# 1

# Introduction and Preliminaries

**Abstract**

In this chapter, we draw motivation from real-world networks and formulate random graph models for them. We focus on some of the models that have received the most attention in the literature, namely, Erdős–Rényi random graphs, inhomogeneous random graphs, configuration models, and preferential attachment models. We follow van der Hofstad (2017), which we refer to as [V1], both for motivation and for the introduction to the random graph models involved.

## Looking Back, and Ahead

In Volume 1 of this pair of books, we 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. Because of their *dynamic* nature, preferential attachment models give us a convincing explanation of the abundance of power-law degree sequences in various applications. We will often refer to Volume 1. When we do so, we write [V1, Theorem 2.17] to signify that we refer to Theorem 2.17 in van der Hofstad (2017).

In [V1, Chapters 6–8], we focussed on the properties of the *degrees* of such graphs. However, we noted in [V1, Chapter 1] that not only do many real-world networks have degree sequences that are rather different from the ones of the Erdős–Rényi random graph, also many examples have a *giant connected component* and are *small worlds*.

In Chapters 3–8, we will return to the models discussed in [V1, Chapters 6–8], and focus on their local structure, and their connected components, as well as on their distance structure. 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 is the *power-law degree exponent*, and the physics literature predicts the behavior in random graph models with similar degree sequences to be similar. This is an example of the notion of *universality*, a central notion in statistical physics. Despite its importance, there are only a few examples of universality that can be rigorously proved. In Chapters 3–8, we investigate the level of universality present in random graph models.

## Organization of this Chapter

This chapter is organized as follows. In Section 1.1 we discuss real-world networks and the inspiration that they provide. In Section 1.2, we then discuss how *graph sequences*, where the size of the involved graphs tends to infinity, aim at describing *large* complex networks. In Section 1.3 we recall the definition of several random graph models, as introduced in Volume 1. In Section 1.4, we discuss *power-law* random variables, as they play an important role in this book. In Section 1.5 we recall some of the standard notation and notions used in this book. We close this chapter with notes and discussion in Section 1.6 and with



https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

exercises in Section 1.7. We give few references to the literature within this chapter, but defer a discussion of the history of the various models to the extensive notes in Section 1.6.

## 1.1 Motivation: Real-World Networks

In the past two decades, an enormous research effort has been performed with regard to modeling various real-world phenomena using networks. Networks arise in various applications ranging from the connections between friends in friendship networks to the connectivity of neurons in the brain, to the relations between companies and countries in economics, and the hyperlinks between webpages in the World-Wide Web. The advent of the computer era has made many network data sets available. Around 1999–2000, various groups started to investigate network data from an empirical perspective. [V1, Chapter 1] gives many examples of real-world networks and the empirical findings from them. Here we give some basics.

### 1.1.1 Graphs and Networks

A graph $G = (V, E)$ consists of a collection $V = V(G)$ of vertices, also called a vertex set, and a collection of edges $E = E(G)$, often called an edge set. The vertices correspond to the objects that we model; the edges indicate some relation between pairs of these objects. In our settings, graphs are usually *undirected*. Thus, an edge is an unordered pair $\{u, v\} \in E$ indicating that $u$ and $v$ with $u, v \in V(G)$ are directly connected. When $G$ is undirected, if $u$ is directly connected to $v$ then also $v$ is directly connected to $u$. Therefore, an edge can be seen as a pair of vertices. When dealing with social networks, the vertices represent the individuals in the population while the edges represent the friendships among them. We sometimes work with *multi-graphs*, which are graphs possibly having *self-loops* or *multiple edges* between vertices, and we will clearly indicate when we do so.

We mainly deal with *finite* graphs and then, for simplicity, we often take $V = [n] := \{1, \ldots, n\}$. The *degree* $d_u^{(G)}$ of a vertex $u \in V(G)$ in the graph $G$ is equal to the number of edges containing $u$, i.e.,

$$d_u^{(G)} = \#\{v \in V(G) \colon \{u, v\} \in E(G)\}. \tag{1.1.1}$$

Often, we deal with the degree of a *random vertex* in $G$. Let $o \in V(G)$ be a vertex chosen uniformly at random (uar) in $V(G)$. The *typical degree* is the random variable $D_n$ given by

$$D_n = d_o^{(G)}. \tag{1.1.2}$$

It is not hard to see that the probability mass function of $D_n$ is given by

$$\mathbb{P}(D_n = k) = \frac{1}{|V(G)|} \sum_{v \in V(G)} \mathbb{1}_{\{d_v^{(G)} = k\}}, \tag{1.1.3}$$

where, for a set $A$, we write $|A|$ for its size. Exercise 1.1 asks you to prove (1.1.3).

The *average degree* in a network is equal to

$$\frac{1}{|V(G)|} \sum_{v \in V(G)} d_v^{(G)} = \frac{2|E(G)|}{|V(G)|}. \tag{1.1.4}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

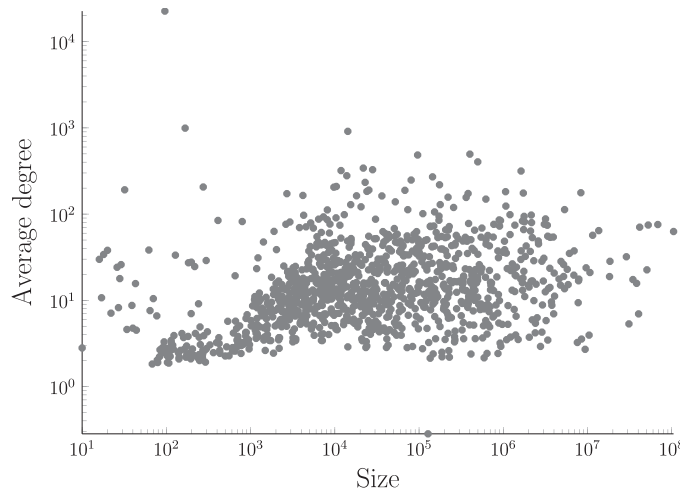


**Figure 1.1** Average degrees in the 727 networks of size larger than 10,000 from the KONECT data base.

We can rewrite (1.1.4) as

$$\frac{1}{|V(G)|}\sum_{v\in V(G)} d_v^{(G)} = \mathbb{E}[D_n], \tag{1.1.5}$$

where the expectation is with respect to the random vertex $o$ in $D_n = d_o^{(G)}$ (recall (1.1.2)). The average degree can take any value in between 0 for an empty graph, and $|V(G)| - 1$ for a complete graph. In reality, however, we see that the average degree of many real-world networks is not very large, i.e., these networks tend to be *sparse*. Figure 1.1 shows the average degrees in the KONECT data base, and we see that the average degree does not seem to grow with the network size.

We next discuss some common features that many real-world networks turn out to have.

### 1.1.2 Scale-Free Phenomenon

The first, maybe quite surprising, fundamental property of many real-world networks is that the number of vertices with degree at least $k$ decays slowly for large $k$. This implies that degrees are highly variable and that, even though the average degree is not particularly large, there exist vertices with extremely high degree. Often, the tail of the empirical degree distribution seems to fall off as an inverse power of $k$. This is called a "power-law degree sequence," and the resulting graphs often go under the name "scale-free graphs." This is visualized for the Autonomous Systems (AS) graph from the Internet in Figure 1.5(a), where the degree distribution of the AS graph is plotted on a 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 exponent $\tau$,

$$n_k \approx c_n k^{-\tau}, \tag{1.1.6}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

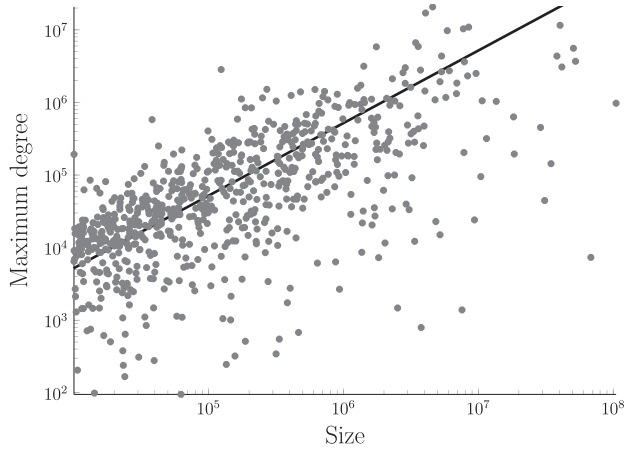


**Figure 1.2** Maximal degrees in the 727 networks of size larger than 10,000 from the KONECT data base. Linear regression gives $\log d_{\max} = 0.742 + 0.519 \log n$.

and thus

$$\log n_k \approx \log c_n - \tau \log k, \tag{1.1.7}$$

so that the plot of $\log k \mapsto \log n_k$ is close to a *straight line*. This is the reason why degree sequences in networks are often depicted in a log–log fashion, rather than in the more customary form of $k \mapsto n_k$. Here, and in the remainder of this section, we write $\approx$ to denote an uncontrolled approximation. The power-law exponent $\tau$ can be estimated by the absolute value of the slope of the line in the log–log plot. Naturally, we must have that

$$\sum_k n_k = |V(G_n)| < \infty, \tag{1.1.8}$$

so that it is reasonable to assume that $\tau > 1$. In fact, many networks are *sparse*, meaning that their average $\sum_k k n_k / |V(G_n)|$ remains uniformly bounded, which in turn suggests that $\tau > 2$ is to be expected. See Figure 1.2 for the maximal degrees in the KONECT data base in log–log scale, which should be compared with Figure 1.1. While there does not seem to be a trend in Figure 1.1, there does seem to be one in Figure 1.2; this indicates that the log of the maximal degree tends to grow linearly with the log of the network size. The latter is consistent with power-law degrees.

Let us define the *degree distribution* by $p_k^{(G_n)} = n_k/|V(G_n)| = \mathbb{P}(D_n = k)$ (recall (1.1.2) and (1.1.3)), so that $p_k^{(G_n)}$ equals the probability that a *uniformly chosen vertex* in a graph $G_n$ with $n$ vertices has degree $k$ (recall (1.1.3)). Then (1.1.6) can be reexpressed as

$$p_k^{(G_n)} \approx c k^{-\tau}, \tag{1.1.9}$$

where again $\approx$ denotes an uncontrolled approximation.

Vertices with extremely high degrees go under various names, indicating their importance in the field. They are often called *hubs*, like the hubs in airport networks. Another name for them is *super-spreader*, indicating the importance of the high-degree vertices in spreading information or diseases. The hubs quantify the level of inhomogeneity in the real-world

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

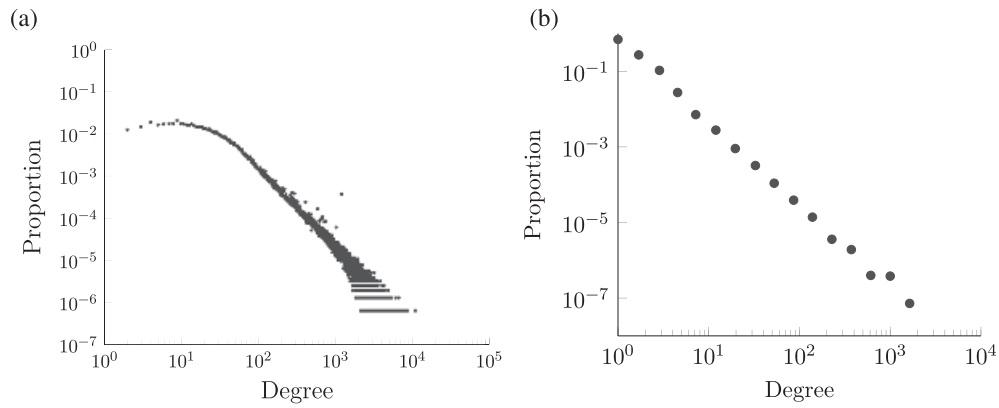


**Figure 1.3** (a) Log–log plot of the degree sequence in the 2007 Internet Movie Data base. (b) Log–log plot of the probability mass function of the Autonomous Systems degree sequence on April 2014, on a log–log scale from Krioukov et al. (2012) (data courtesy of Dmitri Krioukov). This degree distribution looks smoother than others (see e.g., Figure 1.3(a) and 1.4), due to binning of the data.

networks, and a large part of this book is centered around rigorously establishing the effect that the high-degree vertices have on various properties of the graphs involved.

Further, a central topic in network science is how the behavior of stochastic processes on networks is affected by degree inhomogeneities. Such effects are especially significant when the networks are "scale-free," meaning that they can be well approximated by power laws with exponents $\tau$ satisfying $\tau \in (2, 3)$, so that random variables with such degrees have *infinite variance*. Since maximal degrees of networks of size $n$ can be expected to grow as $n^{1/(\tau-1)}$ (see Exercise 1.2 for an illuminating example), Figure 1.2 suggests that, on average, $1/(\tau - 1) \approx 0.519$, so that, again on average, $\tau \approx 2.93$, which is in line with such predictions.

For the Internet, log–log plots of degree sequences first appeared in a paper by the Faloutsos brothers (1999) (see Figure 1.3(b) for the degree sequence in the Autonomous Systems graph, where the degree distribution looks relatively smooth because it is binned). Here, the power-law exponent is estimated as $\tau \approx 2.15$–$2.20$. Figure 1.3(a) displays the degree distribution in the Internet Movie Data base (IMDb), in which the vertices are actors and two actors are connected when they have acted together in a movie. Figure 1.4 displays the degree-sequence for both the in- as well as the out-degrees in various World-Wide Web data bases.

### Recent Discussion on Power-Law Degrees in Real-World Networks

Recently, a vigorous discussion has emerged on how often real-world networks have power-law degree distributions. This discussion was spurred by Broido and Clauset (2019), who claimed (even as the title of their paper) that

*Scale-free networks are rare.*

What did they do to reach this conclusion? Broido and Clauset (2019) performed the first extensive analysis of a large number of real-world network data sets, and compared degree

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Table 1.1 *For comparison, fits of scale-free and alternative distributions to real-world networks taken from (Broido and Clauset, 2019, Table 1). Listed are the percentage of network data sets that favor the power-law model* $\mathrm{M}_{\mathrm{PL}}$*, the alternative model* $\mathrm{M}_{\mathrm{Alt}}$*, or neither, under a likelihood-ratio test, along with the form of the alternative distribution indicated by the alternative density* $x \mapsto f(x)$*.*

| Alternative | $f(x) \propto$ | $\mathrm{M}_{\mathrm{PL}}$ | Inconclusive | $\mathrm{M}_{\mathrm{Alt}}$ |
|---|---|---|---|---|
| Exponential | $\mathrm{e}^{-\lambda x}$ | 33% | 26% | 41% |
| Log-normal | $\frac{1}{x}\mathrm{e}^{-(\log x-\mu)^2/(2\sigma^2)}$ | 12% | 40% | 48% |
| Weibull | $\mathrm{e}^{-(x/b)^a}$ | 33% | 20% | 47% |
| Power law with cutoff | $x^{-\tau}\mathrm{e}^{-Ax}$ | – | 44% | 56% |

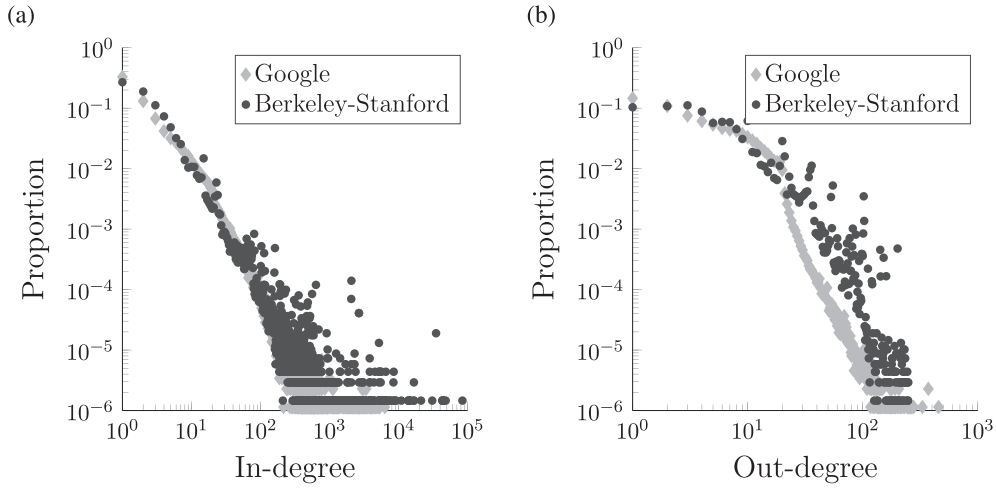


**Figure 1.4** The probability mass function of the in- and out-degree sequences in the Berkeley-Stanford and Google competition graph data sets of the World Wide Web in Leskovec et al. (2009). (a) In-degree; (b) out-degree.

sequences of these real-world networks with power-law, as well as with log-normal, exponential and Weibull distributions. They also made comparisons with power-law distributions having exponential truncation. The main conclusion of Broido and Clauset (2019) was that, in many cases, alternative distributions are preferred over power laws (see also Table 1.1).

Clearly this work caused quite a stir, as the conclusion, if correct, would make about 20 years of network science close to redundant from a practical perspective. Barabási (2018) wrote a blog post containing detailed criticism of the methods and results in Broido and Clauset (2019), see also Voitalov et al. (2019). Holme (2019) summarized the status of the arguments in 2019, reaching an almost philosophical conclusion:

*Still, it often feels like the topic of scale-free networks transcends science – debating them probably has some dimension of collective soul searching as our field slowly gravitates toward data science, away from complexity science.*

So, what did the discussion focus on? Here is a list of questions:

**What are power-law data?** An important question in the discussion on power-law degree distributions is how to interpret the approximation sign in (1.1.9). Most approaches start

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

by assuming that the data are realizations of *independent and identically distributed* (iid) random variables. This can only be an assumption, as degree distributions are mostly *graphical* (meaning that they can arise as degree sequences of graphs without self-loops and multiple edges), which introduces dependencies between them (if only because the sum of the degrees needs to be even). However, without this assumption, virtually any analysis becomes impossible, so let us assume this as well.

Under the above assumption, one needs to infer the degree distribution from the sample of degrees obtained from a real-world network. We denote the asymptotic degree distribution by $p_k$, i.e., the proportion of vertices of degree $k$ in the *infinite-graph limit*. Under this assumption, $p_k^{(G_n)}$ in (1.1.9) is the *empirical probability mass function* corresponding to the true underlying degree distribution $(p_k)_{k\geq 0}$. The question is thus what probability mass functions $(p_k)_{k\geq 0}$ correspond to a power law.

Broido and Clauset (2019) interpreted the power-law assumption as

$$p_k = ck^{-\tau} \qquad \text{for all} \quad k \geq k_{\min}, \tag{1.1.10}$$

and $p_k$ arbitrary for $k \in [k_{\min} - 1]$; here $c > 0$ is chosen appropriately. The inclusion of the $k_{\min}$ parameter is based on the observation that small values of $k$ generally do not satisfy the pure power law (see also Clauset et al. (2009), where (1.1.10) first appeared).

Barabási (2018) instead argued from the perspective of *generative models* (such as the preferential attachment models described in Section 1.3.5, as well as in Chapters 5 and 8):

*In other words, by 2001 it was pretty clear that there is no one-size-fits-all formula for the degree distribution for networks driven by the scale-free mechanism. A pure power law only emerges in simple idealised models, driven by only growth and preferential attachment, and free of any additional effects.*

Bear in mind that this dynamical approach is very different from that of Broido and Clauset (2019), as the degrees in generative models can hardly be expected to be realizations of an iid sample! Barabási (2018) instead advocated a theory that predicts power laws with exponential truncation for many settings, meaning that

$$p_k = ck^{-\tau}\mathrm{e}^{-Ak} \qquad \text{for all} \quad k \geq d_{\min}, \tag{1.1.11}$$

where $d_{\min}$ denotes the minimal degree in the graph and $c, A > 0$ are appropriate constants, but the theory also allows for "additional effects," such as vertex fitnesses that describe intrinsic differences in how likely it is to connect to vertices, and that may be realistic in some real-world networks.

Voitalov et al. (2019) took a static approach related to that of Broido and Clauset (2019), but instead assumed more general power laws of the form

$$1 - F(x) = \sum_{k>x} p_k = x^{-(\tau-1)}L(x) \qquad \text{for all} \quad x \geq 1, \tag{1.1.12}$$

where $x \mapsto L(x)$ is a so-called *slowly varying function*, meaning a function that does not change the power-law exponent, in that it grows or decays more slowly than any power at infinity. See [V1, Definition 1.5], or Definition 1.19 below, for a precise definition. In particular, distributions that satisfy (1.1.10) also satisfy (1.1.12), but not necessarily the other way around.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

The advantage of working with (1.1.12) is that this definition is quite general, yet a large body of work within the *extreme-value statistics* community becomes available. These results, as summarized in Voitalov et al. (2019), allow for the "most accurate" ways of estimating the power-law exponent $\tau$, which brings us to the next question.

**How to estimate the power-law exponent?** Since Broido and Clauset (2019) interpreted the power-law assumption as in (1.1.10), estimating the model parameters then boiled down to estimating $k_{\min}$ and $\tau$. For this, Broido and Clauset (2019) relied on the first paper on estimating power-law exponents in the area of networks, by Clauset et al. (2009), who proposed the *power-law-fit method* (PLFIT). This method chooses the best possible $k_{\min}$ on the basis of the difference between the empirical degree distribution for values above $k_{\min}$ and the power-law distribution function based on (1.1.10) with an appropriately estimated value $\hat{\tau}$ of $\tau$, as proposed by Hill (1975), for realizations above $k_{\min}$.

The estimator $\hat{\tau}_{\text{PLFit}}$ is then the estimator of $\tau$ corresponding to the optimal $k_{\min}$. The PLFIT method was recently proved to be a *consistent* method by Bhattacharya et al. (2020), which means that the estimator will, in the limit, converge in probability to the correct value $\tau$, even under the weaker assumption in (1.1.12). Of course, the question remains whether $\hat{\tau}_{\text{PLFit}}$ is a good estimator, for example in the sense that the rate of convergence of $\hat{\tau}_{\text{PLFit}}$ to $\tau$ is optimal. The results and simulations in Drees et al. (2020) suggest that, even in the case of a pure power law as in (1.1.10) with $k_{\min} = 1$, $\hat{\tau}_{\text{PLFit}}$ is outperformed by more classical estimators (such as the maximum likelihood estimator for $\tau$). Voitalov et al. (2019) rely on the estimators proposed in the extreme-value literature; see e.g. Danielsson et al. (2001); Draisma et al. (1999); Hall and Welsh (1984) for such methods and Resnick (2007); Beirlant et al. (2006) for extensive overviews of extreme-value statistics.

The dynamical approach by Barabási (2018) instead focusses on estimating the parameters in the proposed dynamical models, a highly interesting topic that is beyond the scope of this book.

**How to perform tests?** When confronted with a model, or with two competing models such as in Table 1.1, a statistician would often like to compare the *fit* of these models to the data, so as to be able to choose between them. When both models are *parametric*, meaning that they involve a finite number of parameters, like the models in Table 1.1, this can be done using a so-called *likelihood-ratio test*. For this, one computes the likelihood of the data (basically the probability that the model in question gives rise to exactly what was found in the data) for each of the models, and then takes the ratio of the two likelihoods. In the settings in Table 1.1, this means that the likelihood of the data for the power-law model is divided by that for the alternative model. When this exceeds a certain threshold, the test does not reject the possibility that the data comes from a power law, otherwise it rejects the null hypothesis of a power-law degree distribution. This is done for each of the networks in the data base, and Table 1.1 indicates the percentages for which each of the models is deemed the most likely.

Unfortunately, such likelihood ratio tests can be performed only when one compares *parametric* settings. The setting in (1.1.12) is *non-parametric*, as it involves the unknown slowly varying function $x \mapsto L(x)$, and thus, in that setting, no statistical test can be

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

performed unless one makes parametric assumptions on the shape of $x \mapsto L(x)$ (by assuming, for example, that $L(x)$ is a power of $\log x$). Thus, the parametric choice in (1.1.10) is crucial in that it allows for a testing procedure to be performed. Alternatively, if one does not believe in the "pure" power-law form as in (1.1.10), then tests are no longer feasible. What approach should one then follow? See Artico et al. (2020) for a related testing procedure, in which the authors reached a rather different conclusion than that of Broido and Clauset (2019).

**How to partition networks?** Broido and Clauset (2019) investigated a large body of networks, relying on a data base consisting of 927 real-world networks from the KONECT project; see `http://konect.cc` as well as Kunegis (2013). We are also relying on this data base for graphs showing network properties, such as average and maximal degrees, etc. These networks vary in size, as well as in their properties (directed versus undirected, static versus temporal, etc.). In their paper, Broido and Clauset (2019) report percentages of networks having certain properties; see for example Table 1.1.

A substantial part of the discussion around Broido and Clauset (2019) focusses on whether these percentages are representative. Take the example of a directed network, which has several degree distributions, namely, in-degree, out-degree, and total degree distributions (in the latter, the directions are simply ignored). This "diversity of degree distributions" becomes even more pronounced when the network is *temporal*, meaning that edges come and go as time progresses. When does one say that a temporal network has a power-law degree distribution? When one of these degree distributions is classified as power-law, when a certain percentage of them is, or when all of them are?

**What is our approach in this book?** We prefer to avoid the precise debate about whether power laws in degree distributions are omnipresent or rare. We view power laws as a way to *model* settings where there is a large amount of variability in the data, and where the maximum values of the degrees are several orders of magnitude larger than the average values (compare Figures 1.1 and 1.2). Power laws predict such differences in scale.

There is little debate about the fact that degree distributions in networks tend to be highly inhomogeneous. Power laws are the model of choice to model such inhomogeneities, certainly in settings where empirical moments (for example, empirical variances) are very large. Further, inhomogeneities lead to interesting *differences in structure* of the networks in question, which will be a focal point of this book. All the alternative models in Table 1.1 have tails that are too thin for such differences to emerge. Thus, it is natural to focus on models with power-law degrees to highlight the relation between degree structure and network topology. Therefore, we often consider degree distributions that are either *exactly* described by power laws or are bounded above or below by them. The focus then resides in how the degree power-law exponent $\tau$ changes the network topology.

After this extensive discussion of degrees in graphs, we continue by discussing *graph distances* and their relation to *small-world phenomena*, a topic that is much less heatedly debated.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

### 1.1.3 Small-World Phenomenon

A second fundamental network property observed in many real-world networks is the fact that typical distances between vertices are small. This is called the "small-world" phenomenon (see, e.g., the book by Watts (1999)). In particular, such networks are highly connected: their largest connected component contains a significant proportion of the vertices. Many networks, such as the Internet, even consist of *one* connected component, since otherwise e-mail messages could not be delivered between pairs of vertices in distinct connected components.

Graph distances between pairs of vertices tend to be quite small in most networks. For example, in the Internet, IP packets cannot use more than a threshold of physical links, and if distances in the Internet were larger than this threshold then the e-mail service would simply break down. Thus, the Internet graph has evolved in such a way that typical distances are relatively small, even though the Internet itself is rather large. As seen in Figure 1.5(a), the number of Autonomous Systems (ASs) traversed by an e-mail data set, sometimes referred to as the AS-count, is typically at most 7. In Figure 1.5(b), the proportion of routers traversed by an e-mail message between two uniformly chosen routers, referred to as the *hopcount*, is shown. It shows that the number of routers traversed is at most 27. Figure 1.6 shows typical distances in the IMDb; the distances are quite small despite the fact that the network contains more than one million vertices.

The small-world nature of real-world networks is highly significant. Indeed, in small worlds, news can spread quickly as relatively few people are needed to spread it between two typical individuals. This is quite helpful in the Internet, where e-mail messages hop along the edges of the network. At the other side of the spectrum, it also implies that infectious diseases can spread quite quickly, as just a few infections can carry the disease to a large part of the population. This implies that diseases have a large potential of becoming pandemic, as the corona pandemic has made painfully clear.

Let us continue this discussion by formally introducing graph distances, as displayed in Figures 1.5 and 1.6. For a graph $G = (V(G), E(G))$ and a pair of vertices $u, v \in V(G)$,

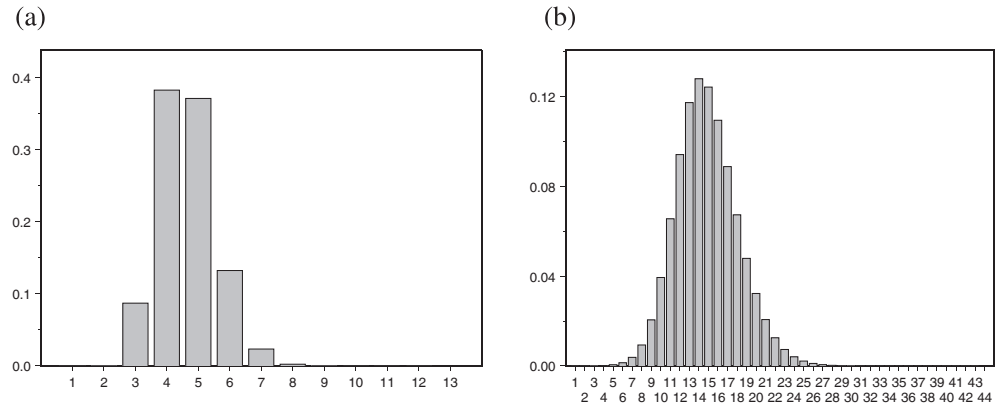


**Figure 1.5** (a) Number of Autonomous Systems traversed in hopcount data. (b) Internet hopcount data (courtesy of Hongsuda Tangmunarunkit).

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

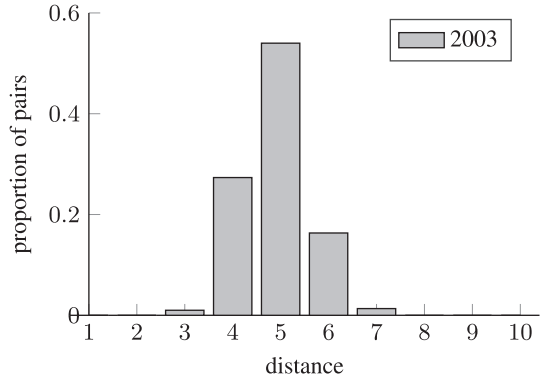


**Figure 1.6** Typical distances in the Internet Movie Data base (IMDb) in 2003.

we let the graph distance $\mathrm{dist}_G(u, v)$ between $u$ and $v$ be equal to the minimal number of edges in a path linking $u$ and $v$. When $u$ and $v$ are not in the same connected component, we set $\mathrm{dist}_G(u, v) = \infty$. We are interested in settings where $G$ has a high amount of connectivity, so that many pairs of vertices are connected to one another by short paths. In order to describe the typical distances between vertices, we draw $o_1$ and $o_2$ independently and uar from $V(G)$, and we investigate the random variable

$$\mathrm{dist}_G(o_1, o_2). \tag{1.1.13}$$

The quantity in (1.1.13) is a random variable even for *deterministic* graphs, owing to the presence of the two uar-chosen vertices $o_1, o_2 \in V(G)$. Figures 1.5 and 1.6 display the probability mass functions of this random variable for some real-world networks.

Often, we consider $\mathrm{dist}_G(o_1, o_2)$ conditional on $\mathrm{dist}_G(o_1, o_2) < \infty$. This means that we consider the typical number of edges between a uniformly chosen pair of *connected* vertices. As a result, $\mathrm{dist}_G(o_1, o_2)$ is sometimes referred to as the *typical distance*.

The nice property of $\mathrm{dist}_G(o_1, o_2)$ is that its distribution tells us something about *all possible* distances in the graph. An alternative and frequently used measure of distance in a graph is the *diameter* of the graph $G$, defined as

$$\mathrm{diam}(G) = \max_{u,v \in V(G)} \mathrm{dist}_G(u, v). \tag{1.1.14}$$

However, the diameter has several disadvantages. First, in many instances, the diameter is algorithmically more difficult to compute than the typical distances (since one has to compute the distances between *all* pairs of vertices and maximize over them). Second, it is a *number* instead of a *distribution of a random variable*, and therefore contains far less information than the distribution $\mathrm{dist}_G(o_1, o_2)$. Finally, the diameter is highly sensitive to relatively small changes in the graph $G$ under consideration. For example, adding a relatively small string of connected vertices to a graph (each of the vertices in the string having degree 2) may drastically change the diameter, while it hardly influences the typical distances.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

### 1.1.4 Related Network Properties

There are many more features that one could take into account when modeling real-world networks. See e.g., [V1, Section 1.5] for a slightly expanded discussion of such features. Other features that many networks share, or, rather, form a way to distinguish between them, are the following:

(a) their *degree correlations*, measuring the extent to which high-degree vertices tend to be connected to high-degree vertices rather than to low-degree vertices (and vice versa);
(b) their *clustering*, measuring the extent to which pairs of neighbors of vertices are neighbors themselves;
(c) their *community structure*, measuring the extent to which the network has more densely-connected subgraphs;
(d) their *spatial structure*, where the spatial component is either describing true vertex locations in real-world networks, or instead some *latent geometry* in them. The spatial structure is such that vertices that are near are more likely to be connected.

See, e.g., the book by Newman (2010) for an extensive discussion of such features, as well as the algorithmic problems that arise from them. We also refer the reader to Chapter 9, where we discuss several related models that focus on these properties.

## 1.2 Random Graphs and Real-World Networks

In this section we discuss how *random graph sequences* can be used to model real-world networks. We start by discussing graph sequences.

### Graph Sequences

Motivated by the previous section, in which empirical evidence was discussed showing that many real-world networks are *scale free* and *small world*, we set about the question of how to model them. Since many networks are quite large, mathematically, we model real-world networks by *graph sequences* $(G_n)_{n\geq 1}$, where $G_n = (V(G_n), E(G_n))$ has size $|V(G_n)| = n$ and we take the limit $n \to \infty$. Since most real-world networks are such that the average degree remains bounded, we will focus on the *sparse* regime. In the sparse regime (recall (1.1.2) and (1.1.3)), it is assumed that

$$\limsup_{n\to\infty} \mathbb{E}[D_n] = \limsup_{n\to\infty} \frac{1}{|V(G_n)|} \sum_{v\in V(G_n)} d_v^{(G_n)} < \infty. \tag{1.2.1}$$

Furthermore, we aim to study graphs that are asymptotically well behaved. For example, we often either assume, or prove, that the typical degree distribution converges, i.e., there exists a limiting degree random variable $D$ such that

$$D_n \xrightarrow{d} D, \tag{1.2.2}$$

where $\xrightarrow{d}$ denotes weak convergence of random variables. Also, we assume that our graphs are *small worlds*, which is often translated in the asymptotic sense that there exists a constant $K < \infty$ such that

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{dist}_G(o_1, o_2) \leq K \log n) = 1, \tag{1.2.3}$$

where $n$ denotes the network size. Sometimes, we even discuss *ultra-small worlds*, for which

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{dist}_G(o_1, o_2) \leq \varepsilon \log n) = 1 \tag{1.2.4}$$

for every $\varepsilon > 0$. In what follows, we discuss random graph models that share these two features.

### Random Graphs as Models for Real-World Networks

Real-world networks tend to be quite complex and unpredictable. This is understandable, since connections often arise rather irregularly. We model such irregular behavior by letting connections arise through a *random process*, thus leading us to study *random graphs*. By the previous discussion, our graphs are large and their sizes $n$ tend to infinity.

In such settings, we can either model the graphs by *fixing* their size to be large, or rather by letting the graphs *grow* to infinite size in a consistent manner. We refer to these two settings as *static* and *dynamic* random graphs. Both are useful viewpoints. Indeed, a static graph is a model for a snapshot of a network at a fixed time, where we do not know how the connections arose over time. Many network data sets are of this form. A dynamic setting, however, may be useful when we know how the network came to be as it is. In the static setting, we can make model assumptions on the degrees such that they are scale free. In the dynamic setting, we can let the evolution of the graphs give rise to power-law degree sequences, so that these settings may provide explanations for the frequent occurrence of power laws in real-world networks.

Most of the random graph models that have been investigated in the (extensive) literature are *caricatures of reality*, in the sense that one cannot confidently argue that they describe any real-world network quantitatively correctly. However, these random graph models *do* provide insight into how any of the above features can influence the global behavior of networks. In this way, they provide possible explanations of the empirical properties of real-world networks that are observed. Also, random graph models can be used as *null models*, where certain aspects of real-world networks are taken into account while others are not. This gives a qualitative way of investigating the importance of such empirical features in the real world. Often, real-world networks are compared with uniform random graphs with certain specified properties, such as their number of edges or even their degree sequence. Below, we will come back to how to generate random graphs uar from the collection of all graphs with these properties.

In the next section we describe four models of random graphs, three of which are static and one is dynamic. Below, we frequently write $f(n) = O(g(n))$ if $|f(n)|/|g(n)|$ is uniformly bounded from above by a positive constant as $n \to \infty$, $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$, and $f(n) = o(g(n))$ if $f(n)/g(n)$ tends to 0 as $n \to \infty$. We say that $f(n) \gg g(n)$ when $g(n) = o(f(n))$.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

## 1.3 Random Graph Models

We start with the most basic and simple random graph model, which has proved to be a source of tremendous inspiration, both for its mathematical beauty, as well as for providing a starting point for the analysis of random graphs.

### 1.3.1 Erdős–Rényi Random Graph

The Erdős–Rényi random graph is the simplest possible random graph. In it, we make every possible edge between a collection of $n$ vertices independently either open or closed with equal probability. This means that the Erdős–Rényi random graph has vertex set $[n] = \{1, \ldots, n\}$, and the edge $uv$ is occupied or present with probability $p$, and vacant or absent otherwise, independently of all the other edges. Here we denote the edge between vertices $u, v \in [n]$ by $uv$. The parameter $p$ is called the *edge probability*. The above random graph is denoted by $\mathrm{ER}_n(p)$. The model is named after Erdős and Rényi, since they made profound contributions in the study of this model. Exercise 1.3 investigates the uniform nature of $\mathrm{ER}_n(p)$ with $p = \frac{1}{2}$. Alternatively speaking, $\mathrm{ER}_n(p)$ with $p = \frac{1}{2}$ is the *null model*, where we take no properties of the network into account except for the total number of edges. The vertices in this model have expected degree $(n-1)/2$, which is quite large. As a result, this model is not sparse at all. Thus, we next make this model *sparse* by making $p$ smaller.

Since each edge is occupied with probability $p$, we obtain that

$$\mathbb{P}(D_n = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} = \mathbb{P}(\mathsf{Bin}(n-1, p) = k), \tag{1.3.1}$$

where $\mathsf{Bin}(m, p)$ is a binomial random variable with $m$ trials and success probability $p$. Note that

$$\mathbb{E}[D_n] = (n-1)p, \tag{1.3.2}$$

so for this model to be sparse, we need that $p$ becomes small with $n$. Thus, we take

$$p = \frac{\lambda}{n}, \tag{1.3.3}$$

and study the graph as $\lambda$ is held fixed while $n \to \infty$. In this regime, we know that

$$D_n \xrightarrow{d} D, \tag{1.3.4}$$

with $D \sim \mathsf{Poi}(\lambda)$, where $\mathsf{Poi}(\lambda)$ is a Poisson random variable with mean $\lambda$. It turns out that this result can be strengthened to the statement that the proportion of vertices with degree $k$ also converges to the probability mass function of a Poisson random variable (see [V1, Section 5.4], and in particular [V1, Theorem 5.12]), i.e., for every $k \geq 0$,

$$P_k^{(n)} = \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{\{d_v = k\}} \xrightarrow{\mathbb{P}} p_k \equiv \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \tag{1.3.5}$$

where $d_v$ denotes the degree of $v \in [n]$.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

It is well known that the Poisson distribution has very thin tails, even thinner than any exponential, as you are requested to prove in Exercise 1.4. We conclude that the Erdős–Rényi random graph is not a good model for real-world networks with their highly variable degree distributions. In the next subsection, we discuss *inhomogeneous* extensions of Erdős–Rényi random graphs which can have highly variable degrees.

### 1.3.2 Inhomogeneous Random Graphs

In inhomogeneous random graphs, we keep the independence of the edges, but make the edge probabilities different for different edges. We will discuss such general inhomogeneous random graphs in Chapter 3 below. Here, we start with one key example, which has attracted most attention in the literature so far and is also discussed in great detail in [V1, Chapter 6].

#### Rank-1 Inhomogeneous Random Graphs

The simplest inhomogeneous random graph models are sometimes referred to as *rank*-1 models, since the edge probabilities are (close to) products of vertex weights (see Remark 1.5 below for more details). This means that the expected number of edges between vertices, when viewed as a matrix, is (close to) a rank-1 matrix. We start by discussing one such model, which is the so-called *generalized random graph*.

In the generalized random graph model, the edge probability of the edge between vertices $u$ and $v$, for $u \neq v$, is equal to

$$p_{uv} = p_{uv}^{(\mathrm{GRG})} = \frac{w_u w_v}{\ell_n + w_u w_v}, \tag{1.3.6}$$

where $\boldsymbol{w} = (w_v)_{v\in[n]}$ are the *vertex weights,* and $\ell_n$ is the total vertex weight, given by

$$\ell_n = \sum_{v\in[n]} w_v. \tag{1.3.7}$$

We denote the resulting graph by $\mathrm{GRG}_n(\boldsymbol{w})$. In many cases, the vertex weights actually depend on $n$, and it would be more appropriate (but also more cumbersome), to write the weights as $\boldsymbol{w}^{(n)} = (w_v^{(n)})_{v\in[n]}$. To keep the notation simple, we refrain from making the dependence on $n$ explicit. A special case of the generalized random graph occurs when we take $w_v \equiv \frac{n\lambda}{n-\lambda}$, in which case $p_{uv} = \lambda/n$ for all $u, v \in [n]$ so that we retrieve the Erdős–Rényi random graph $\mathrm{ER}_n(\lambda/n)$.

The generalized random graph $\mathrm{GRG}_n(\boldsymbol{w})$ is close to many other inhomogeneous random graph models, such as the *random graph with prescribed expected degrees* or Chung–Lu model, denoted by $\mathrm{CL}_n(\boldsymbol{w})$, where instead

$$p_{uv} = p_{uv}^{(\mathrm{CL})} = \min(w_u w_v/\ell_n, 1). \tag{1.3.8}$$

A further adaptation is the so-called *Poissonian random graph* or Norros–Reittu model, denoted by $\mathrm{NR}_n(\boldsymbol{w})$, for which

$$p_{uv} = p_{uv}^{(\mathrm{NR})} = 1 - \exp\left(-w_u w_v/\ell_n\right). \tag{1.3.9}$$

See [V1, Sections 6.7 and 6.8] for conditions under which these random graphs are *asymptotically equivalent*, meaning that all events have equal asymptotic probabilities.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Naturally, the topology of the generalized random graph depends sensitively upon the choice of the vertex weights $\boldsymbol{w} = (w_v)_{v\in[n]}$. These vertex weights can be rather general, and we investigate both settings where the weights are *deterministic* as well as settings where they are *random*. In order to describe the empirical proportions of the weights, we define their *empirical distribution function* to be

$$F_n(x) = \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{w_v\le x\}}, \qquad x \ge 0. \tag{1.3.10}$$

We can interpret $F_n$ as the distribution of the weight of a uniformly chosen vertex in $[n]$ (see Exercise 1.7). We denote the weight of a uniformly chosen vertex $o$ in $[n]$ by $W_n = w_o$, so that, by Exercise 1.7, $W_n$ has distribution function $F_n$.

The degree distribution can converge only when the vertex weights are sufficiently regular. We often assume that the vertex weights satisfy the following *regularity conditions,* which turn out to imply convergence of the degree distribution in the generalized random graph:

**Condition 1.1** (Regularity conditions for vertex weights) *There exists a distribution function $F$ such that, as $n\to\infty$, the following conditions hold:*

**(a) Weak convergence of vertex weights.** *As $n\to\infty$,*

$$W_n \xrightarrow{d} W, \tag{1.3.11}$$

*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,*

$$\lim_{n\to\infty} F_n(x) = F(x). \tag{1.3.12}$$

**(b) Convergence of average vertex weight.** *As $n\to\infty$,*

$$\mathbb{E}[W_n] \to \mathbb{E}[W] \in (0,\infty), \tag{1.3.13}$$

*where $W_n$ and $W$ have distribution functions $F_n$ and $F$ from part (a) above, respectively.*

**(c) Convergence of second moment of vertex weights.** *As $n\to\infty$,*

$$\mathbb{E}[W_n^2] \to \mathbb{E}[W^2] < \infty, \tag{1.3.14}$$

*where $W_n$ and $W$ have distribution functions $F_n$ and $F$ from part (a) above, respectively.*

Condition 1.1 is virtually the same as [V1, Condition 6.4]. Condition 1.1(a) guarantees that the weight of a "typical" vertex is close to a random variable $W$ that is independent of $n$. Condition 1.1(b) implies that the average weight of the vertices in $\mathrm{GRG}_n(\boldsymbol{w})$ converges to the expectation of the limiting weight variable. In turn, this implies that the expectation of the average degree in $\mathrm{GRG}_n(\boldsymbol{w})$ converges to the expectation of this limiting random variable as well. Condition 1.1(c) ensures the convergence of the second moment of the weights to the second moment of the limiting weight variable.

**Remark 1.2** (Regularity for random weights) Sometimes we are interested in cases where the weights of the vertices are *random* themselves. For example, this arises when the weights $\boldsymbol{w} = (w_v)_{v\in[n]}$ are realizations of iid random variables. Then, the function $F_n$ is also a random distribution function. Indeed, in this case $F_n$ is the *empirical distribution function* of the

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

random weights $(w_v)_{v\in[n]}$. We stress that $\mathbb{E}[W_n]$ is then to be interpreted as $\frac{1}{n}\sum_{v\in[n]} w_v$, which is itself random. Therefore, in Condition 1.1 we require random variables to converge, and there are several notions of convergence that may be used. The notion of convergence that we assume is *convergence in probability* (see [V1, Section 6.2]). ◀

Let us now discuss some canonical examples of weight distributions that satisfy the Regularity Condition 1.1.

### Weights Moderated by a Distribution Function

Let $F$ be a distribution function for which $F(0) = 0$ and fix

$$w_v = [1 - F]^{-1}(v/n), \tag{1.3.15}$$

where $[1 - F]^{-1}$ is the generalized inverse function of $1 - F$, defined, for $u \in (0, 1)$, by (recall [V1, (6.2.14) and (6.2.15)])

$$[1 - F]^{-1}(u) = \inf\{x\colon [1 - F](x) \le u\}. \tag{1.3.16}$$

For the choice (1.3.15), we can explicitly compute $F_n$ as (see [V1, (6.2.17)])

$$F_n(x) = \frac{1}{n}\big(\lfloor nF(x)\rfloor + 1\big) \wedge 1, \tag{1.3.17}$$

where $x \wedge y$ denotes the minimum of $x, y \in \mathbb{R}$. It is not hard to see that Condition 1.1(a) holds for $(w_v)_{v\in[n]}$ as in (1.3.15), while Condition 1.1(b) holds when $\mathbb{E}[W] \in (0, \infty)$, and Condition 1.1(c) holds when $\mathbb{E}[W^2] < \infty$, as can be concluded from Exercise 1.9.

### Independent and Identically Distributed Weights

We now discuss the setting where the weights are an independent and identically distributed (iid) sequence of random variables, for which Conditions 1.1(b) and (c) follow from the law of large numbers, and Condition 1.1(a) from the Glivenko–Cantelli Theorem. Since we will often deal with ratios of the form $w_u w_v/(\sum_{k\in[n]} w_k)$, we assume that $\mathbb{P}(w = 0) = 0$ to avoid situations where all weights are zero.

Both settings, i.e., with weights $(w_v)_{v\in[n]}$ as in (1.3.15), and with iid weights $(w_v)_{v\in[n]}$, have their own merits. The great advantage of iid weights is that the vertices in the resulting graph are, in distribution, the same. More precisely, the vertices are completely *exchangeable*, as in the Erdős–Rényi random graph $\mathrm{ER}_n(p)$. Unfortunately, when we take the weights to be iid, in the resulting graph the edges are no longer independent (despite the fact that they are *conditionally* independent *given* the weights). In what follows, we focus on the setting where the weights are *prescribed*. When the weights are deterministic, this changes nothing; when the weights are iid, this means that we are *conditioning on the weights*.

### Degrees in Generalized Random Graphs

We write $d_v$ for the degree of vertex $v$ in $\mathrm{GRG}_n(\boldsymbol{w})$. Thus, $d_v$ is given by

$$d_v = \sum_{u\in[n]} \mathbb{1}_{\{uv\in E(\mathrm{GRG}_n(\boldsymbol{w}))\}}. \tag{1.3.18}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

For $k \geq 0$, we let

$$P_k^{(n)} = \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{\{d_v = k\}} \tag{1.3.19}$$

denote the proportion of vertices with degree $k$ of $\mathrm{GRG}_n(\boldsymbol{w})$. We call $(P_k^{(n)})_{k \geq 0}$ the *degree sequence* of $\mathrm{GRG}_n(\boldsymbol{w})$. We denote the probability mass function of a *mixed-Poisson distribution* by $p_k$, i.e., for $k \geq 0$,

$$p_k = \mathbb{E}\Big[\mathrm{e}^{-W} \frac{W^k}{k!}\Big], \tag{1.3.20}$$

where $W$ is a random variable having distribution function $F$ from Condition 1.1. The main result concerning the vertex degrees is as follows:

**Theorem 1.3** (Degree sequence of $\mathrm{GRG}_n(\boldsymbol{w})$) *Assume that Conditions 1.1(a),(b) hold. Then, for every $\varepsilon > 0$,*

$$\mathbb{P}\Big( \sum_{k=0}^{\infty} |P_k^{(n)} - p_k| \geq \varepsilon \Big) \to 0, \tag{1.3.21}$$

*where $(p_k)_{k \geq 0}$ is given by* (1.3.20).

*Proof* This is given in [V1, Theorem 6.10]. □

Consequently, with $D_n = d_o$ denoting the degree of a random vertex, we obtain

$$D_n \xrightarrow{d} D, \tag{1.3.22}$$

where $\mathbb{P}(D = k) = p_k$, defined in (1.3.20), as shown in Exercise 1.10.

Recall from Section 1.1.2 that we are often interested in scale-free random graphs, i.e., random graphs for which the degree distribution obeys a power law. We see from Theorem 1.3 that this is true precisely when $D$ obeys a power law. This, in turn, occurs precisely when $W$ obeys a power law, for example, when, for $w$ large,

$$\mathbb{P}(W > w) = \frac{c}{w^{\tau - 1}}(1 + o(1)). \tag{1.3.23}$$

Then, for $w$ large,

$$\mathbb{P}(D > w) = \mathbb{P}(W > w)(1 + o(1)). \tag{1.3.24}$$

This follows from Theorem 1.3, in combination with [V1, Exercise 6.12], which shows that the tail behavior of a mixed-Poisson distribution and that of its weight distribution agree for power laws.

### Generalized Random Graph Conditioned on its Degrees

The generalized random graph with edge probabilities as in (1.3.6) is rather special. Indeed, when we *condition on its degree sequence*, the graph has a uniform distribution over the set of all graphs with the same degree sequence. For this, note that $\mathrm{GRG}_n(\boldsymbol{w})$ can be equivalently encoded by $(X_{uv})_{1 \leq u \leq v \leq n}$, where $X_{uv}$ is the indicator that the edge $uv$ is occupied. Then, $(X_{uv})_{1 \leq u \leq v \leq n}$ are independent Bernoulli random variables with edge probabilities as in (1.3.6). By convention, let $X_{vv} = 0$ for every $v \in [n]$, and $X_{vu} = X_{uv}$ for

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

$1 \le u < v \le n$. In terms of the variables $X = (X_{uv})_{1\le u<v\le n}$, let $d_v(X) = \sum_{u\in[n]} X_{uv}$ be the degree of vertex $v$. Then, the uniformity is equivalent to the statement that, for each $x = (x_{uv})_{1\le u<v\le n}$ such that $d_v(x) = d_v$ for every $v \in [n]$,

$$\mathbb{P}(X = x \mid d_v(X) = d_v\ \forall v \in [n]) = \frac{1}{\#\{y\colon d_v(y) = d_v\ \forall v \in [n]\}}, \tag{1.3.25}$$

that is, the distribution is uniform over all graphs with the prescribed degree sequence. This turns out to be rather convenient, and thus we state it formally here:

**Theorem 1.4** (GRG conditioned on degrees has a uniform law) *The generalized random graph* $\mathrm{GRG}_n(\boldsymbol{w})$ *with edge probabilities* $(p_{uv})_{1\le u<v\le n}$ *given by*

$$p_{uv} = \frac{w_u w_v}{\ell_n + w_u w_v}, \tag{1.3.26}$$

*conditioned on* $\{d_v(X) = d_v \forall v \in [n]\}$*, is uniform over all graphs with degrees* $(d_v)_{v\in[n]}$.

*Proof* See [V1, Theorem 6.15]. □

In Chapter 3 below, we discuss a far more general setting of inhomogeneous random graphs. The analysis of such random graphs is substantially more challenging than the rank-1 case. As explained in more detail there, this is due to the fact that these random graphs are no longer locally described by single-type branching processes, but rather by multi-type branching processes.

**Remark 1.5** (What's in a name?) The models discussed here, $\mathrm{GRG}_n(\boldsymbol{w})$ in (1.3.6) as well as $\mathrm{CL}_n(\boldsymbol{w})$ in (1.3.8) and $\mathrm{NR}_n(\boldsymbol{w})$ in (1.3.9), go under various names in the literature. Bollobás et al. (2007) referred to them as a *rank*-1 *random graph*, because $p_{uv} \approx w_u w_v/\ell_n$ and the matrix $(w_u w_v/\ell_n)_{u,v\in[n]}$ has rank one. In the physics literature, they go under the name of *hidden variable models*, where the weights $(w_v)_{v\in[n]}$ are interpreted as the hidden variables (and they are often assumed to be iid). Owing to the uniformity in the conditional distribution given its degrees, $\mathrm{GRG}_n(\boldsymbol{w})$ is also a *maximal entropy* model, as will be explained in more detail in Section 9.4.4. Finally, some researchers call them *soft configuration models*; see Remark 1.6 for further discussion of this phrase. ◀

### 1.3.3 Configuration Models

The configuration model is a model in which the degrees of vertices are fixed beforehand. Such a model is more flexible than the generalized random graph. For example, the generalized random graph always has a positive proportion of vertices of degree 0, 1, 2, etc., as easily follows from Theorem 1.3.

Fix an integer $n$ that denotes the number of vertices in the random graph. Consider a sequence of degrees $\boldsymbol{d} = (d_v)_{v\in[n]}$. Again, it might be more appropriate, but also more cumbersome, to write the degrees as $\boldsymbol{d}^{(n)} = (d_v^{(n)})_{v\in[n]}$, and so we will refrain from this. The aim is to construct an undirected (multi-)graph with $n$ vertices, where vertex $v$ has degree $d_v$. Here a multi-graph is a graph *possibly* having self-loops and multiple edges between pairs of vertices.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Without loss of generality, we assume throughout this chapter that $d_v \geq 1$ for all $v \in [n]$, since, when $d_v = 0$, vertex $v$ is isolated and can be removed from the graph. One possible random graph model takes the uniform measure over such undirected and simple graphs. Here, we call a multi-graph *simple* when it has no self-loops, and no multiple edges exist between any pair of vertices. However, the set of undirected simple graphs with $n$ vertices where vertex $v$ has degree $d_v$ may be empty. For example, in order for such a graph to exist, we must assume that the total degree

$$\ell_n = \sum_{v \in [n]} d_v \tag{1.3.27}$$

is even.

We wish to construct a simple graph such that $\boldsymbol{d} = (d_v)_{v \in [n]}$ are the degrees of the $n$ vertices. Even when $\ell_n = \sum_{v \in [n]} d_v$ *is* even, however, this is not always possible. Therefore, instead, we construct a *multi-graph*. One way of obtaining such a multi-graph 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. Let us explain this in more detail.

To construct the multi-graph where vertex $v$ has degree $d_v$ for all $v \in [n]$, we have $n$ separate vertices and, incident to vertex $v$, we have $d_v$ half-edges. Every half-edge needs to be connected to another half-edge to form an edge, and by forming all edges we build the graph. For this, the half-edges are numbered in an arbitrary order from 1 to $\ell_n$. 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 multi-graph, and these half-edges are removed from the list of half-edges that need to be paired. 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 we call the resulting graph the *configuration model with degree sequence* $\boldsymbol{d}$, abbreviated as $\mathrm{CM}_n(\boldsymbol{d})$. The pairing of the half-edges that induces the configuration model graph is sometimes called a *configuration*l.

A careful reader may worry about the order in which the half-edges are being paired. In fact, this ordering turns out to be irrelevant since the random pairing of half-edges is completely *exchangeable*. It can even be done in a *random* fashion, which will be useful when investigating neighborhoods in the configuration model. See e.g., [V1, Definition 7.5 and Lemma 7.6] for more details on this exchangeability.

Interestingly, one can rather explicitly compute the distribution of $\mathrm{CM}_n(\boldsymbol{d})$. To do so, note that $\mathrm{CM}_n(\boldsymbol{d})$ is characterized by the random vector $(X_{uv})_{1 \leq u \leq v \leq n}$. Here $X_{uv}$ is the number of edges between vertex $u$ and $v$, and $X_{vv}$ is the number of self-loops incident to vertex $v$, so that

$$d_v = X_{vv} + \sum_{u \in [n]} X_{uv}. \tag{1.3.28}$$

Note furthermore that $X_{vv}$ appears *twice* in (1.3.28), which is natural, since a self-loop consists of *two* half-edges. This does not conflict with the definition of $d_v$ for $\mathrm{GRG}_n(\boldsymbol{w})$, since $X_{uu} = 0$ and $X_{u,v} \in \{0, 1\}$ for $\mathrm{GRG}_n(\boldsymbol{w})$.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

In terms of this notation, and writing $G = (x_{uv})_{u,v\in[n]}$ to denote a multi-graph on $[n]$,

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) = G) = \frac{1}{(\ell_n - 1)!!} \frac{\prod_{v\in[n]} d_v!}{\prod_{v\in[n]} 2^{x_{vv}} \prod_{1\le u\le v\le n} x_{uv}!}. \tag{1.3.29}$$

See, e.g., [V1, Proposition 7.7] for this result. In particular, $\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) = G)$ is the *same* for each *simple* $G$, where $G$ is simple when $x_{vv} = 0$ for every $v \in [n]$ and $x_{uv} \in \{0, 1\}$ for every $1 \le u < v \le n$. Thus, the configuration model conditioned on simplicity is a *uniform* random graph with the prescribed degree distribution. This is quite relevant, as it gives a convenient way to *obtain* such a uniform graph, which is a highly non-trivial fact.

**Remark 1.6** (What's in a name continued?) The name *configuration model* was invented by Bollobás (1980), who considered the matching of half-edges to be the *configuration* on which the model is based. The model of study for Bollobás (1980) was the uniform simple random regular graph, where all degrees are the same, as we discuss further below. Molloy and Reed (1995, 1998) extended it to general degrees. As a result, it is sometimes also called the *Molloy–Reed model*. With $X_{uv}$ equal to the number of edges between vertices $u$ and $v$,

$$\mathbb{E}[X_{uv}] = \frac{d_u d_v}{\ell_n - 1}, \tag{1.3.30}$$

since each of the $d_v$ half-edges incident to vertex $v$ has probability $d_u/(\ell_n - 1)$ to be connected to vertex $u$. Since (1.3.30) is close to the edge probability $p_{uv}$ in rank-1 random graphs (recall Remark 1.5), rank-1 random graphs are sometimes called *soft configuration models*. The configuration-model degree constraint is instead viewed as a *hard* constraint.◀

The uniform nature of the configuration model conditioned on simplicity partly explains its popularity, and it has become one of the most highly studied random graph models. It also implies that, conditioned on simplicity, the configuration model is the *null model* for a real-world network where all the degrees are fixed. This allows one to distinguish the relevance of the *degree inhomogeneity* from other features of the network, such as its community structure, clustering, etc.

As for $\mathrm{GRG}_n(\boldsymbol{w})$, we 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 $o$ in $[n]$ by $D_n = d_o$. The random variable $D_n$ has distribution function $F_n$ given by

$$F_n(x) = \frac{1}{n} \sum_{v\in[n]} \mathbb{1}_{\{d_v\le x\}}, \tag{1.3.31}$$

which is the *empirical distribution of the degrees*. We assume that the vertex degrees satisfy the following *regularity conditions*:

**Condition 1.7** (Regularity conditions for vertex degrees)

**(a) Weak convergence of vertex degrees.** *There exists a distribution function $F$ such that, as $n \to \infty$,*

$$D_n \xrightarrow{d} D, \tag{1.3.32}$$

*where $D_n$ and $D$ have distribution functions $F_n$ and $F$, respectively.*

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

*Equivalently, for any* $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_n(x) = F(x). \tag{1.3.33}$$

*Further, we assume that* $F(0) = 0$, *i.e.,* $\mathbb{P}(D \geq 1) = 1$.

**(b) Convergence of average vertex degree.** *As* $n \to \infty$,

$$\mathbb{E}[D_n] \to \mathbb{E}[D] < \infty, \tag{1.3.34}$$

*where* $D_n$ *and* $D$ *have the distribution functions* $F_n$ *and* $F$ *from part (a) above, respectively.*

**(c) Convergence of second moment of vertex degrees.** *As* $n \to \infty$,

$$\mathbb{E}[D_n^2] \to \mathbb{E}[D^2] \in (0, \infty), \tag{1.3.35}$$

*where* $D_n$ *and* $D$ *have distribution functions* $F_n$ *and* $F$ *from part (a) above, respectively.*

The possibility that one will obtain a non-simple graph is a major disadvantage of the configuration model. There are two ways of dealing with this complication, as follows:

### Erased Configuration Model

The first way of dealing with self-loops and multi-edges is to *erase* the problems. This means that we replace $\mathrm{CM}_n(\boldsymbol{d}) = (X_{uv})_{1\leq u\leq v\leq n}$ by its erased version $\mathrm{ECM}_n(\boldsymbol{d}) = (X_{uv}^{(\mathrm{er})})_{1\leq u\leq v\leq n}$, where $X_{vv}^{(\mathrm{er})} \equiv 0$, while $X_{uv}^{(\mathrm{er})} = 1$ precisely when $X_{uv} \geq 1$. In words, we remove the self-loops and merge all multiple edges to a single edge. Of course, this changes the precise degree distribution. However, [V1, Theorem 7.10] (see also Theorem 1.8 below) shows that only a small proportion of the edges is erased, so that the erasing does not change the asymptotic degree distribution. See [V1, Section 7.3] for more details. Of course, the downside of this approach is that the degrees are changed by the procedure, while we would like to keep the degrees *precisely* as specified.

Let us describe the degree distribution in the erased configuration model in more detail, to study the effect of the erasure of self-loops and multiple edges. We denote the degrees in the erased configuration model by $\boldsymbol{D}^{(\mathrm{er})} = (D_v^{(\mathrm{er})})_{v\in[n]}$, so that

$$D_v^{(\mathrm{er})} = d_v - 2s_v - m_v, \tag{1.3.36}$$

where $(d_v)_{v\in[n]}$ are the degrees in $\mathrm{CM}_n(\boldsymbol{d})$, $s_v = x_{vv}$ is the number of self-loops of vertex $v$ in $\mathrm{CM}_n(\boldsymbol{d})$, and

$$m_v = \sum_{u\neq v} (x_{uv} - 1)\mathbb{1}_{\{x_{uv}\geq 2\}} \tag{1.3.37}$$

is the number of multiple edges removed from $v$.

Denote the empirical degree sequence $(p_k^{(n)})_{k\geq 1}$ in $\mathrm{CM}_n(\boldsymbol{d})$ by

$$p_k^{(n)} = \mathbb{P}(D_n = k) = \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{d_v=k\}}, \tag{1.3.38}$$

and denote the related degree sequence in the erased configuration model $(P_k^{(\mathrm{er})})_{k\geq 1}$ by

$$P_k^{(\mathrm{er})} = \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{D_v^{(\mathrm{er})}=k\}}. \tag{1.3.39}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

From the notation it should be clear that $(p_k^{(n)})_{k\geq 1}$ is a *deterministic* sequence when $\boldsymbol{d} = (d_v)_{v\in[n]}$ is deterministic, while $(P_k^{(\mathrm{er})})_{k\geq 1}$ is a *random* sequence, since the erased degrees $(D_v^{(\mathrm{er})})_{v\in[n]}$ form a random vector even when $\boldsymbol{d} = (d_v)_{v\in[n]}$ is deterministic.

Now we are ready to state the main result concerning the degree sequence of the erased configuration model:

**Theorem 1.8** (Degree sequence of erased configuration model with fixed degrees) *For fixed degrees $\boldsymbol{d}$ satisfying Conditions 1.7(a),(b), the degree sequence of the erased configuration model $(P_k^{(\mathrm{er})})_{k\geq 1}$ converges in probability to $(p_k)_{k\geq 1}$. More precisely, for every $\varepsilon > 0$,*

$$\mathbb{P}\Big(\sum_{k=1}^{\infty} |P_k^{(\mathrm{er})} - p_k| \geq \varepsilon\Big) \to 0, \tag{1.3.40}$$

*where $p_k = \mathbb{P}(D = k)$ as in Condition 1.7(a).*

*Proof* See [V1, Theorem 7.10]. □

Theorem 1.8 indeed shows that most of the edges are kept in the erasure procedure; see Exercise 1.17.

### Configuration Model Conditioned on Simplicity

The second solution to the multi-graph problem of the configuration model is to throw away the result when it is not simple, and try again. Therefore, this construction is sometimes called the *repeated configuration model*. It turns out that, when Conditions 1.7(a)–(c) hold (see [V1, Theorem 7.12]),

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ is a simple graph}) = \mathrm{e}^{-\nu/2-\nu^2/4}, \tag{1.3.41}$$

where

$$\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} \tag{1.3.42}$$

is the expected forward degree. This is a realistic option when $\mathbb{E}[D^2] < \infty$. Unfortunately, this is not an option when the asymptotic degrees obey an asymptotic power law with $\tau \in (2,3)$ (as, e.g., in (1.1.12)), since then $\mathbb{E}[D^2] = \infty$. Note that, by (1.3.29), $\mathrm{CM}_n(\boldsymbol{d})$ conditioned on simplicity is a *uniform random graph* with the prescribed degree sequence. We denote this random graph by $\mathrm{UG}_n(\boldsymbol{d})$. We return to the difficulty of generating simple graphs with infinite-variance degrees in Section 1.3.4 below.

### Relation between Generalized Random Graph and Configuration Model

Since $\mathrm{CM}_n(\boldsymbol{d})$ conditioned on simplicity yields a uniform (simple) random graph with these degrees, and, also, by (1.3.25), $\mathrm{GRG}_n(\boldsymbol{w})$ conditioned on its degrees is a uniform (simple) random graph with the given degree distribution, the laws of these (conditioned) random graph models are the same. As a result, one can prove results for $\mathrm{GRG}_n(\boldsymbol{w})$ by proving them for $\mathrm{CM}_n(\boldsymbol{d})$ under the appropriate degree conditions, and then proving that $\mathrm{GRG}_n(\boldsymbol{w})$ satisfies these conditions in probability.

A further useful result in this direction is that the weight regularity conditions in Conditions 1.1(a),(b) imply the degree regularity conditions in Conditions 1.7(a),(b):

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

**Theorem 1.9** (Regularity conditions for weights and degrees) *Let $d_v$ be the degree of vertex $v$ in $\mathrm{GRG}_n(\boldsymbol{w})$, and let $\boldsymbol{d} = (d_v)_{v\in[n]}$. Then, $\boldsymbol{d}$ satisfies Conditions 1.7(a),(b) in probability when $\boldsymbol{w}$ satisfies Conditions 1.1(a),(b), where*

$$\mathbb{P}(D=k) = \mathbb{E}\Big[\frac{W^k}{k!}\mathrm{e}^{-W}\Big] \tag{1.3.43}$$

*denotes the mixed-Poisson distribution with mixing distribution $W$ having distribution function $F$ in Condition 1.1(a). Further, $\boldsymbol{d}$ satisfies Conditions 1.7(a)–(c) in probability when $\boldsymbol{w}$ satisfies Conditions 1.1(a)–(c).*

*Proof* See [V1, Theorem 7.19]. The weak convergence in Condition 1.7(a) follows from Theorem 1.3. □

**Remark 1.10** (Proving results for $\mathrm{GRG}_n(\boldsymbol{w})$ through $\mathrm{CM}_n(\boldsymbol{d})$) Combined with Theorem 1.4, Theorem 1.9 allows us to prove many results for the generalized random graph by first proving them for the configuration model under appropriate conditions on its degrees, and then extending them to the generalized random graph by proving that its degrees satisfy the assumptions made. In particular, any property that holds *in probability* for $\mathrm{CM}_n(\boldsymbol{d})$ can be extended to $\mathrm{GRG}_n(\boldsymbol{w})$ in this way. See [V1, Sections 6.6 and 7.5] for more details. This strategy is also frequently used in the present volume. ◀

### A Useful Degree-Truncation Argument for Heavy-Tailed Degrees

Recall from Section 1.1.2 that many real-world networks have substantial inhomogeneities in their degrees. As a result, we frequently discuss configuration models with power-law degrees, giving rise to degree distributions with maxima that grow as a positive power of $n$. Such large degrees can be inconvenient in technical estimates. We next present a useful *degree-truncation argument* for the configuration model, which allows us to compare such a model with an alternative configuration model with *bounded* degrees. In its statement, we write $x \wedge y$ for the minimum of $x, y \in \mathbb{R}$:

**Theorem 1.11** (Degree truncation for configuration models) *Consider $\mathrm{CM}_n(\boldsymbol{d})$ with general degrees. Fix $b \geq 1$. There exists a related configuration model $\mathrm{CM}_{n'}(\boldsymbol{d}')$ with $n' \geq n$ that is coupled to $\mathrm{CM}_n(\boldsymbol{d})$ and satisfies the following:*

*(a) the degrees in $\mathrm{CM}_{n'}(\boldsymbol{d}')$ are a truncated version of those in $\mathrm{CM}_n(\boldsymbol{d})$, i.e., $d'_v = (d_v \wedge b)$ for $v \in [n]$, and $d'_v = 1$ for $v \in [n'] \setminus [n]$;*

*(b) the total degree in $\mathrm{CM}_{n'}(\boldsymbol{d}')$ is the same as that in $\mathrm{CM}_n(\boldsymbol{d})$, i.e., $\sum_{v\in[n']} d'_v = \sum_{v\in[n]} d_v$;*

*(c) for all $u, v \in [n]$, if $u$ and $v$ are connected in $\mathrm{CM}_{n'}(\boldsymbol{d}')$, then so are $u$ and $v$ in $\mathrm{CM}_n(\boldsymbol{d})$, i.e., $\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u,v) \leq \mathrm{dist}_{\mathrm{CM}_{n'}(\boldsymbol{d}')}(u,v)$ almost surely.*

**Remark 1.12** (Truncation of degrees in range) The construction that proves Theorem 1.11 is highly flexible, and also allows for a degree truncation that maintains restrictions on the minimal degree $d_{\min} = \min_{v\in[n]} d_v$. Indeed, fix $b \geq 2$. There exists a related configuration model $\mathrm{CM}_{n'}(\boldsymbol{d}')$ satisfying (b) and (c) in Theorem 1.11, while (a) is replaced by $d'_v = d_v$ when $d_v < 2b$, by $d'_v = b$ when $d_v \geq 2b$ for $v \in [n]$, and by $b \leq d'_v < 2b$ for $v \in [n'] \setminus [n]$, so that $d'_{\min} = \min_{v\in[n']} d'_v \geq d_{\min} \wedge b$. ◀

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

*Proof* The proof relies on an "explosion" or "fragmentation" of the vertices $[n]$ in $\mathrm{CM}_n(\boldsymbol{d})$. Label the half-edges from 1 to $\ell_n$. We go through the vertices $v \in [n]$ one by one. When $d_v \leq b$, we do nothing. When $d_v > b$, we let $d'_v = b$ and keep the $b$ half-edges with the lowest labels. The remaining $d_v - b$ half-edges are exploded from vertex $v$, in that they are incident to vertices of degree 1 in $\mathrm{CM}_{n'}(\boldsymbol{d}')$, and are given vertex labels above $n$. We give the exploded half-edges the remaining labels of the half-edges incident to $v$. Thus, the half-edges receive labels both in $\mathrm{CM}_n(\boldsymbol{d})$ as well as in $\mathrm{CM}_{n'}(\boldsymbol{d}')$, and the labels of the half-edges incident to $v \in [n]$ in $\mathrm{CM}_{n'}(\boldsymbol{d}')$ are a subset of those in $\mathrm{CM}_n(\boldsymbol{d})$. In total, we thus create an extra $n^+ = \sum_{v\in[n]}(d_v - b) \vee 0$ "exploded" vertices of degree 1, and $n' = n + n^+$, where $x \vee y$ denotes the maximum of $x, y \in \mathbb{R}$.

We then pair the half-edges randomly, in the same way in $\mathrm{CM}_n(\boldsymbol{d})$ as in $\mathrm{CM}_{n'}(\boldsymbol{d}')$. This means that when the half-edge with label $x$ is paired with the half-edge with label $y$ in $\mathrm{CM}_n(\boldsymbol{d})$, then also the half-edge with label $x$ is paired with the half-edge with label $y$ in $\mathrm{CM}_{n'}(\boldsymbol{d}')$, for all $x, y \in [\ell_n]$.

We now check parts (a)–(c). Obviously parts (a) and (b) follow from the construction. For part (c), we note that all exploded vertices in $[n^+] \setminus [n]$ have degree 1. Further, for vertices $u, v \in [n]$, if there exists a path in $\mathrm{CM}_{n'}(\boldsymbol{d}')$ connecting them then the intermediate vertices have degree at least 2, so that they cannot correspond to exploded vertices and must therefore in $\mathrm{CM}_{n'}(\boldsymbol{d}')$ have labels in $[n]$. Thus, the same path of paired half-edges also exists in $\mathrm{CM}_n(\boldsymbol{d})$, so that $u$ and $v$ are also connected in $\mathrm{CM}_n(\boldsymbol{d})$.

We conclude by adapting the construction to prove the statement in Remark 1.12. We again go through the vertices $v \in [n]$ one by one. When $d_v < 2b$, we do nothing. When $d_v \geq 2b$, we let $d'_v = b$ and keep the $b$ half-edges with the lowest labels. The remaining $d_v - b$ half-edges are exploded from vertex $v$, in that they are incident to "exploded" vertices that all have degree $b$ in $\mathrm{CM}_{n'}(\boldsymbol{d}')$ possibly except for one vertex that has degree in $[b, 2b)$, and are given vertex labels above $n$. This means that a vertex of degree $d_v \geq 2b$ is replaced by one vertex in $[n]$ and $\lfloor d_v/b \rfloor - 1$ vertices in $[n'] \setminus [n]$, of which all, possibly except for the last vertex, have degree $b$, and the degree of the last vertex equals $d_v - b(\lfloor d_v/b \rfloor - 1) \in [b, 2b)$. We again give the exploded half-edges the remaining labels of the half-edges incident to $v$. This identifies the desired construction for Remark 1.12. For part (c), we note that the half-edges incident to exploded vertices arise from the same vertex in $[n]$ as before explosion, so a path between vertices $u', v' \in [n']$ in $\mathrm{CM}_{n'}(\boldsymbol{d}')$ implies that a path between the vertices $u, v \in [n]$ that correspond to $u', v'$ exists. This implies that part (c) holds. □

### 1.3.4 Uniform Random Graphs and Switching Algorithms for them

So far, we have focussed on obtaining a uniform random graph with a prescribed degree sequence by conditioning the configuration model on being simple. As explained above, this does not work so well when the degrees have infinite variance. Another setting where this method fails to deliver occurs when the average degree is *large* rather than bounded, so that the graph is no longer *sparse* in the strict sense (recall Section 1.1.1).

An alternative method for producing a sample from the uniform distribution on simple graphs uses a *switching algorithm*. A switching algorithm is a Markov chain on the space of simple graphs, where, in each step, some edges in the graph are rewired while keeping the

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

graph simple. Under mild conditions on the precise switching dynamics, the uniform distribution is the stationary distribution of this Markov chain, so letting the switching algorithm run for an infinitely long time, we obtain a perfect sample from the uniform distribution. The "mild" conditions follow, for example, when the switch chain is *doubly stochastic*.

Switching algorithms can also be used rather effectively to compute probabilities of certain events for uniform random graphs with specified degrees, as we explain later. As such, switching methods form an indispensable tool in studying uniform random graphs with prescribed degrees. We start by explaining the basic switching algorithms and their relation to uniform sampling.

### Switch Markov Chain

The switch Markov chain is a Markov chain on the space of simple graphs with prescribed degrees given by $\boldsymbol{d}$. Fix a simple graph $G = ([n], E(G))$ for which the degree of vertex $v$ equals $d_v$ for all $v \in [n]$. We assume that such a simple graph exists, i.e., we assume that $\boldsymbol{d} = (d_v)_{v\in[n]}$ is *graphical*.

In order to describe the dynamics of the switch chain, choose two edges $\{u, v\}$ and $\{x, y\}$ uar from the edge set $E(G)$, where $G$ is the current simple graph. The possible switches of these two edges are (1) $\{u, x\}$ and $\{v, y\}$; (2) $\{v, x\}$ and $\{u, y\}$; and (3) $\{u, v\}$ and $\{x, y\}$ (so that no change is made). Choose each of these three options with probability equal to $\frac{1}{3}$, and write the chosen edges as $e_1, e_2$. *Accept* the switch when the resulting graph with edges $\{e_1, e_2\} \cup (E(G) \setminus \{\{u, v\}, \{x, y\}\})$ is simple, and *reject* the switch otherwise (so that the graph remains unchanged under the dynamics).

It is not hard to see that the resulting Markov chain is aperiodic and irreducible. Further, the switch chain is doubly stochastic since it is reversible. As a result, its stationary distribution is the uniform random graph with prescribed degree sequence $\boldsymbol{d}$, which we have denoted by $\mathrm{UG}_n(\boldsymbol{d})$, as required.

The above method works rather generally, and, in the limit of infinitely many switches, produces a sample from $\mathrm{UG}_n(\boldsymbol{d})$ for *every* graphical degree sequence, even when the degrees are large. As a result, this chain is the method of choice to produce a sample of $\mathrm{UG}_n(\boldsymbol{d})$ when the probability of simplicity of the configuration model vanishes. However, it is unclear *precisely* how often one needs to switch in order for the Markov chain to be sufficiently close to the uniform (and thus stationary) distribution. See the notes in Section 1.6 for a discussion of the history of the switch chain, as well as the available results about its convergence.

### Switching Methods for Random Graphs with Prescribed Degrees

Switching algorithms can also be used to prove properties about uniform random graphs with prescribed degrees. Here, we explain how switching can be used to estimate the connection probability between vertices of specific degrees in a uniform random graph. Recall that $\ell_n = \sum_{v\in[n]} d_v$. Then, the asymptotics for the edge probabilities for $\mathrm{UG}_n(\boldsymbol{d})$ are given in the following theorem, where $E(\mathrm{UG}_n(\boldsymbol{d}))$ denotes the edge set of $\mathrm{UG}_n(\boldsymbol{d})$:

**Theorem 1.13** (Edge probabilities for uniform random graphs with prescribed degrees) *Assume that the empirical distribution $F_n$ of $\mathbf{d}$ satisfies, for all $x \geq 1$,*

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

$$[1-F_n](x) \le c_F x^{-(\tau-1)}, \tag{1.3.44}$$

*for some $c_F > 0$ and $\tau \in (2,3)$. Let $U$ denote a set of unordered pairs of vertices and let $\mathcal{E}_U = \{\{s,t\} \in E(\mathrm{UG}_n(\boldsymbol{d})) \; \forall \{s,t\} \in U\}$ denote the event that $\{s,t\}$ is an edge for every $\{s,t\} \in U$. Then, assuming that $|U| = O(1)$, for every $\{u,v\} \notin U$,*

$$\mathbb{P}(\{u,v\} \in E(\mathrm{UG}_n(\boldsymbol{d})) \mid \mathcal{E}_U) = (1+o(1))\frac{(d_u - |U_u|)(d_v - |U_v|)}{\ell_n + (d_u - |U_u|)(d_v - |U_v|)}, \tag{1.3.45}$$

*where $U_v$ denotes the set of pairs in $U$ that contain $v \in [n]$.*

**Remark 1.14** (Relation to $\mathrm{ECM}_n(\boldsymbol{d})$ and $\mathrm{GRG}_n(\boldsymbol{w})$) Theorem 1.13 shows that, when $d_u d_v \gg \ell_n$,

$$1 - \mathbb{P}(\{u,v\} \in E(\mathrm{UG}_n(\boldsymbol{d}))) = (1+o(1))\frac{\ell_n}{d_u d_v}. \tag{1.3.46}$$

In the erased configuration model, on the other hand,

$$1 - \mathbb{P}(\{u,v\} \in E(\mathrm{ECM}_n(\boldsymbol{d}))) \le \mathrm{e}^{-d_u d_v/(2\ell_n)}, \tag{1.3.47}$$

as will be crucially used in Chapter 7 below (see Lemma 7.12 for a proof of (1.3.47)). Thus, the probability that two high-degree vertices are not connected is much smaller for $\mathrm{ECM}_n(\boldsymbol{d})$ than for $\mathrm{UG}_n(\boldsymbol{d})$. On a related note, the fact that

$$\mathbb{P}(\{u,v\} \in E(\mathrm{UG}_n(\boldsymbol{d}))) \approx \frac{d_u d_v}{\ell_n + d_u d_v},$$

as in $\mathrm{GRG}_n(\boldsymbol{w})$ when $\boldsymbol{w} = \boldsymbol{d}$, indicates once more that $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{UG}_n(\boldsymbol{d})$ are closely related. ◀

We now proceed with the proof of Theorem 1.13. We first prove a useful lemma about the number of 2-paths starting from a specified vertex, where a 2-path is a path consisting of two edges:

**Lemma 1.15** (The number of 2-paths) *Assume that $\boldsymbol{d}$ satisfies* (1.3.44) *for some $c_F > 0$ and $\tau \in (2,3)$. For any graph $G$ whose degree sequence is $\boldsymbol{d}$, the number of 2-paths starting from any specified vertex is $O(n^{(2\tau-3)/(\tau-1)^2}) = o(n)$.*

*Proof* Without loss of generality we may assume that the degrees are ordered from large to small as $d_1 \ge d_2 \ge \cdots \ge d_n$. Then, for every $v \in [n]$, the number of vertices with degree at least $d_v$ is at least $v$. By (1.3.44), for every $v \in [n]$,

$$c_F n (d_v - 1)^{1-\tau} \ge n[1-F_n](d_v - 1) \ge v. \tag{1.3.48}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Thus, $d_v \le (c_F n/v)^{1/(\tau-1)} + 1$. The number of 2-paths from any vertex is bounded by $\sum_{v=1}^{d_1} d_v$, which is at most

$$\sum_{v=1}^{d_1} \left( \left( \frac{c_F n}{v} \right)^{1/(\tau-1)} + 1 \right) = (c_F n)^{1/(\tau-1)} \sum_{v=1}^{d_1} v^{-1/(\tau-1)} + d_1 \tag{1.3.49}$$
$$= O\left( n^{1/(\tau-1)} \right) d_1^{(\tau-2)/(\tau-1)} = O\left( n^{(2\tau-3)/(\tau-1)^2} \right),$$

since $d_1 \le (c_F n)^{1/(\tau-1)} + 1$. Since $\tau \in (2,3)$, the above is $o(n)$. □

*Proof of Theorem 1.13.* To compute the asymptotics of $\mathbb{P}(\{u,v\} \in E(\mathrm{UG}_n(\boldsymbol{d})) \mid \mathcal{E}_U)$, we switch between two classes of graphs, $\mathcal{S}$ and $\bar{\mathcal{S}}$. Class $\mathcal{S}$ consists of graphs where all edges in $\{u,v\} \cup U$ are present, whereas $\bar{\mathcal{S}}$ consists of all graphs where every $\{s,t\} \in U$ is present, but $\{u,v\}$ is not. Recall that $\mathcal{E}_U = \{\{s,t\} \in E(\mathrm{UG}_n(\boldsymbol{d})) \; \forall \{s,t\} \in U\}$ denotes the event that $\{s,t\}$ is an edge for every $\{s,t\} \in U$. Then, since the law on simple graphs is *uniform* (see also Exercise 1.18),

$$\mathbb{P}(\{u,v\} \in E(\mathrm{UG}_n(\boldsymbol{d})) \mid \mathcal{E}_U) = \frac{|\mathcal{S}|}{|\mathcal{S}| + |\bar{\mathcal{S}}|} = \frac{1}{1 + |\bar{\mathcal{S}}|/|\mathcal{S}|}, \tag{1.3.50}$$

and we are left to compute the asymptotics of $|\bar{\mathcal{S}}|/|\mathcal{S}|$.

For this, we define an operation called a *forward switching* that converts a graph in $G \in \mathcal{S}$ to a graph $G' \in \bar{\mathcal{S}}$. The reverse operation, converting $G'$ to $G$, is called a *backward switching*. Then we estimate $|\bar{\mathcal{S}}|/|\mathcal{S}|$ by counting the number of forward switchings that can be applied to the graph $G \in \mathcal{S}$, and the number of backward switchings that can be applied to the graph $G' \in \bar{\mathcal{S}}$. In our switching, we wish to have control on whether $\{u,v\}$ is present or not, so we tune it to take this restriction into account.

The forward switching on $G \in \mathcal{S}$ is defined by choosing two edges and specifying their ends as $\{x,a\}$ and $\{y,b\}$. We write this as *directed* edges $(x,a)$ since the roles of $x$ and $a$ are different, as indicated in Figure 1.7. We assume that $\mathcal{E}_U$ occurs. The choice must satisfy the following constraints:

(1) none of $\{u,x\}$, $\{v,y\}$, or $\{a,b\}$ is an edge in $G$;
(2) $\{x,a\}, \{y,b\} \notin U$;
(3) all of $u$, $v$, $x$, $y$, $a$, and $b$ must be distinct except that $x = y$ is permitted.

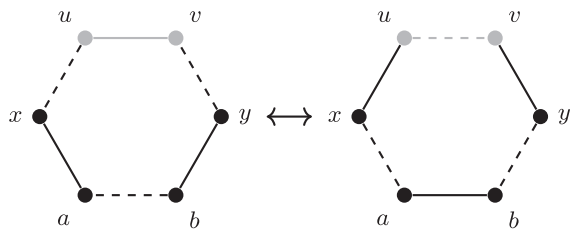


**Figure 1.7** Forward and backward switchings. The edge $\{u,v\}$ is present on the left, but not on the right.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Given a valid choice, forward switching replaces the three edges $\{u, v\}$, $\{x, a\}$, and $\{y, b\}$ by $\{u, x\}$, $\{v, y\}$, and $\{a, b\}$, while ensuring that the graph after switching is simple. Note that forward switching preserves the degree sequence, and converts a graph in $\mathcal{S}$ to a graph in $\bar{\mathcal{S}}$. See Figure 1.7 for an illustration of both the forward and backward switchings.

Next, we estimate the number of ways to perform a forward switching to a graph $G$ in $\mathcal{S}$, denoted by $f(G)$, and the number of ways to perform a backward switching to a graph $G'$ in $\bar{\mathcal{S}}$, denoted by $b(G)$. The number of total switchings between $\mathcal{S}$ and $\bar{\mathcal{S}}$ is equal to (see Exercise 1.19)

$$|\mathcal{S}|\mathbb{E}[f(G)] = |\bar{\mathcal{S}}|\mathbb{E}[b(G')], \tag{1.3.51}$$

where the expectation is over a uniformly random $G \in \mathcal{S}$ on the left-hand side, and over a uniformly random $G' \in \bar{\mathcal{S}}$ on the right-hand side, respectively. Consequently,

$$\frac{|\bar{\mathcal{S}}|}{|\mathcal{S}|} = \frac{\mathbb{E}[f(G)]}{\mathbb{E}[b(G')]}. \tag{1.3.52}$$

We next compute each of these factors.

### The Number of Forward Switchings: Computing $\mathbb{E}[f(G)]$

Given an arbitrary graph $G \in \mathcal{S}$, the number of ways to carry out a forward switching is at most $\ell_n^2$, since there are at most $\ell_n$ ways to choose $(x, a)$, and at most $\ell_n$ ways to choose $(y, b)$. Note that choosing $(x, a)$ for the first directed edge and $(y, b)$ for the second directed edge results in a different switch from vice versa.

To find a lower bound on the number of ways of performing a forward switching, we subtract from $\ell_n^2$ an upper bound on the number of invalid choices for $(x, a)$ and $(y, b)$. Such invalid choices can be categorized as follows:

(a) at least one of $\{u, x\}, \{a, b\}, \{v, y\}$ is an edge in $G$;
(b) at least one of $\{x, a\}$ or $\{y, b\}$ is in $U$;
(c) any vertex overlap other than $x = y$ (i.e., if one of $a$ or $b$ is equal to one of $x$ or $y$, or if $a = b$, or if one of $u$ or $v$ is one of $\{a, b, x, y\}$).

We now bound all these different categories of invalid choices. To find an upper bound for (a), note that any choice in case (a) must involve a single edge, and a 2-path starting from a specified vertex. By Lemma 1.15, the number of choices for (a) is then upper bounded by $3 \times o(\ell_n) \times \ell_n = o(\ell_n^2)$ (noting that $n = \Theta(\ell_n)$). The number of choices for case (b) is $O(\ell_n)$, as $|U| = O(1)$, and there are at most $\ell_n$ ways to choose the other directed edge, which is not restricted to be in $U$.

To bound the number of choices for (c), we investigate each case:

(c1) Either $a$ or $b$ is equal to $x$ or $y$, or $a = b$. In this case, $x, y, a, b$ forms a 2-path in $G$. Thus, there are at most $5 \times n \times o(\ell_n) = o(\ell_n^2)$ choices (noting that $n = \Theta(\ell_n)$), where $n$ is the number of ways to choose a vertex, and $o(\ell_n)$ bounds the number of 2-paths starting from this specified vertex.
(c2) One of $u$ and $v$ is one of $\{a, b, x, y\}$. In this case, there is one 2-path starting from $u$ or $v$, and a single edge. Thus, there are at most $8 \times \ell_n \times d_{\max} = o(\ell_n^2)$ choices, where $d_{\max} = \max_{v \in [n]} d_v$ bounds the number of ways to choose a vertex adjacent to $u$ or $v$ and $\ell_n$ bounds the number of ways to choose a single edge, by Lemma 1.15.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Thus, the number of invalid choices for $(x, a)$ and $(y, b)$ is $o(\ell_n^2)$, so that the number of forward switchings which can be applied to any $G \in \mathcal{S}$ is $(1 + o(1))\ell_n^2$. We conclude that

$$\mathbb{E}[f(G)] = (1 + o(1))\ell_n^2. \tag{1.3.53}$$

### The Number of Backward Switchings: Computing $\mathbb{E}[b(G')]$

Given a graph $G' \in \bar{\mathcal{S}}$, consider the backward switchings that can be applied to $G'$. There are at most $\ell_n(d_u - |U_u|)(d_v - |U_v|)$ ways to do the backward switching, since we are choosing an edge that is adjacent to $u$ but not in $U$, an edge that is adjacent to $v$ but not in $U$, and another directed edge $(a, b)$. For a lower bound, we consider the following forbidden choices:

(a′) at least one of $\{x, a\}$ or $\{y, b\}$ is an edge;

(b′) $\{a, b\} \in U$;

(c′) any vertices overlapping other than $x = y$ (i.e., when $\{a, b\} \cap \{u, v, x, y\} \neq \varnothing$).

We now go through each of these forbidden cases.

For (a′), suppose that $\{x, a\}$ is present, giving the 2-path $\{x, a\}, \{a, b\}$ in $G'$. There are at most $(d_u - |U_u|)(d_v - |U_v|)$ ways to choose $x$ and $y$. Given any choice for $x$ and $y$, by Lemma 1.15, there are at most $o(\ell_n)$ ways to choose a 2-path starting from $x$ in $G'$, and hence $o(\ell_n)$ ways to choose $a, b$. Thus, the total number of choices is at most $o((d_u - |U_u|)(d_v - |U_v|)\ell_n)$. The case where $\{y, b\}$ is an edge is symmetric.

For (b′), there are $O(1)$ choices for choosing $\{a, b\}$ since $|U| = O(1)$, and at most $(d_u - |U_u|)(d_v - |U_v|)$ choices for $x$ and $y$. Thus, the number of choices for case (b′) is $O((d_u - |U_u|)(d_v - |U_v|)) = o((d_u - |U_u|)(d_v - |U_v|)\ell_n)$.

For (c′), the case where $a$ or $b$ is equal to $x$ or $y$ corresponds to a 2-path starting from $u$ or $v$ together with a single edge from $u$ or $v$. Since $o(\ell_n)$ bounds the number of 2-paths starting from $u$ or $v$ and $d_u - |U_u| + d_v - |U_v|$ bounds the number of ways to choose the single edge, there are $o(\ell_n(d_v - |U_v|)) + o(\ell_n(d_u - |U_u|))$ total choices. If $a$ or $b$ is equal to $u$ or $v$, there are $(d_u - |U_u|)(d_v - |U_v|)$ ways to choose $x$ and $y$, and at most $d_u + d_v$ ways to choose the last vertex as a neighbor of $u$ or $v$. Thus, there are $O((d_u - |U_u|)(d_v - |U_v|)d_{\max}) = o((d_u - |U_u|)(d_v - |U_v|)\ell_n)$ total choices, since $d_{\max} = O(n^{1/(\tau-1)}) = o(n) = o(\ell_n)$.

We conclude that the number of backward switchings that can be applied to any graph $G' \in \mathcal{S}'$ is $(d_u - |U_u|)(d_v - |U_v|)\ell_n(1 + o(1))$, so that

$$\mathbb{E}[b(G')] = (d_u - |U_u|)(d_v - |U_v|)\ell_n(1 + o(1)). \tag{1.3.54}$$

### Conclusion

Combining (1.3.52), (1.3.53), and (1.3.54) results in

$$|\bar{\mathcal{S}}|/|\mathcal{S}| = (1 + o(1))\frac{\ell_n^2}{(d_u - |U_u|)(d_v - |U_v|)\ell_n}, \tag{1.3.55}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

and thus (1.3.50) yields

$$\mathbb{P}(\{u,v\} \in E(\mathrm{UG}_n(\boldsymbol{d})) \mid \mathcal{E}_U) = \frac{1}{1+|\bar{\mathcal{S}}|/|\mathcal{S}|} = (1+o(1))\frac{(d_u-|U_u|)(d_v-|U_v|)}{\ell_n+(d_u-|U_u|)(d_v-|U_v|)}. \quad (1.3.56)$$

□

**Remark 1.16** (Uniform random graphs and configuration models) Owing to the close links between uniform random graphs with prescribed degrees and configuration models, we treat the two models together, in Chapters 4 and 7. ◀

### 1.3.5 Preferential Attachment Models

Most networks grow in time. Preferential attachment models describe growing networks, where the numbers of edges and vertices grow with time. Here we give a brief introduction. The model that we investigate produces a *graph sequence* denoted by $(\mathrm{PA}_n^{(m,\delta)}(a))_{n\geq 1}$ and which, for every time $n$, yields a graph of $n$ vertices and $mn$ edges for some $m = 1, 2, \ldots$ This model is denoted by $(\mathrm{PA}_n^{(m,\delta)}(a))_{n\geq 1}$ in [V1, Chapter 8]. Below, we define $(\mathrm{PA}_n^{(m,\delta)}(b))_{n\geq 1}$ and $(\mathrm{PA}_n^{(m,\delta)}(d))_{n\geq 1}$, which are variations of this model.

We start by defining the model for $m = 1$, for which the graph consists of a collection of trees. In this case, $\mathrm{PA}_1^{(1,\delta)}(a)$ consists of a single vertex with a single self-loop. We denote the vertices of $\mathrm{PA}_n^{(1,\delta)}(a)$ by $v_1^{(1)}, \ldots, v_n^{(1)}$. We denote the degree of vertex $v_i^{(1)}$ in $\mathrm{PA}_n^{(1,\delta)}(a)$ by $D_i(n)$, where, by convention, a self-loop increases the degree by 2.

We next describe the evolution of the graph. Conditional on $\mathrm{PA}_n^{(1,\delta)}(a)$, the growth rule to obtain $\mathrm{PA}_{n+1}^{(1,\delta)}(a)$ is as follows. Add a single vertex $v_{n+1}^{(1)}$ having a single edge. This edge is connected to a second vertex (including itself), according to the probabilities

$$\mathbb{P}\big(v_{n+1}^{(1)} \to v_i^{(1)} \mid \mathrm{PA}_n^{(1,\delta)}(a)\big) = \begin{cases} \dfrac{1+\delta}{n(2+\delta)+(1+\delta)} & \text{for } i = n+1, \\ \dfrac{D_i(n)+\delta}{n(2+\delta)+(1+\delta)} & \text{for } i \in [n]. \end{cases} \quad (1.3.57)$$

This preferential attachment mechanism is called *affine*, since the attachment probabilities in (1.3.57) depend in an affine way on the degrees of the random graph $\mathrm{PA}_n^{(1,\delta)}(a)$.

The model with $m > 1$ is defined in terms of the model for $m = 1$ as follows. Fix $\delta \geq -m$. We start with $\mathrm{PA}_{mn}^{(1,\delta/m)}(a)$, and denote the vertices in $\mathrm{PA}_{mn}^{(1,\delta/m)}(a)$ by $v_1^{(1)}, \ldots, v_{mn}^{(1)}$. Then we identify or collapse the $m$ vertices $v_1^{(1)}, \ldots, v_m^{(1)}$ in $\mathrm{PA}_{mn}^{(1,\delta/m)}(a)$ to become vertex $v_1^{(m)}$ in $\mathrm{PA}_n^{(m,\delta)}(a)$. In doing so, we let all the edges that are incident to any of the vertices in $v_1^{(1)}, \ldots, v_m^{(1)}$ be incident to the new vertex $v_1^{(m)}$ in $\mathrm{PA}_n^{(m,\delta)}(a)$. Then, we collapse the $m$ vertices $v_{m+1}^{(1)}, \ldots, v_{2m}^{(1)}$ in $\mathrm{PA}_{mn}^{(1,\delta/m)}(a)$ to become vertex $v_2^{(m)}$ in $\mathrm{PA}_n^{(m,\delta)}(a)$, etc. More generally, we collapse the $m$ vertices $v_{(j-1)m+1}^{(1)}, \ldots, v_{jm}^{(1)}$ in $\mathrm{PA}_{mn}^{(1,\delta/m)}(a)$ to become vertex $v_j^{(m)}$ in $\mathrm{PA}_n^{(m,\delta)}(a)$. This defines the model for general $m \geq 1$.

The resulting graph $\mathrm{PA}_n^{(m,\delta)}(a)$ is a multi-graph with precisely $n$ vertices and $mn$ edges, so that the total degree is equal to $2mn$. The model with $\delta = 0$ is sometimes called the *proportional* model. The inclusion of the extra parameter $\delta > -m$ is relevant, though, as

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

we will see later. It can be useful to think of edges and vertices as carrying *weights*, where a vertex has weight $\delta$ and an edge has weight 1. Then, the vertex $v_{n+1}^{(1)}$ attaches its edges with a probability proportional to the weight of the vertex plus the edges to which it is incident. This, for example, explains why $\mathrm{PA}_{mn}^{(1,\delta/m)}(a)$ needs to be used in the collapsing procedure, rather than $\mathrm{PA}_{mn}^{(1,\delta)}(a)$.

The preferential attachment model $(\mathrm{PA}_n^{(m,\delta)}(a))_{n\geq 1}$ is increasing in time, in the sense that vertices and edges, once they have appeared, remain there forever. Thus, the degrees are monotonically increasing in time. Moreover, vertices with a high degree have a higher chance of attracting further edges of later vertices. Therefore, the model is sometimes called the *rich-get-richer model*. It is not hard to see that $D_i(n) \xrightarrow{a.s.} \infty$ for each fixed $i \geq 1$, as $n \to \infty$ (see Exercise 1.20). As a result, one could also call the preferential attachment model the *old-get-richer model*.

### Degrees of Fixed Vertices

We start by investigating the degrees of fixed vertices as $n \to \infty$, i.e., we will study $D_i(n)$ for fixed $i$ as $n \to \infty$. To formulate our results, we define the *Gamma function* $t \mapsto \Gamma(t)$ for $t > 0$ by

$$\Gamma(t) = \int_0^\infty x^{t-1} \mathrm{e}^{-x} \mathrm{d}x. \tag{1.3.58}$$

The following theorem describes the evolution of the degree of fixed vertices:

**Theorem 1.17** (Degrees of fixed vertices) *Consider* $\mathrm{PA}_n^{(m,\delta)}(a)$ *with* $m \geq 1$ *and* $\delta > -m$. *Then,* $D_i(n)/n^{1/(2+\delta/m)}$ *converges almost surely to a random variable* $\xi_i$ *as* $n \to \infty$.

*Proof* This is to be found in [V1, Theorem 8.2 and (8.3.11)]. □

It turns out that also $n^{-1/(2+\delta/m)} \max_{v\in[n]} D_v(n) \xrightarrow{a.s.} M$ for some limiting positive and finite random variable $M$ (see [V1, Section 8.7]). In analogy to iid random variables, this fact suggests that the degree of a random vertex satisfies a power law with power-law exponent $\tau = 3 + \delta/m$, and this is our next item on the agenda.

### Degree Sequence of the Preferential Attachment Model

We write

$$P_k(n) = \frac{1}{n} \sum_{i\in[n]} \mathbb{1}_{\{D_i(n)=k\}} \tag{1.3.59}$$

for the (random) proportion of vertices with degree $k$ at time $n$. For $m \geq 1$ and $\delta > -m$, we define $(p_k)_{k\geq 0}$ by $p_k = 0$ for $k = 0, \ldots, m-1$ and, for $k \geq m$,

$$p_k = (2 + \delta/m) \frac{\Gamma(k+\delta)\Gamma(m+2+\delta+\delta/m)}{\Gamma(m+\delta)\Gamma(k+3+\delta+\delta/m)}. \tag{1.3.60}$$

It turns out that $(p_k)_{k\geq 0}$ is a probability mass function (see [V1, Section 8.4]). It arises as the limiting degree distribution for $\mathrm{PA}_n^{(m,\delta)}(a)$, as shown in the following theorem:

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

**Theorem 1.18** (Degree sequence in preferential attachment model) *Consider* $\mathrm{PA}_n^{(m,\delta)}(a)$ *with* $m \geq 1$ *and* $\delta > -m$. *There exists a constant* $C = C(m,\delta) > 0$ *such that, as* $n \to \infty$,

$$\mathbb{P}\Big( \max_k |P_k(n) - p_k| \geq C\sqrt{\frac{\log n}{n}} \Big) = o(1). \tag{1.3.61}$$

*Proof* See [V1, Theorem 8.3]. □

We next investigate the scale-free properties of $(p_k)_{k\geq 0}$ by investigating the asymptotics of $p_k$ for $k$ large. By (1.3.60) and Stirling's formula, as $k \to \infty$ we have

$$p_k = c_{m,\delta} k^{-\tau}(1 + O(1/k)), \tag{1.3.62}$$

where

$$\tau = 3 + \delta/m > 2, \qquad \text{and} \qquad c_{m,\delta} = (2 + \delta/m)\frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m+\delta)}. \tag{1.3.63}$$

Therefore, by Theorem 1.18 and (1.3.62), the asymptotic degree sequence of $\mathrm{PA}_n^{(m,\delta)}(a)$ is close to a power law with exponent $\tau = 3 + \delta/m$. We note that any exponent $\tau > 2$ can be obtained by choosing $\delta > -m$ and $m \geq 1$ appropriately.

### Related Preferential Attachment Rules

In this book, we also sometimes investigate the related $(\mathrm{PA}_n^{(m,\delta)}(b))_{n\geq 1}$ model, in which self-loops for $m = 1$ in (1.3.57) are not allowed, so that

$$\mathbb{P}\big(v_{n+1}^{(1)} \to v_i^{(1)} \mid \mathrm{PA}_n^{(1,\delta)}(b)\big) = \frac{D_i(n) + \delta}{n(2+\delta)} \qquad \text{for } i \in [n]. \tag{1.3.64}$$

For $m = 1$, this model starts with two vertices and two edges in between, so that at time $n$, there are precisely $n$ edges. The model for $m \geq 2$ is again defined in terms of the model $(\mathrm{PA}_{nm}^{(1,\delta/m)}(b))_{n\geq 1}$ for $m = 1$ by collapsing blocks of $m$ vertices, so that $\mathrm{PA}_n^{(m,\delta)}(b)$ has $n$ vertices and $mn$ edges. The advantage of $(\mathrm{PA}_n^{(m,\delta)}(b))_{n\geq 1}$ compared to $(\mathrm{PA}_n^{(m,\delta)}(a))_{n\geq 1}$ is that $(\mathrm{PA}_n^{(m,\delta)}(b))_{n\geq 1}$ is naturally *connected*, while $(\mathrm{PA}_n^{(m,\delta)}(a))_{n\geq 1}$ may not be. Note that $\mathrm{PA}_n^{(m,\delta)}(b)$ can contain self-loops when $m \geq 2$, due to the collapsing procedure.

(Model $(\mathrm{PA}_n^{(m,\delta)}(c))_{n\geq 1}$, as formulated in [V1, Section 8.3], is defined by connecting edges with probability $\alpha$ to a uniformly chosen vertex and with probability $1-\alpha$ to a vertex chosen proportionally to its degrees. It turns out to be equivalent to $(\mathrm{PA}_n^{(m,\delta)}(a))_{n\geq 1}$. We will not discuss this model further here.)

Another adaptation of the preferential attachment rule arises when no self-loops are ever allowed, while the degrees are updated when the $m$ edges incident to the new vertex are being attached. We denote this model by $(\mathrm{PA}_n^{(m,\delta)}(d))_{n\geq 1}$. In this case, the model starts at time $n = 2$ with two vertices labeled 1 and 2, and $m$ edges between them. $\mathrm{PA}_n^{(m,\delta)}(d)$ has vertex set $[n]$, and $m(n-1)$ edges. At time $n+1$, for $m \geq 1$ and $j \in \{0,\dots,m-1\}$, we attach the $(j+1)$th edge of vertex $v_{n+1}^{(m)}$ to vertex $v_i^{(m)}$ with probability

$$\mathbb{P}\big(v_{n+1,j+1}^{(m)} \to v_i^{(m)} \mid \mathrm{PA}_{n,j}^{(m,\delta)}(d)\big) = \frac{D_i(n,j) + \delta}{n(2m+\delta)} \qquad \text{for } i \in [n]. \tag{1.3.65}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Here, $D_i(n,j)$ is the degree of vertex $v_i^{(m)}$ after the connection of the edges incident to the first $n+1$ vertices, as well as the first $j$ edges incident to vertex $v_{n+1}^{(m)}$, and $\mathrm{PA}_{n,j}^{(m,\delta)}(d)$ is the graph of the first $n$ vertices, as well as the first $j$ edges incident to vertex $v_{n+1}^{(m)}$. The model is by default connected, and at time $n$ consists of $n+1$ vertices and $mn$ edges. For $m=1$, apart from the different starting graphs, models (b) and (d) are identical. Indeed, $\mathrm{PA}_n^{(1,\delta)}(b)$ for $n=2$ consist of two vertices with two edges between them, while $\mathrm{PA}_n^{(1,\delta)}(d)$ for $n=2$ consists of two vertices with one edge between them for $\mathrm{PA}_n^{(1,\delta)}(d)$.

Many other adaptations are possible and have been investigated in the literature, such as settings where the $m$ edges incident to $v_{n+1}^{(m)}$ are *independently* connected as in (1.3.65) when $j=0$. We refrain from discussing these. It is not hard to verify that Theorem 1.18 holds for all these adaptations, which explains why authors have often opted for the version of the model that is most convenient for them. From the perspective of local convergence, it turns out that $(\mathrm{PA}_n^{(m,\delta)}(d))_{n\geq 1}$ is the most convenient, as we will see in Chapter 5. On the other hand, Theorem 1.17 contains minor adaptations between models, particularly since the limiting random variables $(\xi_i)_{i\geq 1}$ *do* depend on the precise model.

### Bernoulli Preferential Attachment Model

We finally discuss a model that is quite a bit different from the other preferential attachment models discussed above. The main difference is that in this model, the number of edges is *not* fixed, but instead there is *conditional independence* in the edge attachments. We call this model the *Bernoulli preferential attachment model*, as the attachment indicators are all conditionally independent Bernoulli variables. Let us now give the details.

Fix a preferential attachment function $f\colon \mathbb{N}_0 \mapsto (0,\infty)$. Then, the graph evolves as follows. We start with a graph $\mathrm{BPA}_1^{(f)}$ containing one vertex $v_1$ and no edges. At each time $n\geq 2$, we add a vertex $v_n$. Conditional on $\mathrm{BPA}_{n-1}^{(f)}$, and independently for every $v\in[n-1]$, we connect this vertex to $v$ by a directed edge with probability

$$\frac{f(D_v^{(\mathrm{in})}(n-1))}{n-1}, \tag{1.3.66}$$

where $D_v^{(\mathrm{in})}(n-1)$ is the in-degree of vertex $v$ at time $n-1$. This creates the random graph $\mathrm{BPA}_n^{(f)}$. Note that the number of edges in the random graph process $(\mathrm{BPA}_n^{(f)})_{n\geq 1}$ is a random variable, and thus *not* fixed. In particular, it makes a difference whether we use the in-degree in (1.3.66) or the total degree.

We consider functions $f\colon \mathbb{N}\mapsto(0,\infty)$ that satisfy that $f(k+1)-f(k)<1$ for every $k\geq 0$. Under this assumption and when $f(0)\leq 1$, the empirical degree sequence converges as $n\to\infty$, i.e.,

$$P_k(n)\equiv \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{D_i(n)=k\}} \xrightarrow{\mathbb{P}} p_k, \quad \text{where} \quad p_k=\frac{1}{1+f(k)}\prod_{l=0}^{k-1}\frac{f(l)}{1+f(l)}. \tag{1.3.67}$$

In particular, $\log(1/p_k)/\log k \to 1+1/\gamma$ when $f(k)/k\to\gamma\in(0,1)$ (see Exercise 1.23). Remarkably, when $f(k)=\gamma k+\beta$, the power-law exponent of the degree distribution does not depend on $\beta$. The restriction that $f(k+1)-f(k)<1$ is needed to prevent the degrees

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

from exploding. Further, $\log(1/p_k) \sim k^{1-\alpha}/(\gamma(1-\alpha))$ when $f(k) \sim \gamma k^{\alpha}$ for some $\alpha \in (0,1)$ (see Exercise 1.24). Interestingly, there exists a persistent hub, i.e., a vertex that has maximal degree for all but finitely many times, when $\sum_{k\geq 1} 1/f(k)^2 < \infty$. When $\sum_{k\geq 1} 1/f(k)^2 = \infty$, this does not happen.

### 1.3.6 Universality of random graphs

There are many other graph topologies where one can expect results similar to those in the random graphs discussed above. We will discuss many related models in Chapter 9, where we include several settings that are relevant in practice, such as *directed graphs* as well as graphs with *community structure* and *geometry*. The random graph models that we investigate are *inhomogeneous*, and one can expect that the results depend sensitively on the amount of inhomogeneity present. This is reflected in the results that we prove, where the precise asymptotics is different when the vertices have heavy-tailed rather than light-tailed degrees. However, interestingly, what is "heavy tailed" and what is "light tailed" depends on the precise model and setting at hand. Often, as we will see, the distinction depends on how many moments the degree distribution has.

We have proposed many random graph models for real-world networks. Since these models aim at describing similar real-world networks, one would hope that they also give similar answers. Indeed, for a real-world network with a power-law degree sequence, we could model its static structure by the configuration model with the same degree sequence, and its dynamical properties by the preferential attachment model with similar scale-free degrees. How do we interpret implications to the real world when these attempts give completely different predictions?

*Universality* is the phrase physicists use when different models display similar behavior. Models that show similar behavior are then in the same *universality class*. Enormous effort has gone, and is currently going, into deciding whether various random graph models are in the same universality class, or rather in different ones, and why. We will see that the *degree distribution* decides the universality class for a wide range of models, as one might possibly hope. This also explains why the degree distribution plays such a dominant role in the investigation of random graphs. See Chapter 9 for more details.

## 1.4 Power Laws and Their Properties

In this book, we frequently deal with random variables having an (asymptotic) power-law distribution. For such random variables, we often need to investigate *truncated moments*, and we also often deal with their *sized-biased distribution*. In this section, we collect some results concerning power-law random variables. We start by recalling the definition of a power-law distribution:

**Definition 1.19** (Power-law distributions) We say that $X$ has a *power-law distribution* with exponent $\tau$ when there exists a function $x \mapsto L(x)$ that is slowly varying at infinity such that

$$1 - F_X(x) = \mathbb{P}(X > x) = L(x)x^{-(\tau-1)}. \tag{1.4.1}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Here, we recall that a function $x \mapsto L(x)$ is *slowly varying at infinity* when, for every $t > 0$,

$$\lim_{x\to\infty} \frac{L(xt)}{L(x)} = 1. \tag{1.4.2}$$

◀

A crucial result about slowly varying functions is *Potter's Theorem*, which we next recall:

**Theorem 1.20** (Potter's Theorem) *Let $x \mapsto L(x)$ be slowly varying at infinity. For every $\delta$, there exists a constant $C_\delta \geq 1$ such that, for all $x \geq 1$,*

$$x^{-\delta}/C_\delta \leq L(x) \leq C_\delta x^\delta. \tag{1.4.3}$$

Theorem 1.20 implies that the tail of any general power-law distribution, as in Definition 1.19, can be bounded above and below by that of a *pure* power-law distribution (i.e., one without a slowly varying function) with a slightly adapted power-law exponent. As a result, we can often deal with pure power laws instead.

We continue by studying the relation between power-law tails of the empirical degree distribution and bounds on the degrees themselves:

**Lemma 1.21** (Tail and degree bounds) *Let $\boldsymbol{d} = (d_v)_{v\in[n]}$ be a degree distribution, $d_{(1)} \geq d_{(2)} \geq \cdots \geq d_{(n-1)} \geq d_{(n)}$ its non-increasing ordered version, and*

$$F_n(x) = \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{d_v\leq x\}} \tag{1.4.4}$$

*its empirical distribution. Then*

$$[1 - F_n](x) \leq c_F x^{-(\tau-1)} \qquad \forall x \geq 1 \tag{1.4.5}$$

*implies that*

$$d_{(v)} \leq (c_F n/v)^{1/(\tau-1)} + 1 \qquad \forall v \in [n], \tag{1.4.6}$$

*while*

$$d_{(v)} \leq (c_F n/v)^{1/(\tau-1)} \qquad \forall v \in [n] \tag{1.4.7}$$

*implies that*

$$[1 - F_n](x) \leq c_F x^{-(\tau-1)} \qquad \forall x \geq 1. \tag{1.4.8}$$

*Proof* Assume first that (1.4.5) holds. For every $v \in [n]$, the number of vertices with degree at least $d_{(v)}$ is at least $v$. By (1.4.5), for every $v \in [n]$,

$$c_F n(d_{(v)} - 1)^{1-\tau} \geq n[1 - F_n](d_{(v)} - 1) \geq v. \tag{1.4.9}$$

Thus, $d_{(v)} \leq (c_F n/v)^{1/(\tau-1)} + 1$, as required.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Next, assume that (1.4.7) holds. Then

$$\begin{aligned}[1-F_n](x) &= \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{d_v>x\}} = \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{d_{(v)}>x\}} \\ &\le \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{(c_F n/v)^{1/(\tau-1)}>x\}} \\ &= \frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{v<nc_F x^{-(\tau-1)}\}} \le c_F x^{-(\tau-1)}, \end{aligned} \tag{1.4.10}$$

as required. □

We next study truncated moments of random variables whose tail is bounded by that of a power law:

**Lemma 1.22** (Truncated moments) *Let $X$ be a non-negative random variable whose distribution function $F_X(x) = \mathbb{P}(X \le x)$ satisfies, for every $x \ge 1$,*

$$1 - F_X(x) \le C_X x^{-(\tau-1)}. \tag{1.4.11}$$

*Then, for all $a < \tau - 1$, there exists a constant $C_X(a)$ such that, for all $\ell \ge 1$,*

$$\mathbb{E}[X^a \mathbb{1}_{\{X>\ell\}}] \le C_X(a)\ell^{a-(\tau-1)}, \tag{1.4.12}$$

*while, for $a > \tau - 1$ and all $\ell \ge 1$,*

$$\mathbb{E}[X^a \mathbb{1}_{\{X\le\ell\}}] \le C_X(a)\ell^{a-(\tau-1)}. \tag{1.4.13}$$

*Proof* We note that for any cumulative distribution function $x \mapsto F_X(x)$ on the non-negative reals, we have a *partial integration identity*, stating that, for every $f\colon \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}\int_u^\infty f(x)F_X(\mathrm{d}x) &= f(u)[1-F_X(u)] + \int_u^\infty [f(x)-f(u)]F_X(\mathrm{d}x) \\ &= f(u)[1-F_X(u)] + \int_u^\infty \int_u^x f'(y)\mathrm{d}y F_X(\mathrm{d}x) \\ &= f(u)[1-F_X(u)] + \int_u^\infty f'(y)\int_y^\infty F_X(\mathrm{d}x)\mathrm{d}y \\ &= f(u)[1-F_X(u)] + \int_u^\infty f'(y)[1-F_X(y)]\mathrm{d}y, \end{aligned} \tag{1.4.14}$$

provided that either (a) $y \mapsto f'(y)[1-F_X(y)]$ is absolutely integrable, or (b) $x \mapsto f(x)$ is either non-decreasing or non-increasing, so that $f'(y)[1-F_X(y)]$ has a fixed sign. Here, the interchange of the integration order is allowed by Fubini's Theorem for non-negative functions (Halmos, 1950, Section 36, Theorem B) when $x \mapsto f(x)$ is non-decreasing, and by Fubini's Theorem for absolutely–integrable functions (Halmos, 1950, Section 36, Theorem C) when $y \mapsto f'(y)[1-F_X(y)]$ is absolutely integrable. Similarly, for $f$ with $f(0) = 0$,

$$\begin{aligned}\int_0^u f(x)F_X(\mathrm{d}x) &= \int_0^u \int_0^x f'(y)\mathrm{d}y F_X(\mathrm{d}x) = \int_0^u f'(y)\int_y^u F_X(\mathrm{d}x)\mathrm{d}y \\ &= \int_0^u f'(y)[F_X(u)-F_X(y)]\mathrm{d}y. \end{aligned} \tag{1.4.15}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

When $X \geq 0$, using (1.4.11) and (1.4.14), for $a < \tau - 1$ and $\ell > 0$,

$$\mathbb{E}\left[X^a \mathbb{1}_{\{X>\ell\}}\right] = \ell^a \mathbb{P}(X > \ell) + \int_\ell^\infty a x^{a-1} \mathbb{P}(X > x) \mathrm{d}x$$
$$\leq C_X \ell^{a-(\tau-1)} + a C_X \int_\ell^\infty x^{a-1} x^{-(\tau-1)} \mathrm{d}x \leq C_X(a) \ell^{a-(\tau-1)}. \quad (1.4.16)$$

as required. Further, for $a > \tau - 1$ and $\ell > 0$, now using (1.4.15),

$$\mathbb{E}\left[X^a \mathbb{1}_{\{X\leq\ell\}}\right] \leq a C_X \int_0^\ell x^{a-1} x^{-(\tau-1)} \mathrm{d}x \leq C_X(a) \ell^{a-(\tau-1)}. \quad (1.4.17)$$

□

An important notion in many graphs is the *size-biased* version $X^\star$ of a non-negative random variable $X$, which is given by

$$\mathbb{P}(X^\star \leq x) = \frac{\mathbb{E}[X \mathbb{1}_{\{X\leq x\}}]}{\mathbb{E}[X]}. \quad (1.4.18)$$

Exercise 1.25 shows that the size-biased distribution of the degree of a random vertex is the degree of a random vertex in a random edge. Let $F_x^\star$ denote the distribution function of $X^\star$. The following lemma gives bounds on the tail of the distribution function $F_x^\star$:

**Lemma 1.23** (Size-biased tail distribution) *Let $X$ be a non-negative random variable whose distribution function $F_x(x) = \mathbb{P}(X \leq x)$ satisfies that there exists a $C_x$ such that, for every $x \geq 1$,*

$$1 - F_x(x) \leq C_x x^{-(\tau-1)}. \quad (1.4.19)$$

*Assume that $\tau > 2$, so that $\mathbb{E}[X] < \infty$. Further, assume that $\mathbb{E}[X] > 0$. Then, there exists a constant $C_x^\star$ such that*

$$1 - F_x^\star(x) \leq C_x^\star x^{-(\tau-2)}. \quad (1.4.20)$$

*Proof* This follows immediately from (1.4.18), by using (1.4.12) with $a = 1$. □

## 1.5 Notation and Preliminaries

Let us introduce some standard notation used throughout this book, and recall some properties of trees and Poisson processes.

### Abbreviations

We write rhs for *right-hand side*, and lhs for *left-hand side*. Further, we abbreviate *with respect to* by wrt.

### Random variables

We write $X \stackrel{d}{=} Y$ to denote that $X$ and $Y$ have the same distribution. We write $X \sim \mathsf{Be}(p)$ when $X$ has a Bernoulli distribution with success probability $p$, i.e., $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$. We write $X \sim \mathsf{Bin}(n, p)$ when the random variable $X$ has a binomial

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

distribution with parameters $n$ and $p$, and we write $X \sim \mathsf{Poi}(\lambda)$ when $X$ has a Poisson distribution with parameter $\lambda$.

We write $X \sim \mathsf{Exp}(\lambda)$ when $X$ has an exponential distribution with mean $1/\lambda$. We write $X \sim \mathsf{Gam}(r, \lambda)$ when $X$ has a gamma distribution with scale parameter $\lambda$ and shape parameter $r$, for which the density, for $x \geq 0$, is given by

$$f_X(x) = \lambda^r x^{r-1} \mathrm{e}^{-\lambda x} / \Gamma(r), \tag{1.5.1}$$

where $r, \lambda > 0$, and we recall (1.3.58), while $f_X(x) = 0$ for $x < 0$. The random variable $\mathsf{Gam}(r, \lambda)$ has mean $r/\lambda$ and variance $r/\lambda^2$. Finally, we write $X \sim \mathsf{Beta}(\alpha, \beta)$ when $X$ has a beta distribution with parameters $\alpha, \beta > 0$, so that $X$ has density, for $x \in [0, 1]$,

$$f_X(x) = x^{\alpha-1}(1-x)^{\beta-1} / B(\alpha, \beta), \tag{1.5.2}$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{1.5.3}$$

is the Beta-function, while $f_X(x) = 0$ for $x \notin [0, 1]$. We sometimes abuse notation, and write e.g., $\mathbb{P}(\mathsf{Bin}(n, p) = k)$ to denote $\mathbb{P}(X = k)$ when $X \sim \mathsf{Bin}(n, p)$.

We call a sequence of random variables $(X_i)_{i \geq 1}$ *independent and identically distributed* (iid) when they are independent, and $X_i$ has the same distribution as $X_1$ for every $i \geq 1$. For a finite set $\mathcal{X}$, we say that $X \in \mathcal{X}$ is drawn *uniformly at random* (uar) when $X$ has the uniform distribution on $\mathcal{X}$.

### Convergence of Random Variables

We say that a sequence of events $(\mathcal{E}_n)_{n \geq 1}$ occurs *with high probability* (whp) when $\lim_{n \to \infty} \mathbb{P}(\mathcal{E}_n) = 1$.

For sequences of random variables $(X_n)_{n \geq 1}$, $X_n \xrightarrow{d} X$ denotes that $X_n$ converges in distribution to $X$, while $X_n \xrightarrow{\mathbb{P}} X$ denotes that $X_n$ converges in probability to $X$ and $X_n \xrightarrow{a.s.} X$ denotes that $X_n$ converges almost surely to $X$. We write that $X_n = O_{\mathbb{P}}(Y_n)$ when $|X_n|/Y_n$ is a tight sequence of random variables and $X_n = \Theta_{\mathbb{P}}(Y_n)$ when $X_n = O_{\mathbb{P}}(Y_n)$ and $Y_n = O_{\mathbb{P}}(X_n)$. Finally, we write that $X_n = o_{\mathbb{P}}(Y_n)$ when $X_n / Y_n \xrightarrow{\mathbb{P}} 0$.

### Stochastic Domination

We recall that a random variable $X$ is *stochastically dominated* by a random variable $Y$ when $F_X(x) = \mathbb{P}(X \leq x) \geq F_Y(x) = \mathbb{P}(Y \leq x)$ for every $x \in \mathbb{R}$. We write this as $X \preceq Y$. See [V1, Section 2.3] for more details on stochastic ordering.

### Two Useful Martingale Inequalities

Recall [V1, Section 2.5] for the definition of a martingale $(M_n)_{n \geq 0}$. We rely on Doob's martingale inequality, which for a martingale $(M_n)_{n \geq 0}$ states that

$$\mathbb{E}\Big[ \sup_{m \leq n} \big| M_m - \mathbb{E}[M_m] \big|^2 \Big] \leq \mathrm{Var}(M_n). \tag{1.5.4}$$

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

An immediate consequence is Kolmogorov's martingale inequality, which states that

$$\mathbb{P}(\sup_{m\le n} |M_m - \mathbb{E}[M_m]| \ge \varepsilon) \le \varepsilon^{-2} \operatorname{Var}(M_n). \tag{1.5.5}$$

### Densities in Inhomogeneous Poisson Processes

Let $\Pi$ be an inhomogeneous Poisson process with intensity measure $\Lambda\colon \mathcal{X} \to \mathbb{N}$, where $\mathcal{X} \subseteq [0,\infty)$. This means that the number of points $\Pi(\mathcal{A})$ in $\mathcal{A} \subseteq \mathcal{X}$ has a Poisson distribution with parameter $\int_{\mathcal{A}} \Lambda(\mathrm{d}x)$, and the number of points in disjoint sets are independent (see Last and Penrose (2018) for details on Poisson processes).

Let $\mathcal{A}$ be a bounded set. We wish to give a formula for the probability of the event that the points of $\Pi$ in $\mathcal{A}$ (of which there are $\Pi(\mathcal{A})$) are in $(a_1 + \mathrm{d}a_1, \dots, a_k + \mathrm{d}a_k)$, where we assume that $a_1 < a_2 < \cdots < a_k$. Thus, there is one point in $a_1 + \mathrm{d}a_1$, one in $a_2 + \mathrm{d}a_2$, etc. Note that this event is a subset of the event $\Pi(\mathcal{A}) = k$. Denote this event by $\mathcal{P}(a_1 + \mathrm{d}a_1, \dots, a_k + \mathrm{d}a_k)$. Assume that $x \mapsto \Lambda(x)$ is continuous almost everywhere. Then, for all measurable $\mathcal{A} \subseteq \mathcal{X}$ and all ordered elements $a_1, \dots, a_k \in \mathcal{A}$,

$$\mathbb{P}(\mathcal{P}(a_1 + \mathrm{d}a_1, \dots, a_k + \mathrm{d}a_k)) = \mathrm{e}^{-\int_{\mathcal{A}} \Lambda(\mathrm{d}x)} \prod_{i=1}^{k} \Lambda(a_i)\mathrm{d}a_i. \tag{1.5.6}$$

We refer to $\mathrm{e}^{-\int_{\mathcal{A}} \Lambda(\mathrm{d}x)}$ as the *no-further-point probability*, in that it ensures that $\Pi$ has precisely $k$ points in $\mathcal{A}$. We refer to Exercise 1.26 for a proof that (1.5.6) implies that $\Pi(\mathcal{A}) \sim \mathsf{Poi}(\int_{\mathcal{A}} \Lambda(\mathrm{d}x))$, and Exercise 1.27 for a proof that (1.5.6) implies that $\Pi(\mathcal{A})$ and $\Pi(\mathcal{B})$ are independent when $\mathcal{A}$ and $\mathcal{B}$ are disjoint.

### Ordered Trees and Their Exploration

In this book, *trees* play a central role, and it is important to be clear about exactly what we mean by a tree. Trees are rooted and ordered. A tree $\mathbf{t}$ has root $\varnothing$, vertex set $V(\mathbf{t})$ and edge set $E(\mathbf{t})$, and the vertex set will be given an ordering below.

It is convenient to think of a tree $\mathbf{t}$ with root $\varnothing$ as being labeled in the Ulam–Harris way, so that a vertex $v$ in generation $k$ has a label $\varnothing v_1 \cdots v_k$, where $v_i \in \mathbb{N}$. Naturally, there are some restrictions, in that if $\varnothing v_1 \cdots v_k \in V(\mathbf{t})$, then also $\varnothing v_1 \cdots v_{k-1} \in V(\mathbf{t})$, and $\varnothing v_1 \cdots (v_k - 1) \in V(\mathbf{t})$ when $v_k \ge 2$. We refer to [V1, Chapter 3] for details.

It will sometimes also be useful to explore trees in a breadth-first manner. This corresponds to the lexicographical ordering in the Ulam–Harris encoding of the tree. Ulam–Harris trees are also sometimes called *plane trees* (see, e.g., (Drmota, 2009, Chapter 1)). Let us now make the breadth-first ordering of the tree precise:

**Definition 1.24** (Breadth-first order on a tree) For $v \in V(\mathbf{t})$, let $|v|$ be its height. Thus $|v| = k$ when $v = \varnothing v_1 \cdots v_k$ and $|\varnothing| = 0$. Let $u, v \in V(\mathbf{t})$. Then $u < v$ when either $|u| < |v|$ or $|u| = |v|$ and $u = \varnothing u_1 \cdots u_k$ and $v = \varnothing v_1 \cdots v_k$ are such that $(u_1, \dots, u_k) < (v_1, \dots, v_k)$ in the lexicographic sense. ◀

We next explain the breadth-first exploration of $\mathbf{t}$:

**Definition 1.25** (Breadth-first exploration of a tree) For a tree $\mathbf{t}$ of size $|V(\mathbf{t})| = t$, we let $(a_k)_{k=0}^{t}$ be the elements of $V(\mathbf{t})$ ordered according to the breadth-first ordering of $\mathbf{t}$ (recall

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

Definition 1.24). For $i \geq 1$, let $x_i$ denote the number of children of vertex $a_i$. Thus, if $d_v$ denotes the degree of $v \in V(\mathbf{t})$ in the tree $\mathbf{t}$, we have $x_1 = d_{a_0} = d_\varnothing$ and $x_i = d_{a_i} - 1$ for $i \geq 2$. The recursion

$$s_i = s_{i-1} + x_i - 1 \quad \text{for } i \geq 1, \qquad \text{with} \qquad s_0 = 1, \tag{1.5.7}$$

describes the evolution of the number of unexplored vertices in the breadth-first exploration. For a finite tree $\mathbf{t}$ of size $|V(\mathbf{t})| = t$, thus $s_i > 0$ for $i \in \{0, \ldots, t-1\}$ while $s_t = 0$. ◀

The sequence $(x_i)_{i=1}^t$ gives an alternative encoding of the tree $\mathbf{t}$ that is often convenient. Indeed, by Exercise 1.28, the sequence $(x_i)_{i=1}^t$ is in one-to-one correspondence to $\mathbf{t}$.

### Unimodular Branching Process Trees

We next describe one type of random tree that occurs frequently in our analyses, the so-called *unimodular branching process tree*:

**Definition 1.26** (Unimodular branching process tree) Fix a probability distribution $(p_k)_{k\geq 1}$, where $p_k = \mathbb{P}(D = k)$ for some integer-valued random variable $D$. The *unimodular branching process tree* with root-offspring distribution $(p_k)_{k\geq 1}$ is the branching process where the root has offspring distribution $(p_k)_{k\geq 1}$, while all vertices in other generations have offspring distribution $p_k^\star$ given by

$$p_k^\star = \mathbb{P}(D^\star - 1 = k) = \frac{(k+1)}{\mathbb{E}[D]}\mathbb{P}(D = k+1), \tag{1.5.8}$$

where we recall that $D^\star$ denotes the size-biased version of $D$ in (1.4.18). ◀

It turns out that unimodular branching process trees arise as local limits of random graphs, seen from a uniform vertex. The distribution $(p_k)_{k\geq 1}$, where $p_k = \mathbb{P}(D = k)$, describes the *degree distribution* in the graph, while the law (1.5.8) is related to the degree distribution of other vertices that are close to a uniform vertex.

We let $\mathsf{BP}_{\leq r}$ denote the branching process up to and including generation $r$, and write $\mathsf{BP}_r = \mathsf{BP}_{\leq r} \setminus \mathsf{BP}_{\leq r-1}$ for the $r$th generation. It is convenient to think of the branching process tree, denoted as $\mathsf{BP}$, as being labeled in the Ulam–Harris way (recall Definitions 1.24 and 1.25), so that a vertex $v$ in generation $r$ has a label $\varnothing a_1 \cdots a_r$, where $a_i \in \mathbb{N}$. When applied to $\mathsf{BP}$, we denote this process by $(\mathsf{BP}(t))_{t\geq 1}$, where $\mathsf{BP}(t)$ consists of precisely $t+1$ vertices (with $\mathsf{BP}(0)$ equal to the root $\varnothing$).

## 1.6 Notes and Discussion for Chapter 1

Sections 1.1–1.3 are in the majority summaries of chapters in Volume 1, to which we refer for notes and discussion, so we restrict ourselves here to the exceptions.

### Notes on Sections 1.1 and 1.2

See Barabási (2002) and Watts (2003) for expository accounts of the discovery of network properties by Barabási, Watts, and co-authors. Newman et al. (2006) bundles together some of the original papers detailing the empirical findings of real-world networks and the network models invented for them. The introductory book by Newman (2010) lists many of the empirical properties of, and scientific methods for, networks.

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

See also Barabási (2016) for an online book giving an extensive background to the science of networks, and Coscia (2021) for an "atlas to the aspiring network scientist."

The discussion of the scale-free phenomenon in Section 1.1.2 has been substantially extended compared with [V1, Section 1.4.1]. Artico et al. (2020) considered another static definition of the degree distribution, based on that in preferential attachment models (which they call the degree distribution of the *de Solla Price model* in honor of Price (1965), who invented the model for citation networks; see also Section 9.1.1). This can be seen as an interpolation between the static approach of Broido and Clauset (2019) and the dynamic approach advocated by Barabási (2018). Artico et al. (2020) used maximum-likelihood techniques to argue that power-law network degree distributions are not rare, classifying almost 65% of the tested networks as having a power-law tail with at least 80% power. We further refer to Nair et al. (2022) for an extensive discussion on heavy-tailed phenomena.

## Notes on Section 1.3

*Erdős–Rényi random graph.* The seminal papers Erdős and Rényi (1959, 1960, 1961a,b) investigated a related model in which a collection of $m$ edges is chosen uar from the collection of $\binom{n}{2}$ possible edges. The model described here as the Erdős–Rényi random graph was actually *not* invented by Erdős and Rényi, but rather by Gilbert (1959). When adding a fixed number of edges, the proportion of edges is $2m/n(n-1) \approx 2m/n^2$, so we should think of $m \approx 2\lambda n$ for a fair comparison. Note that when we condition the total number of edges in the independent-edge model to be equal to $m$, the law of the Erdős–Rényi random graph is equal to the model where a collection of $m$ uniformly chosen edges is added, which explains the close relation between the two models. Owing to the concentration of the total number of edges, we can indeed roughly exchange the binomial model with $p = \lambda/n$ with the combinatorial model with $m = 2\lambda n$. The combinatorial model has the nice feature that it produces a uniform graph from the collection of all graphs with $m$ edges, and thus could serve as a *null model* for a real-world network in which only the number of edges is fixed. We can also view $\mathrm{ER}_n(\lambda/n)$ as percolation on the complete graph. Percolation is a paradigmatic model in statistical physics describing random failures in networks (see Grimmett (1999) for an extensive overview of percolation theory focussing on $\mathbb{Z}^d$).

*Inhomogeneous random graphs* were first proposed by Söderberg (2002, 2003a,c,b). A general formalism for inhomogeneous random graphs is described in the seminal work of Bollobás et al. (2007). The generalized random graph was first introduced by Britton et al. (2006). The random graph with prescribed expected degrees, or Chung–Lu model, was introduced, and studied intensively, by Chung and Lu (2002a,b, 2003, 2006a,b). The Poissonian random graph or Norros–Reittu model was introduced by Norros and Reittu (2006). The conditions under which these random graphs are asymptotically equivalent [V1, Sections 6.7 and 6.8] were derived by Janson (2010a). Condition 1.1 has been slightly modified compared with [V1, Condition 6.4], in that we now assume that $\mathbb{E}[W] \in (0, \infty)$, which excludes trivial cases where the graph is almost empty.

*Configuration model and uniform random graphs with prescribed degrees.* The configuration model was invented by Bollobás (1980) to study uniform random regular graphs (see also (Bollobás, 2001, Section 2.4)). The introduction was inspired by, and generalized the results in, the work of Bender and Canfield (1978). The original work allowed for a careful computation of the number of regular graphs, 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, as we will see below. The configuration model, as well as uniform random graphs with a prescribed degree sequence, were further studied in greater generality by Molloy and Reed (1995, 1998). This extension is quite relevant to us, as the scale-free nature of many real-world applications encourages us to investigate configuration models with power-law degree sequences. Condition 1.7 is a minor modification of [V1, Condition 7.8]. The terms "erased" and "repeated configuration model" were coined by Britton et al. (2006).

The degree-truncation argument for the configuration model is, to the best of our knowledge, novel. Switching algorithms, as discussed in Section 1.3.4 have a long history, dating back at least to McKay (1981), see also Erdős et al. (2022); Gao and Greenhill (2021); Gao and Wormald (2016); McKay and Wormald (1990), as well as McKay (2011) and the references therein for overviews. The literature on switch chains focusses on two key aspects: first, their rapid mixing (Erdős et al. (2022); Gao and Greenhill (2021), and various related papers, for which we refer to Erdős et al. (2022)), and, second, counting the number of simple graphs using switch chain arguments (as in Gao and Wormald (2016)), which is the approach that we take in this section. Rapid mixing means that the mixing time of the switch chain is bounded by an explicit power of the number of vertices (or number of edges, or both combined). The powers, however, tend to be large, and thus "rapid mixing" may not be rapid enough to give good guarantees when one is

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

trying to sample a uniform random graph of the degree distribution of some real-world network. Theorem 1.13 is adapted from Gao et al. (2020), where it was used to compute the number of triangles in uniform random graphs with power-law degree distributions having infinite variance. See also Janson (2020b) for a relation between the configuration model and uniform random graphs using switchings.

*Preferential attachment models* were first introduced in the context of complex networks by Barabási and Albert (1999). Bollobás et al. (2001) studied the model by Barabási and Albert (1999), and later many other papers followed on this, and related, models. Barabási and Albert (1999) and Bollobás et al. (2001) focussed on the proportional model, for which $\delta = 0$. The affine model was proposed by Dorogovtsev et al. (2000). All these works were pre-dated by Price (1965); Simon (1955); Yule (1925); see [V1, Chapter 8] for more details on the literature. The Bernoulli preferential attachment model was introduced and investigated by Dereich and Mörters (2009, 2011, 2013).

## Notes on Section 1.4

This material is folklore. A lot is known about slowly varying functions; we refer to the classic book on the topic by Bingham et al. (1989) for details.

## Notes on Section 1.5

Our choice of notation was heavily influenced by Janson (2011), to which we refer for further background and equivalent notation.

## 1.7 Exercises for Chapter 1

**Exercise 1.1** (Probability mass function typical degree) *Prove that the probability mass function of the degree of a uniform vertex is given by* (1.1.3).

**Exercise 1.2** (Growth maxima power-law random variables) *Suppose that the non-negative random variable $X$ satisfies that there exists $\tau > 1$,*

$$\mathbb{P}(X > x) = c_X x^{-(\tau-1)}. \tag{1.7.1}$$

*Let $(X_v)_{v\in[n]}$ be a sequence of iid copies of $X$. Show that $M_n = \max_{v\in[n]} X_v$ satisfies $n^{-1/(\tau-1)} M_n \xrightarrow{d} M$ for some limiting random variable $M$.*

**Exercise 1.3** (Uniform random graph) *Consider $\mathrm{ER}_n(p)$ with $p = \frac{1}{2}$. Show that the result is a uniform graph, i.e., it has the same distribution as a uniform choice from all the graphs on $n$ vertices.*

**Exercise 1.4** (Thin-tailed Poisson) *Show that, $\lim_{k\to\infty} \mathrm{e}^{\alpha k} p_k = 0$ for every $\alpha > 0$, where $p_k = \mathrm{e}^{-\lambda}\lambda^k/k!$ denotes the Poisson probability mass function.*

**Exercise 1.5** (A nice power-law distribution) *Let the random variable $X$ have generating function*

$$G_X(s) = \mathbb{E}[s^X] = 1 - (1-s)^\alpha. \tag{1.7.2}$$

*Fix $\alpha \in (0,1)$. Identify the probability mass function $\mathbb{P}(X = k)$ of $X$.*

**Exercise 1.6** (A power-law distribution?) *Consider $G(s) = 1 - (1-s)^\alpha$ as in Exercise 1.5, now for $\alpha > 1$. Is $G(s)$ the generating function of a random variable?*

**Exercise 1.7** (Weight of uniformly chosen vertex) *Let $o$ be a vertex chosen uar from $[n]$. Show that the weight $w_o$ of $o$ has the distribution function $F_n$ given in* (1.3.10).

**Exercise 1.8** (Maximal weight bound) *Assume that Conditions 1.1(a) and (b) hold. Show that $\max_{v\in[n]} w_v = o(n)$. Further, show that $\max_{v\in[n]} w_v = o(\sqrt{n})$ when Conditions 1.1(a)–(c) hold.*

**Exercise 1.9** (Domination weights) *Let $W_n$ have the distribution function $F_n$ from* (1.3.17). *Show that $W_n$ is stochastically dominated by the random variable $W$ having distribution function $F$. Here we recall that $W_n$ is stochastically dominated by $W$ when $\mathbb{P}(W_n \le w) \ge \mathbb{P}(W \le w)$ for all $w \in \mathbb{R}$.*

**Exercise 1.10** (Degree of uniformly chosen vertex in $\mathrm{GRG}_n(\boldsymbol{w})$) *Prove that the asymptotic degree in $\mathrm{GRG}_n(\boldsymbol{w})$ satisfies* (1.3.22) *under the conditions of Theorem 1.3.*

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

**Exercise 1.11** (Power-law degrees in generalized random graphs) *Prove that, under the conditions of Theorem 1.3, the degree power-law tail in* (1.3.24) *for* $\mathrm{GRG}_n(\boldsymbol{w})$ *follows from the weight power-law tail in* (1.3.23)*. Does the converse also hold?*

**Exercise 1.12** (Degree example) *Let the degree sequence* $\boldsymbol{d} = (d_v)_{v\in[n]}$ *be given by*

$$d_v = 1 + (v \bmod 3). \tag{1.7.3}$$

*Show that Conditions 1.7(a)–(c) hold. What is the limiting degree variable* $D$*?*

**Exercise 1.13** (Poisson degree example) *Let the degree sequence* $\boldsymbol{d} = (d_v)_{v\in[n]}$ *satisfy*

$$n_k/n \to \mathrm{e}^{-\lambda}\frac{\lambda^k}{k} \tag{1.7.4}$$

*and*

$$\sum_{k\geq 0} k n_k/n \to \lambda, \qquad \sum_{k\geq 0} k^2 n_k/n \to \lambda(\lambda+1). \tag{1.7.5}$$

*Show that Conditions 1.7(a)–(c) hold. What is the limiting degree variable* $D$*?*

**Exercise 1.14** (Power-law degree example) *Consider the random variable* $D$ *having generating function, for* $\alpha \in (0,1)$*,*

$$G_D(s) = s - (1-s)^{\alpha+1}/(\alpha+1). \tag{1.7.6}$$

*What is the probability mass function of* $D$*?*

**Exercise 1.15** (Power-law degree example) *Consider the random variable* $D$ *having generating function* (1.7.6) *with* $\alpha \in (0,1)$*. Show that* $D$ *has an asymptotic power-law distribution and compute its power-law exponent.*

**Exercise 1.16** (Power-law degree example (cont.)) *Consider the degree sequence* $\boldsymbol{d} = (d_v)_{v\in[n]}$ *with* $d_v = [1-F]^{-1}(v/n)$*, where* $F$ *is the distribution of a random variable* $D$ *having generating function* (1.7.6) *with* $\alpha \in (0,1)$*. Show that Conditions 1.7(a) and (b) hold, but Condition 1.7(c) does not.*

**Exercise 1.17** (Number of erased edges) *Assume that Conditions 1.7(a) and (b) hold. Show that Theorem 1.8 implies that the number of erased edges in* $\mathrm{ECM}_n(\boldsymbol{d})$ *is* $o_{\mathbb{P}}(n)$*.*

**Exercise 1.18** (Edge probability of uniform random graphs with prescribed degrees) *Prove the formula for the (conditional) edge probabilities in uniform random graphs with prescribed degrees in* (1.3.50)*.*

**Exercise 1.19** (Edge probability of uniform random graphs with prescribed degrees (cont.)) *Prove the formula for the number of switches with and without a specific edge in uniform random graphs with prescribed degrees in* (1.3.51)*. Hint: Use an "out-is-in" argument that the number of switches from* $\mathcal{S}$ *to* $\bar{\mathcal{S}}$ *is the same as the number of switches that enter* $\bar{\mathcal{S}}$ *from* $\mathcal{S}$*.*

**Exercise 1.20** (Degrees grow to infinity almost surely) *Consider the preferential attachment model* $\mathrm{PA}_n^{(m,\delta)}(a)$*. Fix* $m = 1$ *and* $i \geq 1$*. Prove that* $D_i(n) \xrightarrow{a.s.} \infty$ *as* $n \to \infty$*, by using* $\sum_{s=i}^{n} I_s \preceq D_i(n)$*, where* $(I_n)_{n\geq i}$ *is a sequence of independent Bernoulli random variables with* $\mathbb{P}(I_n = 1) = (1+\delta)/(n(2+\delta)+1+\delta)$*. What does this imply for* $m > 1$*?*

**Exercise 1.21** (Degrees of fixed vertices) *Consider the preferential attachment model* $\mathrm{PA}_n^{(m,\delta)}(a)$*. Prove Theorem 1.17 for* $m = 1$ *and* $\delta > -1$ *using the martingale convergence theorem and the fact that*

$$M_i(n) = \frac{D_i(n)+\delta}{1+\delta}\prod_{s=i-1}^{n-1}\frac{(2+\delta)s+1+\delta}{(2+\delta)(s+1)} \tag{1.7.7}$$

*is a martingale for every* $i \geq 1$ *and for* $n \geq i$*.*

**Exercise 1.22** (Power-law degree sequence) *Prove that the limiting degree distribution of preferential attachment models in* (1.3.60) *satisfies the power-law asymptotics in* (1.3.63) *by using Stirling's formula.*

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press

**Exercise 1.23** (Degrees distribution of affine Bernoulli preferential attachment model) *Recall the limiting degree distribution* $(p_k)_{k\geq 0}$ *in* (1.3.67)*. Show that* $p_k \sim c_{\gamma,\beta} k^{-(1+1/\gamma)}$ *when* $f(k) = \gamma k + \beta$*. What is* $c_{\gamma,\beta}$*?*

**Exercise 1.24** (Degrees distribution of sublinear Bernoulli preferential attachment model) *Recall the limiting degree distribution* $(p_k)_{k\geq 0}$ *in* (1.3.67)*. Show that* $\log(1/p_k) \sim k^{1-\alpha}/(\gamma(1-\alpha))$ *when* $f(k) \sim \gamma k^{\alpha}$ *for some* $\alpha \in (0, 1)$*.*

**Exercise 1.25** (Size-biased degree distribution and random edges) *Let* $D_n$ *be the degree of a random vertex in a graph* $G_n = (V(G_n), E(G_n))$ *of size* $|V(G_n)| = n$*. Let* $D_n^\star$ *be the degree of a random vertex in an edge drawn uar from* $E(G_n)$*. Show that* $D_n^\star$ *has the size-biased distribution of* $D_n$*, where we recall the definition of the size-biased distribution of a random variable from* (1.4.18)*.*

**Exercise 1.26** (Number of points in an inhomogeneous Poisson process) *Prove that the Poisson density formula in* (1.5.6) *implies that the number of points of the Poisson process in* $\mathcal{A}$ *has the appropriate Poisson distribution, i.e.,* $\Pi(\mathcal{A}) \sim \mathsf{Poi}(\int_{\mathcal{A}} \Lambda(\mathrm{d}x))$*.*

**Exercise 1.27** (Number of points in an inhomogeneous Poisson process (cont.)) *In the setting of Exercise 1.26, show that* (1.5.6) *implies that* $\Pi(\mathcal{A})$ *and* $\Pi(\mathcal{B})$ *are independent when* $\mathcal{A}$ *and* $\mathcal{B}$ *are disjoint.*

**Exercise 1.28** (Breadth-first encoding ordered rooted tree) *Recall Definitions 1.24 and 1.25 for the breadth-first order on, and exploration of, a rooted ordered tree. Show that the sequence* $(x_i)_{i=1}^t$ *is in one-to-one correspondence to the rooted ordered tree* $\mathbf{t}$*.*

https://doi.org/10.1017/9781316795552.002 Published online by Cambridge University Press