aligned} \tag{B.92}$$

which proves the pointwise convergence of F_n to F, as desired. □

Solution to Exercise 6.5. We note that $x \mapsto F(x)$ is non-decreasing, since it is a distribution function. This implies that $x \mapsto 1 - F(x)$ is non-increasing, so that $u \mapsto [1-F]^{-1}(u)$ is non-increasing.

To see (6.1.15), we let U be a uniform random variable, and note that

$$\frac{1}{n}\sum_{i=1}^{n} h(w_i) = \mathbb{E}\Big[h\Big([1-F]^{-1}(\lceil Un\rceil/n)\Big)\Big]. \tag{B.93}$$

Now, $\lceil Un\rceil/n \geq U$ a.s., and since $u \mapsto [1-F]^{-1}(u)$ is non-increasing, we obtain that $[1-F]^{-1}(\lceil Un\rceil/n) \leq [1-F]^{-1}(U)$ a.s. Further, again since $x \mapsto h(x)$ is non-decreasing,

$$h\Big([1-F]^{-1}(\lceil Un\rceil/n)\Big) \leq h\big([1-F]^{-1}(U)\big). \tag{B.94}$$

Thus,

$$\frac{1}{n}\sum_{i=1}^{n} h(w_i) \leq \mathbb{E}\Big[h\Big([1-F]^{-1}(U)\Big)\Big] = \mathbb{E}[h(W)], \tag{B.95}$$

since $[1-F]^{-1}(U)$ has distribution function F when U is uniform on $(0,1)$ (recall the remark below (6.1.12)). □

Solution to Exercise 6.6. Using the non-decreasing function $h(x) = x^\alpha$ in Exercise 6.5, we have that for a uniform random variable U

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} w_i^\alpha &= \int_0^1 [1-F]^{-1}\left(\frac{\lceil un\rceil}{n}\right)\frac{1}{n}du \\ &= \mathbb{E}\Big[\Big([1-F]^{-1}(\lceil Un\rceil/n)\Big)^\alpha\Big]. \end{aligned} \tag{B.96}$$

We also know that $\lceil Un\rceil/n \geq U$ a.s., and since $u \mapsto [1-F]^{-1}(u)$ is non-increasing by Exercise 6.5 and $x \mapsto x^\alpha$ is non-decreasing, we obtain that

$$\frac{1}{n}\left([1-F]^{-1}(\lceil Un\rceil/n)\right)^\alpha \leq \frac{1}{n}\left([1-F]^{-1}(U)\right)^\alpha. \tag{B.97}$$

The right hand side function is integrable with value $\mathbb{E}[W^\alpha]$, by assumption. Therefore, by the dominated convergence theorem (Theorem A.9), we have that the integral of the left hand side converges to the integral of its pointwise limit. Since $\lceil Un \rceil / n$ converges in distribution to U, we get that $[1-F]^{-1}(\lceil Un \rceil / n) \to [1-F]^{-1}(U)$, as desired. □

Solution to Exercise 6.7. By (6.1.10),

$$w_i = [1-F]^{-1}(i/n). \tag{B.98}$$

Now apply the function $[1-F]$ to both sides to get

$$[1-F](w_i) = i/n, \tag{B.99}$$

which, by the assumption, can be bounded from above by

$$i/n = [1-F](w_i) \leq c w_i^{-(\tau-1)}. \tag{B.100}$$

This inequality can be rewritten to

$$i^{-\frac{1}{\tau-1}} (cn)^{\frac{1}{\tau-1}} \geq w_i, \tag{B.101}$$

where the left hand side is a descending function in i for $\tau > 1$. This implies

$$w_i \leq w_1 \leq c^{\frac{1}{\tau-1}} n^{\frac{1}{\tau-1}}, \quad \forall i \in [n], \tag{B.102}$$

giving the $c' = c^{\frac{1}{\tau-1}}$ as desired. □

Solution to Exercise 6.9. A mixed Poisson variable X has the property that $\mathbb{P}(X = 0) = \mathbb{E}[e^{-W}]$ is strictly positive, unless W is infinite whp. Therefore, the random variable Y with $\mathbb{P}(Y = 1) = \frac{1}{2}$ and $\mathbb{P}(Y = 2) = \frac{1}{2}$ cannot be represented by a mixed Poisson variable. □

Solution to Exercise 6.10. By definition, the characteristic function of X is

$$\mathbb{E}[e^{itX}] = \sum_{n=0}^{\infty} e^{itn} \mathbb{P}(X = n) = \sum_{n=0}^{\infty} e^{itn} \left(\int_0^\infty f_W(w) \frac{e^{-w} w^n}{n!} dw \right),$$

where $f_W(w)$ is the density function of W evaluated in w. Since all terms are non-negative we can interchange summation and integration. Rearranging the terms gives

$$\begin{aligned} \mathbb{E}[e^{itX}] &= \int_0^\infty f_W(w) e^{-w} \left(\sum_{n=0}^{\infty} \frac{\left(e^{it} w\right)^n}{n!} \right) dw = \int_0^\infty f_W(w) e^{-w} \exp(e^{it} w) dw \\ &= \int_0^\infty f_W(w) \exp((e^{it} - 1) w) dw. \end{aligned}$$

The latter expression is the moment generating function of W evaluated in $e^{it} - 1$. □

Solution to Exercise 6.11. By the tower rule, we have that $\mathbb{E}[\mathbb{E}[X|W]] = \mathbb{E}[X]$. Computing the expected value on the left hand side gives

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|W]] &= \sum_w \mathbb{E}[X|W=w]\mathbb{P}(W=w)\\ &= \sum_w \mathbb{P}(W=w)\sum_k k e^{-w}\frac{w^k}{k!}\\ &= \sum_w w\cdot\mathbb{P}(W=w)\cdot e^{-w}\sum_k \frac{w^{(k-1)}}{(k-1)!}\\ &= \sum_w w\cdot\mathbb{P}(W=w) = \mathbb{E}[W], \end{aligned} \tag{B.103}$$

so $\mathbb{E}[X] = \mathbb{E}[W]$. For the second moment of X, we consider $\mathbb{E}[\mathbb{E}[X(X-1)|W]] = \mathbb{E}[X(X-1)]$. Computing the expected value on the left hand side gives

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X(X-1)|W]] &= \sum_w \mathbb{E}[X(X-1)|W=w]\mathbb{P}(W=w)\\ &= \sum_w \mathbb{P}(W=w)\sum_k k(k-1) e^{-w}\frac{w^k}{k!}\\ &= \sum_w w^2\cdot\mathbb{P}(W=w)\cdot e^{-w}\sum_k \frac{w^{(k-2)}}{(k-2)!}\\ &= \sum_w w^2\cdot\mathbb{P}(W=w) = \mathbb{E}[W^2]. \end{aligned} \tag{B.104}$$

Now, we have that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[W^2] + \mathbb{E}[W] - \mathbb{E}[W]^2$, which is the sum of the variance and expected value of W. □

Solution to Exercise 6.13. Suppose there exists a $\varepsilon > 0$ such that $\varepsilon \le w_i \le \varepsilon^{-1}$ for every i. Now take the coupling D_i' as in (**??**). Now, by (**??**), we obtain that

$$\begin{aligned}\mathbb{P}\Big((D_1,\dots,D_m) \neq (\hat{D}_1,\dots,\hat{D}_m)\Big) &\le 2\sum_{i,j=1}^m p_{ij}\\ &= 2\sum_{i,j=1}^m \frac{w_i w_j}{l_n + w_i w_j}. \end{aligned} \tag{B.105}$$

Now $l_n = \sum_{i=1}^n w_i \ge n\varepsilon$ and $\varepsilon^2 \le w_i w_j \le \varepsilon^{-2}$. Therefore,

$$2\sum_{i,j=1}^m \frac{w_i w_j}{l_n + w_i w_j} \le 2m^2 \frac{\varepsilon^{-2}}{n\varepsilon + \varepsilon^2} = o(1), \tag{B.106}$$

since $m = o(\sqrt{n})$. □

Solution to Exercise 6.14. We have to prove

$$\max_k |\mathbb{E}[P_k^{(n)}] - p_k| \le \frac{\varepsilon}{2}. \tag{B.107}$$

We have

$$\max_k |\mathbb{E}[P_k^{(n)}] - p_k| \leq \frac{\varepsilon}{2} \Leftrightarrow \forall_k |\mathbb{E}[p_k^{(n)}] - p_k| \leq \frac{\varepsilon}{2}. \tag{B.108}$$

Furthermore the following limit is given

$$\lim_{n\to\infty} \mathbb{E}[P_k^{(n)}] = \lim_{n\to\infty} \mathbb{P}(D_1 = k) = p_k. \tag{B.109}$$

Hence we can write

$$\forall_{\varepsilon>0}\forall_k\exists_{M_k}\forall_{n>M_k} |\mathbb{E}[P_k^{(n)}] - p_k| \leq \frac{\varepsilon}{2} \tag{B.110}$$

Taking $M := \max_k M_k$ we obtain

$$\begin{gathered}\forall_{\varepsilon>0}\exists_M\forall_k\forall_{n>M} |\mathbb{E}[P_k^{(n)}] - p_k| \leq \tfrac{\varepsilon}{2} \\ \Leftrightarrow \\ \forall_{\varepsilon>0}\exists_M\forall_{n>M} \max_k |\mathbb{E}[P_k^{(n)}] - p_k| \leq \tfrac{\varepsilon}{2}.\end{gathered}$$

□

Solution to Exercise 6.15. Using the hint, we get

$$\begin{aligned}\mathbb{P}(\max_{i=1}^n W_i \geq \varepsilon n) &\leq \sum_{i=1}^n \mathbb{P}(W_i \geq \varepsilon n) \\ &= n\mathbb{P}(W_1 \geq \varepsilon n).\end{aligned} \tag{B.111}$$

This probability can be rewritten, and applying the Markov inequality now gives

$$n\mathbb{P}(W_1 \geq \varepsilon n) = n\mathbb{P}(\mathbb{1}_{\{W_1\geq\varepsilon n\}} W_1 \geq \varepsilon n) \leq \mathbb{P}(W_1 \geq \varepsilon n)\mathbb{E}[W_1] \to 0. \tag{B.112}$$

Therefore, $\max_{i=1}^n W_i$ is $o(n)$ whp, and

$$\frac{1}{n^2}\sum_{i=1}^n W_i^2 \leq \frac{1}{n}\max_{i=1}^n W_i^2 \to 0, \tag{B.113}$$

as desired. □

Solution to Exercise 6.17. Using partial integration we obtain for the mean of W_1

$$\begin{aligned}\mathbb{E}[W_1] &= \int_0^\infty x f(x)dx = [xF(x) - x]_{x=0}^\infty - \int_0^\infty F(x) - 1 dx = \Big(\lim_{R\to\infty} RF(R) - R\Big) - 0 + \int_0^\infty 1 - F(x)dx \\ &= \int_0^\infty 1 - F(x)dx\end{aligned}$$

Hence,

$$\mathbb{E}[W_1] = \infty \Leftrightarrow \int_0^\infty [1 - F(x)]dx = \infty. \tag{B.114}$$

□

Solution to Exercise 6.20. It suffices to prove that $\prod_{1\leq i<j\leq n}(u_i u_j)^{x_{ij}} = \prod_{i=1}^n u_i^{d_i(x)}$, where $d_i(x) = \sum_{j=1}^n x_{ij}$.
The proof will be given by a simple counting argument. Consider the powers of u_k in the left hand side, for some $k = 1, \ldots, n$. For $k < j \leq n$, the left hand side contains the terms $u_k^{x_{kj}}$, whereas for $1 \leq i < k$, it contains the terms $u_k^{x_{ik}}$. When combined, and using the fact that $x_{ij} = x_{ji}$ for all i, j, we see that the powers of u_k in the left hand side can be written as $\sum_{j\neq k} x_{kj}$. But since, $x_{ii} = 0$ for all i, this equals $\sum_{j=1}^n x_{ij} = d_i(x)$, as required. □

Solution to Exercise 6.21. We pick $t_k = t$ and $t_i = 1$ for all $i \neq k$. Then,

$$\begin{aligned}\mathbb{E}[t^{D_k}] &= \prod_{1\leq i\leq n: i\neq k} \frac{l_n + w_i w_k t}{l_n + w_i w_k} \\ &= e^{w_k(t-1)\sum_{1\leq i\leq n:i\neq k}\frac{w_i}{l_n}+R_n},\end{aligned} \tag{B.115}$$

where

$$\begin{aligned}R_n &= \sum_{1\leq i\leq n:i\neq k} \log\left(1+\frac{w_i w_k t}{l_n}\right) - \log\left(1+\frac{w_i w_k}{l_n}\right) - w_k(t-1)\sum_{1\leq i\leq n:i\neq k}\frac{w_i}{l_n} \\ &= \sum_{1\leq i\leq n:i\neq k} \log(l_n + w_i w_k t) - \log(l_n + w_i w_k) - w_k(t-1)\sum_{1\leq i\leq n:i\neq k}\frac{w_i}{l_n}.\end{aligned} \tag{B.116}$$

A Taylor expansion of $x \mapsto \log(a+x)$ yields that

$$\log(a+x) = \log(a) + \frac{x}{a} + O(\frac{x^2}{a^2}). \tag{B.117}$$

Therefore, applying the above with $a = l_n$ and $x = w_i w_k$, yields that, for t bounded,

$$R_n = O(w_k^2 \sum_{i=1}^{n} \frac{w_i^2}{l_n^2}) = o(1), \tag{B.118}$$

by (**??**), so that

$$\begin{aligned}\mathbb{E}[t^{D_k}] &= e^{w_k(t-1)\sum_{1\leq i\leq n:i\neq k}\frac{w_i}{l_n}}(1+o(1)) \\ &= e^{w_k(t-1)}(1+o(1)),\end{aligned} \tag{B.119}$$

since w_k is fixed. Since the generating function of the degree converges, the degree of vertex k converges in distribution to a random variable with generating function $e^{w_k(t-1)}$ (recall Theorem 2.3(c)). The probability generating function of a Poisson random variable with mean λ is given by $e^{\lambda(t-1)}$, which completes the proof of Theorem 6.6(a).

For Theorem 6.6(b), we use similar ideas, now taking $t_i = t_i$ for $i \leq m$ and $t_i = 0$ for $i > m$. Then,

$$\mathbb{E}[\prod_{i=1}^{m} t_i^{D_i}] = \prod_{1\leq i\leq m, i<j\leq n} \frac{l_n + w_i w_j t_i}{l_n + w_i w_j} = \prod_{i=1}^{m} e^{w_i(t_i-1)}(1+o(1)), \tag{B.120}$$

so that the claim follows. □

Solution to Exercise 6.22. The degree of vertex k converges in distribution to a random variable with generating function $e^{w_k(t-1)}$. We take $w_i = \frac{\lambda}{1-\lambda/n}$ which yields for the generating function $e^{\frac{\lambda(t-1)}{1-\lambda/n}}$. This gives us for the degree a Poi($\frac{\lambda}{1-\lambda/n}$) random variable, which for large n is close to a Poi(λ) random variable.

Solution to Exercise 6.23. The Erdős-Rényi Random Graph is obtained by taking $W_i \equiv \frac{\lambda}{1-\frac{\lambda}{n}}$. Since $p_{ij} = \lambda/n \to 0$, Theorem 6.6(b) states that the degrees are asymptotically independent.

□

Solution to Exercise 6.24. Let X be a mixed Poisson random variable with mixing distribution $\gamma W^{\tau-1}$. The generating function of X now becomes

$$\begin{aligned} G_X(t) = \mathbb{E}[t^X] &= \sum_{k=0}^{\infty} t^k \mathbb{P}(X=k) \\ &= \sum_{k=0}^{\infty} t^k \mathbb{E}[e^{-\gamma W^{\tau-1}} \frac{(\gamma W^{\tau-1})^k}{k!}] \\ &= \mathbb{E}\left[e^{-\gamma W^{\tau-1}} \sum_{k=0}^{\infty} \frac{(\gamma W^{\tau-1} t)^k}{k!} \right] \\ &= \mathbb{E}[e^{(t-1)\gamma W^{\tau-1}}] \end{aligned} \tag{B.121}$$

□

Solution to Exercise 6.25. By using partial integration we obtain

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_0^{\infty} h(x) f(x) dx \\ &= [h(x)(F(x)) - 1]_{x=0}^{\infty} - \int_0^{\infty} h'(x)[F(x)-1]dx \\ &= \Big(\lim_{R\to\infty} h(R)(1-F(R)) \Big) - h(0)(1-F(0)) + \int_0^{\infty} h'(x)[1-F(x)]dx \\ &= \int_0^{\infty} h'(x)[1-F(x)]dx. \end{aligned}$$

□

Solution to Exercise 6.27. By definition, $p^{(n)}$ and $q^{(n)}$ are asymptotically equivalent if for every sequence (x_n) of events

$$\lim_{n\to\infty} p_{x_n}^{(n)} - q_{x_n}^{(n)} = 0. \tag{B.122}$$

By taking the sequence of events $x_n \equiv x \in \mathcal{X}$ for all n, this means that asymptotical equivalence implies that also

$$\lim_{n\to\infty} \max_{x\in\mathcal{X}} |p_x^{(n)} - q_x^{(n)}| = \lim_{n\to\infty} d_{\mathrm{TV}}(p^{(n)}, q^{(n)}) = 0. \tag{B.123}$$

Conversely, if the total variation distance converges to zero, which means that the maximum over all $x \in \mathcal{X}$ of the difference $p_x^{(n)} - q_x^{(n)}$ converges in absolute value to zero. Since this maximum is taken over all $x \in \mathcal{X}$, it will certainly hold for all $x \in (x_n) \subseteq \mathcal{X}$ as well. Therefore, it follows that for any sequence of events, $p_{x_n}^{(n)} - q_{x_n}^{(n)}$ must converge to zero as well, which implies asymptotical equivalence. /ensol

Solution to Exercise 6.28. We recall that

$$d_{\mathrm{TV}}(M, M') = \sup_{A\subset\mathbb{Z}} |\mathbb{P}(M \in A) - \mathbb{P}(M' \in A)|. \tag{B.124}$$

Now, for binomial random variables with the same m and with success probabilities p and q respectively, we have that

$$\frac{\mathbb{P}(M=k)}{\mathbb{P}(M'=k)} = \big(\frac{p}{q}\big)^k \big(\frac{1-p}{1-q}\big)^{m-k} = \big(\frac{1-p}{1-q}\big)^m \big(\frac{p(1-q)}{q(1-p)}\big)^k, \tag{B.125}$$

which is monotonically increasing or decreasing for $p \neq q$. As a result, we have that the supremum in (B.124) is attained for a set $A = \{0, \dots, j\}$ for some $j \in \mathbb{N}$, i.e.,

$$d_{\mathrm{TV}}(M, M') = \sup_{j \in \mathbb{N}} |\mathbb{P}(M \leq j) - \mathbb{P}(M' \leq j)|. \tag{B.126}$$

Now assume that $\lim_{N\to\infty} m(p-q)/\sqrt{mp} = \alpha \in (-\infty, \infty)$. Then, by Exercise 4.22, $(M - mp)/\sqrt{mp} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$ and $(M' - mp)/\sqrt{mp} \xrightarrow{d} Z' sim \mathcal{N}(\alpha, 1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2. Therefore, we arrive at

$$\begin{aligned} d_{\mathrm{TV}}(M, M') &= \sup_{j \in \mathbb{N}} |\mathbb{P}(M \leq j) - \mathbb{P}(M' \leq j)| = \sup_{x \in \mathbb{R}} |\mathbb{P}(Z \leq x) - \mathbb{P}(Z' \leq x)| + o(1) \\ &\to \Phi(\alpha/2) - \Phi(-\alpha/2), \end{aligned} \tag{B.127}$$

where $x \mapsto \Phi(x)$ is the distribution function of a standard normal random variable. Thus, $d_{\mathrm{TV}}(M, M') = o(1)$ precisely when $\alpha = 0$, which implies that $m(p-q)/\sqrt{mp} = o(1)$.

Solution to Exercise 6.29. We write

$$\begin{aligned} d_{\mathrm{TV}}(p, q) &= \frac{1}{2} \sum_x |p_x - q_x| = \frac{1}{2} \sum_x (\sqrt{p_x} + \sqrt{q_x}) |\sqrt{p_x} - \sqrt{q_x}| \\ &= \frac{1}{2} \sum_x \sqrt{p_x} |\sqrt{p_x} - \sqrt{q_x}| + \frac{1}{2} \sum_x \sqrt{q_x} |\sqrt{p_x} - \sqrt{q_x}|. \end{aligned} \tag{B.128}$$

By the Cauchy-Schwarz inequality, we obtain that

$$\sum_x \sqrt{p_x} |\sqrt{p_x} - \sqrt{q_x}| \leq \sqrt{\sum_x p_x} \sqrt{\sum_x (\sqrt{p_x} - \sqrt{q_x})^2} \leq 2^{-1/2} d_{\mathrm{H}}(p, q). \tag{B.129}$$

The same bound applies to the second sum on the right-hand side of (B.128), which proves the upper bound in (6.6.11).

For the lower bound, we bound

$$d_{\mathrm{H}}(p, q)^2 = \frac{1}{2} \sum_x (\sqrt{p_x} - \sqrt{q_x})^2 \leq \frac{1}{2} \sum_x (\sqrt{p_x} + \sqrt{q_x}) |\sqrt{p_x} - \sqrt{q_x}| = d_{\mathrm{TV}}(p, q). \tag{B.130}$$

Solution to Exercise 6.30. By exercise 6.27, we have that $p^{(n)} = \{p_x^{(n)}\}_{x \in \mathcal{X}}$ and $q^{(n)} = \{q_x^{(n)}\}_{x \in \mathcal{X}}$ are asymptotically equivalent if and only if their total variation distance converges to zero. By exercise 6.29, we know that (6.6.11) holds, and therefore also

$$2^{-1/2} d_{\mathrm{TV}}(p^{(n)}, q^{(n)}) \leq d_{\mathrm{H}}(p^{(n)}, q^{(n)}) \leq \sqrt{d_{\mathrm{TV}}(p^{(n)}, q^{(n)})}. \tag{B.131}$$

Both the left and right hand side of those inequalities converge to zero if $d_{\mathrm{TV}}(p^{(n)}, q^{(n)}) \to 0$, which implies by the sandwich theorem that $d_{\mathrm{H}}(p^{(n)}, q^{(n)}) \to 0$. Conversely, if $d_{\mathrm{H}}(p^{(n)}, q^{(n)}) \to 0$, by (6.6.11) we have that $d_{\mathrm{TV}}(p^{(n)}, q^{(n)}) \to 0$. □

Solution to Exercise 6.31. We bound

$$\rho(p, q) = \big(\sqrt{p} - \sqrt{q}\big)^2 + \big(\sqrt{1-p} - \sqrt{1-q}\big)^2 = (p-q)^2 \big((\sqrt{p} + \sqrt{q})^{-2} + (\sqrt{1-p} + \sqrt{1-q})^{-2}\big). \tag{B.132}$$

Solution to Exercise 6.32. We wish to show that $\mathbb{P}(Y=k) = e^{-\lambda p}\frac{(\lambda p)^k}{k!}$. We will use that in the case of X fixed, Y is simply a $\text{Bin}(X,p)$ random variable. We have

$$\begin{aligned}
\mathbb{P}(Y=k) &= \mathbb{P}\Big(\sum_{i=0}^{X} I_i = k\Big) = \sum_{x=k}^{\infty}\mathbb{P}(X=x)\cdot\mathbb{P}\Big(\sum_{i=0}^{x} I_i = k\Big) \\
&= \sum_{x=k}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}\cdot\binom{x}{k}p^k(1-p)^{x-k} = e^{-\lambda}\sum_{x=k}^{\infty}\frac{\lambda^x}{x!}\cdot\frac{x!}{(x-k)!k!}p^k(1-p)^{x-k} \\
&= e^{-\lambda}\frac{(\lambda p)^k}{k!}\sum_{x=k}^{\infty}\frac{\lambda^{x-k}(1-p)^{x-k}}{(x-k)!} = e^{-\lambda}\frac{(\lambda p)^k}{k!}\sum_{x=0}^{\infty}\frac{(\lambda-\lambda p)^x}{x!} \\
&= e^{-\lambda}e^{\lambda-\lambda p}\frac{(\lambda p)^k}{k!} = e^{-\lambda p}\frac{(\lambda p)^k}{k!}
\end{aligned}$$

If we define Y to be the number of edges between i and j at time t and X the same at time $t-1$. Furthermore we define I_k to be the decision of keeping edge k or not. It is given that $X \sim \text{Poi}(\frac{W_iW_j}{L_{t-1}})$ and $I_k \sim \text{Be}(1-\frac{W_t}{L_t})$. According to what is shown above we now obtain for Y to be a Poisson random variable with parameter

$$\frac{W_iW_j}{L_{t-1}}\cdot(1-\frac{W_t}{L_t}) = W_iW_j\frac{1}{L_{t-1}}\frac{L_t-W_t}{L_t} = W_iW_j\frac{1}{L_{t-1}}\frac{L_{t-1}}{L_T} = \frac{W_iW_j}{L_t} \tag{B.133}$$

□

Solution to Exercise 6.33. A graph is simple when it has no self loops or double edges between vertices. Therefore, the Norros-Reittu random graph is simple at time n if for all i $X_{ii}=0$, and for all $i\neq j$ $X_{ij}=0$ or $X_{ij}=1$. By Exercise 6.32, we know that the number of edges X_{ij} between i and j at time n are Poisson with parameter $\frac{w_iw_j}{\ell_n}$. The probability then becomes

$$\begin{aligned}
\mathbb{P}(\text{NR}_n(\boldsymbol{w}) \text{ simple}) &= \mathbb{P}(0\le X_{ij}\le 1, \forall i\neq j)\mathbb{P}(X_{ii}=0,\forall i) \\
&= \prod_{1\le i<j\le n}(\mathbb{P}(X_{ij}=0)+\mathbb{P}(X_{ij}=1))\prod_{k=1}^{n}\mathbb{P}(X_{kk}=0) \\
&= \prod_{1\le i<j\le n}\mathrm{e}^{-\frac{w_iw_j}{\ell_n}}(1+\frac{w_iw_j}{\ell_n})\prod_{k=1}^{n}\mathrm{e}^{-\frac{w_k^2}{\ell_n}} \\
&= \mathrm{e}^{-\sum_{1\le i\le j\le n}\frac{w_iw_j}{\ell_n}}\prod_{1\le i<j\le n}(1+\frac{w_iw_j}{\ell_n}).
\end{aligned} \tag{B.134}$$

□

Solution to Exercise 6.34. Let $X_{ij}\sim\text{Poi}(\frac{w_iw_j}{\ell_n})$ be the number of edges between vertex i and j at time n. The degree of vertex k at time n becomes $\sum_{j=1}^{n}X_{kj}$, and because X_{kj} is Poisson with mean $\frac{w_kw_j}{L_n}$, the sum will be Poisson with mean $\sum_{j=1}^{n}\frac{w_kw_j}{\ell_n} = w_k\frac{\sum_{j=1}^{n}w_j}{\ell_n} = W_k$. Therefore, since the w_i are i.i.d, the degree at time n has a mixed Poisson distribution with mixing distribution F_w

Solution to Exercise 6.35. Couple $X_n = X(G_n)$ and $X'_n = X(G'_n)$ by coupling the edge occupation statuses X_{ij} of G_n and X'_{ij} of G'_n such that (6.7.11) holds. Let $(\hat{X}_n, \hat{X}'_n)$ be this coupling and let E_n and E'_n be the sets of edges of the coupled versions of G_n and G'_n, respectively. Then, since X is increasing

$$\mathbb{P}(\hat{X}_n \leq \hat{X}'_n) \geq \mathbb{P}(E_n \subseteq E'_n) = \mathbb{P}(X_{ij} \leq X'_{ij} \forall i, j \in [n]) = 1, \tag{B.135}$$

which proves the stochastic domination by Lemma 2.11.

Solutions to the exercises of Chapter 7.

Solution to Exercise 7.1. Consider for instance the graph of size $n = 4$ with degrees $\{d_1, \ldots, d_4\} = \{3, 3, 1, 1\}$ or the graph of size $n = 5$ with degrees $\{d_1, \ldots, d_5\} = \{4, 4, 3, 2, 1\}$. □

Solution to Exercise 7.2. For $2m$ vertices we use m pairing steps, each time pairing two vertices with each other. For step $i + 1$, we have already paired $2i$ vertices. The next vertex can thus be paired with $2m - 2i - 1$ other possible vertices. This gives for all pairing steps the total amount of possibilities to be

$$(2m - 1)(2m - 3) \cdots (2m - (2m - 2) - 1) = (2m - 1)!!. \tag{B.136}$$

□

Solution to Exercise 7.8. We can write

$$\mathbb{P}\Big(L_n \text{ is odd}\Big) \quad = \quad \mathbb{P}\Big((-1)^{L_n} = -1\Big) = \frac{1}{2}\Big(1 - \mathbb{E}[(-1)^{L_n}]\Big). \tag{B.137}$$

To compute $\mathbb{E}[(-1)^{L_n}]$, we use the characteristic function $\phi_{D_1}(t) = \mathbb{E}[e^{itD_1}]$ as follows:

$$\phi_{D_1}(\pi) \quad = \quad \mathbb{E}[(-1)^{D_1}] \tag{B.138}$$

Since $(-1)^{L_n} = (-1)^{\sum D_i}$ where $\{D_i\}_{i=1}^n$ are i.i.d. random variables, we have for the characteristic function of L_n, $\phi_{L_n}(\pi) = (\phi_{D_1}(\pi))^n$. Furthermore, we have

$$\phi_{D_1}(\pi) = -\mathbb{P}(D_1 \text{ is odd}) + \mathbb{P}(D_1 \text{ is even}). \tag{B.139}$$

Now we assume $\mathbb{P}(D_1 \text{ is odd}) \notin \{0, 1\}$. This gives us

$$-1 < \mathbb{P}(D_1 \text{ is even}) - \mathbb{P}(D_1 \text{ is odd}) < 1, \tag{B.140}$$

so that $|\phi_{D_1}(\pi)| < 1$, which by (B.137) leads directly to the statement that $\mathbb{P}(L_n \text{ is odd})$ is exponentially close to $\frac{1}{2}$.

□

Solution to Exercise 7.10. We compute

$$\sum_{k=1}^{\infty} k p_k^{(n)} \quad = \quad \sum_{k=1}^{\infty} k\Big(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\tilde{d}_i = k\}}\Big) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty} k \mathbb{1}_{\{\tilde{d}_i = k\}} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{l_n}{n}$$

□

Solution to Exercise ??. First we shall prove that the degrees $P_k^{(n)}$ converge to some probability distribution $\{p_k\}_{k=1}^{\infty}$. Obviously,

$$P_k^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{D_i=k\}}, \tag{B.141}$$

and the variables $\{\mathbb{1}_{\{D_i=k\}}\}_{i=1}^{n}$ are i.i.d. random variables with a $\mathrm{Be}(p_k)$ distribution. Thus, by the strong law of large numbers, $P_k^{(n)} \xrightarrow{a.s.} p_k$.

To see (**??**), we note that the mean of the degree distribution is finite precisely when $\mathbb{E}[D_i] < \infty$. Since $p_k = \mathbb{P}(D_i = k)$, we have

$$\mu = \sum_{k=0}^{\infty} k p_k. \tag{B.142}$$

Now, by definition, the total degree equals

$$L_n = \sum_{i=1}^{n} D_i, \tag{B.143}$$

where, since the degrees are i.i.d. $\{D_i\}_{i=1}^{n}$ is an i.i.d. sequence. Moreover, we have that $\mu = \mathbb{E}[D_i] < \infty$. Thus, (**??**) follows from the strong law of large numbers, since

$$L_n/n = \frac{1}{n}\sum_{i=1}^{n} D_i \xrightarrow{a.s.} \mathbb{E}[D_i] = \mu. \tag{B.144}$$

□

Solution to Exercise ??. We need to prove that (**??**) and (**??**) imply that

$$\sum_{k=1}^{\infty} k p_k^{(n)} \to \mu = \sum_{k=1}^{\infty} k p_k. \tag{B.145}$$

We note that, as $m \to \infty$,

$$\mu = \sum_{k=1}^{\infty} k p_k = \sum_{k=1}^{m} k p_k + o(1). \tag{B.146}$$

Moreover, by (**??**), we have that

$$\sum_{k=m+1}^{\infty} k p_k^{(n)} \le \frac{1}{m}\sum_{k=m+1}^{\infty} k(k-1) p_k^{(n)} \le \frac{1}{m}\sum_{k=1}^{\infty} k(k-1) p_k^{(n)} = O(1/m). \tag{B.147}$$

Thus,

$$\sum_{k=1}^{\infty} k p_k^{(n)} - \mu = \sum_{k=1}^{m} k(p_k^{(n)} - p_k) + o(1). \tag{B.148}$$

Now, for every m fixed, by (**??**),

$$\lim_{N\to\infty} \sum_{k=1}^{m} k(p_k^{(n)} - p_k) = 0, \tag{B.149}$$

and we conclude that, by first sending $n \to \infty$ followed by $m \to \infty$ that $\sum_{k=1}^{\infty} k p_k^{(n)} \to \mu$. □

Solution to Exercise 7.11. We start by evaluating (7.3.20) from the right- to the left-hand side.

$$\begin{aligned}\mu\mathbb{E}[(X+1)^{r-1}] &= \mu\sum_{k=1}^{\infty}(k+1)^{r-1}\frac{e^{-\mu}\mu^k}{k!} = \sum_{k=1}^{\infty}(k+1)^{r}\frac{e^{-\mu}\mu^{k+1}}{(k+1)!};\\ &= \sum_{n=1}^{\infty}n^{r}\frac{e^{-\mu}\mu^{n}}{n!} = \sum_{x=0}^{\infty}x^{r}\frac{e^{-\mu}\mu^{x}}{x!} = \mathbb{E}[X^r].\end{aligned}$$

Now we can use the independency of the two random variables and the result above for the evaluation of (7.3.21).

$$\mathbb{E}[X^rY^s] = \mathbb{E}[X^r]\mathbb{E}[Y^s] = \mathbb{E}[X^r]\mu_Y\mathbb{E}[(Y+1)^{s-1}] = \mu_Y\mathbb{E}[X^r(Y+1)^{s-1}].$$

□

Solution to Exercise 7.12. We use a two-dimensional extension of Theorem 2.3(e), stating that when the mixed moments $\mathbb{E}[X_n^rY_n^s]$ converge to the moments $\mathbb{E}[X^rY^s]$ for each $r,s=0,1,2,\ldots$, and the moments of X and Y satisfy (2.1.8), then (X_n,Y_n) converges in distribution to (X,Y). See also Theorem 2.6 for the equivalent statement for the factorial moments instead of the normal moments, from which the above claim actually follows. Therefore, we are left to prove the asymptotics of the mixed moments of (S_n,M_n).

To prove that $\mathbb{E}[S_n^rM_n^s]$ converge to the moments $\mathbb{E}[S^rM^s]$, we again make use of induction, now in both r and s.

Proposition 7.9 follows when we prove that

$$\lim_{n\to\infty}\mathbb{E}[S_n^r] = \mathbb{E}[S^r] = \mu_S\mathbb{E}[(S+1)^{r-1}], \tag{B.150}$$

and

$$\lim_{n\to\infty}\mathbb{E}[S_n^rM_n^s] = \mathbb{E}[S^rM^s] = \mu_M\mathbb{E}[S^r(M+1)^{s-1}], \tag{B.151}$$

where the second equalities in (B.150) and (B.151) follow from (7.3.20) and (7.3.21).

To prove (B.150), we use the shape of S_n in (7.2.20), which we restate here as

$$S_n = \sum_{i=1}^{n}\sum_{1\le a<b\le d_i}I_{ab,i}. \tag{B.152}$$

Then, we prove by induction on r that

$$\lim_{n\to\infty}\mathbb{E}[S_n^r] = \mathbb{E}[S^r]. \tag{B.153}$$

The induction hypothesis is that (B.153) is true for all $r'\le r-1$, for $\mathrm{CM}_n(\boldsymbol{d})$ when $n\to\infty$ and for all $\{d_i\}_{i=1}^n$ satisfying (**??**). We prove (B.153) by induction on r. For $r=0$, the statement is trivial, which initializes the induction hypothesis.

To advance the induction hypothesis, we write out

$$\begin{aligned}\mathbb{E}[S_n^r] &= \sum_{i=1}^{n}\sum_{1\le a<b\le d_i}\mathbb{E}[I_{ab,i}S_n^{r-1}]\\ &= \sum_{i=1}^{n}\sum_{1\le a<b\le d_i}\mathbb{P}(I_{ab,i}=1)\mathbb{E}[S_n^{r-1}|I_{ab,i}=1].\end{aligned} \tag{B.154}$$

When $I_{ab,i} = 1$, then the *remaining* stubs need to be paired in a uniform manner. The number of self-loops in the total graph in this pairing has the same distribution as

$$1 + S'_n, \tag{B.155}$$

where S'_n is the number of self-loops in the configuration model where with degrees $\{d'_i\}_{i=1}^n$, where $d'_i = d_i - 2$, and $d'_j = d_j$ for all $j \neq i$. The added 1 in (B.155) originates from $I_{ab,i}$. By construction, the degrees $\{d'_i\}_{i=1}^n$ still satisfy (**??**). By the induction hypothesis, for all $k \leq r - 1$

$$\lim_{n\to\infty} \mathbb{E}[(S'_n)^k] = \mathbb{E}[S^k]. \tag{B.156}$$

As a result,

$$\lim_{n\to\infty} \mathbb{E}[(1 + S'_n)^{r-1}] = \mathbb{E}[(1 + S)^{r-1}]. \tag{B.157}$$

Since the limit does not depend on i, we obtain that

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[S_n^r] &= \mathbb{E}[(1 + S)^{r-1}] \lim_{n\to\infty} \sum_{i=1}^n \sum_{1\leq a<b\leq d_i} \mathbb{P}(I_{ab,i} = 1)\\ &\mathbb{E}[(1 + S)^{r-1}] \lim_{n\to\infty} \sum_{i=1}^n \frac{d_i(d_i - 1)}{2}\\ &= \frac{\nu}{2}\mathbb{E}[(1 + S)^{r-1}] = \mathbb{E}[S^r].\end{aligned} \tag{B.158}$$

This advances the induction hypothesis, and completes the proof of (B.150).

To prove (B.151), we perform a similar induction scheme. Now we prove that, for all $r \geq 0$, $\mathbb{E}[S_n^r M_n^s]$ converges to $\mathbb{E}[S^r M^s]$ by induction on s. The claim for $s = 0$ follows from (B.150), which initializes the induction hypothesis, so we are left to advance the induction hypothesis. We follow the argument for S_n above. It is not hard to see that it suffices to prove that, for every ij,

$$\lim_{n\to\infty} \mathbb{E}[S_n^r M_n^{s-1} | I_{s_1t_1,s_2t_2,ij} = 1] = \mathbb{E}[S^r(1 + M)^{s-1}]. \tag{B.159}$$

Note that when $I_{s_1t_1,s_2t_2,ij} = 1$, then we know that two edges are paired together to form a multiple edge. Removing these two edges leaves us with a graph which is very close to the configuration model with degrees $\{d'_i\}_{i=1}^n$, where $d'_i = d_i - 2$, and $d'_j = d_j - 2$ and $d'_t = d_t$ for all $t \neq i, j$. The only difference is that when a stub connected to i is attached to a stub connected to j, then this creates an additional number of multiple edges. Ignoring this effect creates the lower bound

$$\mathbb{E}[S_n^r M_n^{s-1} | I_{s_1t_1,s_2t_2,ij} = 1] \geq \mathbb{E}[S_n^r (M_n + 1)^{s-1}], \tag{B.160}$$

which, by the induction hypothesis, converges to $\mathbb{E}[S^r(1 + M)^{s-1},]$ as required.

Let $I'_{s_1t_1,s_2t_2,ij}$ denote the indicator that stub s_1 is connected to t_1, s_2 to t_2 and *no other stub* of vertex i is connected to a stub of vertex j. Then,

$$\frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} I'_{s_1t_1,s_2t_2,ij} \leq M_n \leq \frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} I_{s_1t_1,s_2t_2,ij}. \tag{B.161}$$

Hence,

$$\mathbb{E}[S_n^r M_n^s] \leq \frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} \mathbb{P}(I_{s_1t_1,s_2t_2,ij} = 1)\mathbb{E}\big[S_n^r M_n^{s-1} | I_{s_1t_1,s_2t_2,ij} = 1\big], \tag{B.162}$$

and

$$\mathbb{E}[S_n^r M_n^s] \leq \frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} \mathbb{P}(I'_{s_1t_1,s_2t_2,ij}=1)\mathbb{E}\big[S_n^r M_n^{s-1} | I'_{s_1t_1,s_2t_2,ij}=1\big]. \tag{B.163}$$

Now, by the above, $\mathbb{E}\big[S_n^r M_n^{s-1} | I_{s_1t_1,s_2t_2,ij}=1\big]$ and $\mathbb{E}\big[S_n^r M_n^{s-1} | I'_{s_1t_1,s_2t_2,ij}=1\big]$ converge to $\mathbb{E}\big[S^r (M_+1)^{s-1}\big]$, independently of s_1t_1, s_2t_2, ij. Further,

$$\frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} \mathbb{P}(I'_{s_1t_1,s_2t_2,ij}=1) \to \nu^2/2, \tag{B.164}$$

and also

$$\frac{1}{2} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s_1<s_2\leq d_i} \sum_{1\leq t_1\neq t_2\leq d_j} \mathbb{P}(I_{s_1t_1,s_2t_2,ij}=1) \to \nu^2/2. \tag{B.165}$$

This implies that

$$\mathbb{E}[S_n^r M_n^{s-1} | I_{s_1t_1,s_2t_2,ij}=1] = \mathbb{E}[S_{n-1}^r M_{n-1}^{s-1}] + o(1). \tag{B.166}$$

The remainder of the proof is identical to the one leading to (B.158). □

Solution to Exercise 7.13. To obtain a triangle we need to three connected stubs say $(s_1,t_1),(s_2,t_2),(s_3,t_3)$ where s_1 and t_3 belong to some vertex i with degree d_i, s_2 and t_1 to vertex j with degree d_j and s_3, t_2 to some vertex k with degree d_k. Obviously we have

$$\begin{aligned} 1 \leq \quad & s_1 \leq \quad d_i, \\ 1 \leq \quad & t_1 \leq \quad d_j, \\ 1 \leq \quad & s_2 \leq \quad d_j, \\ 1 \leq \quad & t_2 \leq \quad d_k, \\ 1 \leq \quad & s_3 \leq \quad d_k, \\ 1 \leq \quad & t_3 \leq \quad d_i. \end{aligned}$$

The probability of connecting s_1 to t_1 is $1/(l_n-1)$. Furthermore, connecting s_2 to t_2 appears with probability $1/(l_n-3)$ and s_3 to t_3 with probability $1/(l_n-5)$. Of course we can pick all stubs of i to be s_1, and we have d_i-1 vertices left from which we may choose t_3. Hence, for the amount of triangles we obtain

$$\begin{aligned} \sum_{i<j<k} \frac{d_i d_j}{l_n-1} \cdot \frac{(d_j-1)d_k}{l_n-3} \cdot \frac{(d_k-1)(d_i-1)}{l_n-5} &= \sum_{i<j<k} \frac{d_i(d_i-1)}{l_n-1} \cdot \frac{d_j(d_j-1)}{l_n-3} \cdot \frac{d_k(d_k-1)}{l_n-5} \\ &\sim \frac{1}{6}\Big(\sum_{i=1}^n \frac{d_i(d_i-1)}{l_n}\Big)^3. \end{aligned} \tag{B.167}$$

We will show that

$$\sum_{i<j<k} \frac{d_i(d_i-1)}{l_n-1} \cdot \frac{d_j(d_j-1)}{l_n-3} \cdot \frac{d_k(d_k-1)}{l_n-5} \sim \frac{1}{6}\Big(\sum_{i=1}^n \frac{d_i(d_i-1)}{l_n}\Big)^3$$

by expanding the righthand-side. We define

$$S := \Big(\sum_{i=1}^n \frac{d_i(d_i-1)}{l_n}\Big)^3. \tag{B.168}$$

Then, we have

$$S = \sum_{i=1}^{n} \Big(\frac{d_i(d_i-1)}{l_n}\Big)^3 + 3\sum_{i=1}^{\infty}\sum_{j=1,j\neq i}^{\infty} \Big(\frac{d_i(d_i-1)}{l_n}\Big)^2 \Big(\frac{d_j(d_j-1)}{l_n}\Big) \tag{B.169}$$

$$+ \quad \sum_{i\neq j\neq k} \frac{d_i(d_i-1)}{l_n} \cdot \frac{d_j(d_j-1)}{l_n} \cdot \frac{d_k(d_k-1)}{l_n}, \tag{B.170}$$

where the first part contains n terms, the second $n(n-1)$ and the third $n(n-1)(n-2)$. So for large n we can say that

$$S \sim \sum_{i\neq j\neq k} \frac{d_i(d_i-1)}{l_n} \cdot \frac{d_j(d_j-1)}{l_n} \cdot \frac{d_k(d_k-1)}{l_n}. \tag{B.171}$$

Now there are six possible orderings of i, j, k, hence

$$\frac{1}{6}S \sim \sum_{i<j<k} \frac{d_i(d_i-1)}{l_n} \cdot \frac{d_j(d_j-1)}{l_n} \cdot \frac{d_k(d_k-1)}{l_n} \sim \sum_{i<j<k} \frac{d_i(d_i-1)}{l_n-1} \cdot \frac{d_j(d_j-1)}{l_n-3} \cdot \frac{d_k(d_k-1)}{l_n-5}. \tag{B.172}$$

□

Solution to Exercise 7.17. In this case we have $d_i = r$ for all $i \in [n]$. This gives us

$$\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{d_i(d_i-1)}{l_n} = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{r(r-1)}{nr} = r-1. \tag{B.173}$$

Furthermore we obtain

$$\prod_{i=1}^{n} d_i! = \prod_{i=1}^{n} r! = (r!)^n. \tag{B.174}$$

Finally we have for the total number of stubs $l_n = rn$. Substituting these variables in (7.4.1) gives us for the number of simple graphs with constant degree sequence $d_i = r$

$$e^{-\frac{(r-1)}{2} - \frac{(r-1)^2}{4}} \frac{(rn-1)!!}{(r!)^n}(1+o(1)). \tag{B.175}$$

□

Solutions to the exercises of Chapter 8.

Solution to Exercise 8.1. At time t, we add a vertex v_t, and connect it with each vertex v_i, $1 \leq i < t$ with probability p. In the previous chapters, we had the relation $p = \frac{\lambda}{n}$, but since n is increasing over time, using this expression for p will not result in an Erdős-Rényi random graph. We could off course wish to obtain a graph of size N, thus stopping the algorithm at time $t = N$, and using $p = \frac{\lambda}{N}$.

□

Solution to Exercise 8.2. We will use an induction argument over t. For $t = 1$ we have a single vertex v_1 with a self-loop, hence $d_1(1) = 2 \geq 1$.

Now suppose at time t we have $d_i(t) \geq 1 \, \forall_i$.

At time $t+1$ we add a vertex v_{t+1}. We do not remove any edges, so we only have to check whether the newly added vertex has a non-zero degree. Now the algorithm adds the vertex having a single edge, to be connected to itself, in which case $d_{t+1}(t+1) = 2$, or to be connected to another already existing vertex, in which case it's degree is 1. In the latter case, one is added to the degree of the vertex to which v_{t+1} is connected, thus that degree is still greater than zero. Hence we can say that $d_i(t+1) \geq 1 \, \forall_i$
We can now conclude that $d_i(t) \geq 1$ for all i and t. The statement $d_i(t) + \delta \geq 0$ for all $\delta \geq -1$ follows directly. □

Solution to Exercise 8.3. The statement

$$\frac{1+\delta}{t(2+\delta)+(1+\delta)} + \sum_{i=1}^{t} \frac{d_i(t)+\delta}{t(2+\delta)+(1+\delta)} = 1 \tag{B.176}$$

will follow directly if the following equation holds:

$$(1+\delta) + \sum_{i=1}^{t} (d_i(t)+\delta) = t(2+\delta)+(1+\delta). \tag{B.177}$$

Which is in its turn true if

$$\sum_{i=1}^{t} (d_i(t)+\delta) = t(2+\delta). \tag{B.178}$$

But since $\sum_{i=1}^{t} d_i(t) = 2t$ by construction, the latter equation holds. Hence, the upper statement holds and the probabilities do sum up to one.

□

Solution to Exercise 8.6. We will again use an induction argument. At time $t = 1$ we have a single vertex v_1 with a self-loop, and the statement holds. At time $t = 2$ we add a vertex v_2 and connect it with v_1 with the given probability

$$\mathbb{P}\big(v_2 \to v_1 \big| \mathrm{PA}_{1,\delta}(1)\big) = \frac{2-1}{1} = 1. \tag{B.179}$$

Now suppose at time t we have a graph with one vertex v_1 containing a self-loop and $t-1$ other vertices having only one edge which connects it to v_1. In that case $d_1(t) = 2 + (t-1) = t+1$ and all other vertices have degree 1.
At time $t+1$ we add a vertex v_{t+1} having one edge which will be connected to v_1 with probability

$$\mathbb{P}\big(v_{t+1} \to v_1 \big| \mathrm{PA}_{1,\delta}(t)\big) = \frac{t+1-1}{t} = 1. \tag{B.180}$$

Hence, the claim follows by induction. □

Solution to Exercise 8.7. The proof is by induction on $t \geq 1$. For $t = 1$, the statement is correct, since, at time 2, both graphs consist of two vertices with two edges between them. This initializes the induction hypothesis.

To advance the induction hypothesis, we assume that the law of $\{\mathrm{PA}_{1,\alpha}^{(b')}(t)\}_{s=1}^{t}$ is equal to the one of $\{\mathrm{PA}_{1,\delta}^{(b)}(s)\}_{s=1}^{t}$, and, from this, prove that the law of $\{\mathrm{PA}_{1,\alpha}^{(b')}(s)\}_{s=1}^{t}$ is equal to the one of $\{\mathrm{PA}_{1,\delta}^{(b)}(s)\}_{s=1}^{t}$. The only difference between $\mathrm{PA}_{1,\delta}^{(b)}(t+1)$ and $\mathrm{PA}_{1,\delta}^{(b)}(t)$ and between $\mathrm{PA}_{1,\alpha}^{(b')}(t+1)$ and $\mathrm{PA}_{1,\alpha}^{(b')}(t)$ is to what vertex the $(t+1)^{\mathrm{st}}$ edge is attached. For

$\{\mathrm{PA}_{1,\delta}^{(b)}(t)\}_{t=1}^{\infty}$ and conditionally on $\mathrm{PA}_{1,\delta}^{(b)}(t)$, this edge is attached to vertex i with probability

$$\frac{D_i(t)+\delta}{t(2+\delta)}, \tag{B.181}$$

while, for $\{\mathrm{PA}'_{1,\alpha}(t)\}_{t=1}^{\infty}$ and conditionally on $\mathrm{PA}'_{1,\alpha}(t)$, this edge is attached to vertex i with probability

$$\alpha\frac{1}{t}+(1-\alpha)\frac{D_i(t)}{2t}. \tag{B.182}$$

Bringing the terms in (B.182) onto a single denominator yields

$$\frac{D_i(t)+2\frac{\alpha}{1-\alpha}}{\frac{2}{1-\alpha}t}, \tag{B.183}$$

which agrees with (B.181) precisely when $2\frac{\alpha}{1-\alpha}=\delta$, so that

$$\alpha=\frac{\delta}{2+\delta}. \tag{B.184}$$

□

Solution to Exercise 8.9. We write

$$\Gamma(t+1)=\int_0^\infty x^t e^{-x}dx. \tag{B.185}$$

Using partial integration we obtain

$$\Gamma(t+1) \quad = \quad [-x^t e^{-x}]_{x=0}^{\infty}+\int_0^\infty tx^{t-1}e^{-x}dx=0+t\cdot\int_0^\infty x^{t-1}e^{-x}dx=t\Gamma(t).$$

In order to prove that $\Gamma(n)=(n-1)!$ for $n=1,2,\ldots$ we will again use an induction argument. For $n=1$ we have

$$\Gamma(1) \quad = \quad \int_0^\infty x^0 e^{-x}dx=\int_0^\infty e^{-x}dx=1=(0)!.$$

Now the upper result gives us for $n=2$

$$\Gamma(2)=1\cdot\Gamma(1)=1=(2-1)!. \tag{B.186}$$

Suppose now that for some $n\in\mathbb{N}$ we have $\Gamma(n)=(n-1)!$. Again (8.2.2) gives us for $n+1$

$$\Gamma(n+1)=n\Gamma(n)=n(n-1)!=n!. \tag{B.187}$$

Induction yields $\Gamma(n)=(n-1)!$ for $n=1,2,\ldots$. □

Solution to Exercise 8.10. We rewrite (8.2.9) to be

$$\begin{aligned}
e^{-t}t^{t-\frac{1}{2}}\sqrt{2\pi} \quad &\leq \Gamma(t+1)\leq \quad e^{-t}t^t\sqrt{2\pi}\Big(1+\frac{1}{12t}\Big),\\
(\frac{t}{e})^t\sqrt{\frac{2\pi}{t}} \quad &\leq \Gamma(t+1)\leq \quad (\frac{t}{e})^t\sqrt{2\pi}(1+\frac{1}{12t}),\\
(\frac{t}{e})^t\sqrt{\frac{2\pi}{t}} \quad &\leq t\Gamma(t)\leq \quad (\frac{t}{e})^t\sqrt{2\pi}(1+\frac{1}{12t}),\\
(\frac{t}{e})^t\sqrt{\frac{2\pi}{t}}\frac{1}{t} \quad &\leq \Gamma(t)\leq \quad (\frac{t}{e})^t\sqrt{\frac{2\pi}{t}}\sqrt{t}(1+\frac{1}{12t}).
\end{aligned}$$

Using this inequality in the left-hand side of (8.2.8) we obtain

$$\begin{aligned}\frac{(\frac{t}{e})^t\sqrt{\frac{2\pi}{t}}\frac{1}{t}}{(\frac{t-a}{e})^{t-a}\sqrt{\frac{2\pi}{t-a}}\sqrt{t-a}(1+\frac{1}{12(t-a)})} &\leq \tfrac{\Gamma(t)}{\Gamma(t-a)} \leq \frac{(\frac{t}{e})^t\sqrt{\frac{2\pi}{t}}\sqrt{t}(1+\frac{1}{12t})}{(\frac{t-a}{e})^{t-a}\sqrt{\frac{2\pi}{t-a}}\frac{1}{t-a}}\\ \frac{t^t}{(t-a)^{t-a}}\frac{e^{-a}}{t\sqrt{t}(1+12/(t-a))} &\leq \tfrac{\Gamma(t)}{\Gamma(t-a)} \leq \frac{t^t}{(t-a)^{t-a}}\frac{e^{-a}(1+1/12t)}{\sqrt{t-a}}.\end{aligned}$$

We complete the proof by noting that $t-a = t(1+O(1/t))$ and $1+1/12t = 1+O(1/t)$. □

Solution to Exercise 8.11. This result is immediate from the collapsing of the vertices in the definition of $\mathrm{PA}_{m,\delta}(t)$, which implies that the degree of vertex $v_i^{(m)}$ in $\mathrm{PA}_{m,\delta}(t)$ is equal to the sum of the degrees of the vertices $v^{(1)}_{m(i-1)+1}, \ldots, v^{(1)}_{mi}$ in $\mathrm{PA}_{mt}(1, \delta/m)$. □

Solution to Exercise 8.16. We wish to prove

$$\mathbb{P}\Big(|P_{\geq k}(t) - \mathbb{E}[P_{\geq k}(t)]| \geq C\sqrt{t\log t}\Big) = o(t^{-1}). \tag{B.188}$$

First of all we have $P_{\geq k}(t) = 0$ for $k > mt$. We define, similarly to the proof of Proposition 8.3 the martingale

$$M_n = \mathbb{E}\big[P_{\geq k}(t)|\mathrm{PA}_{m,\delta}(n)\big]. \tag{B.189}$$

We have

$$\mathbb{E}[M_{n+1}|\mathrm{PA}_{m,\delta}(n)] = \mathbb{E}\Big[\mathbb{E}\big[P_{\geq k}(t)|\mathrm{PA}_{m,\delta}(n+1)\big]\Big|\mathrm{PA}_{m,\delta}(n)\Big] = \mathbb{E}\big[P_{\geq k}(t)|\mathrm{PA}_{m,\delta}(n)\big] = M_n. \tag{B.190}$$

Hence M_n is a martingale. Furthermore, M_n satisfies the moment condition, since

$$\mathbb{E}\big[M_n\big] = \mathbb{E}\big[P_{\geq k}(t)\big] \leq t < \infty. \tag{B.191}$$

Clearly, $\mathrm{PA}_{m,\delta}(0)$ is the empty graph, hence for M_0 we obtain

$$M_0 = \mathbb{E}\big[P_{\geq k}(t)|\mathrm{PA}_{m,\delta}(0)\big] = \mathbb{E}\big[P_{\geq k}(t)]. \tag{B.192}$$

We obtain for M_t

$$M_t = \mathbb{E}\big[P_{\geq k}(t)|\mathrm{PA}_{m,\delta}(t)\big] = \big[P_{\geq k}(t), \tag{B.193}$$

since $P_{\geq k}(t)$ can be determined when $\mathrm{PA}_{m,\delta}(t)$ is known. Therefore, we have

$$P_{\geq k}(t) - \mathbb{E}[P_{\geq k}(t)] = M_t - M_0. \tag{B.194}$$

To apply the Azuma-Hoeffding inequality, Theorem 2.23, we have to bound $|M_n - M_{n-1}|$. In step n, m edges are added to the graph. Now $P_{\geq k}$ only changes is an edge is added to a vertex with degree $k-1$. Now m edges have influence on the degree of at most $2m$ vertices, hence, the maximum amount of vertices of which de degree is increased to k is at most $2m$. So we have $|M_n - M_{n-1}| \leq 2m$. The Azuma-Hoeffding inequality now gives us

$$\mathbb{P}\Big(|P_{\geq k}(t) - \mathbb{E}[P_{\geq k}(t)]| \geq a\Big) \leq 2e^{-\frac{a^2}{8m^2t}}. \tag{B.195}$$

Taking $a = C\sqrt{t\log t}$, $C^2 \geq 8m$, we obtain

$$\mathbb{P}\Big(|P_{\geq k}(t) - \mathbb{E}[P_{\geq k}(t)]| \geq C\sqrt{t\log t}\Big) = o(t^{-1}). \tag{B.196}$$

□

Solution to Exercise 8.18. We have for $\kappa_k(t)$ and $\gamma_k(t)$ the following equation.

$$\begin{aligned}\kappa_k(t) &= \Big(\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big)(k-1+\delta)p_{k-1} - \Big(\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big)(k+\delta)p_k,\\ \gamma_k(t) &= -\mathbb{1}\{k=1\}\frac{1+\delta}{t(2+\delta)+(1+\delta)} + \mathbb{1}\{k=2\}\frac{1+\delta}{t(2+\delta)+(1+\delta)}.\end{aligned}$$

We start with C_γ. We have

$$|\gamma_k(t)| \le \frac{1+\delta}{t(2+\delta)+(1+\delta)} \le \frac{1}{t(\frac{2+\delta}{1+\delta})+1} \le \frac{1}{t+1}. \tag{B.197}$$

So indeed $C_\gamma = 1$ does the job. For $\kappa_k(t)$ we have

$$\kappa_k(t) = \Big(\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big)\Big((k-1+\delta)p_{k-1} - (k+\delta)p_k\Big). \tag{B.198}$$

This gives us

$$\begin{aligned}|\kappa_k(t)| &\le \Big|\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big| \cdot \Big|(k-1+\delta)p_{k-1} - (k+\delta)p_k\Big|,\\ &\le \Big|\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+(1+\delta)}\Big| \cdot \sup_{k\ge1}(k+\delta)p_k,\\ &= \Big|\frac{t(2+\delta)+(1+\delta)-(2+\delta)t}{t(2+\delta)^2+(1+\delta)(2+\delta)}\Big| \cdot \sup_{k\ge1}(k+\delta)p_k,\\ &= \Big|\frac{1+\delta}{t(2+\delta)^2+(1+\delta)(2+\delta)}\Big| \cdot \sup_{k\ge1}(k+\delta)p_k,\\ &= \Big|\frac{1}{2+\delta}\cdot\frac{1}{t(\frac{2+\delta}{1+\delta})+1}\Big| \cdot \sup_{k\ge1}(k+\delta)p_k,\\ &\le \Big|\frac{1}{t(\frac{2+\delta}{1+\delta})+1}\Big| \cdot \sup_{k\ge1}(k+\delta)p_k,\\ &\le \frac{1}{t+1}\cdot\sup_{k\ge1}(k+\delta)p_k.\end{aligned}$$

Hence, $C_\kappa = \sup_{k\ge1}(k+\delta)p_k$

□

Solution to Exercise 8.17. We note that

$$\sum_{i:D_i(t)\ge l} D_i(t) \ge l N_{\ge l}(t), \tag{B.199}$$

where we recall that $N_{\ge l}(t) = \#\{i \le t : D_i(t) \ge l\}$ is the number of vertices with degree at least l.

By the proof of Proposition 8.3 (see also Exercise 8.16), there exists C_1 such that uniformly for all l,

$$\mathbb{P}\Big(|N_{\ge l}(t) - \mathbb{E}[N_{\ge l}(t)]| \ge C_1\sqrt{t\log t}\Big) = o(t^{-1}). \tag{B.200}$$

By Proposition 8.4, there exists a constant C_2 such that

$$\sup_{l\ge1}|\mathbb{E}[P_l(t)] - tp_l| \le C_2. \tag{B.201}$$

Therefore, we obtain that, with probability exceeding $1-o(t^{-1})$,

$$N_{\geq l}(t) \geq \mathbb{E}[N_{\geq l}(t)] - C_1\sqrt{t\log t} \geq \mathbb{E}[N_{\geq l}(t)] - \mathbb{E}[N_{\geq 2l}(t)] - C_1\sqrt{t\log t}$$
$$\geq \sum_{l=l}^{2l-1}[tp_l - C_2] - C_1\sqrt{t\log t} \geq C_3 t l^{1-\tau} - C_2 l - C_1\sqrt{t\log t} \geq B t l^{2-\tau}, \quad \text{(B.202)}$$

whenever l is such that

$$t l^{1-\tau} \gg l, \qquad \text{and} \qquad t l_t^{1-\tau} \gg \sqrt{t\log t}. \quad \text{(B.203)}$$

The first condition is equivalent to $l \ll t^{\frac{1}{\tau}}$, and the second to $l \ll t^{\frac{1}{2(\tau-1)}}(\log t)^{-\frac{1}{2(\tau-1)}}$. Note that $\frac{1}{\tau} \geq \frac{1}{2(\tau-1)}$ for all $\tau > 2$, so the second condition is the strongest, and follows when $t l^{2-\tau} \geq K\sqrt{t\log t}$ for some K sufficiently large.

Then, for l satisfying $t l^{2-\tau} \geq K\sqrt{t\log t}$, we have with probability exceeding $1-o(t^{-1})$,

$$\sum_{i:D_i(t)\geq l} D_i(t) \geq B t l^{2-\tau}. \quad \text{(B.204)}$$

Also, with probability exceeding $1-o(t^{-1})$, for all such l, $N_{\geq l}(t) \gg \sqrt{t}$. □

Solution to Exercise 8.19. We prove (8.6.3) by induction on $j \geq 1$. Clearly, for every $t \geq i$,

$$\mathbb{P}(D_i(t)=1) = \prod_{s=i+1}^{t}\Big(1-\frac{1+\delta}{(2+\delta)(s-1)+(1+\delta)}\Big) = \prod_{s=i+1}^{t}\Big(\frac{s-1}{s-1+\frac{1+\delta}{2+\delta}}\Big) = \frac{\Gamma(t)\Gamma(i+\frac{1+\delta}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})\Gamma(i)}, \quad \text{(B.205)}$$

which initializes the induction hypothesis, since $C_1 = 1$.

To advance the induction, we let $s \leq t$ be the last time at which a vertex is added to i. Then we have that

$$\mathbb{P}(D_i(t)=j) = \sum_{s=i+j-1}^{t}\mathbb{P}\big(D_i(s-1)=j-1\big)\frac{j-1+\delta}{(2+\delta)(s-1)+1+\delta}\mathbb{P}\big(D_i(t)=j|D_i(s)=j\big). \quad \text{(B.206)}$$

By the induction hypothesis, we have that

$$\mathbb{P}\big(D_i(s-1)=j-1\big) \leq C_{j-1}\frac{\Gamma(s-1)\Gamma(i+\frac{1+\delta}{2+\delta})}{\Gamma(s-1+\frac{1+\delta}{2+\delta})\Gamma(i)}. \quad \text{(B.207)}$$

Moreover, analogously to (B.205), we have that

$$\mathbb{P}(D_i(t)=j|D_i(s)=j) = \prod_{q=s+1}^{t}\Big(1-\frac{j+\delta}{(2+\delta)(q-1)+(1+\delta)}\Big) \quad \text{(B.208)}$$
$$= \prod_{q=s+1}^{t}\Big(\frac{q-1-\frac{j-1}{2+\delta}}{q-1+\frac{1+\delta}{2+\delta}}\Big) = \frac{\Gamma(t-\frac{j-1}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})\Gamma(s-\frac{j-1}{2+\delta})}.$$

Combining (B.207) and (B.208), we arrive at

$$\mathbb{P}(D_i(t)=j) \leq \sum_{s=i+j-1}^{t}\Big(C_{j-1}\frac{\Gamma(s-1)\Gamma(i+\frac{1+\delta}{2+\delta})}{\Gamma(s-1+\frac{1+\delta}{2+\delta})\Gamma(i)}\Big)\Big(\frac{j-1+\delta}{(2+\delta)(s-1)+(1+\delta)}\Big)$$
$$\times\Big(\frac{\Gamma(t-\frac{j-1}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})\Gamma(s-\frac{j-1}{2+\delta})}\Big). \quad \text{(B.209)}$$

We next use that

$$\Gamma(s-1+\frac{1+\delta}{2+\delta})((2+\delta)(s-1)+(1+\delta))=(2+\delta)\Gamma(s+\frac{1+\delta}{2+\delta}), \tag{B.210}$$

to arrive at

$$\mathbb{P}(D_i(t)=j)\le C_{j-1}\frac{j-1+\delta}{2+\delta}\frac{\Gamma(i+\frac{1+\delta}{2+\delta})}{\Gamma(i)}\frac{\Gamma(t-\frac{j-1}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})}\sum_{s=i+j-1}^{t}\frac{\Gamma(s-1)}{\Gamma(s-\frac{j-1}{2+\delta})}. \tag{B.211}$$

We note that, whenever $l+b, l+1+a>0$ and $a-b+1>0$,

$$\sum_{s=l}^{t}\frac{\Gamma(s+a)}{\Gamma(s+b)}=\frac{1}{a-b+1}\Big[\frac{\Gamma(t+1+a)}{\Gamma(t+b)}-\frac{\Gamma(l+1+a)}{\Gamma(l+b)}\Big]\le\frac{1}{a-b+1}\frac{\Gamma(t+1+a)}{\Gamma(t+b)}. \tag{B.212}$$

Application of (B.212) for $a=-1, b=-\frac{j-1}{2+\delta}, l=i+j-1$, so that $a-b+1=\frac{j-1}{2+\delta}>0$ when $j>1$, leads to

$$\begin{aligned}\mathbb{P}(D_i(t)=j)&\le C_{j-1}\frac{j-1+\delta}{2+\delta}\frac{\Gamma(i+\frac{1+\delta}{2+\delta})}{\Gamma(i)}\frac{\Gamma(t-\frac{j-1}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})}\frac{1}{\frac{j-1}{2+\delta}}\frac{\Gamma(t)}{\Gamma(t-\frac{j-1}{2+\delta})}\\&=C_{j-1}\frac{j-1+\delta}{j-1}\frac{\Gamma(i+\frac{1+\delta}{2+\delta})}{\Gamma(i)}\frac{\Gamma(t)}{\Gamma(t+\frac{1+\delta}{2+\delta})}.\end{aligned} \tag{B.213}$$

Equation (B.213) advances the induction by (8.6.4). □

Solution to Exercise 8.24. Suppose $\alpha\delta_{\text{in}}+\gamma=0$, then, since all non-negative, we have $\gamma=0$ and either $\alpha=0$ or $\delta_{\text{in}}=0$.
Since $\gamma=0$, no new vertices are added with non zero in-degree.
In case of $\alpha=0$ we have $\beta=1$, and thus we only create edges in G_0. Hence, no vertices exist outside G_0 and thus there cannot exist vertices outside G_0 with in-degree non zero.
In case of $\delta_{\text{in}}=0$ (and $\gamma=0$ still), vertices can be created outside G_0, but in in it's creation phase we will only give it an outgoïng edge. And this edge will be connected to a vertex inside G_0, since $\delta_{\text{in}}=0$ and the possibility to is thus zero to create an ingoing edge to a vertex with $d_i(t)=0$. Similarly, in case edges are created within the existing graphs, all ingoing edges will be in G_0 for the same reason. So, during all stages all vertices outside G_0 will have in-degree zero.

Now suppose $\gamma=1$. Then the only edges being created during the process are those from inside the existing graph to the newly created vertex. So once a vertex is created and connected to the graph, it will only be able to gain out-going edges. Hence, the in-degree remains one for all vertices outside G_0 at all times.

Solutions to the exercises of Chapter 9.

Solution to Exercise ??. Let $\kappa(x,y)=[1-F]^{-1}(x)[1-F]^{-1}(y)/\mathbb{E}[W]$, which consists of the product of two functions that are continuous a.e. on $\mathcal{S}$, and therefore satisfies 9.27(a). To see 9.27 is satisfied, consider the double integral

$$\begin{aligned}\iint_{\mathcal{S}^2}\kappa(x,y)d\mu(x)d\mu(y)&=\frac{1}{\mathbb{E}[W]}\int_{\mathcal{S}}[1-F]^{-1}(y)\int_{\mathcal{S}}[1-F]^{-1}(x)d\mu(x)d\mu(y)\\&=\frac{1}{\mathbb{E}[W]}\int_{\mathcal{S}}[1-F]^{-1}(y)\mathbb{E}[W]d\mu(y)=\mathbb{E}[W],\end{aligned} \tag{B.214}$$

which is finite if and only if $\mathbb{E}[W]$ is finite.
The expected number of edges in the graph is half of the sum of expected degrees over the vertices, due to the handshaking lemma, which is $\frac{n\mu}{2}$ in this case. Therefore,

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))] \to \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)d\mu(x)d\mu(y) = \frac{1}{2}\mu, \tag{B.215}$$

which proves 9.27(c), and thus the kernel is graphical exactly when $\mathbb{E}[W]$ is finite.

For irreducibility, note that $[1-F]^{-1}(x)[1-F]^{-1}(y) = 0$ if and only if $x = 1$ or $y = 1$, which gives rise to two sets that both have measure 0 in $\mathcal{S}^2$. □

Solution to Exercise 9.30. Denote by $\tilde{\mathbf{p}}$ and $\mathbf{p}$ the edge occupation probabilities for $\mathrm{CL}_n(\tilde{\boldsymbol{w}})$ and $\mathrm{CL}_n(\boldsymbol{w})$, respectively, i.e.,

$$\tilde{p}_{ij} = \min\{1, \tilde{w}_i\tilde{w}_j/l_n\} = \min\{1, w_iw_j/(n\mu)\} \qquad p_{ij} = \min\{1, w_iw_j/l_n\}. \tag{B.216}$$

We assume without loss of generality that $l_n \geq \mu n$, when $l_n \leq \mu n$ the argument can easily be adapted. Then, $p_{ij} \leq \tilde{p}_{ij}$, and

$$\tilde{p}_{ij} - p_{ij} \leq (\frac{\mu n}{l_n} - 1)\tilde{p}_{ij}. \tag{B.217}$$

We note that

$$\sum_{1\leq i<j\leq n} (\tilde{p}_{ij} - p_{ij})^2/\tilde{p}_{ij} \leq (\frac{\mu n}{l_n} - 1)^2 \sum_{1\leq i<j\leq n} \tilde{p}_{ij} \leq (\frac{\mu n}{l_n} - 1)^2 l_n = o(1), \tag{B.218}$$

whenever $(\frac{\mu n}{l_n} - 1)^2 = o(n)$. Theorem 6.17 then proves the claim. □

Solution to Exercise ??. Assume (9.5.13) holds. Note that this means that for each i and j, $\kappa(x_i, x_j)/n \to 0$ as well. Asymptotically, when considering $\min\{\kappa(x,y)/n, 1\}$, we get that $\kappa(x,y)/n < 1$, and we will write $\kappa_{ij} = \kappa(x_i, x_j)$. We now apply Theorem 6.17 with p_{ij} as in (9.5.7) and $q_{ij} = p_{ij}^{(\mathrm{NR})}(\kappa)$. To approximate $q_{ij} = 1 - \mathrm{e}^{-\kappa_{ij}/n}$ we use the Taylor approximation $\kappa_{ij}/n - \kappa_{ij}^2/2n^2 + O(\kappa_{ij}^3/n^3)$.

$$\begin{aligned}\sum_{i<j} \frac{(p_{ij} - q_{ij})^2}{p_{ij}} &= \sum_{i<j} \frac{(\kappa_{ij}/n - \kappa_{ij}/n + \kappa_{ij}^2/2n^2 + O(\kappa_{ij}^3/n^3))^2}{\kappa_{ij}/n} \\ &= \sum_{i<j} \frac{\kappa_{ij}^4/4n^4 + O(\kappa_{ij}^3/n^3)}{\kappa_{ij}/n} \to 0,\end{aligned} \tag{B.219}$$

as $n \to \infty$, by (9.5.13). Therefore, the random graphs are asymptotically equivalent by Theorem 6.17.
Now we apply Theorem 6.17 with p_{ij} as in (9.5.7) and $q_{ij} = p_{ij}^{(\mathrm{GRG})}(\kappa)$.

$$\begin{aligned}\sum_{i<j} \frac{(p_{ij} - q_{ij})^2}{p_{ij}} &= \sum_{i<j} \frac{\left(\kappa_{ij}/n - \frac{\kappa_{ij}}{n+\kappa_{ij}}\right)^2}{\kappa_{ij}/n} \\ &= \sum_{i<j} \frac{\left(\frac{\kappa_{ij}^2}{n^2+n\kappa_{ij}}\right)^2}{\kappa_{ij}/n} \\ &= \sum_{i<j} \frac{\kappa_{ij}^3}{n(n+\kappa_{ij})^2} \to 0,\end{aligned} \tag{B.220}$$

as $n \to \infty$, by (9.5.13). Therefore, the random graphs are asymptotically equivalent by Theorem 6.17, as desired. □

Solution to Exercise 9.32. Definition 9.26(i) holds because $\mathcal{S} = \{1, 2\}$ is finite and therefore a separable metric space, and because μ given by $\mu(\{1\}) = \mu(\{2\}) = 1/2$ is a Borel probability measure. Since $\boldsymbol{x}_n$ consists of $n/2$ vertices of each type, $\nu_n(\{1\}) = \nu_n(\{2\}) = (n/2)/n = 1/2 = \mu(\{1\}) = \mu(\{2\})$, and Definition 9.26(ii) holds. Now, κ can be written as an $|\mathcal{S}| \times |\mathcal{S}|$ matrix, which makes it a linear and therefore continuous function. Since continuity implies Borel measurability, Definition 9.26(iii) holds as well.

We have seen that κ is continuous, so κ satisfies Definition 9.27(i)(a). As for Definition 9.27(i)(b) and (c),

$$\iint_{\mathcal{S}^2} \kappa(x, y) d\mu(x) d\mu(y) = \lambda/2 < \infty \tag{B.221}$$

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))] = (\frac{n}{2})^2 \frac{\lambda}{n} = \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x, y) d\mu(x) d\mu(y). \tag{B.222}$$

There are only two subsets $A \subseteq S$ such that $0 < \mu(A) < 1$: $\{1\}$ and $\{2\}$, and $\kappa = \lambda \neq 0$ on $\{1\} \times \{2\}$. Therefore κ is irreducible.

Solution to Exercise 9.34. The proof goes by contraposition. Assume κ is reducible. This means there is some nonempty proper subset $A \subset \mathcal{S}$ such that $\kappa = 0$ a.e. on $A \times (\mathcal{S} \setminus A)$. Since $\mathcal{S}$ is finite, κ is an $r \times r$ matrix such that for each $x \in A$ and $y \in (\mathcal{S} \setminus A)$, the matrix element κ_{xy} is zero. Therefore, the submatrix of κ restricted to A and the submatrix of κ restricted to $(\mathcal{S} \setminus A)$ contain amongst them all nonzero elements of κ, the other elements must be zero. Under the right permutation of rows and columns, κ will be of the form

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{B.223}$$

where A is the $|A| \times |A|$ submatrix of κ restricted to A, D is the $|(\mathcal{S} \setminus A)| \times |(\mathcal{S} \setminus A)|$ submatrix of κ restricted to $(\mathcal{S} \setminus A)$, and B and C are zero matrices. Repeatedly multiplying this matrix with itself will not change the form of the matrix, because the blocks B and C will always remain zero. Therefore, there is no m such that κ^m contains no zeros, as desired. □

Solution to Exercise 9.33. Consider the finite types case where $r = 2$, $s_1 = 1$, $s_2 = 2$, and $\mu(\{i\}) = 1/2$ for $i = 1, 2$. Now take n even and take $n_1 = n/2$ vertices of type 1 and $n_2 = n/2$ vertices of type 2. Then $n_i/n = 1/2$ for both $i = 1, 2$ and for all n. Let

$$\kappa = \begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix}, \tag{B.224}$$

be the 2×2 matrix, such that vertices of the same type are not connected by an edge with high probability, and vertices of differing types are connected by an edge with probability λ/n, independently of the other edges. This is exactly the homogeneous bipartite random graph, as desired. □

Solution to Exercise 9.42. Assume that κ is irreducible, $\sup_{x,y} \kappa(x, y) < \infty$ and $\nu = \|\mathbf{T}_\kappa\| > 1$. Now, by Theorem 9.31 and the irreducibility of κ, we have $\mathbb{P}(H_n < \infty) = \zeta_\kappa^2 + o(1)$. Furthermore, since $\sup_{x,y} \kappa(x, y) < \infty$, Theorem 9.32(i) gives $\mathbb{P}(H_n \leq (1 -$

$\varepsilon)\log_\nu n) = o(1)$. Finally, since κ is irreducible and $\nu = \|\mathbf{T}_\kappa\| > 1$, Theorem 9.32(ii) gives that $\mathbb{P}(H_n \le (1+\varepsilon)\log_\nu n) = \zeta_\kappa^2 + o(1)$. We now rewrite the probability

$$\begin{aligned}\mathbb{P}(1-\varepsilon \le \frac{H_n}{\log_\nu n} \le 1+\varepsilon | H_n \le \infty) &= \mathbb{P}(\frac{H_n}{\log_\nu n} \le 1+\varepsilon | H_n \le \infty) - \mathbb{P}(\frac{H_n}{\log_\nu n} < 1-\varepsilon | H_n \le \infty) \\ &= \frac{\mathbb{P}(\frac{H_n}{\log_\nu n} \le 1+\varepsilon)}{\mathbb{P}(H_n < \infty)} - \frac{\mathbb{P}(\frac{H_n}{\log_\nu n} < 1-\varepsilon)}{\mathbb{P}(H_n < \infty)} \\ &= \frac{\zeta_\kappa^2 + o(1)}{\zeta_\kappa^2 + o(1)} - \frac{o(1)}{\zeta_\kappa^2 + o(1)} \to 1, \end{aligned} \tag{B.225}$$

as $n \to \infty$. Because $\log_a b = \log b / \log a$, we now have that conditionally on $H_n < \infty$, $\frac{H_n}{\log n/\log \nu} \xrightarrow{\mathbb{P}} 1$, or equivalently $H_n/\log n \xrightarrow{\mathbb{P}} 1/\log \nu$, as desired. □

Solution to Exercise 9.43. Under the assumptions, and using Theorem 9.31 and the irreducibility of κ, we once again get that $\mathbb{P}(H_n < \infty) = \zeta_\kappa^2 + o(1)$. Furthermore, we have that there exists a function $f(n) = o(n)$ such that $\mathbb{P}(H_n \le f(n)) = \zeta_\kappa^2 + o(1)$. When rewriting this, and conditioning on $H_n < \infty$, we get

$$\begin{aligned}\mathbb{P}(H_n/\log n \le f(n)/\log n | H_n < \infty) &= \mathbb{P}(H_n \le f(n) | H_n < \infty) \\ &= \frac{\zeta_\kappa^2 + o(1)}{\zeta_\kappa^2 + o(1)} \to 1, \end{aligned} \tag{B.226}$$

as $n \to \infty$. Since $f(n) = o(n)$, we get that $H_n/\log n \xrightarrow{\mathbb{P}} 0$, as desired.

Solutions to the exercises of Chapter 10.

Solution to Exercise ??. We note that $p_k \le p_k(n)$ when $k \le n^a$, while $p_k(n) = 0$ for $k > n^a$. Therefore, we obtain that

$$d_{\mathrm{TV}}(p, p(n)) = \frac{1}{2}\sum_{k=1}^{n^a} p_k\Big(\frac{1}{F(n^a)} - 1\Big) + \frac{1}{2}\sum_{k>n^a} p_k = \frac{1}{2}F(n^a)\Big(\frac{1}{F(n^a)} - 1\Big) + \frac{1}{2}[1 - F(n^a)] = 1 - F(n^a).$$

Solution to Exercise ??. We use Exercise **??** and a coupling argument. We take $\{D_i\}_{i=1}^n$ and $\{D_i^{(n)}\}_{i=1}^n$ to be two i.i.d. sequences of random variables, where D_i has probability mass function $\{p_k\}_{k=1}^\infty$, while $D_i^{(n)}$ has probability mass function $\{p_k(n)\}_{k=1}^\infty$.

By the coupling in Section 2.2, we can couple each of the $D_i, D_i^{(n)}$ with $(\hat{D}_i, \hat{D}_i^{(n)})$ such that

$$\mathbb{P}(\hat{D}_i \neq \hat{D}_i^{(n)}) = d_{\mathrm{TV}}(p, p(n)).$$

By Exercise **??**, and (2.6.8), we obtain that

$$\mathbb{P}(\hat{D}_i \neq \hat{D}_i^{(n)}) \le n^{a(1-\tau)}.$$

Therefore, by Boole's inequality,

$$\begin{aligned}\mathbb{P}(\{\hat{D}_i^{(n)}\}_{i=1}^n \neq \{\hat{D}_i\}_{i=1}^n) &\le \sum_{i=1}^n \mathbb{P}(\hat{D}_i \neq \hat{D}_i^{(n)}) \\ &= n\mathbb{P}(\hat{D}_1 \neq \hat{D}_1^{(n)}) \le n^{1+a(1-\tau)}. \end{aligned}$$

Solutions to the exercises of Chapter 11.

References

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling or, power-law degree distributions in regular graphs. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 694–703, New York, (2005). ACM.

[2] L. A. Adamic. The small world web. In *Lecture Notes in Computer Science*, volume **1696**, pages 443–454. Springer, (1999).

[3] L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science*, **287**:2115, (2000).

[4] W. Aiello, F. Chung, and L. Lu. Random evolution in massive graphs. In *Handbook of massive data sets*, volume **4** of *Massive Comput.*, pages 97–122. Kluwer Acad. Publ., Dordrecht, (2002).

[5] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.*, **74**(1):47–97, (2002).

[6] R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world-wide web. *Nature*, **401**:130–131, (1999).

[7] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, **406**:378–382, (2001).

[8] D. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume **1117** of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, (1985).

[9] D. Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, **1**(2):228–266, (1991).

[10] D. Aldous. Tree-based models for random distribution of mass. *J. Stat. Phys.*, **73**:625–641, (1993).

[11] D. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, **25**(2):812–854, (1997).

[12] N. Alon and J. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, second edition, (2000).

[13] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA*, **97**:11149–11152, (2000).

[14] R. Arratia, AD Barbour, and S. Tavaré. *Logarithmic combinatorial structures: a probabilistic approach.* European Mathematical Society, (2003).

[15] Richard Arratia and Thomas M. Liggett. How likely is an i.i.d. degree sequence to be graphical? *Ann. Appl. Probab.*, **15**(1B):652–670, (2005).

[16] K. Athreya and P. Ney. *Branching processes.* Springer-Verlag, New York, (1972). Die Grundlehren der mathematischen Wissenschaften, Band 196.

[17] T. L. Austin, R. E. Fagen, W. F. Penney, and J. Riordan. The number of components in random linear graphs. *Ann. Math. Statist*, **30**:747–754, (1959).

[18] P. Bak. *How Nature Works: The Science of Self-Organized Criticality*. Copernicus, New York, (1996).

[19] A.-L. Barabási. *Linked: The New Science of Networks*. Perseus Publishing, Cambridge, Massachusetts, (2002).

[20] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, **286**(5439):509–512, (1999).

[21] A.-L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A.*, **311**:69–77, (2000).

[22] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Phys. A*, **311**(3-4):590–614, (2002).

[23] D. Barraez, S. Boucheron, and W. Fernandez de la Vega. On the fluctuations of the giant component. *Combin. Probab. Comput.*, **9**(4):287–304, (2000).

[24] V. Batagelj and A. Mrvar. Some analyses of Erdős collaboration graph. *Social Networks*, **22**(2):173–186, (2000).

[25] E.A. Bender and E.R. Canfield. The asymptotic number of labelled graphs with a given degree sequences. *Journal of Combinatorial Theory (A)*, **24**:296–307, (1978).

[26] G. Bennet. Probability inequaltities for the sum of independent random variables. *J. Amer. Statist. Assoc.*, **57**:33–45, (1962).

[27] N. Berger, B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Degree distribution of the FKP network model. In *Automata, languages and programming*, volume **2719** of *Lecture Notes in Comput. Sci.*, pages 725–738. Springer, Berlin, (2003).

[28] N. Berger, C. Borgs, J. T. Chayes, R. M. D'Souza, and R. D. Kleinberg. Competition-induced preferential attachment. In *Automata, languages and programming*, volume **3142** of *Lecture Notes in Comput. Sci.*, pages 208–221. Springer, Berlin, (2004).

[29] N. Berger, C. Borgs, J. T. Chayes, R. M. D'Souza, and R. D. Kleinberg. Degree distribution of competition-induced preferential attachment graphs. *Combin. Probab. Comput.*, **14**(5-6):697–721, (2005).

[30] J. Bertoin. *Random fragmentation and coagulation processes*, volume **102** of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, (2006).

[31] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. Available from `http://www.unc.edu/ bhamidi/preferent.pdf`, In preparation.

[32] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. First passage percolation on random graphs with finite mean degrees. (2009).

[33] S. Bhamidi, R. van der Hofstad, and J. van Leeuwaarden. Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electronic Journal of Probability*, **15**:1682–1702, (2010).

[34] S. Bhamidi, R. van der Hofstad, and J. van Leeuwaarden. Novel scaling limits for critical inhomogeneous random graphs. Preprint (2009).

[35] G. Bianconi and A.-L. Barabási. Bose-Einstein condensation in complex networks. *Physical Review Letters*, **86**(24):5632–5635, (2001).

[36] G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. *Europhys. Lett.*, **54**:436–442, (2001).

[37] P. Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, New York, (1968).

[38] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, (1995). A Wiley-Interscience Publication.

[39] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume **27** of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[40] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, **1**(4):311–316, (1980).

[41] B. Bollobás. Degree sequences of random graphs. *Discrete Math.*, **33**(1):1–19, (1981).

[42] B. Bollobás. The evolution of random graphs. *Trans. Amer. Math. Soc.*, **286**(1):257–274, (1984).

[43] B. Bollobás. The evolution of sparse graphs. In *Graph theory and combinatorics (Cambridge, 1983)*, pages 35–57. Academic Press, London, (1984).

[44] B. Bollobás. *Random graphs*, volume **73** of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, (2001).

[45] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)*, pages 132–139, New York, (2003). ACM.

[46] B. Bollobás and W. Fernandez de la Vega. The diameter of random regular graphs. *Combinatorica*, **2**(2):125–134, (1982).

[47] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, **31**(1):3–122, (2007).

[48] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, **24**(1):5–34, (2004).

[49] B. Bollobás and O. Riordan. Shortest paths and load scaling in scale-free trees. *Phys. Rev. E.*, **69**:036114, (2004).

[50] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, **18**(3):279–290, (2001).

[51] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. I. The scaling window under the triangle condition. *Random Structures Algorithms*, **27**(2):137–184, (2005).

[52] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. II. The lace expansion and the triangle condition. *Ann. Probab.*, **33**(5):1886–1944, (2005).

[53] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. III. The phase transition for the n-cube. *Combinatorica*, **26**(4):395–410, (2006).

[54] C. Borgs, J. T. Chayes, C. Daskalis, and S. Roch. First to market is not everything: an analysis of preferential attachment with fitness. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 135–144, New York, NY, USA, (2007). ACM Press.

[55] C. Borgs, J. T. Chayes, H. Kesten, and J. Spencer. Uniform boundedness of critical crossing probabilities implies hyperscaling. *Random Structures Algorithms*, **15**(3-4):368–413, (1999).

[56] C. Borgs, J. T. Chayes, H. Kesten, and J. Spencer. The birth of the infinite cluster: finite-size scaling in percolation. *Comm. Math. Phys.*, **224**(1):153–204, (2001). Dedicated to Joel L. Lebowitz.

[57] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Computer Networks and ISDN Systems*, volume **33**, pages 107–117, (1998).

[58] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, **124**(6):1377–1397, (2006).

[59] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, **33**:309–320, (2000).

[60] P. G. Buckley and D. Osthus. Popularity based random graph models leading to a scale-free degree sequence. *Discrete Math.*, **282**(1-3):53–68, (2004).

[61] A. Cayley. A theorem on trees. *Q. J. Pure Appl. Math.*, **23**:376–378, (1889).

[62] D.G. Champernowne. A model of income distribution. *Econ. J.*, bf 63:318, (1953).

[63] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, **23**:493–507, (1952).

[64] S. A. Choudum. A simple proof of the Erdős-Gallai theorem on graph sequences. *Bull. Austral. Math. Soc.*, **33**(1):67–70, (1986).

[65] F. Chung and L. Lu. The diameter of sparse random graphs. *Adv. in Appl. Math.*, **26**(4):257–279, (2001).

[66] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, **99**(25):15879–15882 (electronic), (2002).

[67] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, **6**(2):125–145, (2002).

[68] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Math.*, **1**(1):91–113, (2003).

[69] F. Chung and L. Lu. Coupling online and offline analyses for random power law graphs. *Internet Math.*, **1**(4):409–461, (2004).

[70] F. Chung and L. Lu. *Complex graphs and networks*, volume **107** of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, (2006).

[71] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Math.*, **3**(1):79–127, (2006).

[72] F. Chung and L. Lu. The volume of the giant component of a random graph with given expected degrees. *SIAM J. Discrete Math.*, **20**:395–411, (2006).

[73] A. Clauset and C. Moore. Accuracy and scaling phenomena in internet mapping. *Phys. Rev. Lett.*, **94**:018701: 1–4, (2005).

[74] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Letters*, **85**:4626, (2000).

[75] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Letters*, **86**:3682, (2001).

[76] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures Algorithms*, **22**(3):311–335, (2003).

[77] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth. On the Lambert W function. *Adv. Comput. Math.*, **5**:329–359, (1996).

[78] D. A. Darling. The Galton-Watson process with infinite mean. *J. Appl. Probability*, **7**:455–456, (1970).

[79] P. L. Davies. The simple branching process: a note on convergence when the mean is infinite. *J. Appl. Probab.*, **15**(3):466–480, (1978).

[80] R. De Castro and J.W. Grossman. Famous trails to Paul Erdős. *Rev. Acad. Colombiana Cienc. Exact. Fís. Natur.*, **23**(89):563–582, (1999). Translated and revised from the English.

[81] R. De Castro and J.W. Grossman. Famous trails to Paul Erdős. *Math. Intelligencer*, **21**(3):51–63, (1999). With a sidebar by Paul M. B. Vitanyi.

[82] M. Deijfen, H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. A preferential attachment model with random initial degrees. *Ark. Mat.*, **47**(1):41–72, (2009).

[83] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume **38** of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, (1998).

[84] S. Dereich, C. M
"onch, and P. M
"orters. Typical distances in ultrasmall random networks. *Arxiv preprint arXiv:1102.5680*, Preprint (2011).

[85] S. Dommers, R. van der Hofstad, and G. Hooghiemstra. Diameters in preferential attachment graphs. *Journ. Stat. Phys.*, **139**:72–107, (2010).

[86] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Advances in Physics*, **51**:1079–1187, (2002).

[87] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, (2002). Revised reprint of the 1989 original.

[88] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, (2007).

[89] M. Dwass. A fluctuation theorem for cyclic random variables. *Ann. Math. Statist.*, **33**:1450–1454, (1962).

[90] M. Dwass. A theorem about infinitely divisible distributions. *Z. Wahrscheinleikheitsth.*, **9**:206–224, (1968).

[91] M. Dwass. The total progeny in a branching process and a related random walk. *J. Appl. Prob.*, **6**:682–686, (1969).

[92] H. Ebel, L.-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, **66**:035103, (2002).

[93] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume **33** of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, (1997). For insurance and finance.

[94] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, **53**:292–294, (1947).

[95] P. Erdős and T. Gallai. Graphs with points of prescribed degrees. (Hungarian). *Mat. Lapok*, **11**:264–274, (1960).

[96] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, **6**:290–297, (1959).

[97] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **5**:17–61, (1960).

[98] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, **38**:343–347, (1961).

[99] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, **12**:261–267, (1961).

[100] P. Erdős and R. J. Wilson. On the chromatic index of almost all graphs. *J. Combinatorial Theory Ser. B*, **23**(2–3):255–257, (1977).

[101] G. Ergün and G. J. Rodgers. Growing random networks with fitness. *Physica A*, **303**:261–272, (2002).

[102] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs. *J. Stat. Phys.*, **133**(1):169–202, (2008).

[103] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, **8**:111–140, (2006).

[104] C. Faloutsos, P. Faloutsos, and M. Faloutsos. On power-law relationships of the internet topology. *Computer Communications Rev.*, **29**:251–262, (1999).

[105] W. Feller. *An Introduction to Probability Theory and Its Applications,* Volume I. Wiley, New York, 3rd edition, (1968).

[106] W. Feller. *An Introduction to Probability Theory and Its Applications,* Volume II. Wiley, New York, 2nd edition, (1971).

[107] D. Fernholz and V. Ramachandran. The diameter of sparse random graphs. *Random Structures Algorithms*, **31**(4):482–516, (2007).

[108] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, **30**:1141–1144, (1959).

[109] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Fourth edition prepared by Ju. V. Geronimus and M. Ju. Ceĭtlin. Translated from the Russian by Scripta Technica, Inc. Translation edited by Alan Jeffrey. Academic Press, New York, (1965).

[110] G. Grimmett. *Percolation.* Springer, Berlin, 2nd edition, (1999).

[111] G.R. Grimmett and D.R. Stirzaker. *Probability and random processes.* Oxford University Press, New York, third edition, (2001).

[112] O. Hagberg and C. Wiuf. Convergence properties of the degree distribution of some growing network models. *Bull. Math. Biol.*, **68**:1275–1291, (2006).

[113] P. Hall. Order of magnitude of moments of sums of random variables. *J. London Math Soc.*, **24**(2):562–568, (1981).

[114] P. Halmos. *Measure Theory.* D. Van Nostrand Company, Inc., New York, N. Y., (1950).

[115] T. Harris. *The theory of branching processes.* Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, (1963).

[116] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, **58**:13–30, (1963).

[117] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures Algorithms*, **27**(1):76–123, (2005).

[118] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, **12**(25):703–766 (electronic), (2007).

[119] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Random graphs with arbitrary i.i.d. degrees. Preprint Mittag-Leffer Institute, (2004).

[120] R. van der Hofstad and M. Keane. An elementary proof of the hitting time theorem. *Amer. Math. Monthly*, **115**(8):753–756, (2008).

[121] R. van der Hofstad and J. Spencer. Counting connected graphs asymptotically. *European J. Combin.*, **27**(8):1294–1320, (2006).

[122] F. den Hollander. *Large deviations*, volume **14** of *Fields Institute Monographs.* American Mathematical Society, Providence, RI, (2000).

[123] P. Jagers. *Branching processes with biological applications.* Wiley-Interscience [John Wiley & Sons], London, (1975). Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics.

[124] P. Jagers and O. Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, **16**(2):221–259, (1984).

[125] P. Jagers and O. Nerman. The asymptotic composition of supercritical multi-type branching populations. In *Séminaire de Probabilités, XXX*, volume **1626** of *Lecture Notes in Math.*, pages 40–54. Springer, Berlin, (1996).

[126] S. Janson. Asymptotic degree distribution in random recursive trees. *Random Structures Algorithms*, **26**(1-2):69–83, (2005).

[127] S. Janson. Monotonicity, asymptotic normality and vertex degrees in random graphs. *Bernoulli*, **13**(4):952–965, (2007).

[128] S. Janson. The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.*, **18**(4):1651–1668, (2008).

[129] S. Janson. Standard representation of multivariate functions on a general probability space. *Electron. Commun. Probab.*, **14**:343–346, (2009).

[130] S. Janson. The probability that a random multigraph is simple. *Combinatorics, Probability and Computing*, **18**(1-2):205–225, (2009).

[131] S. Janson. Asymptotic equivalence and contiguity of some random graphs. *Random Structures Algorithms*, **36**(1):26–45, (2010).

[132] S. Janson, D.E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Structures Algorithms*, **4**(3):231–358, (1993). With an introduction by the editors.

[133] S. Janson and M. Luczak. A new approach to the giant component problem. *Random Structures Algorithms*, **34**(2):197–216, (2009).

[134] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, (2000).

[135] S. Janson and J. Spencer. A point process describing the component sizes in the critical window of the random graph evolution. *Combin. Probab. Comput.*, **16**(4):631–658, (2007).

[136] S. Jin and A. Bestavros. Small-world characteristics of Internet topologies and implications on multicast scaling. *Computer Networks*, **50**:648–666, (2006).

[137] J. Jordan. The degree sequences and spectra of scale-free random graphs. *Random Structures Algorithms*, **29**(2):226–242, (2006).

[138] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, (2002).

[139] F. Karinthy. Chains. In *Everything is different*. Publisher unknown, (1929).

[140] Z. Katona and T. Móri. A new class of scale free random graphs. *Statist. Probab. Lett.*, **76**(15):1587–1593, (2006).

[141] J. H. B. Kemperman. *The passage problem for a stationary Markov chain*. Statistical Research Monographs, Vol. I. The University of Chicago Press, Chicago, Ill., (1961).

[142] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Statist.*, **37**:1463–1481, (1966).

[143] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Statist.*, **37**:1211–1223, (1966).

[144] H. Kesten and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.*, **17**:309–338, (1967).

[145] J. F. C. Kingman. The first birth problem for an age-dependent branching process. *Ann. Probability*, **3**(5):790–801, (1975).

[146] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, **46**(5):604–632, (1999).

[147] J. M. Kleinberg. Navigation in a small world. *Nature*, **406**:845, (2000).

[148] J. M. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proc. of the twenty-third annual ACM symposium on Principles of distributed computing*, pages 163–170, May (2000).

[149] J.M. Kleinberg, R. Kumar, P. Raghavan, S Rajagopalan, and A. Tomkins. The web as a graph: measurements, models, and methods. In *Computing and Combinatorics: 5th Annual International Conference, COCOON'99, Tokyo, Japan, July 1999. Proceedings*, Lecture Notes in Computer Science, pages 1–17, (1999).

[150] T. Konstantopoulos. Ballot theorems revisited. *Statist. Probab. Lett.*, **24**(4):331–338, (1995).

[151] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, **63**:066123, (2001).

[152] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, **85**:4629, (2000).

[153] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *42st Annual IEEE Symposium on Foundations of Computer Science*, pages 57–65, (2000).

[154] R. Kumar, P. Raghavan, S Rajagopalan, and A. Tomkins. Trawling the web for emerging cyber communities. *Computer Networks*, **31**:1481–1493, (1999).

[155] A. Lakhina, J.W. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *Proceedings of IEEE INFOCOM 1*, pages 332–341, (2003).

[156] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York, (1983).

[157] R. LePage, M Woodroofe, and J. Zinn. Convergence to a stable distribution via order statistics. *Ann. Probab.*, **9**:624–632, (1981).

[158] F. Liljeros, C. R. Edling, L. A. N. Amaral, and H. E. Stanley. The web of human sexual contacts. *Nature*, **411**:907, (2001).

[159] J. H. van Lint and R. M. Wilson. *A course in combinatorics.* Cambridge University Press, Cambridge, second edition, (2001).

[160] A.J. Lotka. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, **16**(12):317–323, (1926).

[161] L. Lu. *Probabilistic methods in massive graphs and Internet Computing*. PhD thesis, University of California, San Diego, (2002). Available at `http://math.ucsd.edu/~llu/thesis.pdf`.

[162] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures Algorithms*, **1**(3):287–310, (1990).

[163] T. Łuczak. On the number of sparse connected graphs. *Random Structures Algorithms*, **1**(2):171–173, (1990).

[164] T. Łuczak, B. Pittel, and J. Wierman. The structure of a random graph at the point of the phase transition. *Trans. Amer. Math. Soc.*, **341**(2):721–748, (1994).

[165] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, **23**(3):1125–1138, (1995).

[166] A. Martin-Löf. The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *J. Appl. Probab.*, **35**(3):671–682, (1998).

[167] S. Milgram. The small world problem. *Psychology Today*, May:60–67, (1967).

[168] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Math.*, **1**(2):226–251, (2004).

[169] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, **6**(2-3):161–179, (1995).

[170] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, **7**(3):295–305, (1998).

[171] J. W. Moon. The distance between nodes in recursive trees. In *Combinatorics (Proc. British Combinatorial Conf., Univ. Coll. Wales, Aberystwyth, 1973)*, pages 125–132. London Math. Soc. Lecture Note Ser., No. 13. Cambridge Univ. Press, London, (1974).

[172] T. F. Móri. On random trees. *Studia Sci. Math. Hungar.*, **39**(1-2):143–155, (2002).

[173] T. F. Móri. The maximum degree of the Barabási-Albert random tree. *Combin. Probab. Comput.*, **14**(3):339–348, (2005).

[174] A. Nachmias and Y. Peres. Component sizes of the random graph outside the scaling window. *ALEA Lat. Am. J. Probab. Math. Stat.*, **3**:133–142, (2007).

[175] O. Nerman and P. Jagers. The stable double infinite pedigree process of supercritical branching populations. *Z. Wahrsch. Verw. Gebiete*, **65**(3):445–460, (1984).

[176] M. E. J. Newman. Models of the small world. *J. Stat. Phys.*, **101**:819–841, (2000).

[177] M. E. J. Newman. The structure of scientific collaboration networks. *Proc.Natl.Acad.Sci.USA*, **98**:404, (2001).

[178] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, **45**(2):167–256 (electronic), (2003).

[179] M. E. J. Newman, S. Strogatz, and D. Watts. Random graphs with arbitrary degree distribution and their application. *Phys. Rev. E*, **64**:026118, 1–17, (2000).

[180] M. E. J. Newman, S. Strogatz, and D. Watts. Random graph models of social networks. *Proc. Nat. Acad. Sci.*, **99**:2566–2572, (2002).

[181] M. E. J. Newman, D. J. Watts, and A.-L. Barabási. *The Structure and Dynamics of Networks*. Princeton Studies in Complexity. Princeton University Press, (2006).

[182] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Adv. in Appl. Probab.*, **38**(1):59–75, (2006).

[183] M. Okamoto. Some inequalities relating to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.*, **10**:29–35, (1958).

[184] R. Oliveira and J. Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.*, **2**(2):121–163, (2005).

[185] E. Olivieri and M.E. Vares. *Large deviations and metastability*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, (2005).

[186] R. Otter. The multiplicative process. *Ann. Math. Statist.*, **20**:206–224, (1949).

[187] V. Pareto. *Cours d'Economie Politique*. Droz, Geneva, Switserland, (1896).

[188] R. Pemantle. A survey of random processes with reinforcement. *Probab. Surv.*, **4**:1–79 (electronic), (2007).

[189] J. Pickands III. Moment convergence of sample extremes. *Ann. Math. Statistics*, **39**:881–889, (1968).

[190] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, **25**(2):855–900, (1997).

[191] B. Pittel. On tree census and the giant component in sparse random graphs. *Random Structures Algorithms*, **1**(3):311–342, (1990).

[192] B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, **5**(2):337–347, (1994).

[193] B. Pittel. On the largest component of the random graph at a nearcritical stage. *J. Combin. Theory Ser. B*, **82**(2):237–269, (2001).

[194] I. de S. Pool and M. Kochen. Contacts and influence. *Social Networks*, **1**:5–51, (1978).

[195] H. Reittu and I. Norros. On the power law random graph model of massive data networks. *Performance Evaluation*, **55**(1-2):3–23, (2004).

[196] S. M. Ross. *Stochastic processes*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, (1996).

[197] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures Algorithms*, **31**(2):186–202, (2007).

[198] H.-J. Schuh and A. D. Barbour. On the asymptotic behaviour of branching processes with infinite mean. *Advances in Appl. Probability*, **9**(4):681–723, (1977).

[199] E. Seneta. Functional equations and the Galton-Watson process. *Advances in Appl. Probability*, **1**:1–42, (1969).

[200] E. Seneta. The simple branching process with infinite mean. I. *J. Appl. Probability*, **10**:206–212, (1973).

[201] G. Sierksma and H. Hoogeveen. Seven criteria for integer sequences being graphic. *J. Graph Theory*, **15**(2):223–231, (1991).

[202] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos. Power laws and the AS-level internet topology. *IEEE/ACM Trans. Netw.*, **11**(4):514–524, (2003).

[203] H. A. Simon. On a class of skew distribution functions. *Biometrika*, **42**:425–440, (1955).

[204] R. Smythe and H. Mahmoud. A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.*, (51):1–29, (1994).

[205] B. Söderberg. General formalism for inhomogeneous random graphs. *Phys. Rev. E (3)*, **66**(6):066121, 6, (2002).

[206] B. Söderberg. Properties of random graphs with hidden color. *Phys. Rev. E (3)*, **68**(2):026107, 12, (2003).

[207] B. Söderberg. Random graph models with hidden color. *Acta Physica Polonica B*, **34**:5085–5102, (2003).

[208] B. Söderberg. Random graphs with hidden color. *Phys. Rev. E (3)*, **68**(1):015102, 4, (2003).

[209] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bull. Math. Biophys.*, **13**:107–117, (1951).

[210] J. Spencer. Enumerating graphs and Brownian motion. *Comm. Pure Appl. Math.*, **50**(3):291–294, (1997).

[211] F. Spitzer. *Principles of Random Walk*. Springer, New York, 2nd edition, (1976).

[212] S. Strogatz. Exploring complex networks. *Nature*, **410**(8):268–276, (2001).

[213] J. Szymański. Concentration of vertex degrees in a scale-free random graph process. *Random Structures Algorithms*, **26**(1-2):224–236, (2005).

[214] H. Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, (2000).

[215] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, **32**:425–443, (1969).

[216] T.S. Turova. The largest component in subcritical inhomogeneous random graphs. To appear in *Combinatorics, Probability and Computing*, Preprint (2007).

[217] T.S. Turova. Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1. Preprint (2009).

[218] T.S. Turova and T. Vallier. Merging percolation on $\mathbf{Z}^d$ and classical random graphs: Phase transition. *Random Structures Algorithms*, **36**(2):185–217, (2010).

[219] D. J. Watts. *Small worlds. The dynamics of networks between order and randomness*. Princeton Studies in Complexity. Princeton University Press, Princeton, NJ, (1999).

[220] D. J. Watts. *Six degrees. The science of a connected age.* W. W. Norton & Co. Inc., New York, (2003).

[221] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, **393**:440–442, (1998).

[222] J. G. Wendel. Left-continuous random walk and the Lagrange expansion. *Amer. Math. Monthly*, **82**:494–499, (1975).

[223] D. Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, (1991).

[224] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker. Scaling phenomena in the internet: Critically examining criticality. *Proc. Natl. Acad. Sci.*, **99**:2573–2580, (2002).

[225] E. M. Wright. The number of connected sparsely edged graphs. *J. Graph Theory*, **1**(4):317–330, (1977).

[226] E. M. Wright. The number of connected sparsely edged graphs. III. Asymptotic results. *J. Graph Theory*, **4**(4):393–407, (1980).

[227] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the internet's large-scale topology. *PNAS*, **99**(22):13382–13386, (2002).

[228] G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Phil. Trans. Roy. Soc. London, B*, **213**:21–87, (1925).

[229] G.K. Zipf. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, **15**:1–95, (1929).

Index