# Random graphs:

from static to dynamic random graphs

Henri van den Esker

Random graphs:
from static to dynamic random graphs

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op

21 mei 2008 om 15:00 uur

door

Hendrik VAN DEN ESKER
wiskundig ingenieur

geboren te Zwolle

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr. F.M. Dekking

Toegevoegd promotor: Dr. G. Hooghiemstra

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr. F.M. Dekking,	Technische Universiteit Delft, promotor
Dr. G. Hooghiemstra,	Technische Universiteit Delft, toegevoegd promotor
Prof.dr. J. van den Berg,	Vrije Universiteit Amsterdam
Prof.dr. R. van der Hofstad,	Universiteit Eindhoven
Prof.dr.ir. G. Jongbloed,	Technische Universiteit Delft
Prof.dr. R.W.J. Meester,	Vrije Universiteit Amsterdam
Prof. B. Tóth,	Technical University Budapest
Prof.dr.ir. P.F.A. Van Mieghem,	Technische Universiteit Delft, reservelid

Het Stieltjes Instituut heeft bijgedragen in de drukkosten van het proefschrift.

THOMAS STIELTJES INSTITUTE
FOR MATHEMATICS

To the memory of my father, W. van den Esker

Contents

Contents

Chapter 1

Introduction

> Empirical studies on real-life networks, such as the Internet, the World-Wide Web, social networks, and various types of technological and biological networks, show fascinating similarities. Many of the networks are *small worlds*, meaning that typical distances in the network are small, and many of them have *power-law degree sequences*, meaning that the number of vertices with degree k falls off as $k^{-\tau}$ for some exponent $\tau > 1$. (...) Incited by these empirical findings, random graphs have been proposed to model and/or explain these phenomena.
>
> M. Deijfen, H. van den Esker, R. van der Hofstad and G. Hooghiemstra; [29]

Random graphs are mathematical objects that have been proposed to model real-life networks. A random graph produces *graphs*, a *graph* consists of *nodes* which are connected by each other by *edges*. A real-life network that is abundantly considered in the literature is the Internet. We can define the graph corresponding to the Internet as the set of routers (the nodes) and the wires (the edges) between them.

In 1959 Erdős and Rényi [34] introduced the first random graph, which we refer as the *Erdős and Rényi* (ER) random graph. The study of this random graph is to be considered the start of *random graph theory*. At first only *static random graphs* are considered in the literature. In these random graphs the number of nodes is fixed, and edges are added between the nodes using *a connection rule*. The ER random graph is an example of a static random graph, since in this model the number of nodes is fixed and between each pair of nodes an edge is added with some fixed probability. By appropriately choosing the connection rule of the edges we can produce graphs with power-law degree sequences. An example of a static random graph that produces power-law degree sequences is the *configuration model*, which is one of the random graphs used in this thesis, see Chapter 2.

Most real-life networks have the property in common that they grow over time. For example, the Internet is growing due to the addition of routers. A static random graph cannot reveal the growth of a network over time. We need *dynamic random graphs* that are able to mimic the growth of a network. In 1999, Barabási and Albert introduced the first dynamic random graph. Barabási

and Albert show in [4], that the model produces graphs with power-law degree sequences.

At this moment, the research interest in random graph theory is shifting from static random graphs to dynamic random graphs. In the past static random graphs were a logical choice to start with, since they are more tractable from a mathematical point of view. Nowadays, it is clear that dynamic random graphs are a more natural choice, although these random graphs are more complex and more difficult to analyze.

1.1   Outline of thesis

In this section we will give an overview of the articles in this thesis. In total there are four papers, Chapter 2 up to Chapter 5, which are preceded by this introductionary chapter. After giving an overview of the articles, we end this section with an overview of the remainder of the sections in this chapter. The overview will be brief and technicalities will be avoided, since at this point no formal introduction to *graph theory* has been given.

In Chapter 2 we consider the *configuration model*. In this model we start with a fixed number of nodes, where from each node a random number of stubs or half-edges emanates. We construct the graph by randomly connecting pairs of stubs, where two different stubs combined give precisely one edge. Observe that the total number of stubs should be even, otherwise we end up with one free stub. The configuration model can be used to model the network of routers of the Internet, where we ignore the underlying geometry. In Chapter 2, we will study the case that for each node the expected number of stubs of a node is infinite. We will show that if the number of nodes goes to infinity, then one needs either 2 or 3 edges to connect two randomly chosen nodes, or in other words the distance between two randomly chosen nodes is either 2 or 3. It is possible that the number of stubs of a single node exceeds the number of available nodes, since its expectation is infinite. This is in real networks not very likely. Consider for example the Internet, where connections between routers are identified with edges. Therefore, we consider the same model with the following restriction: if the total number of nodes is n, then the number of stubs of a single node is at most n^α, where α is some value between 0 and 1. Under this restriction, we will show that, in the limiting case, the distance between two randomly chosen nodes is with high probability constant and that this constant depends only on the value of α.

In Chapter 3, we introduce a *preferential attachment model with random initial degrees*. Conveniently, we call it the *PARID* random graph. Initially, we start off with two nodes which are connected by a random number of edges. Then at each discrete time step, we add a new node with a random number of stubs. Each stub is randomly connected to one of the old nodes by forming an edge, where nodes with a high number of outgoing edges are preferred, which is called *preferential attachment*. By repeating this process, we can construct a graph with any number of nodes. In Chapter 3 we describe the degree distribution of the PARID random graph, i.e., the distribution of the number of edges of a node, as the number of nodes goes to infinity.

In Chapter 4 we consider random graphs consisting of n nodes, where each

node has a random weight. The number of edges between each pair of nodes v and w does only depend on the weights of the nodes v and w. The classical example is the Erdős and Rényi random graph, where each node has the same weight. In this model the number of edges between pairs of nodes is at most one, and each of the $\binom{n}{2}$ possible edges occurs independently with probability p, which depends on the weights. In Chapter 4 we will consider the *Poissonian random graph* and derive distance results for this random graph. Furthermore, we show that these distance results, also, do hold for other random graphs found in the literature.

In the models above the underlying geometry is ignored. However, in for example wireless ad-hoc networks the geometry is of great importance, since in these networks nodes are spread over some surface and nodes can only communicate with neighbors within a certain range, depending on the geometry. The random graph, which we present in Chapter 5, could serve as a model for wireless ad-hoc networks.

The remainder of this chapter consists of 5 subsections. In Section 1.2, we list the 4 publications. In Section 1.3, we give some preliminaries of *graph theory*. Then we give the history from the first definition of a *graph* up to the introduction of *scale-free graphs* in Section 1.4, and, in Section 1.5, we discuss the properties of these graphs. After that, we introduce two types of random graphs. In Section 1.5.3, we define two type of random graphs: *static* and *dynamic* random graphs. For both types we give examples in Section 1.6.

1.2   List of publications

In this section we give the list of all papers presented in this thesis.

- *Distances in random graphs with infinite mean degree sequence* (2006)
  H. van den Esker, R. van der Hofstad, G. Hooghiemstra and D. Znamenski; published in *Extremes* **8**, pages 111-140.

- *A preferential attachment model with random initial degrees* (2007)
  M. Deijfen, H. van den Esker, R. van der Hofstad and G. Hooghiemstra; accepted by *Arkiv för Matematik*.

- *Universality for the distance in finite variance random graphs* (2007)
  H. van den Esker, R. van der Hofstad, and G. Hooghiemstra; submitted.

- *A geometric preferential attachment model with fitness* (2008)
  H. van den Esker; submitted.

1.3   Preliminaries of graph theory

In this thesis it is assumed that the reader has some familiarity with basic probability theory.

A *graph* is the basic element of graph theory. A graph consists of *vertices* or *nodes*, which are connected to each other by directed or undirected *edges*. We will introduce the *simple graph* and the *multigraph*.

A *simple graph* is a pair $G = (V, E)$, where V and E are sets. An element of V is called a *vertex* or a *node* and an element of E is called an *edge* or a *link*. Each element in E is an unordered pair (x, y), where x and y are elements of V such that $x \neq y$. Furthermore, each element in E is unique. Thus, in a simple graph it is assumed that there are no multiple edges and no self-loops, since the elements in E are unique and $(x, x) \notin E$ for all $x \in V$. The left picture in Figure 1.1 is an example of a simple graph.

In a multigraph we allow for multiple edges between vertices and self-loops. A multigraph is a pair $G = (V, E)$, where V and E are sets. As before, V is the set of vertices. Each element in E is an ordered pair (x, y), where x and y are elements of V and are not necessarily different. An edge $(x, y) \in E$ represents a directed edge from the vertex x to y. In literature this is often the definition of a *multigraph*, but there are ambiguities concerning this definition. In this thesis, we will use the definition as given here.

Figure 1.1: Some examples of graphs.

(a) A simple graph consisting of 5 vertices and 4 edges.

(b) A multigraph consisting of 6 directed edges; one of the edges is a self-loop.

(c) A graph with 3 components.
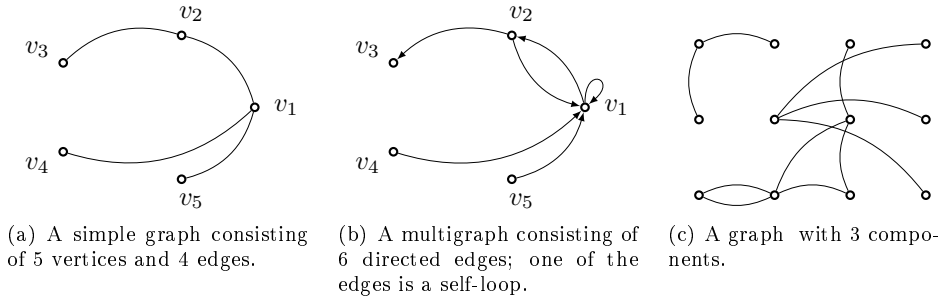
A *path* of length n in the graph $G = (V, E)$ from vertex $v \in V$ to vertex $w \in V$ is a sequence $x_0 = v, x_1, \ldots, x_n = w$ of distinct elements of V such that $(x_{i-1}, x_i) \in E$ for $i = 1, \ldots, n$.

We call $v, w \in V$ connected if there exists a path between the vertices v and w. The *graph distance*, or *hopcount*, between the vertices $v, w \in V$ is the length of the shortest path between the vertices v and w. We denote by $d(v, w)$ the graph distance if the nodes v and w are connected, otherwise we set $d(v, w) = \infty$.

The *typical distance* of the graph G is the graph distance between two connected randomly chosen vertices, which we denote by

$$\text{Dist}(G). \tag{1.3.1}$$

The diameter of a graph G is the maximum finite distance between any pair of vertices in G, which we denote by

$$\text{diam}(G) = \max_{\substack{v, w \in V \\ d(v,w) < \infty}} d(v, w). \tag{1.3.2}$$

Observe that the diameter is a real number, but the typical distance is a random variable.

The graph G is *connected* if for any two vertices v, w in V there exist a path between the vertices v and w. A graph that is not connected can always be divided into connected *components*, i.e., disjoint connected subgraphs. For example, the right most graph in Figure 1.1 consists of three connected components.

1.4 History

"The origins of graph theory are humble, even frivolous. Whereas many branches of mathematics were motivated by fundamental problems of calculation, motion, and measurement, the problems which led to the development of graph theory were often little more than puzzles, designed to test the ingenuity rather than to stimulate the imagination. But despite the apparent triviality of such puzzles, they captured the interest of mathematicians, with the result that graph theory has become a subject rich in theoretical results of a surprising variety and depth."

<div align="right">From 'Graph Theory/1736-1963'; [13]</div>

In this section we will give a brief overview of the history of random scale-free graphs. We will consider, in chronological order, the most influential persons in the history. We start with the roots of graph theory, founded by Leonhard Euler. Then, we will consider Frigyes Karinthy, who published the short story titled *Chains* in 1929. In this story, Karinthy claims that the modern world is shrinking due to the ever-increasing connectedness of human beings. We will continue this argument, and we will end with scale-free random graphs.

1.4.1 Euler

From a mathematical point of view, Leonhard Euler's solution of the *Königsberg bridge problem* is to be considered the first theorem of graph theory [39].

The Königsberg bridge problem is based on an actual place and situation. The city of Königsberg, now Kaliningrad in Russia, was situated at the Pregel River. It included two large islands which were connected to each other and the mainland by seven bridges, see Figure 1.2. It is said, that the people of Königsberg used to entertain themselves by trying to find a route around the city which crosses each of the seven bridges just once and where the starting point and the endpoint coincide. As their attempts were in vain, it was believed that finding such a route is impossible.

In 1736, Euler proved that it is not possible to give a route which crosses each bridge only once and where the starting point and the endpoint coincide. In proving the result, Euler formulated the problem in terms of graph theory, i.e., by abstracting all features of the problem. The landmasses are substituted by vertices, and the bridges are substituted by edges which connect the vertices. In this way, the first graph was defined, see Figure 1.2. Observe that replacing the bridges of Königsberg by a simple mathematical object, i.e., the graph, is perhaps more important than actually solving the problem.

The paper by Leonhard Euler on the seven bridges of Königsberg was published in 1736. It is regarded as the first paper in the history of graph theory. The paper

Figure 1.2: On the left we see, in gray, the engravings illustrating the 1736 paper of Euler about the bridges of Königsberg. The bridges are labeled a, b, c, d, e, f and g. On the right they correspond to the edges a, b, c, d, e, f and g. Similarly, the landmasses are are labeled A, B, C and D, corresponding to the vertices A, B, C and D in the graph.
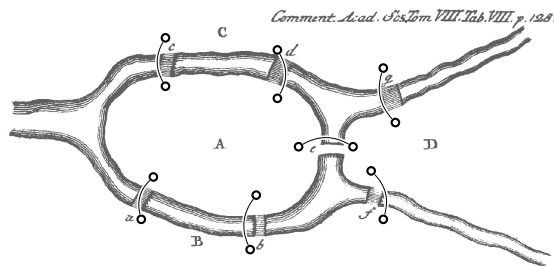
(a) The engraving illustrating the paper [39].
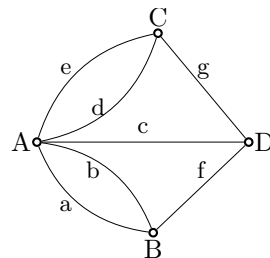
(b) A graph representing the bridges of Königsberg.

demonstrates that there is no possibility for a route which crosses all bridges only once starting and ending at the same point. Euler gave an intuitive argument to explain why there is no such route. In order that such a route exists, one needs that the number of edges that is incident to a vertex, the so-called degree of vertex, is even. The idea is simple, each time you enter a vertex (landmass) by an edge (bridge) and depart from it the degree is augmented by 2. Therefore all the degrees should be even. In Figure 1.2 we see that the degree of any node is odd, therefore no such route can exist.

1.4.2   Karinthy

In the previous section, we discussed the very early roots of graph theory. In this thesis we are interested in properties of large graphs, i.e, questions related to the diameter or typical distance of graphs when the number of vertices goes to infinity. The first paper that addressed these questions dates back to 1929. This publication is not a scientific paper at all, but a short story titled *Chains*. The author Frigyes Karinthy published this story in a volume of short stories titled *Everything is Different*. The author of *Chains* described the *small-world effect* or *six degrees of separation*, although without proof. In *Chains* Karinthy argued that the world is getting smaller. Due to technological developments in communications and travelling, social networks grow larger and span greater distances in the same amount of time. As a result, Karinthy believed that any two individuals can be connected through at most five acquaintances.

1.4.3   Erdős and Rényi

Generally speaking, a *random graph* is a graph that is 'constructed' by some 'random procedure'. In probability theory the 'random procedure' is represented by a probability space and the 'construction' by a function from the probability space into a suitable family of graphs. Thus, the theory of random graphs is a

mix of graph theory and probability theory. The study of random graphs began a few decades ago, independently and almost simultaneously, by several authors, namely Ford and Uhlenbeck [45], Gilbert [46], Austin, Fagen, Penney and Riordan [7], and Erdős and Rényi [34]. From these authors, Erdős and Rényi are considered to be the founding fathers of the theory of random graphs. They were the first to introduce methods which underlie the probabilistic treatment of random graphs, while the other authors were more concerned with enumeration problems.

In the paper [34], Erdős and Rényi introduce the random graph $G(n, m)$, where n and m are positive integers. In the $G(n, m)$ model a graph is chosen uniformly at random from the collection of all graphs with n vertices and m edges. A second model of interest is the random graph $G(n, p)$, which is often referred to as the Erdős and Rényi (ER) random graph, although the random graph was introduced by Gilbert [46]. In this model each graph consists of n vertices and each of the $\binom{n}{2}$ possible edges occurs independently with probability p. The two random graph models are closely related, see for example [58, §1.4] or [16, §2.2].

1.4.4 Milgram

Until the experiments of Stanley Milgram, the challenge of linking two individuals to each other by at most five other individuals had been a brainteaser. Stanley Milgram, a famous experimental social psychologist, started to experiment in 1967 at Harvard University in Cambridge, Massachusetts, USA, to test the hypothesis of the six degrees of separation. Like Karinthy, Milgram was intrigued by the increasing social interconnectedness among individuals over time. However, it is not known to me whether Milgram knew the work of Karinthy.

Milgram's experiments were designed to measure the typical number of intermediaries between two randomly selected people. In one of the experiments, which is described in [81], Milgram determined how many intermediaries it took to send a booklet from selected individuals in Nebraska and Boston to a specific target in Massachusetts. Participants were asked to try to get the booklet to the target by passing the booklet to someone they knew personally, who they believed either knew the target, or might know someone who did. These acquaintances were asked to do the same, repeating the process until the booklet had reached the target.

A significant problem was that people refused to participate. As a consequence, the majority of the booklets never reached their destination. In the above described experiment 232 out of the 296 booklets never reached the destination, see [81]. However, 64 of the booklets did reach their target, and the mean number of intermediaries between the starters and the target was 5.2. Milgram deducted that people in the US are separated by about six people on average. Although Milgram himself never used the term six degrees of separation, these findings attributed to the widespread acceptance of six degrees of separation.

1.4.5 Watts and Strogatz

The paper of Watts and Strogatz [82] was probably the first paper to compare networks from different sciences and to show that their diameters were considerably smaller than those of regularly constructed graphs, i.e., lattices or grid graphs,

with the same number of vertices and edges. In the literature, the existence of a small diameters is called the *small-world effect*. Watts and Strogatz considered the collaboration graph of actors in featured films, the electrical power grid of the western United States and the neural network of the worm *C. Elegans*.

Watts and Strogatz introduced the so-called Watts-Strogatz (WS) random graph, which generates highly clustered graphs with a relatively small diameter. Essentially, the model is a rewiring model. One starts with a fixed set of vertices on a circle and one connects each vertex only to its k neighbors on the right and to its k neighbors on the left. Then, with given probability, edges are rewired. The graphs obtained by rewiring edges on a circle do not resemble real networks, although they share some of their characteristics. To model real networks, it was desirable to define new families of graphs models.

1.4.6 Barabási and Albert

As indicated in the previous subsection there was a need for introducing new random networks that both explain the small world effect and the occurrence of so called power-law degree distributions, which we will explain now. In [66, 67, 76] it was observed that in many real-life networks the number of nodes with degree k is proportional to an inverse power of k. More precisely, if N_k is the number of nodes in the network having degree k, then

$$N_k \approx Ck^{-\tau} \tag{1.4.1}$$

where C is a constant and $\tau > 1$ a parameter of the network which is called the power-law exponent. See Figure 1.3 or [66, Table II] for a more extensive overview of examples of networks with the corresponding power-law exponent.

Figure 1.3: Some real-life networks from [76, Table 1].

Network	nodes	links	τ	Brief description
Gnutella	1.026	3.752	1.4	A file sharing network. The users are the nodes and the upload and download links are the edges.
PIN	2.115	2.240	2.4	Protein Interaction Networks.
Citation	783.339	67.161.198	3.0	Each node is an author, where authors are connected to one and another if they have published a paper together.
Actors	449.913	255.516.482	2.3	Each node is an actor, where two actors are connected if they featured together in the same moving picture.
Internet	10.697	31.992	2.5	The nodes are (large) routers and the links are the connection between the routers.

The paper by Barabási and Albert [4] was the first paper that gave the cause of the existence of these power-law degree distributions. Barabási and Albert introduced a random graph model, which is nowadays referred to as the *Barabási-Albert* (BA) random graph model. The BA model incorporates two important

general concepts, which are present in most real networks, namely: growth and preferential attachment. In the BA model nodes are entering the model one by one at discrete time steps. At time n a new node enters the network and from this node there are emanating a fixed number $m \geq 1$ edges. These edges are then connected to the old nodes $1, 2, \ldots, n - 1$ of the network, where the attachment probabilities are taken proportional to the degrees of the old nodes. Barabási and Albert showed by simulation that the above set-up leads to a random graph having power-law degree sequences. A rigorous proof of the fact that the BA-model has power law degree sequence was given by Bollobás and Riordan [19]. The small world effect was also proved by Bollobás and Riordan. They showed for the BA model that the diameter scales for $m = 1$ as $\log n$, i.e, the logarithm of the number of nodes. For $m > 1$, the diameter scales as $\log n / \log \log n$.

1.5   Scale-free graphs and modeling

> "Nature normally hates power laws. In ordinary systems all quantities follow bell curves, and correlations decay rapidly, obeying exponential laws. But all that changes if the system is forced to undergo a phase transition. Then power laws emerge – Nature's unmistakable sign that chaos is departing in favor of order."
>
> Albert-László Barabási

To study a real-life network one could identify the nodes and edges and construct the corresponding graph. For example the Internet consists of routers (nodes) which are connected to each other by wires (edges), or, the World-Wide Web (WWW) consists of web pages (nodes) which are linked to each other by hyper links (directed edges). In practice this is not very useful, since the networks of interest are large and, as a consequence, this procedure is very time consuming and the amount of memory to store the corresponding graph is huge. Therefore, it is practically impossible to give a full description of the network.

The solution is to model a real-life network under consideration using a random graph, where random rules describe how vertices are connected to one another. In this thesis we consider *scale-free networks*, which will be explained in the next subsection. We close this section by introducing two types of random graphs.

1.5.1   Scale-free graphs and their properties

The random graphs, which are used to model real-life networks, have in common that the degree distribution follows a power law, i.e., the probability that a vertex has degree k is proportional to $k^{-\tau}$, for some constant $\tau > 1$, the so-called *power-law exponent*, as already mentioned in Section 1.4.6. It turns out the asymptotic behavior of the degree sequence is independent of the size of the graph, and depends only on the constant τ. We call a graph with this property *scale-free*.

In the literature, we observe that the word *network* and *graph* are used as synonyms. We use the terminology *real-life network* for a network existing in the real world. As a consequence, a scale-free network is a scale-free graph. In this thesis we prefer to use the word *graph* over the word *network*.

Figure 1.4: In the picture, we see the degree sequence of a realization G of the *configuration model* with 2500 vertices and parameter 3, this model will be introduced in §1.6.1. The theoretical power-law exponent is 3. Observe, if $p_k \approx Ck^{-3}$, for some positive constant C, then $\log p_k \approx \log C - 3 \log k$. Therefore, we expect that if the values p_k are plotted against k on a log-log scale, then the points lie, approximately, on a straight line with slope -3. This is indeed the case, as shown by the dotted line. The slope of this line on the log-log scale is -3 and the intercept is $10^{0.3214}$, which is obtained by using least squares estimation on the log-log data. In the right picture, we see the largest component of the simulated graph.
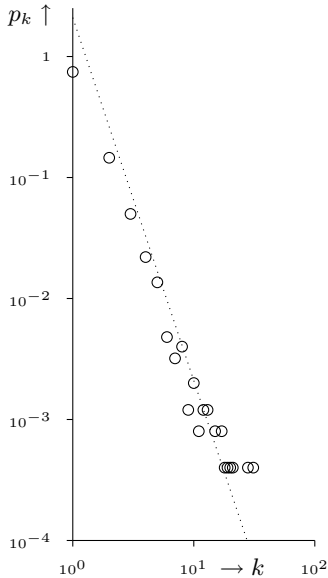
(a) The degree sequence of the graph G on a log-log scale, where p_k is the fraction of vertices with degree k.
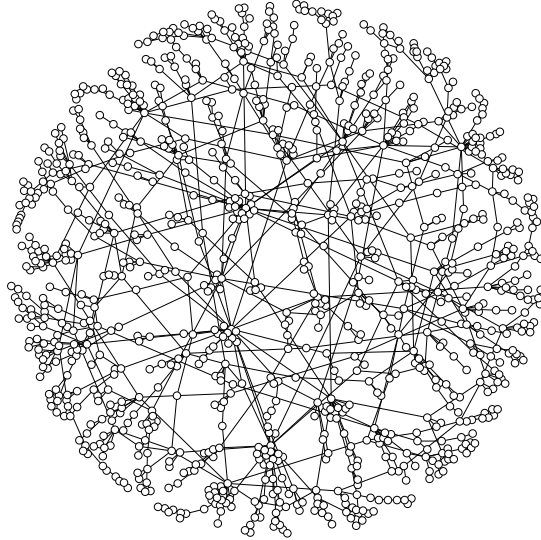
(b) A visual representation of the largest component of the graph G. The component consists of 1051 vertices and 1126 edges.

In Figure 1.3 a table is given of some scale-free networks and their corresponding power-law exponents as found by experimental data. We, also, included the number of nodes and links.

Another property, that is seen in real-life networks is the existence of *hubs*. We identify hubs in graphs with vertices with a high degree. If one models a network by a scale-free random graph, then this property is a consequence of the fact that the degree distribution follows a power law, since this is a skewed and heavy tailed distribution. The skewness explains that most vertices are of low degree, and the heavy tailness explains the existence of vertices with a very high degree (the hubs).

1.5.2 Characteristics of the power-law exponent

We defined a scale-free graph by the existence of a power-law exponent $\tau > 1$. The value of the power-law exponent determines the characteristics of the random

graph. Depending on the value of τ, we classify the following three categories: the *infinite mean case*, *the finite mean and infinite variance case*, and the *finite variance case*, which corresponds to $\tau \in (1,2)$, $\tau \in (2,3)$ and $\tau > 3$, respectively.

These categories are of interest, since it influences the properties of the graph, e.g., the existence of hubs or the size of the typical distance.

Results in the literature show that if $\tau > 3$, then the graph is homogeneous or flat in the sense that all vertices have roughly the same kind of neighborhood. If $\tau \in (2,3)$, then there will appear vertices with a high degree. These vertices act as hubs, and their impact on the connectivity cannot be neglected. If $\tau \in (1,2)$, then a finite number of vertices act as super hubs for all the other vertices, the graph has a star-shaped structure, see [71].

It should be clear that the power-law exponent τ directly influences the typical distance. For $\tau \in (1,2)$ the typical distance is bounded by some constant, because of the star-shaped structure, see [38]. If $\tau \in (2,3)$ then the hubs acts as shortcuts which results in typical distances of order $\log \log n$, see [53, 56]. If $\tau > 3$ then the typical distance is of order $\log n$, because it lacks hubs, see [53, 54, 71]. Although these properties are only proven for certain families of random graphs, we conjecture that these results do hold in general, i.e., that they only depend on the value of τ, for any random graph where τ is well-defined.

Observe that we ignored the boundary cases $\tau \in \{1,2,3\}$. The classification in the three above mentioned cases depends on the model, which is illustrated in Chapter 2 by means of examples. In this chapter, we consider the configuration model, and using this model we generate random graphs with $\tau \in (1,2)$. One of the main results, Theorem 2.1, states that the typical distance is concentrated *with high probability* on the values 2 and 3, where an sequence of events $\{A_i\}_{i \geq 1}$ occurs with high probability (**whp**) if the probability of the event A_n tends to one if n tends to infinity. It is shown that for the boundary case $\tau = 1$, see Theorem 2.3, the typical distance is **whp** concentrated only on 2, which implies that one node acts as a hub for every other node. In Section 2.4.2 two examples are presented and in both examples the resulting power-law exponent is $\tau = 2$. In one example the typical distance is **whp** concentrated on the values 2 and 3, while in the other example the typical distance drifts away to $+\infty$ if the number of nodes tends to infinity. Thus, the categorization of the boundary cases, $\tau \in \{1,2,3\}$, depends on the used model.

1.5.3   Dynamic verus static random graph models

Most large networks that we encounter in nature are growing over time. If one wants to model such a network than one can use either a *static* or a *dynamic* random graph.

Static random graphs aim to describe a network and its topology at a given time instant, and it tries to mimic the network under consideration. The Erdős-Rényi graph is an example of a static random graph, see Section 1.4.3. On the other hand, a dynamical random graph aims to explain how the network came to be as it is. Dynamic random graphs often focus on the growth of the network under consideration as a way to explain the power-law degree sequences by means of *preferential attachment* growth rules, where added vertices and links are more

Figure 1.5: In the three figures we see the typical structure of the network for the *infinite mean*$(1 < \tau < 2)$, *finite mean and infinite variance*$(2 < \tau < 3)$ and *finite variance*$(\tau > 3)$ case.

(a) $1 < \tau < 2$.   (b) $2 < \tau < 3$.

(c) $\tau > 3$.
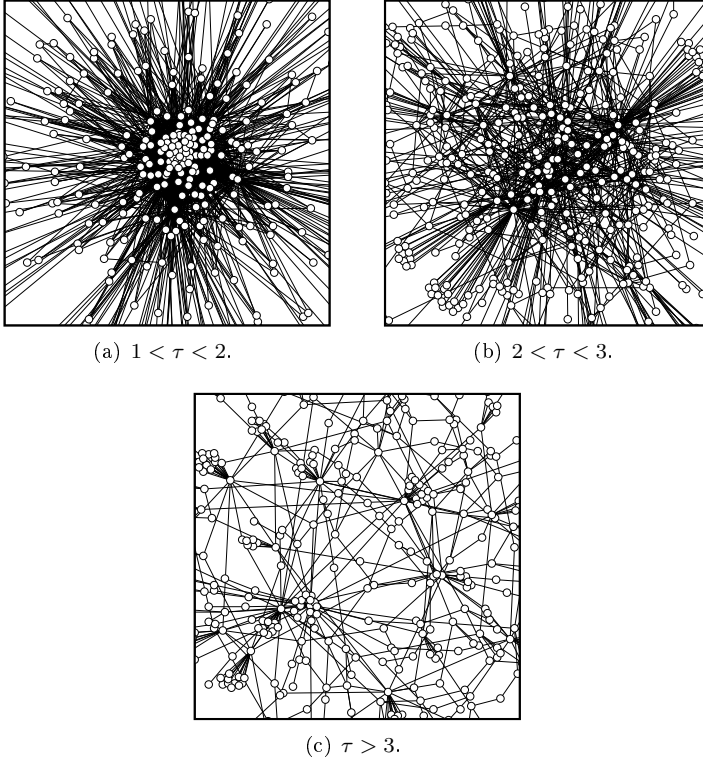
likely to be attached to vertices that already have large degrees. The Barabási-Albert random is an example of a dynamic random graph, see Section 1.4.5.

In the literature, the first random graphs introduced used to model real-life networks are static ones. For example, the first paper that describes the Erdős-Rényi random graph is published in 1959, c.f. [34], while the first paper about the Barabási-Albert random graph is published in 1999, c.f. [4]. In general, static random graphs are from a mathematical viewpoint more tractable than dynamic ones. Therefore, static random graphs were the first random graphs to be studied. Nevertheless, over time it became clear that static random graphs are not well suitable to model real-life networks. Dynamic random graphs are more appropriate as a choice to model real-life networks, since static random graphs lack the ability to minic the growth of network. Unfortunately, dynamic random graphs are more complex and more difficult to analyze.

1.6   Models for complex networks

> "I think I've been contacted by someone from just about every field out-side of English literature. I've had letters from mathematicians, physicists, biochemists, neurophysiologists, epidemiologists, economists, sociologists; from people in marketing, information systems, civil engineering, and from a business enterprise that uses the concept of the small world for networking purposes on the Internet."
>
> Duncan Watts

Over the years, many random graph models have been introduced in the lit-erature. Often, each newly introduced random graph tries to mimic a property of a real network. In the field of physics, we see often that a network is modeled by a random graph, and that subsequently some properties of the random graph are derived using simulations or mean-field theory. One often starts with the be-havior of the degree sequence. Next, the researchers consider the diameter or the expected value of the typical distance. At this stage these results are sometimes picked up by mathematicians and provided with rigorous proofs.

1.6.1   Static models for complex networks

In this section, we will consider some static random graphs. These graphs aim to describe networks and their topology at a fixed time instant.

The Watts-Strogatz random graph

Watts and Strogatz defined a random graph, which generates highly clustered graphs with a small diameter. Essentially, their model is a rewiring model, see Section 1.4.5. It is difficult to derive results rigorously from the definition given by Watts and Strogatz, since both the order of updating the edges is of impor-tance, as are the restrictions that Watts and Strogatz impose in order to keep the random graph simple. Newman and Watts [69] considered a variant of this random graph, which we call the *Watts-Strogatz* (WS) random graph. We will consider the random graph on a circle, with given parameters n, k and $p = p(n)$. To construct the initial graph, one distributes n vertices evenly on a circle and connects each vertex to its k neighbors to the left and to its k neighbors to the right. For example, the leftmost graph in Figure 1.6 shows the initial WS graph for the parameters $n = 12$ and $k = 2$. Instead of rewiring, we add a random number of shortcuts (edges). More precisely, the number of shortcuts is Poisson distributed with mean $pkn/2$, where $p \in [0,1]$. The begin and end-points of each shortcut are chosen, independently, at random from the n vertices and one allows for multiple edges and self-loops. From a mathematical viewpoint, this random graph is more tractable than the random graph defined by Watts and Strogatz.

The behavior of the typical distance depends on the value pkn. If pkn is small compared to 1, that is $pkn = o(1)$, $n \to \infty$, then the distance scales with the number of vertices n:

$$\frac{\text{Dist}(G_n)}{n} \xrightarrow{P} \frac{1}{4k}.$$

Figure 1.6: Three realizations of the Watts and Strogatz random graph with $n = 12$, $k = 2$ and $p \in \{0.00, 0.15, 0.85\}$.

(a) $p = 0.00$ (b) $p = 0.15$ (c) $p = 0.85$
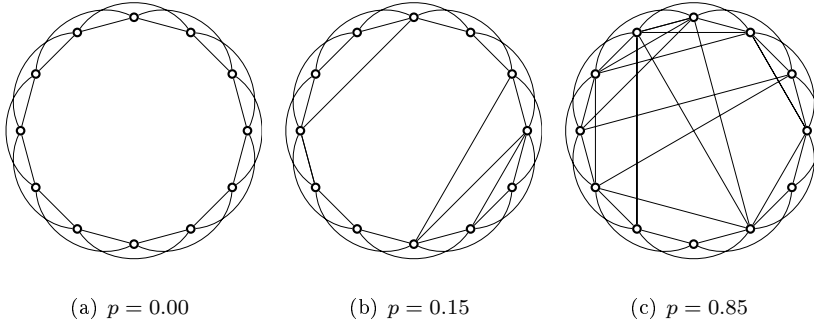
On the other hand, if pkn is large compared to 1, $n \to \infty$, then the typical distance scales as $\log n$:

$$\frac{\text{Dist}(G_n)}{\log n} \xrightarrow{P} \frac{1}{k^2 p}. \tag{1.6.1}$$

We refer to [5, VI.A] or [8] for a more in-depth analysis of the scaling behavior of the distance.

The Erdős-Rényi random graph

The study of random graphs began independently and almost simultaneously by several researchers, see Section 1.4.3. In 1959, Erdős and Rényi [34] introduced the $G(n, m)$, a random graph with n vertices and m edges. The random graph $G(n, m)$ is taken at random from the collection of all graphs with n vertices and m edges. A second random graph is called $G(n, p)$, where, again, n is the number of vertices and p is the probability that there exists an edge between any pair of vertices. In the literature, the graph $G(n, p)$ is often referred to the Erdős-Rényi (ER) random graph, although, the random graph was introduced by Gilbert [46]. The graph $G(n, p)$ consists of n vertices and between each pair of vertices there is possibly an edge, where each of the $\binom{n}{2}$ possible edges occurs independently with probability p.

To compare two different graphs, we need to compare properties of the graphs. Properties are for example the number of triangles, the chromatic number, the existence of a large component. Intuitively, it seems clear that the two random graphs $G(n, m)$ and $G(n, p)$ are similar if $p = m\binom{n}{2}^{-1} = \frac{2m}{(n+1)n}$, since the expected number of edges in the random graph $G(n, p)$ is $\binom{n}{2}p$. It turns out that it depends on the property under consideration. We refer to [58, §1.4] or [16, §2.2], where conditions are stated for different properties when the two random graphs are similar, or not. That is, when one can replace $G(n, m)$ by $G(n, p)$ or the other way around. From a probabilistic viewpoint, it is more pleasant to work with the random graph $G(n, p)$ than with $G(n, m)$, because of the independence of the edges.

We close with some properties of the ER random graph. We will consider the distribution of the degree sequence, the typical distance and the diameter. We start with the behavior of the degree sequence distribution. For $p = \lambda/n$, the degree of any vertex has a binomial distribution with parameters $n - 1$, the number of possible neighbors, and $p = \lambda/n$. Hence, the degree of any vertex is asymptotically Poisson distributed with parameter λ as n tends to infinity.

Before considering the typical distance or the diameter, we first discuss the existence of a giant component. We again take $p = \lambda/n$. If $\lambda \in (0, 1)$, then **whp** every component of the graph $G(n, \lambda/n)$ has at most $\mathcal{O}(\log n)$ vertices. If $\lambda > 1$, then $G(n, \lambda/n)$ contains, **whp**, an unique giant component with $c_\lambda n + o(n)$, $0 < c_\lambda < 1$, vertices. All other components have at most $\mathcal{O}(\log n)$ vertices.

We now turn to the typical distance, assuming $\lambda > 1$. The typical distance, $\mathrm{Dist}(G(n, p))$, normalized by $\log n$ converges in probability to the reciprocal value of $\log \lambda$, i.e.,

$$\frac{\mathrm{Dist}(G(n, p))}{\log n} \xrightarrow{P} \frac{1}{\log \lambda}. \qquad (1.6.2)$$

The proof of this result can be found in [31, Section 2.4].

Similarly, for $\lambda > 1$, the diameter, $\mathrm{Diam}(G(n, p))$, satisfies

$$\frac{\mathrm{Diam}(G(n, p))}{\log n} \xrightarrow{P} \frac{1}{\log \lambda} + \frac{2}{|\log \mu_\lambda|}, \qquad (1.6.3)$$

where μ_λ is the unique solution of $\mu_\lambda e^{-\mu_\lambda} = \lambda e^{-\lambda}$ such that $\mu_\lambda < 1$, see [42].

Intuitively, the scaling of the typical distance and the diameter can be explained as follow: each vertex has on the average $(n - 1)p \approx \lambda$ neighbors. Thus, the number of neighbors up to distance k is approximately bounded from below by order λ^k. Solving $\lambda^k = n$, that is $k = \log_\lambda n$, gives a clue about the expected typical distance between two randomly chosen vertices. This, does not explain the second term in (1.6.3), i.e., $2/|\log \mu_\lambda|$. This term accounts for long isolated paths which can cause the distance between particular pair of vertices to be significantly longer than the average distance. Intuitively, the length of the diameter is the length of the average path ($\log n / \log \lambda$) extended by two strands each having length $\log n / |\log \mu_\lambda|$. For more details, we refer to [42].

The configuration model

The *configuration model* (CM) is introduced by N.C. Wormald [83] and B. Bollobás [15], independently of each other. Given a probability distribution F of a random variable with support $\{0, 1, 2, \ldots\}$, the model describes a way to construct an undirected multigraph on n vertices, whose degree distribution has probability distribution F. We allow for multiple edges and self-loops.

Consider an i.i.d. sequence $\{D_i\}_{i=1}^n$, where each D_i is distributed as D, with $\mathbb{P}(D \leq k) = F(k)$, for nonnegative integer valued k. The value of D_i, $1 \leq i \leq n$, is the degree of vertex v_i. The random graph is constructed as follow:

- Create a list in which the label i, for $1 \leq i \leq n$, of each vertex appears exactly D_i times. The list consists of L_n labels.

- Choose at random a label l_1 from the list and remove it from the list. Then, again, we choose at random a label l_2 from the list and remove it from the list. The labels l_1 and l_2 form a pair. Repeat this process, until all labels in the list are paired.

- Take n vertices and join the pairs above by edges.

Figure 1.7: A realization G of the configuration model with parameters $n = 15$ and $F(x) = 1 - x^{-1.8}$ for $x \geq 1$.

(a) The graph G.

(b) The graph G, where the vertices are put on a circle.
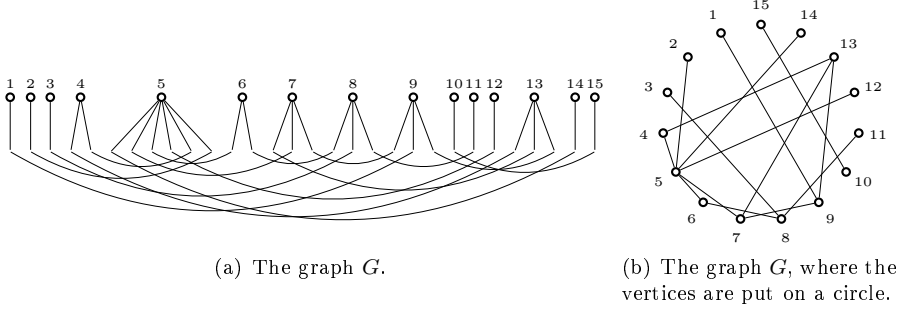
The construction fails if $L_n = \sum_{j=1}^n D_j$ is odd, because after pairing $L_n - 1$ labels we end with a list of one label and we need at least 2 to make a pair. We solve this problem by increasing D_n by 1, i.e., we replace D_n by $D_n + 1$. This change will make hardly any difference in what follows, and we will ignore it in the sequel.

Our main assumption will be that

$$x^{\tau-1} \left[1 - F(x) \right] \tag{1.6.4}$$

is slowly varying at infinity for some given $\tau > 1$.

We consider the distribution of the degree sequence and the typical distance for different values of τ.

By construction, the degree distribution of a vertex v_i, $1 \leq i \leq n$, equals the distribution function F of the random variable D_i. By picking the weight distribution F, we can obtain any degree distribution.

The typical distance depends solely on the tail behavior of the distribution function F, which, by (1.6.4), determines the power-law exponent τ. We will consider the cases $\tau \in (1,2)$, $\tau \in (2,3)$ and $\tau > 3$, separately:

- $1 < \tau < 2$: The typical distance takes only 2 values as n tends to infinity. In Chapter 2 we will consider this case more in-depth. In this thesis we will show that there exists some probability $p_F \in (0,1)$, depending on F, such that

$$\lim_{n \to \infty} \mathbb{P}\left(\mathrm{Dist}(G_n) = 2\right) = 1 - \lim_{n \to \infty} \mathbb{P}\left(\mathrm{Dist}(G_n) = 3\right) = p_F, \tag{1.6.5}$$

see Theorem 2.1. In Chapter 2, we also consider the case where we condition the degrees to be at most n^α for some $\alpha > 0$. For fixed $k \in \{0, 1, 2, \ldots\}$

and α such that $(\tau + k)^{-1} < \alpha < (\tau + k - 1)^{-1}$, the hopcount converges to $k + 3$ in probability, while for $\alpha > (\tau - 1)^{-1}$, the hopcount converges to the same limit as for the unconditioned degrees, see (1.6.5). These results are the contents of Theorem 2.2.

- $2 < \tau < 3$: In [56] it is shown that, under some mild conditions for F,

$$\frac{\mathrm{Dist}(G_n)}{\log \log n} \xrightarrow{P} \frac{2}{|\log(\tau - 2)|}.$$

- $\tau > 3$: Define ν as $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$. If $\nu > 1$, which implies that there is **whp** a giant component, then we can show that

$$\frac{\mathrm{Dist}(G_n)}{\log n} \xrightarrow{P} \frac{1}{\log \nu}. \tag{1.6.6}$$

The above result is proven in [54]. The key ideas behind this proof are also used in Chapter 4, where we derive the fluctuations of the typical distance for another random graph model, which will be introduced in the next paragraph.

Furthermore, for the diameter, a similar, result can be given

$$\frac{\mathrm{Diam}(G_n)}{\log n} \xrightarrow{P} \frac{1}{\log \nu} + \frac{2 - \mathbf{1}_{\{p_1=0\}} - \mathbf{1}_{\{p_2=0\}}}{\log \nu'}, \tag{1.6.7}$$

where $p_k = F(k)$ and ν' is some constant depending on F [42]. The second term in (1.6.7) is due to isolated paths, compare to (1.6.3).

The inhomogeneous random graph

In the Erdős-Rényi random graph $G(n, p)$, the existence of an edge between any pair of vertices is independent of all the other edges. More precisely, the edges are defined by independent Bernoulli random variables $\{X_{ij}\}_{1 \le i < j \le n}$, the *occupation statuses*, with $\mathbb{P}(X_{ij} = 1) = p_{ij}$ and $p_{ij} = p$ for all edges. The event $\{X_{ij} = 1\}$ signifies the presence of an edge between the vertices v_i and v_j. For convenience, we set $X_{ji} = X_{ij}$, for $1 \le i < j \le n$ and $X_{ii} = 0$, for $1 \le i \le n$.

In Chapter 4, we generalize the above definition, which results in the definition of the *inhomogeneous random graph* (IRG). We obtain the inhomogeneous random graph by replacing the probabilities of the occupation statuses in the ER random graph. To that end, we associate with each vertex v_i a weight Λ_i, for $1 \le i \le n$. These weights can be taken random or fixed, and both options have been considered in the literature. We assume that the weights $\{\Lambda_i\}_{i=1}^n$ are independent and identically distributed (i.i.d) random variables. The occupation probabilities will be defined conditionally on the weights. Denote by $L_n = \sum_{i=1}^n \Lambda_i$ the total weight. In the IRG, as used in Chapter 4, we assume that the occupation probabilities are given by

$$p_{ij} = f(\Lambda_i \Lambda_j / L_n),\ 1 \le i < j \le n,$$

for some given function f, satisfying

$$f(x) = x + \mathcal{O}(x^2), \text{ for } x \downarrow 0.$$

Observe that the IRG involves double randomness: first the weights of the vertices are sampled, and then, in turn, the occupation statuses are sampled. As already mentioned, one could investigate the IRG for fixed weights, since the occupation statuses are taken conditionally on the weights. The advantage is that the results are more general, because often the results for random weights can be concluded from the ones where the weights are fixed. The disadvantage, is that one has to make detailed assumptions concerning the precise structure of the deterministic weights, see [25]. The use of i.i.d. weights implies exchangebility of the vertices, which is a nice property to work with.

The Erdős-Rényi random graph is a special case of the IRG. We obtain the ER random graph $G(n, \lambda/n)$ by taking $f(x) = x$ and $\Lambda_i \equiv \lambda$, for $1 \leq i \leq n$, which implies that $p_{ij} = \lambda/n$, for $1 \leq i < j \leq n$. Furthermore, the random graphs introduced in [23, 25, 71], i.e., the generalized random graph (GRG), the expected degree random graph and the Poissonian random graph (PRG) are special cases of the IRG. This will be shown in Section 4.1.2.

We now consider the distribution of the degree sequence and the typical distance, under the condition that the first two moments of Λ_1 do exist. We start with the degree sequence distribution.

If the first two moments of Λ_1 exist, then the degree of a vertex in the IRG is asymptotically a mixed Poisson random variable with mean the weight Λ_1. This contrasts with the configuration model, where the degree distribution of a vertex is, always, the distribution of the random weight Λ_1.

Using results from Chapter 4, the scaling of the typical distance is given by

$$\frac{\text{Dist}(G_n)}{\log n} \xrightarrow{P} \frac{1}{\log \nu}, \tag{1.6.8}$$

where

$$\nu = \frac{\mathbb{E}[\Lambda^2]}{\mathbb{E}[\Lambda]}, \tag{1.6.9}$$

and where we assume that $\nu > 1$, which guarantees that there is, **whp**, a giant component.

Observe that, by construction, in the CM the degrees of the vertices are independent, and the edges are not, whereas in the IRG model precisely the opposite is true, i.e., in the IRG model the edges are independent and the degrees are not.

The IRG is a special case of the random graph introduced in [18]. The 120 pages paper [18] uses kernels to define the construction of the graphs. Instead of giving the formal, and lengthy, definition of this random graph, we only mention that the IRG is a *rank 1 case*, see [18, §16.4]. The results of [18] are impressive, the authors show for the *finite variance* case that the expected typical distance scales with $\log n$, see Lemma 14.3 in [18]. In Chapter 4 we extend this result by giving the fluctuations around the mean of the typical distance.

Next, we show that the distribution of the fluctuations of the typical distance in IRG model does not converge to a limiting distribution for the *finite mean*

case, but depends on the number of nodes. This statement has been proved for the configuration model [54] and therefore the result is not new. Nevertheless, it is nice to illustrate this phenomenon using simulations. Below we will use the *Poissonian random graph* (PRG), which is a special case of the IRG.

Using Theorem 4.1 in Chapter 4, one can show that:

Corollary *Let G_n be the Poissonian random graph on n vertices. Assume that*

$$1 - F(x) \leq cx^{1-\tau}, \; x > 0, \tag{1.6.10}$$

for some given constant $c > 0$ and fixed $\tau > 3$, such that

$$\nu = \frac{\mathbb{E}\left[\Lambda^2\right]}{\mathbb{E}[\Lambda]} > 1. \tag{1.6.11}$$

For $k \geq 1$, let

$$\sigma_k = \lfloor \log_\nu k \rfloor \qquad and \qquad a_k = \sigma_k - \log_\nu k, \tag{1.6.12}$$

then there exists random variables $(R_a)_{a \in (-1,0]}$ such that as $n \to \infty$,

$$\mathbb{P}\left(\mathrm{Dist}(G_n) = \sigma_n + l\right) = \mathbb{P}\left(R_{a_n} = l\right) + o(1), \tag{1.6.13}$$

where l is a positive integer.

Figure 1.8: In the picture the empirical survival functions of the graph distance for 4 IRGs are plotted. Each plot is averaged over 1000 samples. The parameters are given by $\tau = 3.5$ and $\nu \approx 2.231381$.
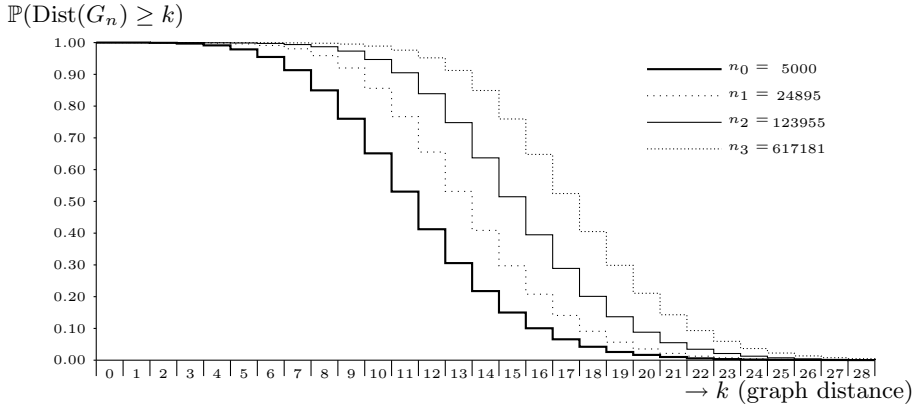
From (1.6.13) it follows that the limiting variable depends only the value a_n. Using simulations we will indicate that the above given corollary is indeed true. To this end, we pick $1 - F(x) = cx^{1-\tau}\mathbf{1}_{\{x>x_0\}}$ with $\tau = 3.5$, $c = 2.5981847$ and $x_0 = 0.7437937$, then

$$\nu = \frac{\mathbb{E}\left[\Lambda^2\right]}{\mathbb{E}[\Lambda]} \approx 2.231381,$$

where the distribution of Λ is given by F. We consider four PRG's, where the number of nodes is given by respectively $n_0 = 5000$, $n_1 = 24.895$, $n_2 = 123.955$ and $n_3 = 617.181$. Observe that $n_i \approx 5000\nu^{2i}$, $0 \le i \le 3$, and

$$a_{n_0} = 0.6117\dots \quad a_{n_1} = 0.6117\dots \quad a_{n_2} = 0.6117\dots \quad \text{and} \quad a_{n_3} = 0.6117\dots.$$

Thus, we expect that the distribution of the typical distance for these sizes, up to a shift, are the same, since, for $0 \le i \le 3$,

$$R_{a_{n_i}} = R_{0.6117\dots} \quad \text{and} \quad \sigma_{n_i} = \lfloor \log_\nu n_i \rfloor = \lfloor \log_\nu (5000\nu^{2i}) \rfloor = \lfloor \log_\nu 5000 + 2i \rfloor,$$

respectively. These conclusions are illustrated by Figure 1.8.

1.6.2   Dynamic models for complex networks

In this section, we will discuss some examples of dynamic random graphs. A dynamic random graph aims to explain how a network came to be as it is by modeling the growth of the network.

Most dynamic random graphs are based on preferential attachment. In these random graphs one chooses the node v proportional to its degree d_v, but one could also consider *non-linear* preferential attachment by choosing the node v proportional to its degree to some power, for example d_v^γ for some fixed $\gamma > 0$, see [24, 75]. The authors of [24] observe that this generalization cannot generate power-law distributions, except if $\gamma = 1$, and, as a consequence, if $\gamma \ne 1$, then the resulting random graphs are not scale-free.

It was once believed that preferential attachment is the underlying mechanism, which generates scale-free random graphs. In [10] a random graph is defined, where the attachment rule of the edges is essentially a minimalization problem. The authors of [10] prove that the resulting graphs are scale-free, and, therefore, show that preferential attachment is not the only way to obtain scale-free graphs.

The Barabási-Albert random graph

The *Barabási-Albert* (BA) random graph is introduced in [4]. The description given in this paper is not well-defined, but is shown in detail in [19]. Therefore, we present here the definition of the BA random graph as given in [19]. The random graph is defined by two parameters: n and m, where n, as usual, is the number of vertices and m is the number of edges emanating from each newly added vertex. We define the random graph for $m = 1$, and using this definition we give the definition for general $m > 1$. Let $G_1^{(1)}$ be a graph consisting of one vertex and one self-loop. Given the graph $G_t^{(1)}$, we construct the graph $G_{t+1}^{(1)}$ by adding the vertex $v_{t+1}^{(1)}$ together with a single edge directed from $v_{t+1}^{(1)}$ to w, where w is chosen randomly with probabilities

$$\mathbb{P}\left(w = v_s^{(1)} \,\big|\, G_t^{(1)}\right) = \begin{cases} \frac{d_t^{(1)}(s)}{2t+1} & \text{if } 1 \le s \le t; \\ \frac{1}{2t+1} & \text{if } s = t+1, \end{cases} \qquad (1.6.14)$$

where $d_t(s)^{(1)}$ is the degree of vertex $v_s^{(1)}$ in the graph $G_t^{(1)}$. Observe that the BA random graph is a preferential attachment random graph.

Figure 1.9: In the figure we see two graphs. The upper graph is a realization of the BA random graph $G_{18}^{(1)}$, whereas the lower one is a realization of $G_6^{(3)}$, which is constructed from the first graph $G_{18}^{(1)}$. The vertex $v_i^{(3)}$, for $1 \leq i \leq 6$, is obtained by merging the vertices $v_{3(i-1)+1}^{(1)}, v_{3(i-1)+1}^{(1)}, v_{3(i-1)+3}^{(1)}$ into a single vertex.
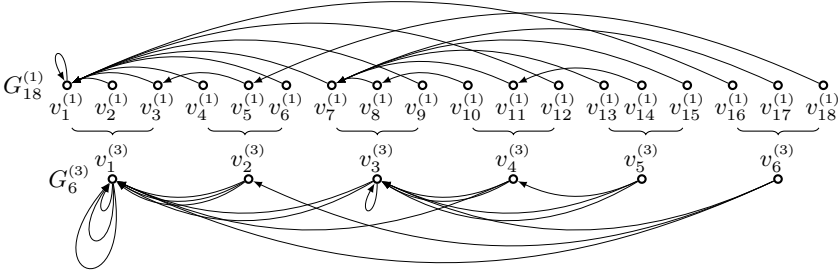
To construct $G_n^{(m)}$, for $m > 1$, we first construct $G_{mn}^{(1)}$. Then $G_n^{(m)}$ is obtained from $G_{mn}^{(1)}$ by merging the vertices in the latter graph. The graph $G_n^{(m)}$ consists of the vertices $\{v_i^{(m)}\}_{i=1}^n$ and the vertex $v_i^{(m)}$, $1 \leq i \leq n$, is obtained by merging the vertices $v_{m(i-1)+1}^{(1)}, v_{m(i-1)+2}^{(1)}, \dots, v_{m(i-1)+m}^{(1)}$. In Figure 1.9 two realizations of BA random graphs are given. The figure also illustrates the merging procedure how $G_6^{(3)}$ is obtained from $G_{18}^{(1)}$.

Now we consider the diameter, where we neglect the direction of the edges. For $m = 1$ the diameter scales as $\log n$, [74]. Bollobás and Riordan, [19, Theorem 1], show that for $m \geq 2$ **whp** the graph $G_n^{(m)}$ is connected and that for any fixed $\varepsilon > 0$, **whp**,

$$(1 - \varepsilon)\frac{\log n}{\log \log n} \leq \text{Diam}(G_n^{(m)}) \leq (1 + \varepsilon)\frac{\log n}{\log \log n}.$$

In the literature, most dynamic random graphs have the BA random graph as a special case, which is also true in this thesis. The PARID model and the GPAF model are such examples, which are introduced in this thesis in Chapter 3 and Chapter 5, respectively. Later on, in this section, a brief description of these models are given.

The Poissonian random graph; revisited

Norros and Reittu observed that in large real-life networks the number of edges joining two disjoint sets of vertices are approximately Poisson distributed given their degree sums. This was the starting point for Norros and Reittu to introduce a random graph in which the number of edges joining two disjoint sets of vertices is Poisson distributed: the *Poissonian Random Graph* (PRG), see [71].

The PRG was already introduced in Section 1.6.1 as a special case of the inhomogeneous random graph using a static description of the model. However, the PRG can also be described as a dynamic random graph. This description will be given next.

Fix a positive integer n, which is the number of nodes of the final graph. Let $\{\Lambda_n\}_{i=1}^n$ be an i.i.d sequence, where Λ_i, for $1 \leq i \leq n$, is the weight of

the vertex v_i and denote by L_t, $1 \leq t \leq n$, the *total weight* of the graph G_t:
$L_t = \Lambda_1 + \Lambda_2 + \cdots + \Lambda_t$.

First step: Start with the graph G_1 consisting of the vertex v_1 and a Poisson distributed number of loops with mean Λ_1.

Induction step: Start with the graph G_{t+1} as an identical copy of the graph G_t and add the vertex v_{t+1} with weight Λ_{t+1}. Now we perform the following two operations:

(i) **Thinning of the edges:** each edge of G_{t+1} is removed independently with probability
$$p_{t+1} = 1 - \frac{L_t}{L_{t+1}}.$$

(ii) **Adding edges:** we add new edges between the vertex v_{t+1} and all existing nodes v_n, $1 \leq n \leq t+1$. The number of edges we add between vertices v_n and v_{t+1} is $E_{t+1}(n, t+1)$ and this is a Poisson variable with mean
$$\mathbb{E}\big[E_{t+1}(n, t+1) \,|\, \{\Lambda_s\}_{s=1}^{t+1}\big] = \frac{\Lambda_n \Lambda_{t+1}}{L_{t+1}}.$$

The preferential attachment model with fitness

The BA random graph is a pure preferential attachment model, i.e., vertices with a high degree are preferred over others. Although the BA model has two parameters n and m, the resulting power-law exponent of the degree sequence is always equal to 3.

As a consequence of the preferential attachment, the degrees of the oldest vertices will be the highest, since they had the most opportunities to acquire edges. Numerous examples convincingly indicate that the connectivity of nodes in real-life networks does not depend on their age. For example, in social networks, not everybody makes friends at the same rate: some individuals are better in turning a random meeting into a lasting social link than others. On the World-Wide Web some pages, through a combination of good content and marketing, acquire a large number of links in a very short time, easily overtaking websites that have been around for much longer time, see [1].

We tend to associate these differences with some intrinsic quality of the nodes, such as the social skills of an individual, or the content of the web page. We call this the *fitness* of a node, which is similar by associating a weight to each vertex in the random graphs we have seen in Section 1.6.1.

In [12] *multiplicative fitness* is introduced. Each vertex v_i, $i \geq 1$, has a corresponding random fitness η_i, the multiplicative fitness. At time $t+1$ a new vertex v_{t+1} is added to the old graph, and from this vertex emanates an edge. The endpoint of this edge is connected to one of the old vertices. One chooses vertex v_i, $1 \leq i \leq t$, proportional to the degree of vertex v_i times its fitness: $\eta_i d_i(t)$. This model is extended in [35] by adding *additive fitness*, each vertex has also a random additive fitness ζ_i and one chooses a vertex v_i proportional to degree in G_t times the multiplicative constant plus the additive constant: $\eta_i d_i(t) + \xi_i$.

As an example, we consider the page rank of a newly added web page by Google. Imagine that we create a web page about the Honda VFR 800 motorcycle. It can be argued that if the page rank is high, then it is more likely that other (new) web pages will link to our web page, instead of other ones. Initially, the page rank of our web page is determined by its keywords and links (the additive fitness). Over time, other web pages will cite out web page. As a consequence the rank of our web page is determined by the web page itself and the number of other web pages citing this web page times some 'factor' (the multiplicative fitness).

Unfortunately, the papers [12, 35] are not rigorous. In this thesis, we consider two random graphs that use preferential attachment with fitness. The first model is the GPAF model, and this model is introduced in the next paragraph and considered more in-depth in Chapter 5. This model uses random multiplicative fitness and additive fitness. The second model is the PARID model. In this model the multiplicative fitness is constant, but the additive fitness is not. The additive fitness is some positive integer-valued random value (the initial degree of a node) plus some constant. We refer to Chapter 3 for a precise definition of the PARID model.

The geometric preferential attachment random graph

A large number of graph models have been introduced to describe complex networks, but often the underlying geometry is ignored. In general it is difficult to get rigorous results for properties like the degree distribution, typical distances or diameter, even if one disregards the geometry. However, in wireless ad-hoc networks the geometry is of great importance, since in these networks nodes are spread over some surface and nodes can only communicate with neighbors within a certain range, depending on the geometry. The *geometric preferential attachment* (GPA) random graph, introduced in [43], could be a model for wireless ad-hoc networks. Essentially, the model is a preferential attachment model with fitness, multiplicative and additive. The multiplicative fitness depends on the position of the vertices, but the additive constant is just a fixed value. The random graphs used in [43] or [44] are restricted to the finite variance case, since the power-law exponent is at least 3.

In Chapter 5 we introduce a random graph that uses the GPA random graph as a starting point. By introducing additive fitness, we obtain a random graph with geometric preferential attachment in which the power-law exponent can take any value larger than 2.

Chapter 2

Distances in random graphs with infinite mean degree sequence

Joint work with R. van der Hofstad, G. Hooghiemstra and D. Znamenski

Abstract

We study random graphs with an i.i.d. degree sequence of which the tail of the distribution function F is regularly varying with exponent $\tau \in [1,2]$. In particular, the degrees have infinite mean. Such random graphs can serve as models for complex networks where degree power laws are observed.

The minimal number of edges between two arbitrary nodes, also called the graph distance or the hopcount, is investigated when the size of the graph tends to infinity. This chapter is as a paper part of a sequel of three papers. The other two papers study the case where $\tau \in (2,3)$, and $\tau \in (3,\infty)$, respectively.

The main result of this paper is that the graph distance for $\tau \in (1,2)$ converges in distribution to a random variable with probability mass exclusively on the points 2 and 3. We also consider the case where we condition the degrees to be at most N^α for some $\alpha > 0$, where N is the size of the graph. For fixed $k \in \{0,1,2,\ldots\}$ and α such that $(\tau+k)^{-1} < \alpha < (\tau+k-1)^{-1}$, the hopcount converges to $k+3$ in probability, while for $\alpha > (\tau-1)^{-1}$, the hopcount converges to the same limit as for the unconditioned degrees. The proofs use extreme value theory.

2.1 Introduction

The study of complex networks has attracted considerable attention in the past decade. There are numerous examples of complex networks, such as co-authorship and citation networks of scientists, the World-Wide Web and Internet, metabolic networks, etc. The topological structure of networks affects the performance in

those networks. For instance, the topology of social networks affects the spread of information and disease (see e.g., [66, 78]), while the performance of traffic in Internet depends heavily on the topology of the Internet.

Measurements on complex networks have shown that many real networks have similar properties. A first example of such a fundamental network property is the fact that typical distances between nodes are small. This is called the 'small world' phenomenon, see the pioneering work of Watts [33], and the references therein. In Internet, for example, e-mail messages cannot use more than a threshold of physical links, and if the distances in Internet would be large, then e-mail service would simply break down. Thus, the graph of the Internet has evolved in such a way that typical distances are relatively small, even though the Internet is rather large.

A second, maybe more surprising, property of many networks is that the number of nodes with degree n falls off as an inverse power of n. This is called a 'power law degree sequence'. In Internet, the power law degree sequence was first observed in [40]. The observation that many real networks have the above properties has incited a burst of activity in network modelling. Most of the models use random graphs as a way to model the uncertainty and the lack of regularity in real networks. See [5, 66] and the references therein for an introduction to complex networks and many examples where the above two properties hold.

The current chapter presents a rigorous derivation for the random fluctuations of the graph distance between two arbitrary nodes (also called the hopcount) in a graph with i.i.d. degrees having *infinite* mean. The model with i.i.d. degrees is a variation of the *configuration model*, which was originally proposed by Newman, Strogatz and Watts [68], where the degrees originate from a given deterministic sequence. The observed power exponents are in the range from $\tau = 1.5$ to $\tau = 3.2$ (see [5, Table II] or [66, Table 3.1]). In a previous paper [54], the case $\tau > 3$ was investigated, while the case $\tau \in (2,3)$ was studied in [56]. Here we focus on the case $\tau \in (1,2)$, and study the typical distances between arbitrary connected nodes. In a forthcoming paper [55], we will survey the results from the different cases for τ, and investigate the connected components of the random graphs.

This section is organized as follows. In Section 2.1.1, we start by introducing the model, and in Section 2.1.2, we state our main results. In Section 2.1.3, we explain heuristically how the results are obtained. Finally, we describe related work in Section 2.1.4.

2.1.1   The model

Consider an i.i.d. sequence $D_1, D_2, \ldots, D_N$. Assume that $L_N = \sum_{j=1}^{N} D_j$ is even. When L_N is odd, then we increase D_N by 1, i.e., we replace D_N by $D_N + 1$. This change will make hardly any difference in what follows, and we will ignore it in the sequel.

We will construct a graph in which node j has degree D_j for all $1 \le j \le N$. We will later specify the distribution of D_j. We start with N separate nodes and incident to node j, we have D_j stubs which still need to be connected to build the graph.

The stubs are numbered in an arbitrary order from 1 to L_N. We continue by

matching at random the first stub with one of the $L_N - 1$ remaining stubs. Once paired, two stubs form an edge of the graph. Hence, a stub can be seen as the left or the right half of an edge. We continue the procedure of randomly choosing and pairing the next stub and so on, until all stubs are connected.

The probability mass function and the distribution function of the nodal degree are denoted by

$$\mathbb{P}(D_1 = j) = f_j, \quad j = 1, 2, \ldots, \quad \text{and} \quad F(x) = \sum_{j=1}^{\lfloor x \rfloor} f_j, \qquad (2.1.1)$$

where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. Our main assumption will be that

$$x^{\tau - 1} \left[1 - F(x) \right] \qquad (2.1.2)$$

is slowly varying at infinity for some $\tau \in (1, 2)$. This means that the random variables D_j obey a power law with infinite mean.

2.1.2   Main results

We define the graph distance H_N between the nodes 1 and 2 as the minimum number of edges that form a path from 1 to 2, where, by convention, this distance equals ∞ if 1 and 2 are not connected. Observe that the graph distance between two randomly chosen nodes is equal in distribution to H_N, because the nodes are exchangeable.

In this chapter, we will present two separate theorems for the case $\tau \in (1, 2)$. We also consider the boundary cases $\tau = 1$ (Theorem 2.3) and $\tau = 2$ (Remark 2.1.2). In Theorem 2.1, we take the sequence $D_1, D_2, \ldots, D_N$ of i.i.d. copies of D with distribution F, satisfying (2.1.2), with $\tau \in (1, 2)$. The result is that the graph distance or hopcount converges in distribution to a limit random variable with mass $p = p_F$, $1 - p$, on the values 2, 3, respectively. In this chapter the abbreviation **whp**, means that the involved event happens with probability converging to 1, as $N \to \infty$.

Theorem 2.1 *Fix $\tau \in (1, 2)$ in (2.1.2) and let $D_1, D_2, \ldots, D_N$ be a sequence of i.i.d. copies of D. Then,*

$$\lim_{N \to \infty} \mathbb{P}(H_N = 2) = 1 - \lim_{N \to \infty} \mathbb{P}(H_N = 3) = p_F \in (0, 1). \qquad (2.1.3)$$

One might argue that including degrees larger than $N - 1$ is artificial in a network with N nodes. In fact, in many real networks, the degree is bounded by a physical constant. Therefore, we also consider the case where the degrees are conditioned to be smaller than N^α, where α is an arbitrary positive number. Of course, we can condition on the degrees to be at most M, where M is fixed and independent on N, since in this case, the degrees are uniformly bounded, but this case is treated in [54]. Therefore, we consider cases where the degrees are conditioned to be at most a given power of N.

The result with conditioned degrees appears in the Theorem 2.2. It turns out that for $\alpha > 1/(\tau - 1)$, the conditioning has no influence in the sense that

the limit random variable is the same as that for the unconditioned case. This is not so strange, since the maximal degree is of order $N^{1/(\tau-1)}$, so that the conditioning does nothing in this case. However, for fixed $k \in \mathbb{N} \cup \{0\}$ and α such that $1/(\tau + k) < \alpha < 1/(\tau + k - 1)$, the graph distance converges to a degenerate limit random variable with mass 1 on the value $k + 3$. It would be of interest to extend the possible conditioning schemes, but we will not elaborate further on it in this chapter.

In the theorem below, we write $D^{(N)}$ for the random variable D conditioned on $D < N^\alpha$. Thus,

$$\mathbb{P}(D^{(N)} = j) = \frac{f_j}{\mathbb{P}(D < N^\alpha)}, \qquad 0 \le j < N^\alpha. \qquad (2.1.4)$$

Theorem 2.2 *Fix $\tau \in (1, 2)$ in (2.1.2) and let $D_1^{(N)}, D_2^{(N)}, \ldots, D_N^{(N)}$ be a sequence of i.i.d. copies of $D^{(N)}$.*

(i) *For $k \in \mathbb{N} \cup \{0\}$ and α such that $1/(\tau + k) < \alpha < 1/(\tau + k - 1)$,*

$$\lim_{N \to \infty} \mathbb{P}(H_N = k + 3) = 1. \qquad (2.1.5)$$

(ii) *If $\alpha > 1/(\tau - 1)$, then*

$$\lim_{N \to \infty} \mathbb{P}(H_N = 2) = 1 - \lim_{N \to \infty} \mathbb{P}(H_N = 3) = p_F, \qquad (2.1.6)$$

where $p_F \in (0, 1)$ is defined in Theorem 2.1.

The boundary case $\tau = 1$ and $\tau = 2$ are treated in Theorem 2.3 and Remark 2.1.2, below. We will prove that for $\tau = 1$, the hopcount converges to the value 2. For $\tau = 2$, we show by presenting two examples, that the limit behavior depends on the behavior of the slowly varying tail $x[1 - F(x)]$.

Theorem 2.3 *For $\tau = 1$ in (2.1.2) and let $D_1, D_2, \ldots, D_N$ be a sequence of i.i.d. copies of D. Then,*

$$\lim_{N \to \infty} \mathbb{P}(H_N = 2) = 1. \qquad (2.1.7)$$

Remark. For $\tau = 2$ in (2.1.2) and with $D_1, D_2, \ldots, D_N$ a sequence of i.i.d. copies of D, the limit behavior of H_N depends on the slowly varying tail $x[1 - F(x)]$, $x \to \infty$. In Section 2.4.2, we present two examples, where we have, depending on the slowly varying function $x[1 - F(x)]$, different limit behavior for H_N. We present an example with $\lim_{N \to \infty} \mathbb{P}(H_N \le k) = 0$, for all fixed integers k, as $N \to \infty$, and a second example where $H_N \in \{2, 3\}$, **whp**, as $N \to \infty$.

2.1.3   Heuristics

When $\tau \in (1, 2)$, we consider two different cases. In Theorem 2.1, the degrees are not conditioned, while in Theorem 2.2 we condition on each node having a degree smaller than N^α. We now give a heuristic explanation of our results.

In two previous papers [54, 56], the cases $\tau \in (2,3)$ and $\tau > 3$ have been treated. In these cases, the probability mass function $\{f_j\}$ introduced in (2.1.1) has a finite mean, and the number of nodes on graph distance n from node 1 can be coupled to the number of individuals in the nth generation of a branching process with offspring distribution $\{g_j\}$ given by

$$g_j = \frac{j+1}{\mu} f_j, \qquad (2.1.8)$$

where $\mu = \mathbb{E}[D_1]$. For $\tau \in (1,2)$, as we are currently investigating, we have $\mu = \infty$, and the branching process used in [54, 56] does not exist.

When we do not condition on D_j being smaller than N^α, then L_N is the i.i.d. sum of N random variables $D_1, D_2, \ldots, D_N$, with infinite mean. It is well known that in this case the bulk of the contribution to $L_N = N^{1/(\tau-1)+o(1)}$ comes from a *finite* number of nodes which have giant degrees (the so-called *giant nodes*). A basic fact in the configuration model is that two sets of stubs of sizes n and m are connected **whp** when nm is at least of order L_N. Since the giant nodes have degree roughly $N^{1/(\tau-1)}$, which is much larger than $\sqrt{L_N}$, they are all attached to each other, thus forming a complete graph of giant nodes. Each stub is **whp** attached to a giant node, and, therefore, the distance between any two nodes is, **whp**, at most 3. In fact, this distance equals 2 when the two nodes are connected to the *same* giant node, and is 3 otherwise. In particular, for $\tau = 1$, the quotient $D_{(N)}/L_N$ converges to 1 in probability, and consequently the hopcount converges to 2, in probability.

When we truncate the distribution as in (2.1.4), with $\alpha > 1/(\tau-1)$, we hardly change anything since without truncation **whp** all degrees are below N^α. On the other hand, if $\alpha < 1/(\tau-1)$, then, with truncation, the largest nodes have degree of order N^α, and $L_N \sim N^{1+\alpha(2-\tau)}$. Again, the bulk of the total degree L_N comes from nodes with degree of the order N^α, so that now these are the giant nodes. Hence, for $1/\tau < \alpha < 1/(\tau-1)$, the largest nodes have degrees much larger than $\sqrt{L_N}$, and thus, **whp,** still constitute a complete graph. The number of giant nodes converges to infinity, as $N \to \infty$. Therefore, the probability that two arbitrary nodes are connected to the *same* giant node converges to 0. Consequently, the hopcount equals 3, **whp**. If $\alpha < 1/\tau$, then the giant nodes no longer constitute a complete graph, so that the hopcount can be greater than 3. For almost every $\alpha < 1/\tau$, the hopcount coverges to a *single* value. The behavior of the hopcount for the cases that $\alpha = 1/(\tau+k)$ for $k \in \mathbb{N} \cup \{0\}$, will be dependent on the slowly varying function in (1.2), as is the case for $\tau = 2$. We do expect that the hopcount converges to at most 2 values in these cases.

The proof in this chapter is based on detailed asymptotics of the sum of N i.i.d. random variables with infinite mean, as well as on the scaling of the order statistics of such random variables. The scaling of these order statistics is crucial in the definition of the giant nodes which are described above. The above considerations are the basic idea in the proof of Theorem 2.1. In the proof of Theorem 2.2, we need to investigate what the conditioning does to the scaling of both the total degree L_N, as well as to the largest degrees.

2.1.4   Related work

The above model is a variation of the configuration model. In the usual configuration model one often starts from a given *deterministic* degree sequence. In our model, the degree sequence is an i.i.d. sequence $D_1, \ldots, D_N$ with distribution equal to a power law. The reason for this choice is that we are interested in models for which all nodes are exchangeable, and this is not the case when the degrees are fixed. The study of this variation of the configuration model was started in [68] for the case $\tau > 3$ and studied by Norros and Reittu [70] in case $\tau \in (2, 3)$.

For a survey of complex networks, power law degree sequences and random graph models for such networks, see [5] and [66]. There, a heuristic is given why the hopcount scales proportionally to $\log N$, which is originally from [68]. The argument uses a variation of the power law degree model, namely, a model where an exponential cut off is present. An example of such a degree distribution is

$$f_j = C j^{-\tau} e^{-j/\kappa} \qquad (2.1.9)$$

for some $\kappa > 0$. The size of κ indicates up to what degree the power law still holds, and where the exponential cut off starts to set in. The above model is treated in [54] for any $\kappa < \infty$, but, for $\kappa = \infty$, falls within the regimes where $\tau \in (2, 3)$ in [56] and within the regime in this chapter for $\tau \in (1, 2)$. In [68], the authors conclude that since the limit as $\kappa \to \infty$ does not seem to converge, the 'average distance is not well defined when $\tau < 3$'. In this chapter, as well as in [56], we show that the average distance *is* well defined, but it scales differently from the case where $\tau > 3$.

In [55], we give a survey to the results for the hopcount in the three different re gimes $\tau \in (1, 2)$, $\tau \in (2, 3)$ and $\tau > 3$. There, we also prove results for the connectivity properties of the random graph in these cases. These results assume that the expected degree is larger than 2. This is always the case when $\tau \in (1, 2)$, and stronger results have been shown there. We prove that the largest connected component has **whp** size $N(1 + o(1))$. When $\tau \in (1, \frac{3}{2})$ we even prove that the graph is **whp** connected. When $\tau > \frac{3}{2}$ this is not true, and we investigate the structure of the remaining 'dust' that does not belong to the largest connected component. The analysis makes use of the results obtained in this chapter for $\tau \in (1, 2)$. For instance, it will be crucial that the probability that two arbitrary nodes are connected converges to 1.

There is substantial related work on the configuration model for the cases $\tau \in (2, 3)$ and $\tau > 3$. References are included in the paper [56] for the case $\tau \in (2, 3)$, and in [54] for $\tau > 3$. We again refer to the references in [55] and [5, 66] for more details. The graph distance for $\tau \in (1, 2)$, that we study here, has, to our best knowledge, not been studied before. Values of $\tau \in (1, 2)$ have been observed in networks of e-mail messages and networks where the nodes consist of software packages (see [66, Table II]), for which our configuration model with $\tau \in (1, 2)$ can possibly give a good model.

In [2], random graphs are considered with a degree sequence that is *precisely* equal to a power law, meaning that the number of nodes with degree n is precisely proportional to $n^{-\tau}$. Aiello *et al.* [2] show that the largest connected component is of the order of the size of the graph when $\tau < \tau_0 = 3.47875\ldots$, where τ_0 is the solution of $\zeta(\tau - 2) - 2\zeta(\tau - 1) = 0$, and where ζ is the Riemann zeta function.

When $\tau > \tau_0$, the largest connected component is of smaller order than the size of the graph and more precise bounds are given for the largest connected component. When $\tau \in (1, 2)$, the graph is **whp** connected. The proofs of these facts use couplings with branching processes and strengthen previous results due to Molloy and Reed [64, 65]. See also [2] for a history of the problem and references predating [64, 65]. See [3] for an introduction to the mathematical results of various models for complex networks (also called massive graphs), as well as a detailed account of the results in [2].

A detailed account for a related model can be found in [25] and [26], where links between nodes i and j are present with probability equal to $w_i w_j$ divided by $\sum_l w_l$ for some 'expected degree vector' $w = (w_1, \ldots, w_N)$. Chung and Lu [25] show that when w_i is proportional to $i^{-\frac{1}{\tau-1}}$, the average distance between pairs of nodes is proportional $\log N(1 + o(1))$ when $\tau > 3$, and equal to $2\frac{\log \log N}{|\log(\tau-2)|}(1 + o(1))$ when $\tau \in (2, 3)$. In their model, also $\tau \in (1, 2)$ is possible, and in this case, similarly to $\tau \in (1, \frac{3}{2})$ in this chapter, the graph is connected **whp**.

The difference between this model and ours is that the nodes are not exchangeable in [25], but the observed phenomena are similar. This can be understood as follows. Firstly, the actual degree vector in [25] should be close to the expected degree vector. Secondly, for the expected degree vector, we can compute that the number of nodes for which the degree is at least n equals

$$|\{i : w_i \geq n\}| = |\{i : ci^{-\frac{1}{\tau-1}} \geq n\}| \propto n^{-\tau+1},$$

where the proportionality constant depends on N. Thus, one expects that the number of nodes with degree at least n decreases as $n^{-\tau+1}$, similarly as in our model. In [26], Chung and Lu study the sizes of the connected components in the above model. The advantage of working with the 'expected degree model' is that different links are present independently of each other, which makes this model closer to the classical random graph $G(p, N)$.

2.1.5   Organization of the chapter

The main body of the chapter consists of the proofs of Theorem 2.1 in Section 2.2 and the proof of Theorem 2.2 in Section 2.3. Both proofs contain a technical lemma and in order to make the argument more transparent, we have postponed the proofs of these lemmas to the appendix. Section 4 contains the proof of Theorem 2.3 and two examples for the case $\tau = 2$. Section 2.5 contains simulation results, conclusions and open problems.

2.2   Proof of Theorem 2.1

In this section, we prove Theorem 2.1, which states that the hopcount between two arbitrary nodes has **whp** a non-trivial distribution on 2 and 3. We start with an outline of our proof.

Below, we introduce an event $\mathcal{A}_{\varepsilon,N}$, such that when $\mathcal{A}_{\varepsilon,N}$ occurs, the hopcount between two arbitrary nodes is either 2 or 3. We then prove that $\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) < \varepsilon$, for

$N \geq N_\varepsilon$ (see Lemma 2.5 below). For this we need a modification of the extreme value theorem for the k largest degrees, for all $k \in \mathbb{N}$.

We introduce

$$D_{(1)} \leq D_{(2)} \leq \cdots \leq D_{(N)},$$

to be the order statistics of $D_1, \ldots, D_N$, so that $D_{(1)} = \min\{D_1, \ldots, D_N\}$, $D_{(2)}$ is the second smallest degree, etc. Let (u_N) be an increasing sequence such that

$$\lim_{N \to \infty} N\left[1 - F(u_N)\right] = 1. \tag{2.2.1}$$

It is well known that the order statistics of the degrees, as well as the total degree, are governed by u_N in the case that $\tau \in (1, 2)$. The following lemma shows this in detail. In the lemma $E_1, E_2, \ldots$ is an i.i.d. sequence of exponential random variables with unit mean and $\Gamma_j = E_1 + E_2 + \ldots + E_j$, hence Γ_j has a gamma distribution with parameters j and 1. Throughout this chapter, equality in distribution is denoted by the symbol $\stackrel{d}{=}$, whereas $\stackrel{d}{\to}$ denotes convergence in distribution. .

Lemma 2.4 (Convergence in distribution of order statistics)
For any $k \in \mathbb{N}$,

$$\left(\frac{L_N}{u_N}, \frac{D_{(N)}}{u_N}, \ldots, \frac{D_{(N-k+1)}}{u_N}\right) \stackrel{d}{\to} (\eta, \xi_1, \ldots, \xi_k), \text{ as } N \to \infty, \tag{2.2.2}$$

where $(\eta, \xi_1, \ldots, \xi_k)$ is a random vector which can be represented by

$$(\eta, \xi_1, \ldots, \xi_k) \stackrel{d}{=} \left(\sum_{j=1}^{\infty} \Gamma_j^{-1/(\tau-1)}, \Gamma_1^{-1/(\tau-1)}, \ldots, \Gamma_k^{-1/(\tau-1)}\right). \tag{2.2.3}$$

Moreover,

$$\xi_k \to 0 \text{ in probability, as } k \to \infty. \tag{2.2.4}$$

Proof.   Because $\tau - 1 \in (0, 1)$, the proof is a direct consequence of [63, Theorem 1'], and the continuous mapping theorem [14, Theorem 5.1], which together yield that on $\mathbb{R} \times \mathbb{R}^\infty$, equipped with the product topology, we have

$$(S_N^\#, Z^{(N)}) \stackrel{d}{\to} (S^\#, Z), \tag{2.2.5}$$

where $S_N^\# = u_N^{-1} L_N$, $Z^{(N)} = u_N^{-1}(D_{(N)}, \ldots, D_{(1)}, 0, 0, \ldots)$, and $Z_j = \Gamma_j^{-1/(\tau-1)}$, $j \geq 1$.

If we subsequently take the projection from $\mathbb{R} \times \mathbb{R}^\infty \mapsto \mathbb{R}^{k+1}$, defined by

$$\pi(s, z) = (s, z_1, \ldots, z_k), \tag{2.2.6}$$

i.e., we keep the sum and the k largest order statistics, then we obtain (2.2.2) and (2.2.3) from, again, the continuous mapping theorem. Finally, (2.2.4) follows because the series $\sum_{j=1}^{\infty} Z_j$ converges almost surely. $\square$

We need some additional notation. In this section, we define the *giant nodes* as the k_ε largest nodes, i.e., those nodes with degrees $D_{(N)}, \ldots, D_{(N-k_\varepsilon+1)}$, where k_ε is some function of ε, to be chosen below. We define

$$\mathcal{A}_{\varepsilon,N} = \mathcal{B}_{\varepsilon,N} \cap \mathcal{C}_{\varepsilon,N} \cap \mathcal{D}_{\varepsilon,N}, \tag{2.2.7}$$

where

(i) $\mathcal{B}_{\varepsilon,N}$ is the event that the stubs of node 1 and node 2 are attached exclusively to stubs of giant nodes;

(ii) $\mathcal{C}_{\varepsilon,N}$ is the event that any two giant nodes are attached to each other; and

(iii) $\mathcal{D}_{\varepsilon,N}$ is defined as

$$\mathcal{D}_{\varepsilon,N} = \{D_1 \le q_\varepsilon, D_2 \le q_\varepsilon\},$$

where $q_\varepsilon = \min\{n : 1 - F(n) < \varepsilon/8\}$.

The reason for introducing the above events is that on $\mathcal{A}_{\varepsilon,N}$, the hopcount or graph distance is either 2 or 3. Indeed, on $\mathcal{B}_{\varepsilon,N}$, both node 1 and node 2 are attached exclusively to giant nodes. On the event $\mathcal{C}_{\varepsilon,N}$, giant nodes have mutual graph distance 1. Hence, on the intersection $\mathcal{B}_{\varepsilon,N} \cap \mathcal{C}_{\varepsilon,N}$, the hopcount between node 1 and node 2 is at most 3. The event $\mathcal{D}_{\varepsilon,N}$ prevents that the hopcount can be equal to 1, because the probability on the intersection of $\{H_N = 1\}$ with $\mathcal{D}_{\varepsilon,N}$ can be bounded by $q_\varepsilon^2/N \to 0$ (see the first part of the proof of Theorem 2.1 for details). Observe that the expected number of stubs of node 1 is not bounded, since the expectation of a random variable with distribution (2.1.2) equals $+\infty$. Putting things together we see that if we can show that $\mathcal{A}_{\varepsilon,N}$ happens **whp**, then the hopcount is either 2 or 3. The fact that $\mathcal{A}_{\varepsilon,N}$ happens **whp** is the content of Lemma 2.5, where we show that $\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) < \varepsilon$, for $N \ge N_\varepsilon$. Finally, we observe that the hopcount between node 1 and 2 is precisely equal to 2, if at least one stub of node 1 and at least one stub of node 2 is attached to the same giant node, and equal to 3 otherwise.

The events $\mathcal{B}_{\varepsilon,N}$ and $\mathcal{C}_{\varepsilon,N}$ depend on the integer k_ε, which we will take to be large for ε small, and will be defined now. The choice of the index k_ε is rather technical, and depends on the distributional limits of Lemma 2.4. Since $L_N/u_N = (D_1 + D_2 + \ldots + D_N)/u_N$ converges in distribution to the random variable η with support $(0, \infty)$, we can find a_ε, such that

$$\mathbb{P}(L_N < a_\varepsilon u_N) < \varepsilon/36, \quad \forall N. \tag{2.2.8}$$

This follows since convergence in distribution implies tightness of the sequence L_N/u_N ([14, p. 9]), so that we can find a closed subinterval $I \subset (0, \infty)$, with

$$\mathbb{P}(L_N/u_N \in I) > 1 - \varepsilon, \quad \forall N.$$

We next define b_ε, which is rather involved. It depends on ε, the quantile q_ε, the value a_ε defined above and the value of $\tau \in (1, 2)$ and reads

$$b_\varepsilon = \left(\frac{\varepsilon^2 a_\varepsilon}{2304 q_\varepsilon} \right)^{\frac{1}{2-\tau}}, \tag{2.2.9}$$

where the peculiar integer 2304 is the product of 8^2 and 36. Given b_ε, we take k_ε equal to the minimal k such that

$$\mathbb{P}(\xi_k \geq b_\varepsilon/2) \leq \varepsilon/72. \qquad (2.2.10)$$

It follows from (2.2.4) that such a number k exists. We have now defined the constants that we will use in the proof, and we next claim that the probability of $\mathcal{A}_{\varepsilon,N}^c$ is at most ε:

Lemma 2.5 (The good event has high probability)
For each $\varepsilon > 0$, there exists N_ε, such that

$$\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) < \varepsilon, \quad N \geq N_\varepsilon. \qquad (2.2.11)$$

The proof of this lemma is rather technical and can be found in appendix 2.A.1. We will now complete the proof of Theorem 2.1 subject to Lemma 2.5.

Proof of Theorem 2.1. As seen in the discussion following the introduction of the event $\mathcal{A}_{\varepsilon,N}$, this event implies the event $\{H_N \leq 3\}$, so that $\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) < \varepsilon$ induces that the event $\{H_N \leq 3\}$ occurs with probability at least $1 - \varepsilon$.

The remainder of the proof consist of two parts. In the first part we show that $\mathbb{P}(\{H_N = 1\} \cap \mathcal{A}_{\varepsilon,N}) < \varepsilon$. In the second part we prove that

$$\lim_{N \to \infty} \mathbb{P}(H_N = 2) = p_F,$$

for some $0 < p_F < 1$. Since ε is an arbitrary positive number, the above statements yield the content of the theorem.

We turn to the first part. The event $\{H_N = 1\}$ occurs *iff* at least one stub of node 1 connects to a stub of node 2. For $j \leq D_1$, we denote by $\{[1.j] \to [2]\}$ the event that j^{th} stub of node 1 connects to a stub of node 2. Then, with $\mathbb{P}_N$ the conditional probability given the degrees $D_1, D_2, \ldots, D_N$,

$$\mathbb{P}(\{H_N = 1\} \cap \mathcal{A}_{\varepsilon,N}) \leq \mathbb{E}\left[\sum_{j=1}^{D_1} \mathbb{P}_N(\{[1.j] \to [2]\} \cap \mathcal{A}_{\varepsilon,N})\right]$$

$$\leq \mathbb{E}\left[\sum_{j=1}^{D_1} \frac{D_2}{L_N - 1}\mathbf{1}_{\{\mathcal{A}_{\varepsilon,N}\}}\right] \leq \frac{q_\varepsilon^2}{N - 1} < \varepsilon, \qquad (2.2.12)$$

for large enough N, since $L_N \geq N$.

We next prove that $\lim_{N \to \infty} \mathbb{P}(H_N = 2) = p$, for some $0 < p < 1$. Since by definition for any $\varepsilon > 0$,

$$\max\{\mathbb{P}(\mathcal{B}_{\varepsilon,N}^c), \mathbb{P}(\mathcal{D}_{\varepsilon,N}^c)\} \leq \mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) \leq \varepsilon,$$

we have that

$$|\mathbb{P}(H_N = 2) - \mathbb{P}(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N})| \leq \left|\mathbb{P}(H_N = 2)\left(1 - \frac{1}{\mathbb{P}(\mathcal{B}_{\varepsilon,N})}\right)\right|$$

$$+ \left|\frac{\mathbb{P}(H_N = 2) - \mathbb{P}(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \cap \mathcal{B}_{\varepsilon,N})}{\mathbb{P}(\mathcal{B}_{\varepsilon,N})}\right|$$

$$\leq \frac{2\mathbb{P}(\mathcal{B}_{\varepsilon,N}^c) + \mathbb{P}(\mathcal{D}_{\varepsilon,N}^c)}{\mathbb{P}(\mathcal{B}_{\varepsilon,N})} \leq \frac{3\varepsilon}{1 - \varepsilon},$$

uniformly in N, for N sufficiently large. If we show that

$$\lim_{N \to \infty} \mathbb{P}\left(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}\right) = r(\varepsilon), \qquad (2.2.13)$$

then there exists a double limit

$$p_F = \lim_{\varepsilon \downarrow 0} \lim_{N \to \infty} \mathbb{P}\left(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}\right) = \lim_{N \to \infty} \mathbb{P}\left(H_N = 2\right).$$

Moreover, if we can bound $r(\varepsilon)$ away 0 and 1, uniformly in ε, for ε small enough, then we also obtain that $0 < p_F < 1$.

In order to prove the existence of the limit in (2.2.13) we claim that

$$\mathbb{P}_N\left(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}\right)$$

can be written as the ratio of two polynomials, where each polynomial only involves components of the vector

$$\left(\frac{D_{(N)}}{u_N}, \ldots, \frac{D_{(N-k_\varepsilon+1)}}{u_N}, \frac{1}{u_N}\right). \qquad (2.2.14)$$

Due to (2.2.2), this vector converges in distribution to $(\xi_1, \ldots, \xi_{k_\varepsilon}, 0)$. Hence, by the continuous mapping theorem [14, Theorem 5.1, p. 30], we have the existence of the limit (2.2.13). We now prove the above claim.

Indeed, the hopcount between nodes 1 and 2 is 2 *iff* both nodes are connected to the same giant node. For any $0 \leq i \leq D_1$, $0 \leq j \leq D_2$ and $0 \leq k < k_\varepsilon$, let $\mathcal{F}_{i,j,k}$ be the event that both the ith stub of node 1 and the jth stub of node 2 are connected to the node with the $(N-k)^{\text{th}}$ largest degree. Then, conditionally on the degrees $D_1, D_2, \ldots, D_N$,

$$\mathbb{P}_N\left(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}\right) = \mathbb{P}_N\left(\bigcup_{i=1}^{D_1} \bigcup_{j=1}^{D_2} \bigcup_{k=0}^{k_\varepsilon-1} \mathcal{F}_{i,j,k} \,\bigg|\, \mathcal{B}_{\varepsilon,N}\right),$$

where the right-hand side can be written by the inclusion-exclusion formula, as a linear combination of terms

$$\mathbb{P}_N\left(\mathcal{F}_{i_1,j_1,k_1} \cap \cdots \cap \mathcal{F}_{i_n,j_n,k_n} \,|\, \mathcal{B}_{\varepsilon,N}\right). \qquad (2.2.15)$$

It is not difficult to see that these probabilities are ratios of polynomials of components of (2.2.14). For example,

$$\mathbb{P}_N\left(\mathcal{F}_{i,j,k} \,|\, \mathcal{B}_{\varepsilon,N}\right) = \frac{D_{(N-k)}(D_{(N-k)} - 1)}{(D_{(N-k_\varepsilon+1)} + \cdots + D_{(N)})(D_{(N-k_\varepsilon+1)} + \cdots + D_{(N)} - 1)}, \qquad (2.2.16)$$

so that dividing both the numerator and the denominator of (2.2.16) by u_N^2, we obtain that the right-hand side of (2.2.16) is indeed a ratio of two polynomials of the vector given in (2.2.14). Similar arguments hold for general terms of the form in (2.2.15). Hence, $\mathbb{P}_N(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N})$ itself can be written as a ratio of two polynomials where the polynomial in the denominator is strictly positive. Therefore, the limit in (2.2.13) exists.

We finally bound $r(\varepsilon)$ from 0 and 1 uniformly in ε, for any $\varepsilon < 1/2$. Since the hopcount between nodes 1 and 2 is 2, given $\mathcal{B}_{\varepsilon,N}$, if they are both connected to the node with largest degree, then

$$\mathbb{P}(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}) \geq \mathbb{E}[\mathbb{P}_N(\mathcal{F}_{1,1,0} \,|\, \mathcal{B}_{\varepsilon,N})],$$

and by (2.2.16) we have

$$
\begin{aligned}
r(\varepsilon) \quad &= \lim_{N\to\infty} \mathbb{P}\left(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}\right) \geq \lim_{N\to\infty} \mathbb{E}\left[\frac{D_{(N)}(D_{(N)}-1)}{(D_{(N)}+\cdots+D_{(N-k_\varepsilon+1)}-1)^2}\right] \\
&= \mathbb{E}\left[\left(\frac{\xi_1}{\xi_1+\cdots+\xi_{k_\varepsilon}}\right)^2\right] \geq \mathbb{E}\left[\left(\frac{\xi_1}{\eta}\right)^2\right].
\end{aligned}
$$

On the other hand, conditionally on $\mathcal{B}_{\varepsilon,N}$, the hopcount between nodes 1 and 2 is at least 3, when all stubs of the node 1 are connected to the node with largest degree, and all stubs of the node 2 are connected to the node with the one but largest degree. Hence, for any $\varepsilon < 1/2$ and similarly to (2.2.16), we have

$$
\begin{aligned}
r(\varepsilon) &= \lim_{N\to\infty} \mathbb{P}\left(\{H_N = 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}\right) \leq 1 - \lim_{N\to\infty} \mathbb{P}\left(\{H_N > 2\} \cap \mathcal{D}_{\varepsilon,N} \,|\, \mathcal{B}_{\varepsilon,N}\right) \\
&\leq 1 - \lim_{N\to\infty} \mathbb{P}\left(\{H_N > 2\} \cap \mathcal{D}_{\frac{1}{2},N} \,|\, \mathcal{B}_{\varepsilon,N}\right) \\
&\leq 1 - \lim_{N\to\infty} \mathbb{E}\left[\left(\prod_{i=0}^{D_1} \frac{D_{(N)}-2i}{D_{(N)}+\cdots+D_{(N-k_\varepsilon+1)}-D_1}\right)\right. \\
&\qquad\qquad\qquad \left.\times \left(\prod_{i=0}^{D_2} \frac{D_{(N-1)}-2i}{D_{(N)}+\cdots+D_{(N-k_\varepsilon+1)}-D_2}\right) \mathbf{1}_{\{\mathcal{D}_{\frac{1}{2},N}\}}\right] \\
&\leq 1 - \mathbb{E}\left[\left(\frac{\xi_1\xi_2}{\eta^2}\right)^{q_{\frac{1}{2}}}\right],
\end{aligned}
$$

because: (i) the event $\mathcal{D}_{\frac{1}{2},N}$ implies that both $D_1 \leq q_{\frac{1}{2}}$ and $D_2 \leq q_{\frac{1}{2}}$, (ii) the event $\mathcal{B}_{\varepsilon,N}$ implies that all stubs of the normal nodes 1 and 2 are connected to stubs of giant nodes, (iii) Lemma 2.4 implies

$$\mathbb{E}\left[\left(\prod_{i=0}^{q_{\frac{1}{2}}} \frac{D_{(N)}-2i}{D_{(N)}+\cdots+D_{(N-k_\varepsilon+1)}-D_1}\right)\left(\prod_{i=0}^{q_{\frac{1}{2}}} \frac{D_{(N-1)}-2i}{D_{(N)}+\cdots+D_{(N-k_\varepsilon+1)}-D_2}\right)\right]$$

is equal to

$$\mathbb{E}\left[\left(\frac{\xi_1}{\xi_1+\cdots+\xi_{k_\varepsilon}}\right)^{q_{\frac{1}{2}}}\left(\frac{\xi_2}{\xi_1+\cdots+\xi_{k_\varepsilon}}\right)^{q_{\frac{1}{2}}}\right]$$

if $N \to \infty$, and (iv) $\xi_1 + \cdots + \xi_{k_\varepsilon} \leq \eta$.

Both expectations

$$\mathbb{E}\left[\left(\frac{\xi_1}{\eta}\right)^2\right] \qquad \text{and} \qquad \mathbb{E}\left[\left(\frac{\xi_1\xi_2}{\eta^2}\right)^{q_{\frac{1}{2}}}\right], \qquad\qquad (2.2.17)$$

are strictly positive and independent of ε. Hence, for any $\varepsilon < 1/2$, the quantity $r(\varepsilon)$ is bounded away from 0 and 1, where the bounds are *independent of* ε, and thus $0 < p_F < 1$. This completes the proof of Theorem 2.1 subject to Lemma 2.5. $\square$

2.3   Proof of Theorem 2.2

In Theorem 2.2, we consider the hopcount in the configuration model with degrees an i.i.d. sequence with a truncated distribution given by (2.1.4), where D has distribution F satisfying (2.1.2). We distinguish two cases: (i) $\alpha < 1/(\tau - 1)$ and (ii) $\alpha > 1/(\tau - 1)$. Since part (ii) is simpler to prove than part (i), we start with part (ii).

Proof of Theorem 2.2(ii). We have to prove that the limit distribution of H_N is a mixed distribution with probability mass p_F on 2 and probability mass $1 - p_F$ on 3, where p_F is given by Theorem 2.1.

As before, we denote by $D_1, D_2, \ldots, D_N$ the i.i.d. sequence without conditioning. We bound the probability that for at least one index $i \in \{1, 2, \ldots, N\}$ the degree D_i exceeds N^α, by

$$\mathbb{P}\left(\bigcup_{i=1}^N \{D_i > N^\alpha\} \right) \leq \sum_{i=1}^N \mathbb{P}(D_i > N^\alpha) = N\left[1 - F\left(N^\alpha\right)\right] \leq N^{-\varepsilon},$$

for some positive ε, because $\alpha > 1/(\tau - 1)$. We can therefore couple the i.i.d. sequence $\vec{D}^{(N)} = (D_1^{(N)}, D_2^{(N)}, \ldots, D_N^{(N)})$ to the sequence $\vec{D} = (D_1, D_2, \ldots, D_N)$, where the probability of a miscoupling, i.e., a coupling such that $\vec{D}^{(N)} \neq \vec{D}$, is at most $N^{-\varepsilon}$. Therefore, the result of Theorem 2.1 carries over to case (ii) in Theorem 2.2. $\qquad\square$

Proof of Theorem 2.2(i). This proof is more involved. We start with an outline of the proof. Fix

$$\frac{1}{\tau + k} < \alpha < \frac{1}{\tau + k - 1}, \tag{2.3.1}$$

with $k \in \mathbb{N} \cup \{0\}$ and define

$$M_N = \sum_{n=1}^N D_n^{(N)}. \tag{2.3.2}$$

From [41, Theorem 1, p. 281], the expected value of M_N is given by

$$\mathbb{E}[M_N] = \frac{N}{F(N^\alpha)} \sum_{i=0}^{N^\alpha - 1} \mathbb{P}(D > i) = N^{1 + \alpha(2 - \tau)} \ell(N), \tag{2.3.3}$$

where $N \mapsto \ell(N)$ is slowly varying at infinity. In the sequel, we will use the same $\ell(N)$, for different slowly varying functions, so that the value of $\ell(N)$ may change from line to line.

For the outline, we assume that M_N has roughly the same size as $\mathbb{E}[M_N]$ in (2.3.3). The proof consists of showing that $\mathbb{P}(H_N \leq k + 2) = o(1)$ and $\mathbb{P}(H_N > k + 3) = o(1)$. We will sketch the proof of each of these results. To prove that $\mathbb{P}(H_N \leq k + 2) = o(1)$, note that **whp** the degrees of nodes 1 and 2 are bounded by q_ε for some large q_ε. Therefore, on this event, the number of nodes that can be reached from node 1 in $l - 1$ steps is at most $q_\varepsilon N^{(l-2)\alpha}$, and the number of stubs attached to nodes at distance $l - 1$ is at most $q_\varepsilon N^{(l-1)\alpha}$. The probability that one of these stubs is attached to a stub of node 2, making H_N at most l, is of

the order $q_\varepsilon^2 N^{(l-1)\alpha}/M_N$. By (2.3.3) and the assumed concentration of M_N, this is at most $q_\varepsilon^2 \ell(N) N^{(l-3+\tau)\alpha-1} = o(1)$, whenever $\alpha < 1/(l-3+\tau)$. Applying this to $l = k + 2$, we see that this probability is $o(1)$ if $\alpha < 1/(k+\tau-1)$.

To prove that $\mathbb{P}(H_N > k + 3) = o(1)$, we use the notion of giant nodes in a similar way as in the proof of Theorem 2.1. Due to the conditioning on the degree, Lemma 2.4 no longer holds, so that we need to adapt the definition of a giant node. In this section, a giant node h is a node with degree $D_h^{(N)}$, satisfying that, for an appropriate choice of β,

$$N^\beta < D_h^{(N)} \le N^\alpha. \tag{2.3.4}$$

Nodes with degree at most N^β will be called *normal* nodes, and we will denote by K_N the total number of stubs of the normal nodes, i.e.,

$$K_N = \sum_{n=1}^N D_n^{(N)} \mathbf{1}_{\{D_n^{(N)} \le N^\beta\}}. \tag{2.3.5}$$

Similarly to (2.3.3), we see that

$$\mathbb{E}[K_N] = N^{1+\beta(2-\tau)}\ell(N). \tag{2.3.6}$$

To motivate our choice of β, which depends on the value of k, observe that a node with (at least) N^β stubs, which chooses exclusively other *giant* nodes, in $k + 1$ steps can reach approximately $N^{(k+1)\beta}$ other nodes. The number of stubs of $N^{(k+1)\beta}$ *giant* nodes is by definition at least $N^{(k+2)\beta}$. Hence, if we take β such that $M_N \sim N^{(k+2)\beta}$, or equivalently, by (2.3.3), $1 + \alpha(2 - \tau) \le (k + 2)\beta$, then we basically have all giant nodes on mutual distance at most $k + 1$, so that (the non-giant) nodes 1 and 2, given that they both connect to at least one giant node, are on distance at most $k + 3$. In the proof, we will see that we can pick any β such that

$$\frac{1 + \alpha(2 - \tau)}{k + 2} < \beta < \alpha,$$

where we use that $\frac{1+\alpha(2-\tau)}{k+2} < \alpha$, precisely when $\alpha > \frac{1}{\tau+k}$. Having this in mind, we choose

$$\beta = \frac{1}{2}\left(\frac{1 + \alpha(2 - \tau)}{k + 2} + \alpha\right). \tag{2.3.7}$$

Here ends the outline of the proof.

We now turn to the definition of the events involved. This part is similar, but not identical, to the introduction of $\mathcal{A}_{\varepsilon,N}$ in (2.2.7), because giant nodes no longer are on mutual distance 1. We keep the same notation for the event $\mathcal{B}_{\varepsilon,N}$, the event that the stubs of node 1 and 2 are attached exclusively to stubs of giant nodes, although the definition of a giant node has been changed. We take this slight abuse of notation for granted. The event $\mathcal{D}_{\varepsilon,N} = \{D_1 \le q_\varepsilon, D_2 \le q_\varepsilon\}$, where $q_\varepsilon = \min\{k : 1 - F(k) < \varepsilon/8\}$, is identical to the definition in Section 2 (below (2.2.7)). We define

$$\mathcal{G}_{\varepsilon,N} = \mathcal{B}_{\varepsilon,N} \cap \mathcal{D}_{\varepsilon,N} \cap \mathcal{H}_{\varepsilon,N}, \tag{2.3.8}$$

where

$$\mathcal{H}_{\varepsilon,N} = \left\{ N^{1+\alpha(2-\tau)}\underline{\ell}(N) \leq M_N \leq N^{1+\alpha(2-\tau)}\bar{\ell}(N) \right\} \cup \{K_N \leq N^{1+\beta(2-\tau)}\ell(N)\}, \tag{2.3.9}$$

where $\ell(N), \underline{\ell}(N), \bar{\ell}(N)$ are slowly varying at infinity. The event $\mathcal{H}_{\varepsilon,N}$ will enable us to control the distance between any pair of giant nodes, as sketched in the outline.

The following lemma is the counterpart of Lemma 2.5 in Section 2.2.

Lemma 2.6 (The good event has high probability) *For each $\varepsilon > 0$, there exists N_ε, such that, for all $N \geq N_\varepsilon$,*

$$\mathbb{P}\big(\mathcal{G}_{\varepsilon,N}^c\big) < \varepsilon. \tag{2.3.10}$$

The proof of Lemma 2.6 is rather technical and can be found in Appendix 2.A.2.

The remainder of the proof of Theorem 2.2 is divided into two parts, namely, the proofs of

$$\mathbb{P}(\{H_N \leq k+2\} \cap \mathcal{G}_{\varepsilon,N}) < \varepsilon/2, \tag{2.3.11}$$

and

$$\mathbb{P}(\{H_N > k+3\} \cap \mathcal{G}_{\varepsilon,N}) < \varepsilon/2. \tag{2.3.12}$$

Indeed, if we combine the statements (2.3.11) and (2.3.12), then

$$\begin{aligned}
\mathbb{P}(H_N = k+3) &= \mathbb{P}(\{H_N = k+3\} \cap \mathcal{G}_{\varepsilon,N}) + \mathbb{P}\big(\{H_N = k+3\} \cap \mathcal{G}_{\varepsilon,N}^c\big) \\
&\geq \mathbb{P}(\{H_N = k+3\} \cap \mathcal{G}_{\varepsilon,N}) - \varepsilon \\
&= 1 - \mathbb{P}(\{H_N > k+3\} \cap \mathcal{G}_{\varepsilon,N}) - \mathbb{P}(\{H_N < k+3\} \cap \mathcal{G}_{\varepsilon,N}) - \varepsilon \\
&> 1 - 2\varepsilon,
\end{aligned}$$

and the conclusion of Theorem 2.2(i) is reached. We will prove (2.3.11), (2.3.12) in two lemmas.

Lemma 2.7 (The distance is at least $k+3$ on the good event) *For a non-negative integer k, and α as in (2.3.1), for each $\varepsilon > 0$, there exists an integer N_ε, such that*

$$\mathbb{P}(\{H_N \leq k+2\} \cap \mathcal{G}_{\varepsilon,N}) < \varepsilon/2, \qquad N \geq N_\varepsilon.$$

Proof.   The inequality of the lemma is proved by a counting argument. We will show that for each $l \in \{1, 2, 3, \dots, k+2\}$

$$\mathbb{P}(\{H_N = l\} \cap \mathcal{G}_{\varepsilon,N}) < N^{-\delta_l}, \tag{2.3.13}$$

for some $\delta_l > 0$. Since

$$\mathbb{P}(\{H_N \leq k+2\} \cap \mathcal{G}_{\varepsilon,N}) \leq \sum_{l=1}^{k+2} \mathbb{P}(\{H_N = l\} \cap \mathcal{G}_{\varepsilon,N}) \leq (k+2)N^{-\delta},$$

where $\delta = \min\{\delta_1, \ldots, \delta_{k+2}\} > 0$, the claim of the lemma follows if we choose N_ε, such that $(k+2)N_\varepsilon^{-\delta} \leq \varepsilon/2$.

To prove that $\mathbb{P}(\{H_N = l\} \cap \mathcal{G}_{\varepsilon,N}) < N^{-\delta_l}$ for any $l \leq k+2$, we note that on $\mathcal{G}_{\varepsilon,N}$, the degrees of nodes 1 and 2 are bounded by q_ε. Therefore, on $\mathcal{G}_{\varepsilon,N}$ and using that all degrees are bounded by N^α, the number of nodes that can be reached from node 1 in $l-1$ steps is at most $q_\varepsilon N^{(l-2)\alpha}$, and the number of stubs incident to nodes at distance $l-1$ from node 1 is at most $q_\varepsilon N^{(l-1)\alpha}$. When $H_N = l$, then one of these stubs should be attached to one of the at most q_ε stubs incident to node 2.

Denote by $M_N^{(l)}$ the number of stubs that are not part of an edge incident to a node at distance at most $l-1$ from node 1. Then, conditionally on $M_N^{(l)}$ and the fact that node 2 is at distance at least $l-1$ from node 1, the stubs of node 2 will be connected to one of these $M_N^{(l)}$ stubs uniformly at random. More precisely, conditionally on $M_N^{(l)}$ and the fact that node 2 is at distance at least $l-1$ from node 1, the event $\{H_N = l\}$ occurs precisely when a stub of node 2 is paired with a stub attached to a node at distance $l-1$ from node 1.

We note that, on $\mathcal{G}_{\varepsilon,N}$,

$$M_N^{(l)} \geq M_N - 2q_\varepsilon N^{(l-2)\alpha} = M_N(1 + o(1)) \geq \ell(N)N^{1+(2-\tau)\alpha}, \qquad (2.3.14)$$

when $(l-2)\alpha < 1 + (2-\tau)\alpha$, i.e., when $\alpha < 1/(l+\tau-4)$. Since $l \leq k+2$ and $\alpha < 1/(k+\tau-1)$, the latter is always satisfied.

The probability that one of the at most q_ε stubs of node 2 is paired with one of the stubs attached to nodes at distance $l-1$ from node 1 is, on $\mathcal{G}_{\varepsilon,N}$ and conditionally on $M_N^{(l)}$ and the fact that node 2 is at distance at least $l-1$ from node 1, bounded from above by

$$\frac{q_\varepsilon^2 N^{(l-1)\alpha}}{M_N^{(l)}} = \frac{q_\varepsilon^2 N^{(l-1)\alpha}}{M_N}(1 + o(1)) \leq \ell(N)N^{(l-3+\tau)\alpha-1} < N^{-\delta_l}, \qquad (2.3.15)$$

for all $\delta_l < 1 - (l-3+\tau)\alpha$ and N sufficiently large. Here, we use the lower bound on M_N in (2.3.9). Applying this to $l = k+2$, which gives the worst possible value of δ_l, we see that this probability is bounded from above by $N^{-\delta}$ for any $\delta < 1 - (k+\tau-1)\alpha$. Since $\alpha < 1/(k+\tau-1)$, we have that $1 - (k+\tau-1)\alpha > 0$, so that we can also choose $\delta > 0$.   □

We turn to the proof (2.3.12), which we also formulate as a lemma:

Lemma 2.8 (Distance is at most $k+3$ on the good event) *Fix a nonnegative integer k, and α as in (2.3.1). For each $\varepsilon > 0$ there exists an integer N_ε, such that,*

$$\mathbb{P}(\{H_N > k+3\} \cap \mathcal{G}_{\varepsilon,N}) < \varepsilon/2, \qquad N \geq N_\varepsilon.$$

In the proof of Lemma 2.8, we need that the number of giant nodes reachable from an arbitrary giant node h in at most l steps, has a lower bound proportional to $N^{l\beta}$. We denote by $Z_h^{(l)}$ the set of all nodes which are reachable in exactly l steps from a node h:

$$Z_h^{(l)} = \{n = 1, 2, \ldots, N : d(h, n) = l\} \text{ for } l \in \{0, 1, \ldots\},$$

where $d(h, n)$ denotes the graph-distance between the nodes h and n. The number of giant nodes in $Z_h^{(l)}$ is denoted by $E_h^{(l)}$.

Lemma 2.9 (Growth of the number of giant nodes) *For each $\varepsilon > 0$, $\alpha < 1/(\tau + k - 1)$, $l \in \{0, 1, \ldots, k\}$ and β given by (2.3.7),*

$$\mathbb{P}\left(\bigcap_{h=1}^{N} \left\{(1-\varepsilon)^l N^{l\beta} \le E_h^{(l)} < N^{l\alpha}\right\} \cap \{h \text{ is giant}\} \cap \mathcal{G}_{\varepsilon,N}\right)$$
$$> 1 - N\sum_{s=1}^{k} e^{-3\varepsilon(1-\varepsilon)^s N^{s\beta}/16},$$

for sufficiently large N.

Proof.   The upper bound $N^{l\alpha}$ on $E_h^{(l)}$ is trivial, because each node has less than N^α stubs. We will prove by induction with respect to l, that for $l \in \{0, 1, \ldots, k\}$, and k fixed,

$$\mathbb{P}\left(\bigcup_{h=1}^{N} \left\{E_h^{(l)} < (1-\varepsilon)^l N^{l\beta}\right\} \cap \{h \text{ is giant}\} \cap \mathcal{G}_{\varepsilon,N}\right) \le N\sum_{s=1}^{l} e^{-3\varepsilon(1-\varepsilon)^s N^{s\beta}/16}. \tag{2.3.16}$$

Denote

$$\mathcal{F}_{\varepsilon,N}^{(l)} = \bigcap_{h=1}^{N} \left\{\left\{E_h^{(l)} \ge (1-\varepsilon)^l N^{l\beta}\right\} \cap \{h \text{ is giant}\}\right\}, \tag{2.3.17}$$

then it suffices to prove that

$$\mathbb{P}\left((\mathcal{F}_{\varepsilon,N}^{(l)})^c \cap \mathcal{F}_{\varepsilon,N}^{(l-1)} \cap \mathcal{G}_{\varepsilon,N}\right) \le Ne^{-3\varepsilon(1-\varepsilon)^l N^{l\beta}/16}. \tag{2.3.18}$$

Indeed, if (2.3.18) holds, then (2.3.16) follows, by the induction hypothesis, as follows:

$$\mathbb{P}\left(\bigcup_{h=1}^{N} \left\{E_h^{(l)} < (1-\varepsilon)^l N^{l\beta}\right\} \cap \{h \text{ is giant}\} \cap \mathcal{G}_{\varepsilon,N}\right) \tag{2.3.19}$$
$$= \mathbb{P}(\mathcal{G}_{\varepsilon,N} \cap (\mathcal{F}_{\varepsilon,N}^{(l)})^c) \le \mathbb{P}(\mathcal{G}_{\varepsilon,N} \cap (\mathcal{F}_{\varepsilon,N}^{(l)})^c \cap \mathcal{F}_{\varepsilon,N}^{(l-1)}) + \mathbb{P}(\mathcal{G}_{\varepsilon,N} \cap (\mathcal{F}_{\varepsilon,N}^{(l-1)})^c)$$
$$\le Ne^{-3\varepsilon(1-\varepsilon)^l N^{l\beta}/16} + N\sum_{s=1}^{l-1} e^{-3\varepsilon(1-\varepsilon)^s N^{s\beta}/16}.$$

For $l = 0$, (2.3.18) trivially holds. We therefore assume that (2.3.18) is valid for $l = m - 1$ and we will prove that (2.3.18) holds for $l = m$.

In this paragraph we will work conditionally given the degrees $D_1, \ldots, D_N$. For h a giant node, we consider only $A_N = E_h^{(m-1)}\lfloor N^\beta \rfloor$ stubs of the nodes in $Z_h^{(m-1)}$. To be more precise: we consider $\lfloor N^\beta \rfloor$ stubs of each of the $E_h^{(m-1)}$ giant nodes in $Z_h^{(m-1)}$. We number these stubs by $i \in \{1, 2, \ldots, A_N\}$ and stub i will connect to a stub of a node n_i. Then we denote by $r_{N,i}$, for $i \in \{1, 2, \ldots, A_N\}$, the probability that stub i does not connect to a stub of a normal node. We denote by $s_{N,i}$ the probability that stub i does not connect to a stub of a node in $Z_h^{(m-1)}$ (and the total number of stubs of this set is at most $N^{m\alpha}$), and finally,

we denote by $t_{N,i,j}$ the probability that stub i does not connect to the giant node h_j previously selected by the stubs $j \in \{1, 2, \ldots, i-1\}$ (for each j there are are most $D_{h_j}^{(N)} \leq N^\alpha$ of such stubs). If none of the above attachments happens, then we have a match with a not previously found giant node, and we denote by $q_{N,i}$ the probability of such a match of stub i, i.e.,

$$q_{N,i} = 1 - r_{N,i} - s_{N,i} - \sum_{j=1}^{i-1} t_{N,i,j}.$$

From the number of stubs mentioned between the parenthesis, we can bound this probability from below by

$$q_{N,i} \geq 1 - \frac{K_N}{M_N} - \frac{N^{m\alpha}}{M_N} - \sum_{j=1}^{i-1} \frac{N^\alpha}{M_N}. \tag{2.3.20}$$

Since,
$$i - 1 \leq E_h^{(m-1)} \lfloor N^\beta \rfloor \leq N^{\alpha(m-1)} \lfloor N^\beta \rfloor \leq N^{\alpha(m-1)+\beta},$$

and $K_N \leq N^{1+\beta(2-\tau)} \ell(N)$, $M_N > \underline{\ell}(N) N^{1+\alpha(2-\tau)}$ on $\mathcal{G}_{\varepsilon,N}$, we can bound $1 - q_{N,i}$ on $\mathcal{G}_{\varepsilon,N}$ from above by

$$1 - q_{N,i} \leq \frac{\ell(N) N^{1+\beta(2-\tau)} + N^{\alpha m} + N^{\alpha(m-1)+\beta+\alpha}}{\underline{\ell}(N) N^{1+\alpha(2-\tau)}}.$$

For sufficiently large N and uniformly in i, we have that $1 - q_{N,i} < \varepsilon/2$, because $\beta < \alpha$, and $m\alpha + \beta \leq k\alpha + \beta < (k+1)\alpha < 1 + \alpha(2-\tau)$.

Introduce the binomially distributed random variable Y_N with parameters B_N and $\varepsilon/2$, where $B_N = \lceil (1-\varepsilon)^m N^{m\beta} \rceil$. On $\mathcal{F}_{\varepsilon,N}^{(m-1)}$, we have that $A_N = E_h^{m-1} \lfloor N^\beta \rfloor \geq B_N$, so that the number of mismatches will be stochastically dominated by Y_N. We need at least $(1-\varepsilon)B_N$ matches, so that

$$\mathbb{P}(\{E_h^{(m)} \geq (1-\varepsilon)^m N^{m\beta}\} \cap \{A_N \geq B_N\} \cap \mathcal{G}_{\varepsilon,N}) \geq \mathbb{P}(Y_N < \varepsilon B_N). \tag{2.3.21}$$

We will now use the Janson inequality [57], which states that for any $t > 0$,

$$\mathbb{P}(|Y_N - \mathbb{E}[Y_N]| \geq t) \leq 2\exp\left(-\frac{t^2}{2(\mathbb{E}[Y_N] + t/3)}\right). \tag{2.3.22}$$

Since $\mathbb{E}[Y_N] = \varepsilon B_N/2$, we obtain, with $t = \varepsilon B_N/2$,

$$\mathbb{P}(Y_N < \varepsilon B_N) \leq \mathbb{P}(|Y_N - \mathbb{E}[Y_N]| > \varepsilon B_N/2) \leq 2\exp\left(-\frac{3\varepsilon B_N}{16}\right).$$

Combining this with (2.3.21), and since there are at most N giant nodes:

$$\mathbb{P}((\mathcal{F}_{\varepsilon,N}^{(m)})^c \cap \mathcal{F}_{\varepsilon,N}^{(m-1)} \cap \mathcal{G}_{\varepsilon,N}) \leq N\mathbb{P}(Y_N \leq \varepsilon B_N) \leq 2N\exp\left(-\frac{3\varepsilon(1-\varepsilon)^m N^{m\beta}}{16}\right). \tag{2.3.23}$$

$\square$

Proof of Lemma 2.8. We start with an outline. On the event $\mathcal{G}_{\varepsilon,N}$, each stub of the nodes 1 and 2 is attached to a stub of some giant node. The idea is to show that **whp** the distance between any two giant nodes is at most $k+1$. This implies that the graph distance between nodes 1 and 2, intersected with the event $\mathcal{G}_{\varepsilon,N}$ is **whp** at most $k+3$, and hence Lemma 2.8.

We will extend the event $\mathcal{G}_{\varepsilon,N}$ to include the main event in Lemma 2.9:

$$\mathcal{I}_{\varepsilon,N} = \mathcal{G}_{\varepsilon,N} \cap \mathcal{F}_{\varepsilon,N}^{(k)}, \tag{2.3.24}$$

where $\mathcal{F}_{\varepsilon,N}^{(k)}$ was defined in (2.3.17). Then

$$\mathbb{P}(\{H_N > k+3\} \cap \mathcal{G}_{\varepsilon,N}) \leq \mathbb{P}(\{H_N > k+3\} \cap \mathcal{I}_{\varepsilon,N}) + \mathbb{P}\big(\mathcal{G}_{\varepsilon,N} \cap (\mathcal{F}_{\varepsilon,N}^{(k)})^c\big), \tag{2.3.25}$$

and the second term on the right hand side of (2.3.25) can be bounded by $\varepsilon/4$ using Lemma 2.9. We use as indicated in the outline of the proof given above, that

$$\mathbb{P}(\{H_N > k+3\} \cap \mathcal{I}_{\varepsilon,N})$$
$$\leq \mathbb{P}\left(\bigcup_{h_1,h_2} \{h_1, h_2 \text{ are giant}\} \cap \{d(h_1,h_2) > k+1\} \cap \mathcal{I}_{\varepsilon,N}\right)$$
$$\leq \sum_{h_1,h_2} \mathbb{P}(\{h_1, h_2 \text{ are giant}\} \cap \{d(h_1,h_2) > k+1\} \cap \mathcal{I}_{\varepsilon,N}), \tag{2.3.26}$$

where the sum is taken over all pairs of nodes, and where, as before, $d(h_1, h_2)$ denotes the graph-distance between h_1 and h_2. Indeed, on $\mathcal{I}_{\varepsilon,N}$, the nodes 1 and 2 are connected to giant nodes, so that when $H_N > k+3$, there must be giant nodes h_1, h_2 at mutual distance at least $k+1$.

Clearly for any pair of nodes h_1 and h_2,

$$\{d(h_1, h_2) > k+1\} \subseteq \{d(h_1, h_2) > k\},$$

which implies that for any pair of nodes h_1 and h_2,

$$\mathbb{P}_N\left(\{d(h_1,h_2) > k+1\} \cap \{h_1, h_2 \text{ are giant}\} \cap \mathcal{I}_{\varepsilon,N}\right)$$
$$\leq \mathbb{P}_N\big(\{d(h_1,h_2) > k+1\} \cap \{h_1, h_2 \text{ are giant}\} \cap \mathcal{I}_{\varepsilon,N} \,|\, d(h_1,h_2) > k\big).$$

On the event $\{d(h_1, h_2) > k\} \cap \{h_1, h_2 \text{ are giant}\}$, the giant node h_2 is not attached to one of the nodes at distance k from the node h_1. More precisely, the giant node h_2 is not attached to one of the $\cup_{l=0}^{k-1} Z_{h_1}^{(l)}$ nodes. We have less than $M_N - \sum_{l=0}^{k-1} E_{h_1}^{(l)} N^\beta$ stubs to choose from, and the event $\{d(h_1,h_2) > k+1\}$ conditioned on $\{d(h_1,h_2) > k\}$ implies that no stubs of the giant node h_2 will attach to one of the at least $E_{h_1}^{(k)} N^\beta$ free stubs of $Z_{h_1}^{(k)}$. Therefore, we have, almost

surely,

$$\mathbb{P}_N\big(\{h_1, h_2 \text{ are giant}\} \cap \{d(h_1, h_2) > k + 1\} \cap \mathcal{G}^{(N)}_{\varepsilon,N} \,|\, d(h_1, h_2) > k\big)$$

$$\leq \prod_{i=0}^{D^{(N)}_{h_2}-1} \left(1 - \frac{E^{(k)}_{h_1} N^\beta}{M_N - \sum_{j=0}^{k-1} E^{(j)}_{h_1} N^\beta - 2i + 1}\right) \mathbf{1}_{\{\mathcal{I}_{\varepsilon,N}\}}$$

$$\leq \left(1 - \frac{E^{(k)}_{h_1} N^\beta}{M_N}\right)^{D^{(N)}_{h_2}} \mathbf{1}_{\{\mathcal{I}_{\varepsilon,N}\}} \leq \left(1 - \frac{\varepsilon(1-\varepsilon)^k N^{\beta(k+1)}}{N^{1+\alpha(2-\tau)}\bar{\ell}(N)}\right)^{N^\beta}$$

$$\leq \exp\left\{-\frac{\varepsilon(1-\varepsilon)^k N^{\beta(k+2)}}{N^{1+\alpha(2-\tau)}\bar{\ell}(N)}\right\} \leq \exp\left\{-\varepsilon(1-\varepsilon)^k N^\delta\right\}, \qquad (2.3.27)$$

where we used the inequality $1 - x \leq e^{-x}$, $x \geq 0$, in the one but last inequality, and where $0 < \delta < \beta(k+2) - (1 + \alpha(2 - \tau))$. If we substitute this upper bound in the right hand side of (2.3.26), then we end up with

$$\mathbb{P}(\{H_N > k + 3\} \cap \mathcal{I}_{\varepsilon,N}) \leq N^2 \exp\left(-\varepsilon(1-\varepsilon)^k N^\delta\right) < \varepsilon/2.$$

This completes the proof of Lemma 2.8 and hence of Theorem 2.2. $\qquad\square$

2.4   The cases $\tau = 1$ and $\tau = 2$

2.4.1   Proof of Theorem 2.3

It is well known, see e.g. [32, 8.2.4], that when $1 - F(x)$ is slowly varying, the quotient of the maximum and the sum of N i.i.d. random variables with distribution F, converges to 1 in probability, i.e.,

$$\frac{D_{(N)}}{L_N} \to 1, \quad \text{in probability.} \qquad (2.4.1)$$

Therefore, we obtain that **whp**, both node 1 and node 2 are connected to the node with maximal degree, which gives the stated result.

$\qquad\square$

2.4.2   Two examples with $\tau = 2$

In the following two examples we show that for $\tau = 2$, the limit hopcount distribution is sensitive to the slowly varying function.
Example 1. Let, for $x \geq 2$,

$$1 - F(x) = \frac{2(\log 2)^2}{(\lfloor x \rfloor)(\log \lfloor x \rfloor)^2}. \qquad (2.4.2)$$

Then we show that for all k fixed,

$$\mathbb{P}\left(H_N > k\right) = 1 + o(1), \qquad \text{as } N \to \infty. \qquad (2.4.3)$$

We first prove (2.4.3) for $k = 2$. We show this in two steps. In the first step we show that for any $\varepsilon > 0$, there exists $v_\varepsilon \in \mathbb{N}$ such that with probability at least

$1 - \varepsilon$ all nodes at distance at most 1 from nodes 1 and 2 have degrees at most v_ε. In the second step we show that there exists $N_v \in \mathbb{N}$, such that for any $N \geq N_v$, with probability at least $1 - \varepsilon$, any two given nodes with degrees at most v_ε, are disconnected. Both steps together clearly imply (2.4.3).

The second step is similar to (2.2.12), and is omitted here.

To obtain the first step we consider the event $\mathcal{D}_{\varepsilon,N}$, defined below (2.2.7). Then, for any $v \in \mathbb{N}$, the probability that within the first q_ε stubs of node 1 or node 2 there is a stub connected to a stub of node with degree at least $v + 1$ is at most

$$\mathbb{E}\left[\frac{2q_\varepsilon}{L_N} \sum_{i=1}^{N} D_i \mathbf{1}_{\{D_i > v\}}\right].$$

It remains to show that the above expectation is at most $\varepsilon/2$ for some $v = v_\varepsilon$ large enough. For this, we need that the first moment of the degree distribution for this example is finite. Indeed, from (2.4.2)

$$\mathbb{E}[D_1] = 1 + \sum_{x=2}^{\infty} \frac{2(\log 2)^2}{x(\log x)^2} \leq 2 + \int_2^{\infty} \frac{2(\log 2)^2}{u(\log u)^2} \, du = 1 + 2(\log 2)^2 \int_{\log 2}^{\infty} \frac{dy}{y^2} < \infty.$$
$$(2.4.4)$$

Then, from the Law of Large Numbers applied to $L_N = D_1 + \cdots + D_N$, we obtain

$$\mathbb{P}\left(L_N \geq \mu_\varepsilon N\right) \leq \frac{\varepsilon}{12q_\varepsilon}, \qquad (2.4.5)$$

for $\mu_\varepsilon > \mathbb{E}[D_1]$. Due to (2.4.4), (2.4.5) and the Markov inequality

$$\mathbb{E}\left[\frac{2q_\varepsilon}{L_N} \sum_{i=1}^{N} D_i \mathbf{1}_{\{D_i > v\}}\right] \leq \frac{\varepsilon}{6} + 2q_\varepsilon \mathbb{P}\left(2q_\varepsilon \sum_{i=1}^{N} D_i \mathbf{1}_{\{D_i > v\}} \geq \frac{\varepsilon L_N}{6}\right)$$

$$\leq \frac{\varepsilon}{6} + 2q_\varepsilon \mathbb{P}\left(L_N \geq \mu_\varepsilon\right) + 2q_\varepsilon \mathbb{P}\left(2q_\varepsilon \sum_{i=1}^{N} D_i \mathbf{1}_{\{D_i > v\}} \geq \frac{\varepsilon}{6}\mu_\varepsilon N\right)$$

$$\leq \frac{\varepsilon}{3} + \frac{24q_\varepsilon^2}{\varepsilon \mu_\varepsilon} \mathbb{E}\left(D_i \mathbf{1}_{\{D_i > v\}}\right) \leq \frac{\varepsilon}{2},$$

for large enough v, and hence we have the second step, since $\mathbb{P}(\mathcal{D}_{\varepsilon,N}^c) \leq 2\mathbb{P}(D_1 > q_\varepsilon) \leq \varepsilon/4$.

In a similar way we can show that, for any $\varepsilon > 0$, there exists $v_\varepsilon \in \mathbb{N}$ such that with probability at least $1 - \varepsilon$ all nodes at distance at most 2 from nodes 1 and 2 have degrees at most v_ε. This statement implies that $\mathbb{P}(H_N > 4) \to 1$. Similarly, we obtain that for any $\varepsilon > 0$ there exists $v_\varepsilon \in \mathbb{N}$ such that with probability at least $1 - \varepsilon$ all nodes at distance at most k from nodes 1 and 2 have degrees at most v_ε, which implies that for any fixed integer k,

$$\lim_{N \to \infty} \mathbb{P}(H_N > 2k) = 1, \qquad (2.4.6)$$

i.e., the probability mass of H_N drifts away to $+\infty$ as $N \to \infty$. This behavior of H_N for $\tau = 2$, is in agreement with the behavior of H_N for the case $\tau \in (2,3)$, (see [56]), where we show, among other things, tightness of the sequence

$$H_N - \frac{\log \log N}{|\log(\tau - 2)|}. \qquad (2.4.7)$$

□

Example 2. Let

$$1 - F(x) = c\frac{(\log x)^{\log \log x - 1} \log \log x}{x}, \qquad x \geq x^*, x \in \mathbb{N}. \tag{2.4.8}$$

where x^* is chosen such that for $x \geq x^*$, the right side of (2.4.8) is a non-increasing function, and c is such that $1 - F(x^*) = 1$. We will show that

$$\mathbb{P}(H_N \in \{2, 3\}) = 1 + o(1), \qquad \text{as } N \to \infty. \tag{2.4.9}$$

Thus, we see entirely different behavior as in the first example.

Define *giant* nodes as nodes with degree at least $N^{\frac{1}{2}+\delta}$, for some $\delta > 0$, to be determined later on. The nodes with degree at most $N^{\frac{1}{2}+\delta} - 1$ we call *normal*. Define the event $\mathcal{A}_{\varepsilon,N}$ as in (2.2.7), where, in the definition of $\mathcal{B}_{\varepsilon,N}$, we use the above definition of the giant node. In Appendix 2.A.3, we will prove the following lemma, which is similar to Lemma 2.5:

Lemma 2.10 *For each $\varepsilon > 0$, there exists N_ε, such that for all $N \geq N_\varepsilon$,*

$$\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) < \varepsilon. \tag{2.4.10}$$

We now complete the proof of (2.4.3) subject to Lemma 2.10, which is straightforward. By (2.2.12), we obtain that $\mathbb{P}(\{H_N = 1\} \cap \mathcal{A}_{\varepsilon,N}) = o(1)$. Moreover, when $\mathcal{A}_{\varepsilon,N}$ occurs, all stubs of nodes 1 and 2 are connected to giant nodes due to $\mathcal{B}_{\varepsilon,N}$, and the giant nodes form a complete graph due to $\mathcal{C}_{\varepsilon,N}$, so that

$$\mathbb{P}(\{H_N > 3\} \cap \mathcal{A}_{\varepsilon,N}) = 0.$$

This proves (2.4.3).   □

2.5   Simulation and conclusions

To illustrate Theorems 2.1 and 2.2, we have simulated our random graph with degree distribution $D = \lceil U^{-\frac{1}{\tau-1}} \rceil$, where U is uniformly distributed over $(0,1)$. Thus,

$$1 - F(x) = \mathbb{P}(U^{-\frac{1}{\tau-1}} > x) = x^{1-\tau}, \quad x = 1, 2, 3, \ldots$$

In Figure 2.1, we have simulated the graph distance or hopcount with $\tau = 1.8$ and the values of $N = 10^3, 10^4, 10^5$. The histogram is in accordance with Theorem 2.1: for increasing values of N we see that the probability mass is divided over the values $H_N = 2$ and $H_N = 3$, where the probability $\mathbb{P}(H_N = 2)$ converges.

As an illustration of Theorem 2.2, we again take $\tau = 1.8$, but now condition the degrees to be less than N, so that $\alpha = 1$. Since in this case $(\tau - 1)^{-1} = \frac{5}{4}$, we expect from Theorem 2.2 case (i), that in the limit the hopcount will concentrate on the value $H_N = 3$. This is indeed the case as is shown in Figure 2.2.

Our results give convincing asymptotic for the hopcount when the mean degree is infinite, using extreme value theory. Some details remain open:
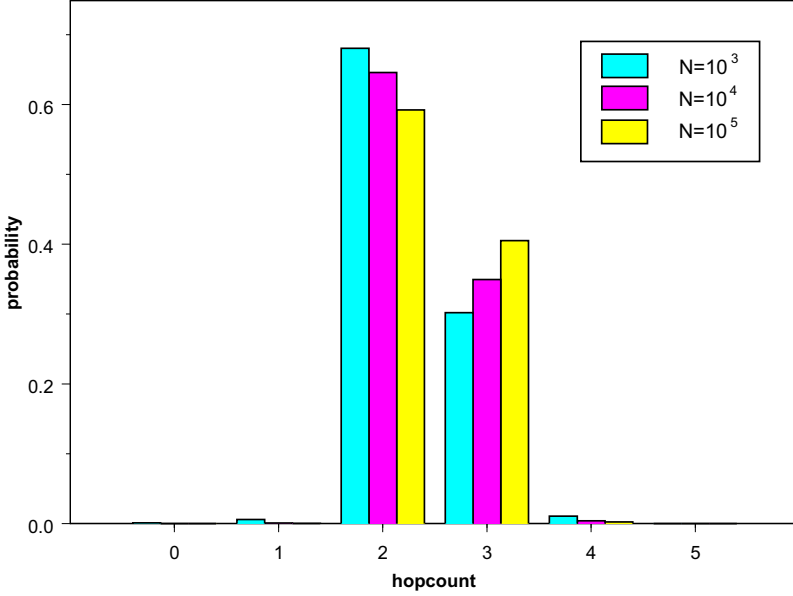
Figure 2.1: Empirical probability mass function of the hopcount for $\tau = 1.8$ and $N = 10^3, 10^4, 10^5$, for the unconditioned degrees.

(i) It is possible to compute upper and lower bounds on the value p_F, based on Lemma 2.4. We presented two such bounds in (2.2.17). These bounds can be obtained from simulating the random variables $\Gamma_1, \Gamma_2, \ldots$ in (2.2.3). It should be possible to obtain much sharper upper and lower bounds, and possibly even numerical values, depending on the specific degree distribution F.

(ii) In the boundary cases $\alpha = 1/(\tau + k)$, $k \in \mathbb{N} \cup \{0\}$, it is natural to conjecture that the specific limit behavior of H_N will depend on the slowly varying function, as is the case for $\tau = 2$ and $\alpha > \frac{1}{\tau - 1} = 1$ as described in Section 2.4.2.

2.A   Appendix.

In the appendix we prove Lemma 2.5, Lemma 2.6 and Lemma 2.10. The proofs of Lemma 2.6 and 2.10 are both adaptations of the proof of Lemma 2.5 in Section A.1 below.
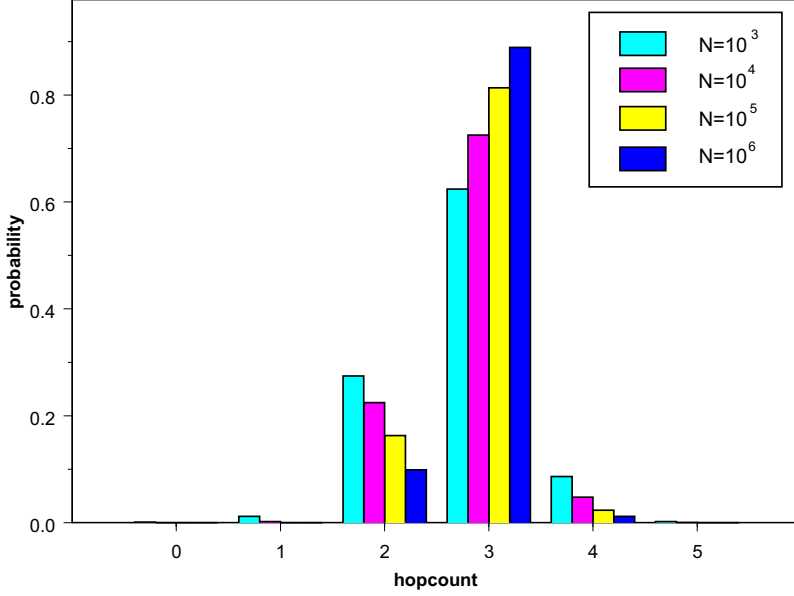
Figure 2.2: Empirical probability mass function of the hopcount for $\tau = 1.8$ and $N = 10^3, 10^4, 10^5, 10^6$, where the degrees are conditioned to be less than N, ($\frac{1}{\tau} < \alpha = 1 < \frac{1}{\tau - 1}$).

2.A.1   Proof of Lemma 2.5

In this section we restate Lemma 2.5 and then give a proof.

Lemma 2.11 *For each $\varepsilon > 0$, there exists N_ε such that*

$$\mathbb{P}(\mathcal{A}^c_{\varepsilon,N}) < \varepsilon, \quad N \geq N_\varepsilon. \tag{2.A.1}$$

Proof. We start with an outline of the proof. By (2.2.7),

$$\mathbb{P}(\mathcal{A}^c_{\varepsilon,N}) \leq \mathbb{P}(\mathcal{B}^c_{\varepsilon,N}) + \mathbb{P}(\mathcal{C}^c_{\varepsilon,N}) + \mathbb{P}(\mathcal{D}^c_{\varepsilon,N}), \tag{2.A.2}$$

and an obvious way to prove result (2.A.1) would be to show that each of the three terms on the right-hand side of (2.A.2) is smaller than $\varepsilon/3$. This direct approach is somewhat difficult and instead we introduce an additional event $\mathcal{E}_{\varepsilon,N}$, which controls the total degree L_N in part (c), the degree of the giant nodes in part (b), and the total degree of all normal (non-giant) nodes in part (a):

$$\mathcal{E}_{\varepsilon,N} = \begin{cases} \left\{ \sum_{n=1}^{N-k_\varepsilon} D_{(n)} \leq \frac{\varepsilon}{8q_\varepsilon} L_N \right\} & (a) \\ \cap \left\{ D_{(N-k_\varepsilon+1)} \geq c_\varepsilon u_N \right\} & (b) \\ \cap \left\{ L_N \leq d_\varepsilon u_N \right\}, & (c) \end{cases} \tag{2.A.3}$$

where q_ε is the ε-quantile of F used in the definition of $\mathcal{D}_{\varepsilon,N}$ and where $c_\varepsilon, d_\varepsilon > 0$ are defined by

$$\mathbb{P}\left(\xi_{k_\varepsilon} < c_\varepsilon\right) < \varepsilon/24 \quad \text{and} \quad \mathbb{P}(\eta > d_\varepsilon) < \varepsilon/24,$$

respectively. Observe that c_ε is a lower quantile of ξ_{k_ε}, whereas b_ε defined in (2.2.9) and (2.2.10) is an upper quantile of ξ_{k_ε}. Furthermore, d_ε is an upper quantile of η, whereas a_ε defined in (2.2.8) is a lower quantile of η. Intersection with the additional event $\mathcal{E}_{\varepsilon,N}$, facilitates the bounding of both $\mathcal{B}_{\varepsilon,N}^c$ and $\mathcal{C}_{\varepsilon,N}^c$. Therefore, we write

$$\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) \leq \mathbb{P}(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N} \cap \mathcal{E}_{\varepsilon,N}) + \mathbb{P}(\mathcal{C}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N} \cap \mathcal{E}_{\varepsilon,N}) + \mathbb{P}(\mathcal{D}_{\varepsilon,N}^c) + \mathbb{P}(\mathcal{E}_{\varepsilon,N}^c), \quad (2.A.4)$$

and our strategy to prove the lemma is that we show that each of the four terms on the right-hand side of (2.A.4) is at most $\varepsilon/4$.

Nodes 1 and 2 are connected to giant nodes only. On $\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N}$ at least one of the $2q_\varepsilon$ stubs is attached to a stub of the nodes $D_{(1)}, \ldots, D_{(N-k_\varepsilon)}$. Hence, the first term on the right side of (2.A.4) satisfies

$$\mathbb{P}(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N} \cap \mathcal{E}_{\varepsilon,N}) \leq 2q_\varepsilon \mathbb{E}\left[\frac{1}{L_N} \sum_{n=1}^{N-k_\varepsilon} D_n \mathbf{1}_{\{\mathcal{E}_{\varepsilon,N}\}}\right] \leq \varepsilon/4,$$

due to point (a) of $\mathcal{E}_{\varepsilon,N}$.

The giant nodes form a complete graph. We turn to the second term of (2.A.4). Recall that $\mathcal{C}_{\varepsilon,N}^c$ induces that no stubs of at least two giant nodes are attached to one another. Since we have at most N^2 pairs of giant nodes h_1 and h_2, the items (b), (c) of $\mathcal{E}_{\varepsilon,N}$ imply

$$\mathbb{P}(\mathcal{C}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N} \cap \mathcal{E}_{\varepsilon,N}) \leq \mathbb{E}\left[N^2 \prod_{i=0}^{\lfloor D_{h_1}/2 \rfloor - 1}\left(1 - \frac{D_{h_2}}{L_N - 2i - 1}\right) \mathbf{1}_{\{h_1, h_2 \text{ giant}\}}\right]$$

$$\leq N^2 \left(1 - \frac{c_\varepsilon u_N}{d_\varepsilon u_N}\right)^{c_\varepsilon u_N/2} \leq N^2 \exp\left(-\frac{c_\varepsilon^2 u_N}{2d_\varepsilon}\right) \leq \varepsilon/4,$$
$$(2.A.5)$$

for large enough N, because $u_N = N^{1/(\tau-1)+o(1)}$.

Nodes 1 and 2 have small degree. The third term on the right-hand side of equation (2.A.4) is at most $\varepsilon/4$, because

$$\mathbb{P}(\mathcal{D}_{\varepsilon,N}^c) \leq 2\mathbb{P}(D_1 > q_\varepsilon) \leq 2\varepsilon/8 = \varepsilon/4. \quad (2.A.6)$$

The order statistics.   It remains to estimate the last term on the right side of (2.A.4). Clearly,

$$
\begin{aligned}
\mathbb{P}\left(\mathcal{E}_{\varepsilon,N}^c\right) \leq \quad &\mathbb{P}\left(\sum_{n=1}^{N-k_\varepsilon} D_{(n)} > \frac{\varepsilon}{8q_\varepsilon} L_N\right) && (a) \\
&+\mathbb{P}\left(D_{(N-k_\varepsilon+1)} < c_\varepsilon u_N\right) && (b) \\
&+\mathbb{P}\left(L_N > d_\varepsilon u_N\right). && (c)
\end{aligned}
\tag{2.A.7}
$$

We will consequently show that each term in the above expression is at most $\varepsilon/12$. Let a_ε and $b_\varepsilon > 0$ be as in (2.2.8) and (2.2.9), then we can decompose the first term on the right-hand side of (2.A.7) as

$$
\mathbb{P}\left(\sum_{n=1}^{N-k_\varepsilon} D_{(n)} > \frac{\varepsilon}{8q_\varepsilon} L_N\right) \leq \mathbb{P}\left(L_N < a_\varepsilon u_N\right) + \mathbb{P}\left(\sum_{n=1}^{N-k_\varepsilon} D_{(n)} > \frac{\varepsilon}{8q_\varepsilon} a_\varepsilon u_N\right) \tag{2.A.8}
$$

$$
\leq \mathbb{P}\left(L_N < a_\varepsilon u_N\right) + \mathbb{P}\left(D_{(N-k_\varepsilon+1)} > b_\varepsilon u_N\right)
$$

$$
+ \mathbb{P}\left(\sum_{i=1}^{N} D_i \mathbf{1}_{\{D_i < b_\varepsilon u_N\}} > \frac{\varepsilon}{8q_\varepsilon} a_\varepsilon u_N\right).
$$

From the Markov inequality,

$$
\mathbb{P}\left(\sum_{i=1}^{N} D_i \mathbf{1}_{\{D_i < b_\varepsilon u_N\}} > \frac{\varepsilon}{8q_\varepsilon} a_\varepsilon u_N\right) \leq \frac{8q_\varepsilon N \mathbb{E}\left[D\mathbf{1}_{\{D < b_\varepsilon u_N\}}\right]}{\varepsilon a_\varepsilon u_N}. \tag{2.A.9}
$$

Since $1 - F(x)$ varies regularly with exponent $\tau - 1$, we have, by [41, Theorem 1(b), p. 281],

$$
\mathbb{E}\left[D\mathbf{1}_{\{D < b_\varepsilon u_N\}}\right] = \sum_{k=0}^{\lfloor b_\varepsilon u_N \rfloor} \left[1 - F(k)\right] \leq 2(2-\tau) b_\varepsilon u_N \left[1 - F(b_\varepsilon u_N)\right], \tag{2.A.10}
$$

for large enough N. Due to (2.2.1), for large enough N, we have also

$$
N\left[1 - F(u_N)\right] \leq 2. \tag{2.A.11}
$$

Substituting (2.A.10) and (2.A.11) in (2.A.9), we obtain

$$
\mathbb{P}\left(\sum_{i=1}^{N} D_i \mathbf{1}_{\{D_i < b_\varepsilon u_N\}} > \frac{\varepsilon}{8q_\varepsilon} a_\varepsilon u_N\right) \leq \frac{16 q_\varepsilon N(2-\tau) b_\varepsilon u_N \left[1 - F(b_\varepsilon u_N)\right]}{\varepsilon u_N a_\varepsilon}
$$

$$
\leq \frac{32 q_\varepsilon (2-\tau) b_\varepsilon \left[1 - F(b_\varepsilon u_N)\right]}{\varepsilon a_\varepsilon \left[1 - F(u_N)\right]}, \tag{2.A.12}
$$

for large enough N. From the regular variation of $1 - F(x)$,

$$
\lim_{N\to\infty} \frac{1 - F(b_\varepsilon u_N)}{1 - F(u_N)} = (b_\varepsilon)^{1-\tau}.
$$

Hence the right-hand side of (2.A.12) is at most

$$
\frac{64 q_\varepsilon (2-\tau) (b_\varepsilon)^{2-\tau}}{\varepsilon a_\varepsilon} \leq \varepsilon/36,
$$

for sufficiently large N, by the definition of b_ε in (2.2.9). We now show that the second term on the right side of (2.A.8) is at most $\varepsilon/36$. Since $D_{(N-k_\varepsilon+1)}/u_N$ converges in distribution to ξ_{k_ε}, we find from (2.2.10),

$$\mathbb{P}\left(D_{(N-k_\varepsilon+1)} > b_\varepsilon u_N\right) \le \mathbb{P}\left(\xi_{k_\varepsilon} > b_\varepsilon/2\right) + \varepsilon/72 \le \varepsilon/36,$$

for large enough N. Similarly, by the definition of a_ε, in (2.2.8), we have

$$\mathbb{P}\left(L_N < a_\varepsilon u_N\right) \le \varepsilon/36.$$

Each of the three terms on the right side of (2.A.8) is at most $\varepsilon/36$, so that the term (2.A.7)(a) is at most $\varepsilon/12$.

The upper bound for (2.A.7)(b), i.e., the bound

$$\mathbb{P}\left(D_{(N-k_\varepsilon+1)} < c_\varepsilon u_N\right) < \varepsilon/12,$$

is an easy consequence of the distributional convergence of $D_{(N-k_\varepsilon+1)}/u_N$ to ξ_{k_ε} and the definition of c_ε. Similarly, we obtain the upper bound for the term in (2.A.7)(c), i.e.,

$$\mathbb{P}\left(L_N > d_\varepsilon u_N\right) < \varepsilon/12,$$

from the convergence in distribution of L_N/u_N to η and the definition of d_ε.

Thus we have shown that $\mathbb{P}(\mathcal{E}_{\varepsilon,N}^c) < \varepsilon/4$. This completes the proof of Lemma 2.5. $\qquad\square$

2.A.2   Proof of Lemma 2.6

In this section we restate Lemma 2.6 and give a proof.

Lemma 2.12 *For each $\varepsilon > 0$, there exists N_ε such that for all $N \ge N_\varepsilon$,*

$$\mathbb{P}(\mathcal{G}_{\varepsilon,N}^c) < \varepsilon. \tag{2.A.13}$$

Proof. From (2.3.8),

$$\mathbb{P}\left(\mathcal{G}_{\varepsilon,N}^c\right) < \mathbb{P}\left(\mathcal{D}_{\varepsilon,N}^c\right) + \mathbb{P}\left(\mathcal{H}_{\varepsilon,N}^c\right) + \mathbb{P}\left(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{H}_{\varepsilon,N} \cap \mathcal{D}_{\varepsilon,N}\right). \tag{2.A.14}$$

We will bound each term on the right hand side of (2.A.14) separately.

From (2.A.6), and because the definition of $\mathcal{D}_{\varepsilon,N}$ is unaltered, the bound $\mathbb{P}\left(\mathcal{D}_{\varepsilon,N}^c\right) < \varepsilon/4$ is immediate.

For $\mathbb{P}\left(\mathcal{H}_{\varepsilon,N}^c\right)$, we will show that the total number of stubs M_N is of the order $\ell(N)N^{1+\alpha(2-\tau)}$ and that the total number K_N of stubs attached to normal nodes is of order $\ell(N)N^{1+\beta(2-\tau)}$. We start with the first statement. Bound

$$M_N = \sum_{i=1}^{N} D_i^{(N)} \ge \frac{1}{2}N^\alpha \sum_{i=1}^{N} \mathbf{1}_{\{D_i^{(N)} > \frac{1}{2}N^\alpha\}}.$$

The sum of indicators is distributed as a binomial random variable V_N with parameters N and $N^{\alpha(1-\tau)}\ell(N)$, because

$$\mathbb{P}\left(D^{(N)} > \frac{1}{2}N^\alpha\right) = N^{\alpha(1-\tau)}\ell(N).$$

We use Janson's inequality (compare (2.3.22)), on the binomial random variable V_N, with expectation $N^{1+\alpha(1-\tau)}\ell(N)$, and with $t = N^{1+\alpha(1-\tau)}\ell(N)/2$, to obtain:

$$\mathbb{P}(|V_N - \mathbb{E}[V_N]| \geq t) \leq 2\exp\left(-\frac{t^2}{2\left(\mathbb{E}[V_N] + t/3\right)}\right)$$
$$= 2\exp\left(-\frac{3}{28}\ell(N)N^{1+\alpha(1-\tau)}\right) < \varepsilon/8,$$

for N sufficiently large. Therefore, with probability at least $1 - \varepsilon/8$, we have, for all $\delta > 0$,

$$M_N > \frac{1}{2}N^\alpha V_N \geq N^{1+\alpha(2-\tau)}\underline{\ell}(N), \qquad (2.A.15)$$

for N sufficiently large and some slowly varying function $\underline{\ell}(N)$.

The mean degree $\mathbb{E}[M_N]$ is given by (2.3.3). Thus, by the Markov inequality,

$$\mathbb{P}\left(\sum_{n=1}^N D_n^{(N)} > \frac{8}{\varepsilon}\ell(N)N^{1+\alpha(2-\tau)}\right) \leq \frac{\varepsilon}{8},$$

so that with probability at least $1 - \varepsilon/8$, we have that

$$M_N \leq N^{1+\alpha(2-\tau)}\bar{\ell}(N), \qquad (2.A.16)$$

for some slowly varying function $\bar{\ell}(N)$. Similarly, the mean degree of a normal node is

$$\mathbb{E}\left[D^{(N)}\mathbf{1}_{\{D\leq N^\beta\}}\right] = \sum_{n=1}^{\lfloor N^\beta\rfloor} \mathbb{P}(D \geq n | D < N^\alpha) = N^{\beta(2-\tau)}\ell(N),$$

so that in exactly the same way, we find from the Markov inequality, that with probability at least $1 - \varepsilon/8$,

$$K_N \leq N^{1+\beta(2-\tau)}\ell(N). \qquad (2.A.17)$$

The inequalities (2.A.15),(2.A.16) and (2.A.17) together imply that

$$\mathbb{P}\left(\mathcal{H}_{\varepsilon,N}^c\right) \leq 3\varepsilon/8.$$

We finally turn to $\mathbb{P}\left(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{H}_{\varepsilon,N}\right)$. From the derivation above, we find that, on $\mathcal{H}_{\varepsilon,N}$, the fraction of the contribution of stubs from normal nodes and giant nodes is at most

$$\frac{\ell(N)N^{1+\beta(2-\tau)}}{\bar{\ell}(N)N^{1+\alpha(2-\tau)}} = \frac{\ell(N)}{\bar{\ell}(N)}N^{(2-\tau)(\beta-\alpha)}.$$

Since $\beta < \alpha$ and $\tau \in (1,2)$ the above ratio tends to 0, as $N \to \infty$. Thus the total number K_N of stubs of the normal nodes is negligible with respect to M_N on the event $\mathcal{H}_{\varepsilon,N}$. This implies that, with probability at least $1-\varepsilon/4$, each stub of nodes 1 and 2 is attached to a stub of a giant node on the event $\mathcal{H}_{\varepsilon,N}$. Therefore, we have showed

$$\mathbb{P}\left(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{H}_{\varepsilon,N} \cap \mathcal{D}_{\varepsilon,N}\right) < \varepsilon/4. \qquad (2.A.18)$$

Since $2\varepsilon/4 + 3\varepsilon/8 < \varepsilon$, the lemma is proved. $\qquad\qquad\square$

2.A.3 Proof of Lemma 2.10

In this section we restate Lemma 2.10 and give a proof.

Lemma 2.13 *For each $\varepsilon > 0$, there exists N_ε such that for all $N \geq N_\varepsilon$,*

$$\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) < \varepsilon. \tag{2.A.19}$$

Proof. The proof is a slight adaptation of the proof of Lemma 2.5 in Section A.1. We use that

$$\mathbb{P}(\mathcal{A}_{\varepsilon,N}^c) \leq \mathbb{P}(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N}) + \mathbb{P}(\mathcal{C}_{\varepsilon,N}^c) + \mathbb{P}(\mathcal{D}_{\varepsilon,N}^c), \tag{2.A.20}$$

and bound each of the three terms. The bound on $\mathbb{P}(\mathcal{D}_{\varepsilon,N}^c)$ is identical to the one in (2.A.6), and will be omitted here.

We next show that $\mathbb{P}(\mathcal{C}_{\varepsilon,N}^c) \leq \frac{\varepsilon}{3}$. First observe that since $\tau = 2$,

$$\mathbb{P}(L_N \geq N^{1+\delta}) \leq \varepsilon/6.$$

Recall that $\mathcal{C}_{\varepsilon,N}^c$ implies that no stubs of at least two giant nodes are attached to one another. Since there are at most N^2 pairs of giant nodes h_1 and h_2, we can use a similar bound as in (2.A.5), to obtain

$$\mathbb{P}(\mathcal{C}_{\varepsilon,N}^c) \leq \mathbb{E}\left(N^2 \prod_{i=0}^{\lfloor D_{h_1}/2 \rfloor - 1} \left(1 - \frac{D_{h_2}}{L_N - 2i - 1} \right) \right)$$

$$\leq N^2 \left(1 - \frac{N^{\frac{1}{2}+\delta}}{N^{1+\delta}} \right)^{\frac{1}{2} N^{\frac{1}{2}+\delta}} + \frac{\varepsilon}{6} \leq N^2 e^{-\frac{1}{2} N^\delta} + \frac{\varepsilon}{6} \leq \frac{\varepsilon}{3},$$

for large enough N.

We finally show that $\mathbb{P}(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N}) \leq \frac{\varepsilon}{3}$. The event $\mathcal{B}_{\varepsilon,N}^c$ occurs if there exists a stub at node 1 or node 2 which is connected to a stub of a normal node. For $i = 1, 2$ and $j \leq D_i$, let $\{[i.j] \rightarrow [n]\}$ denote the event that the jth stub of the ith node is connected to a stub of a normal node. Let K_N denote the total number of stubs of the normal nodes. Then, clearly,

$$\mathbb{P}(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N}) \leq 2\mathbb{P}\left(\mathcal{D}_{\varepsilon,N} \cap \bigcup_{j=1}^{D_1} \{[1.j] \rightarrow [n]\} \right)$$

$$\leq 2\mathbb{E}\left[D_1 \frac{K_N}{L_N} \mathbf{1}_{\mathcal{D}_{\varepsilon,N}} \right] \leq 2q_\varepsilon \mathbb{E}\left[\frac{K_N}{L_N} \right].$$

Therefore, it suffices to prove that $\mathbb{E}\left[\frac{K_N}{L_N} \right] \rightarrow 0$. This is what we will do in the remainder of this proof. We first bound

$$L_N \geq D_{(N)} \geq \varepsilon_N u_N, \tag{2.A.21}$$

where u_N is such that (2.2.1) holds, and $\varepsilon_N \downarrow 0$ will be determined later on. To compute u_N, we use (2.4.8) to obtain

$$N[1 - F(u_N)] = N \frac{\ell(u_N)}{u_N} = 1 + o(1). \tag{2.A.22}$$

A tedious computation using

$$\ell(u_N) = \frac{(\log u_N)^{\log \log u_N - 1} \log \log u_N}{u_N},$$

yields

$$u_N = Ne^{(\log \log N)^2 - \log \log N + o(\log \log N)}. \tag{2.A.23}$$

Furthermore, since $K_N \le L_N$,

$$\mathbb{E}\left[\frac{K_N}{L_N}\right] \le (\varepsilon_N u_N)^{-1}\mathbb{E}[K_N] + \mathbb{P}(L_N \le \varepsilon_N u_N). \tag{2.A.24}$$

The second term is $o(1)$ for any $\varepsilon_N \downarrow 0$, and for the first term, we compute

$$\mathbb{E}[K_N] \le N \sum_{i=1}^{N^{\frac{1}{2}+\delta}} [1 - F(i)]. \tag{2.A.25}$$

We now use that for any $y > x^*$,

$$\begin{aligned}
\sum_{i=x^*}^{y} [1 - F(i)] &= c \sum_{i=x^*}^{y} \frac{(\log i)^{\log \log i - 1} \log \log i}{i} \\
&\le c \int_{x^*-1}^{y} \frac{(\log x)^{\log \log x - 1} \log \log x}{x} dx \\
&= c \int_{\log(x^*-1)}^{\log y} (\log y) y^{\log y - 1} dy \le c' e^{(\log \log y)^2} + \mathcal{O}(1).
\end{aligned} \tag{2.A.26}$$

Applying this to $y = N^{\frac{1}{2}+\delta}$, we obtain

$$\mathbb{E}[K_N] \le c'Ne^{(\log \log(N^{\frac{1}{2}+\delta}))^2} + \mathcal{O}(N) = Ne^{(\log \log N)^2 + 2\log(\frac{1}{2}+\delta)\log \log N + \mathcal{O}(1)}, \tag{2.A.27}$$

so that, using (2.A.23),

$$(\varepsilon_N u_N)^{-1}\mathbb{E}[K_N] = \varepsilon_N^{-1} \exp\left[\left(2\log\left(\frac{1}{2}+\delta\right)+1\right)\log \log N + o(\log \log N)\right] = o(1), \tag{2.A.28}$$

when $\delta < 1$ is so small that $2\log(\frac{1}{2}+\delta)+1 < 0$ and we take

$$\varepsilon_N = \exp\left[\frac{1}{2}\left(2\log\left(\frac{1}{2}+\delta\right)+1\right)\log \log N\right] \to 0. \tag{2.A.29}$$

This completes the proof that $\mathbb{P}(\mathcal{B}_{\varepsilon,N}^c \cap \mathcal{D}_{\varepsilon,N}) = o(1)$, and thus the proof of Lemma 2.13. $\qquad\Box$

Acknowledgement

Chapter 3

A preferential attachment model with random initial degrees

Joint work with M. Deijfen, R. van der Hofstad and G. Hooghiemstra

article [29]

Abstract

In this chapter, a random graph process $\{G(t)\}_{t \geq 1}$ is studied and its degree sequence is analyzed. Let $\{W_t\}_{t \geq 1}$ be an i.i.d. sequence. The graph process is defined so that, at each integer time t, a new vertex with W_t edges attached to it, is added to the graph. The new edges added at time t are then preferentially connected to older vertices, i.e., conditionally on $G(t-1)$, the probability that a given edge of vertex t is connected to vertex i is proportional to $d_i(t-1) + \delta$, where $d_i(t-1)$ is the degree of vertex i at time $t-1$, independently of the other edges. The main result is that the asymptotical degree sequence for this process is a power law with exponent $\tau = \min\{\tau_{\mathrm{W}}, \tau_{\mathrm{P}}\}$, where τ_{W} is the power-law exponent of the initial degrees $\{W_t\}_{t \geq 1}$ and τ_{P} the exponent predicted by pure preferential attachment. This result extends previous work by Cooper and Frieze.

3.1 Introduction

Empirical studies on real-life networks, such as the Internet, the World-Wide Web, social networks, and various types of technological and biological networks, show fascinating similarities. Many of the networks are *small worlds*, meaning that typical distances in the network are small, and many of them have *power-law degree sequences*, meaning that the number of vertices with degree k falls off as $k^{-\tau}$ for some exponent $\tau > 1$. See [40] for an example of these phenomena in the Internet, and [61, 62] for an example on the World-Wide Web. Also, Table 3.1 in [66] gives an overview of a large number of networks and their properties.

Incited by these empirical findings, random graphs have been proposed to model and/or explain these phenomena – see [31] for an introduction to random graph models for complex networks. Two particular classes of models that have been studied from a mathematical viewpoint are (i) graphs where the edge probabilities depend on certain weights associated with the vertices, see e.g. [18, 23, 25, 26, 77], and (ii) so-called preferential attachment models, see e.g. [4, 17, 20, 21, 28]. The first class can be viewed as generalizations of the classical Erdős-Rényi graph allowing for power-law degrees. Typically, the degree of a vertex is determined by its weight. Preferential attachment models are *dynamic* in the sense that a new vertex is added to the graph at each integer time. Each new vertex comes with a number of edges attached to it which are connected to the old vertices in such a way that vertices with high degree are more likely to be attached to. This has been shown to lead to graphs with power-law degree sequences, and these results are extended in the current chapter.

In preferential attachment models, the degree of a vertex increases over time, implying that the oldest vertices tend to have the largest degrees. Indeed, vertices with large degrees are the most likely vertices to obtain even larger degrees. This is sometimes called the *rich-get-richer* effect. Models where the vertex degrees are determined by associated weights, on the other hand, gives rise to something which could be referred to as *rich-by-birth* effect (a vertex is *born* with a weight which controls its degree). In reality, both these effects could play a role.

The aim of the current chapter is to formulate and analyze a model that combines the rich-get-richer and rich-by-birth effects. The model is a preferential attachment model where the number of edges added upon the addition of a new vertex is a random variable associated to the vertex. For bounded initial degrees, the model is included in the very general class of preferential attachment models treated in [28], but the novelty of the model lies in that the initial degrees can have an arbitrary distribution. In particular, we can take the weight distribution to be a power law, which gives a model with two "competing" power laws: the power law caused by the preferential attachment mechanism and the power law of the initial degrees. In such a situation it is indeed not clear which of the power laws will dominate in the resulting degrees of the graph. Our main result implies that the most heavy-tailed power law wins, that is, the degrees in the resulting graph will follow a power law with the same exponent as the initial degrees in case this is smaller than the exponent induced by the preferential attachment, and with an exponent determined by the preferential attachment in case this is smaller.

The proof of our main result requires finite moment of order $1 + \varepsilon$ for the initial degrees. However, we believe that the conclusion is true also in the infinite mean case. More specifically, we conjecture that, when the distribution of the initial degrees is a power law with infinite mean, the degree sequence in the graph will obey a power law with the same exponent as the one of the initial degrees. Indeed, the power law of the initial degrees will always be the "strongest" in this case, since preferential attachment mechanisms only seem to be able to produce power laws with finite mean. In reality, power laws with infinite mean are not uncommon, see e.g. [66, Table 3.1] for some examples, and hence it is desirable to find a model that can capture this. We have not been able to give a full proof for the infinite mean case, but we present partial results in Section 3.1.2.

3.1.1 Definition of the model

The model that we consider is described by a graph process $\{G(t)\}_{t \geq 1}$. To define it, let $\{W_i\}_{i \geq 1}$ be an independent identically distributed (i.i.d.) sequence of positive integer-valued random variables and let $G(1)$ be a graph consisting of two vertices v_0 and v_1 with W_1 edges joining them. For $t \geq 2$, the graph $G(t)$ is constructed from $G(t-1)$ in such a way that a vertex v_t, with associated weight W_t, is added to the graph $G(t-1)$, and the edge set is updated by adding W_t edges between the vertex v_t and the vertices $\{v_0, v_1, \ldots, v_{t-1}\}$. Thus, W_t is the *random initial degree* of vertex v_t. Write $d_0(s), \ldots, d_{t-1}(s)$ for the degrees of the vertices $\{v_0, v_1, \ldots, v_{t-1}\}$ at time $s \geq t-1$. The endpoints of the W_t edges emanating from vertex v_t are chosen independently (with replacement) from $\{v_0, \ldots, v_{t-1}\}$, and the probability that v_i is chosen as the endpoint of a fixed edge is equal to

$$\frac{d_i(t-1) + \delta}{\sum_{j=0}^{t-1}(d_j(t-1) + \delta)} = \frac{d_i(t-1) + \delta}{2L_{t-1} + t\delta}, \qquad 0 \leq i \leq t-1, \tag{3.1.1}$$

where $L_t = \sum_{i=1}^{t} W_i$, and δ is a fixed parameter of the model. Write S_{w} for the support of the distribution of the initial degrees. To ensure that the above expression defines a probability, we require that

$$\delta + \min\{x : x \in S_{\mathrm{w}}\} > 0. \tag{3.1.2}$$

This model will be referred to as the PARID-model (Preferential Attachment with Random Initial Degrees). Note that, when $W_i \equiv 1$ and $\delta = 0$, we retrieve the original preferential attachment model from Barabási-Albert [4].

Remark 3.1 We shall give special attention to the case where $\mathbb{P}(W_i = m) = 1$ for some integer $m \geq 1$, since it turns out that sharper error bounds are possible in this case. These sharper bounds are needed in [53], where the diameter in preferential attachment models is studied.

3.1.2 Main result

Our main result concerns the degree sequence in the graph $G(t)$. To formulate it, let $N_k(t)$ be the number of vertices with degree k in $G(t)$ and define $p_k(t) = N_k(t)/(t+1)$ as the fraction of vertices with degree k. Furthermore, let $\{r_k\}_{k \geq 1}$ be the probabilities associated with the weight distribution, that is,

$$r_k = \mathbb{P}(W_1 = k), \qquad k \geq 1. \tag{3.1.3}$$

Finally, assume that the weights have finite mean $\mu > 0$ and define $\theta = 2 + \delta/\mu$. We are interested in the limiting distribution of $p_k(t)$ as $t \to \infty$. This distribution, denoted by $\{p_k\}_{k \geq 1}$, is obtained as the solution of the recurrence relation

$$p_k = \frac{k-1+\delta}{\theta} p_{k-1} - \frac{k+\delta}{\theta} p_k + r_k. \tag{3.1.4}$$

Roughly, this relation is derived by analyzing how the number of vertices with degree k is changed upon the addition of a new vertex; see e.g. [28] for some

heuristic explanation. By iteration, it can be seen that the recursion is solved by

$$p_k = \frac{\theta}{k + \delta + \theta} \sum_{i=0}^{k-1} r_{k-i} \prod_{j=1}^{i} \frac{k - j + \delta}{k - j + \delta + \theta}, \qquad k \geq 1, \qquad (3.1.5)$$

where the empty product, arising when $i = 0$, is defined to be equal to one. Since $\{p_k\}_{k\geq1}$ satisfies (3.1.4) with $p_0 = 0$, we have that $\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} r_k = 1$. Hence, $\{p_k\}_{k\geq1}$ defines a probability distribution. Our main result states that the limiting degree distribution in the PARID-model is given by $\{p_k\}_{k\geq1}$.

Theorem 3.2 *If the initial degrees $\{W_i\}_{i\geq1}$ have finite moment of order $1 + \varepsilon$ for some $\varepsilon > 0$, then there exists a constant $\gamma \in (0, \frac{1}{2})$ such that*

$$\lim_{t\to\infty} \mathbb{P}\left(\max_{k\geq1} |p_k(t) - p_k| \geq t^{-\gamma}\right) = 0,$$

where $\{p_k\}_{k\geq1}$ is defined in (3.1.5). When $r_m = 1$ for some integer $m \geq 1$, then $t^{-\gamma}$ can be replaced by $C\sqrt{\frac{\log t}{t}}$ for some sufficiently large constant C.

To analyze the distribution $\{p_k\}_{k\geq1}$, first consider the case when the initial degrees are almost surely constant, that is, when $r_m = 1$ for some positive integer m. Then $r_j = 0$ for all $j \neq m$, and (3.1.5) reduces to

$$p_k = \begin{cases} \frac{\theta\Gamma(k+\delta)\Gamma(m+\delta+\theta)}{\Gamma(m+\delta)\Gamma(k+1+\delta+\theta)} & \text{for } k \geq m; \\ 0 & \text{for } k < m, \end{cases}$$

where $\Gamma(\cdot)$ denotes the gamma-function. By Stirling's formula, we have that $\Gamma(s + a)/\Gamma(s) \sim s^a$ as $s \to \infty$, and from this it follows that $p_k \sim ck^{-(1+\theta)}$ for some constant $c > 0$. Hence, the degree sequence obeys a power law with exponent $1 + \theta = 3 + \delta/m$. Note that, by choosing $\delta > -m$ appropriately, any value of the exponent larger than 2 can be obtained. For other choices of $\{r_k\}_{k\geq1}$, the behavior of $\{p_k\}_{k\geq1}$ is less transparent. The following proposition asserts that, if $\{r_k\}_{k\geq1}$ is a power law, then $\{p_k\}_{k\geq1}$ is a power law as well. It also gives the aforementioned characterization of the exponent as the minimum of the exponent of the r_k's and an exponent induced by the preferential attachment mechanism.

Proposition 3.3 *Assume that $r_k = \mathbb{P}(W_1 = k) = k^{-\tau_{\mathrm{W}}} L(k)$ for some $\tau_{\mathrm{W}} > 2$ and some function $k \mapsto L(k)$ which is slowly varying at infinity. Then $p_k = k^{-\tau} \hat{L}(k)$ for some slowly varying function $k \mapsto \hat{L}(k)$ and with power-law exponent τ given by*

$$\tau = \min\{\tau_{\mathrm{W}}, \tau_{\mathrm{P}}\}, \qquad (3.1.6)$$

where τ_{P} is the power-law exponent of the pure preferential attachment model given by $\tau_{\mathrm{P}} = 3 + \delta/\mu$.
When r_k decays faster than a power law, then (3.1.6) remains true with the convention that $\tau_{\mathrm{W}} = \infty$.

Now assume that the mean of the initial degrees $\{W_i\}_{i\geq1}$ is infinite. More specifically, suppose that $\{r_k\}_{k\geq1}$ is a power law with exponent $\tau_{\mathrm{W}} \in [1, 2]$. Then, we conjecture that the main result above remains true.

Conjecture 3.4 *When $\{r_k\}_{k\geq 1}$ is a power law distribution with exponent $\tau_w \in [1,2]$, then the degree sequence in the PARID-model obeys a power law with the same exponent τ_w.*

Unfortunately, we cannot quite prove Conjecture 3.4. However, we shall prove a slightly weaker version of it. To this end, write $N_{\geq k}(t)$ for the number of vertices with degree larger than or equal to k at time t, that is, $N_{\geq k}(t) = \sum_{i=0}^{t} \mathbf{1}_{\{d_i(t)\geq k\}}$, and let $p_{\geq k}(t) = N_{\geq k}(t)/(t+1)$. Since $d_i(t) \geq W_i$, obviously

$$\mathbb{E}[p_{\geq k}(t)] = \frac{\mathbb{E}[N_{\geq k}(t)]}{t+1} \geq \frac{\mathbb{E}[\sum_{i=1}^{t}\mathbf{1}_{\{W_i\geq k\}}]}{t+1}$$

$$= \mathbb{P}(W_1 \geq k)\frac{t}{t+1} = \mathbb{P}(W_1 \geq k)(1 + o(1)), \qquad (3.1.7)$$

that is, the expected degree sequence in the PARID-model is always bounded from below by the weight distribution. In order to prove a related upper bound, we start by investigating the expectation of the degrees.

Theorem 3.5 *Suppose that $\sum_{k>x} r_k = \mathbb{P}(W_1 > x) = x^{1-\tau_w}L(x)$, where $\tau_w \in (1,2)$ and $x \mapsto L(x)$ is slowly varying at infinity. Then, for every $s < \tau_w - 1$, there exists a constant $C > 0$ and a slowly varying function $x \mapsto l(x)$ such that, for $i \in \{0,\dots,t\}$,*

$$\mathbb{E}[d_i(t)^s] \leq C\Big(\frac{t}{i\vee 1}\Big)^{s/(\tau_w-1)}\Big(\frac{l(t)}{l(i)}\Big)^s,$$

where $x \vee y = \max\{x,y\}$.

As a consequence of Theorem 3.5 , we obtain:

Corollary 3.6 *If $\sum_{k>x} r_k = \mathbb{P}(W_1 > x) = x^{1-\tau_w}L(x)$, where $\tau_w \in (1,2)$ and $x \mapsto L(x)$ is a slowly varying function at infinity, then, for every $s < \tau_w - 1$, there exists an M (independent of t) such that*

$$\mathbb{E}[p_{\geq k}(t)] \leq Mk^{-s}.$$

Proof. For $s < \tau_w - 1$, it follows from Theorem 3.5 and Markov's inequality that

$$\mathbb{E}[p_{\geq k}(t)] = \frac{1}{t+1}\sum_{i=0}^{t}\mathbb{P}(d_i(t) \geq k) = \frac{1}{t+1}\sum_{i=0}^{t}\mathbb{P}(d_i(t)^s \geq k^s)$$

$$\leq \frac{1}{t+1}\sum_{i=0}^{t}k^{-s}\mathbb{E}[d_i(t)^s] \leq k^{-s}\frac{C}{t+1}\sum_{i=0}^{t}\Big(\frac{t}{i\vee 1}\Big)^{s/(\tau_w-1)}\Big(\frac{l(t)}{l(i)}\Big)^s \leq Mk^{-s},$$

$$(3.1.8)$$

since, for $s < \tau - 1$ and using [41, Theorem 2, p. 283], there exists a constant $c > 0$ such that

$$\sum_{i=0}^{t}(i\vee 1)^{-s/(\tau_w-1)}l(i)^{-s} = ct^{1-\frac{s}{\tau_w-1}}l(t)^{-s}(1 + o(1)). \qquad \square$$

Combining Corollary 3.6 with (3.1.7) yields that, when the weight distribution is a power law with exponent $\tau_w \in (1, 2)$, the only possible power law for the degrees has exponent equal to τ_w. This statement is obviously not as strong as Theorem 3.2, but it does offer convincing evidence for Conjecture 3.4. We prove Theorem 3.5 in Section 3.3.

3.1.3 Related work

Before proceeding with the proofs, we discuss how the proof of our main result is related to other proofs of similar results in the literature and describe some related work.

Virtually all proofs of asymptotic power laws in preferential attachment models consist of two steps: one step where it is proved that the degree sequence is concentrated around its mean, and one where the mean degree sequence is identified. In this chapter, these two results are formulated in Propositions 3.7 and 3.8 below, respectively. For bounded support of W_i, the concentration result and its proof are identical in all proofs. To handle the case where W_i has unbounded support, we make use of an additional coupling argument. The main differences however arise in the statement and proof of the part where the expected degree sequence is characterized. In our Proposition 3.8, a stronger result is proved than the ones for $\delta = 0$ appearing in [21] for the fixed number of edges case, and in [59] and [28] for the random number of edges case with bounded support and exponential moments respectively. More precisely, Proposition 3.8 is valid for a wider range of k values and the error term is smaller. The model in [28] − which is much more general than the model discussed here − and the model in [59] indeed also allow for a random i.i.d. number of edges $\{W_i\}_{i \geq 1}$. However, as mentioned, there W_i is assumed to have bounded support and exponential moment respectively, and hence, in those models, the competition of the exponents in (3.1.6) do not arise.

A related model related which also tries to combine the rich-get-richer and the rich-by-birth effect is the so-called *fitness model*, formulated by Barabási and Bianconi [11, 12], and later generalized by Ergün and Rodgers [35]. There the vertices are equipped with weights, referred to as *fitnesses*, which determine their ability to compete for edges. The number of edges emanating from each vertex however is fixed. Recently, the degree sequence in this model has been analyzed in [22]. Results similar to ours for various other random graph processes where a fixed number of edges is emanates from each vertex can be found in [49]. Furthermore, in [17], a *directed* preferential attachment model is investigated, and it is proved that the degrees obey a power law similar to the one in [21]. In [3], the error bound in our concentration result (Proposition 3.7) is proved for $m = 1$ for several models. For related references, see [49] and [79]. Finally, we mention [60], where a graph process is studied in which, conditionally on $G(t)$, edges to different vertices are added *independently* with probability proportional to the degree of the vertex. In this case, as in [21], the power-law exponent can only take the value $\tau = 3$, but it can be expected that by incorporating an additive δ-term as in (3.1.1), the model can be generalized to $\tau \geq 3$. However, since $\delta < 0$ is not allowed in this model (by the independence of the edges to different vertices, the degree of any vertex is zero with positive probability), we expect that $\tau < 3$ is not possible.

3.2　Proof of Theorem 3.2 and Proposition 3.3

In this section, we prove Theorem 3.2 and Proposition 3.3. We start by proving Proposition 3.3, since the proof of Theorem 3.2 makes use of it.

3.2.1　Proof of Proposition 3.3

Recall the definition (3.1.5) of p_k. Assume that $\{r_k\}_{k \geq 1}$ is a power law distribution with exponent $\tau_{\mathrm{w}} > 2$, that is, assume that $r_k = L(k)k^{-\tau_{\mathrm{w}}}$, for some slowly varying function $k \mapsto L(k)$. We want to show that then p_k is a power law distribution as well, more precisely, we want to show that $p_k = \hat{L}(k)k^{-\tau}$, where $\tau = \min\{\tau_{\mathrm{w}}, 1+\theta\}$ and $k \mapsto \hat{L}(k)$ is again a slowly varying function. To this end, first note that the expression for p_k can be rewritten in terms of gamma-functions as

$$p_k = \frac{\theta \cdot \Gamma(k+\delta)}{\Gamma(k+\delta+1+\theta)} \sum_{m=1}^{k} \frac{\Gamma(m+\delta+\theta)}{\Gamma(m+\delta)} r_m. \tag{3.2.1}$$

By Stirling's formula, we have that

$$\frac{\Gamma(k+\delta)}{\Gamma(k+\delta+1+\theta)} = k^{-(1+\theta)}\left(1 + O\big(k^{-1}\big)\right), \quad k \to \infty, \tag{3.2.2}$$

and

$$\frac{\Gamma(m+\delta+\theta)}{\Gamma(m+\delta)} = m^{\theta}\left(1 + O\big(m^{-1}\big)\right), \quad m \to \infty. \tag{3.2.3}$$

Furthermore, by the assumption, $r_m = L(m)m^{-\tau_{\mathrm{w}}}$. It follows that

$$\sum_{m=1}^{k} \frac{\Gamma(m+\delta+\theta)}{\Gamma(m+\delta)} r_m \tag{3.2.4}$$

is convergent as $k \to \infty$ if $\theta - \tau_{\mathrm{w}} < -1$, that is, if $\tau_{\mathrm{w}} > 1 + \theta$. For such values of τ_{w}, the distribution p_k is hence a power law with exponent $\tau_{\mathrm{P}} = 1 + \theta$. When $\theta - \tau_{\mathrm{w}} \geq -1$, that is, when $\tau_{\mathrm{w}} \leq \tau_{\mathrm{P}}$, the series in (3.2.4) diverges and, by [41, Lemma, p. 280], it can be seen that

$$k \mapsto \sum_{m=1}^{k} \frac{\Gamma(m+\delta+\theta)}{\Gamma(m+\delta)} r_m$$

varies regularly with exponent $\theta - \tau_{\mathrm{w}} + 1$. Combining this with (3.2.2) yields that p_k (compare (3.2.1)) varies regularly with exponent τ_{w}, as desired. $\qquad \square$

3.2.2　Proof of Theorem 3.2

As mentioned in Section 3.1.3, the proof of Theorem 3.2 consists of two parts: in the first part, we prove that the degree sequence is concentrated around its mean, and in the second part, the mean degree sequence is identified. These results are proved in two separate propositions − Proposition 3.7 and 3.8 − which are proved in Section 3.2.3 and 3.2.4 respectively.

The result on the concentration of the degree sequence is as follows:

Proposition 3.7 *If the initial degrees $\{W_i\}_{i \geq 1}$ in the PARID-model have finite moment of order $1 + \varepsilon$, for some $\varepsilon > 0$, then there exists a constant $\alpha \in (\frac{1}{2}, 1)$ such that*

$$\lim_{t \to \infty} \mathbb{P}\left(\max_{k \geq 1} \left| N_k(t) - \mathbb{E}[N_k(t)] \right| \geq t^\alpha \right) = 0.$$

When $r_m = 1$ for some $m \geq 1$, then t^α can be replaced by $C\sqrt{t \log t}$ for some sufficiently large C. Identical concentration estimates hold for $N_{\geq k}(t)$.

As for the identification of the mean degree sequence, the following proposition says that the expected number of vertices with degree k is close to $(t+1)p_k$ for large t. More precisely, it asserts that the difference between $\mathbb{E}[N_k(t)]$ and $(t+1)p_k$ is bounded, uniformly in k, by a constant times t^β, for some $\beta \in [0, 1)$.

Proposition 3.8 *Assume that the initial degrees $\{W_i\}_{i \geq 1}$ in the PARID-model have finite moment of order $1 + \varepsilon$ for some $\varepsilon > 0$, and let $\{p_k\}_{k \geq 1}$ be defined as in (3.1.5). Then there exist constants $c > 0$ and $\beta \in [0, 1)$ such that*

$$\max_{k \geq 1} |\mathbb{E}[N_k(t)] - (t+1)p_k| \leq ct^\beta. \tag{3.2.5}$$

When $r_m = 1$ for some $m \geq 1$, then the above estimate holds with $\beta = 0$.

With Propositions 3.7 and 3.8 at hand it is not hard to prove Theorem 3.2:

Proof of Theorem 3.2: Combining (3.2.5) with the triangle inequality, it follows that

$$\mathbb{P}\left(\max_{k \geq 1} \left| N_k(t) - (t+1)p_k \right| \geq ct^\beta + t^\alpha \right) \leq \mathbb{P}\left(\max_{k \geq 1} \left| N_k(t) - \mathbb{E}[N_k(t)] \right| \geq t^\alpha \right).$$

By Proposition 3.7, the right side tends to 0 as $t \to \infty$ and hence, since $p_k(t) = N_k(t)/(t+1)$, we have that

$$\lim_{t \to \infty} \mathbb{P}\left(\max_{k \geq 1} |p_k(t) - p_k| \geq \frac{ct^\beta + t^\alpha}{t+1} \right) = 0.$$

The theorem follows from this by picking $0 < \gamma < 1 - \max\{\alpha, \beta\}$. Note that, since $0 \leq \beta < 1$ and $\frac{1}{2} < \alpha < 1$, we have $0 < \gamma < \frac{1}{2}$. The proof for $r_m = 1$ is analogous. $\square$

3.2.3   Proof of Proposition 3.7

This proof is an adaption of a martingale argument, which first appeared in [21], and has been used for all proofs of power-law degree sequences since. The idea is to express the difference $N_k(t) - \mathbb{E}[N_k(t)]$ in terms of a Doob martingale. After bounding the martingale differences, which are bounded in terms of the random number of edges $\{W_i\}_{i \geq 1}$, the Azuma-Hoeffding inequality can be applied to conclude that the probability of observing large deviations is suitably small, at least when the initial number of edges has bounded support. When the initial degrees $\{W_i\}_{i \geq 1}$ are unbounded, an extra coupling step is required. The argument for $N_{\geq k}(t)$ is identical, so we focus on $N_k(t)$.

We start by giving an argument when $W_i \leq t^a$ for all $i \leq t$ and some $a \in (0, \frac{1}{2})$. First note that

$$N_k(t) \leq \frac{1}{k} \sum_{l=k}^{\infty} l N_l(t) \leq \frac{1}{k} \sum_{l=1}^{\infty} l N_l(t) = \frac{L_t}{k}. \qquad (3.2.6)$$

Thus, $\mathbb{E}[N_k(t)] \leq \mu t/k$. For $\alpha \in (\frac{1}{2}, 1)$, let $\eta > 0$ be such that $\eta + \alpha > 1$ (the choice of α will be specified in more detail below). Then, for any $k > t^\eta$, the event $|N_k(t) - \mathbb{E}[N_k(t)]| \geq t^\alpha$ implies that $N_k(t) \geq t^\alpha$, and hence that $L_t \geq k N_k(t) > t^{\eta + \alpha}$. It follows from Boole's inequality that

$$\mathbb{P}\left(\max_{k \geq 1} |N_k(t) - \mathbb{E}[N_k(t)]| \geq t^\alpha \right)$$

can be bounded from above by

$$\sum_{k=1}^{t^\eta} \mathbb{P}\left(|N_k(t) - \mathbb{E}[N_k(t)]| \geq t^\alpha \right) + \mathbb{P}(L_t > t^{\eta + \alpha}).$$

Since $\eta + \alpha > 1$ and $L_t/t \to \mu$ almost surely, the event $L_t > t^{\eta + \alpha}$ has small probability. To estimate the probability $\mathbb{P}(|N_k(t) - \mathbb{E}[N_k(t)]| \geq t^\alpha)$, introduce

$$M_n = \mathbb{E}[N_k(t)|G(n)], \qquad n = 0, \ldots, t,$$

where $G(0)$ is defined as the empty graph. Since $\mathbb{E}[M_n] < \infty$, the process is a Doob martingale with respect to $\{G(n)\}_{n=0}^t$. Furthermore, we have that $M_t = N_k(t)$ and $M_0 = \mathbb{E}[N_k(t)]$, so that

$$N_k(t) - \mathbb{E}[N_k(t)] = M_t - M_0.$$

Also, conditionally on the initial degrees $\{W_i\}_{i=1}^t$, the increments satisfy $|M_n - M_{n-1}| \leq 2W_n$. To see this, note that the additional information contained in $G(n)$ compared to $G(n-1)$ consists in how the W_n edges emanating from v_n are attached. This can affect the degrees of at most $2W_n$ vertices. By the assumption that $W_i \leq t^a$ for all $i = 1, \ldots, t$, we obtain that $|M_n - M_{n-1}| \leq 2t^a$. Combining all this, it follows from the Azuma-Hoeffding inequality – see e.g. [47, Section 12.2] – that, conditionally on $W_i \leq t^a$ for all $i = 1, \ldots, t$,

$$\mathbb{P}\left(|N_k(t) - \mathbb{E}[N_k(t)]| \geq t^\alpha \right) \leq 2 \exp\left\{ -\frac{t^{2\alpha}}{8 \sum_{i=1}^t t^{2a}} \right\} = 2 \exp\left\{ -t^{2\alpha - 1 - 2a}/8 \right\},$$

so that we end up with the estimate, again conditionally on $W_i \leq t^a$ for all $i = 1, \ldots, t$,

$$\mathbb{P}\left(\max_{k \geq 1} |N_k(t) - \mathbb{E}[N_k(t)]| \geq t^\alpha \right) \leq 2t^\eta \exp\left\{ -t^{2\alpha - 1 - 2a}/8 \right\} + \mathbb{P}(L_t > t^{\eta + \alpha}).$$
$$(3.2.7)$$

Since $a < 1/2$, the above exponential tends to 0 for any $\alpha < 1$ satisfying that $\alpha > a + 1/2$. When the initial degrees are *bounded*, the above argument can be adapted to yield that the probability that $\max_{k \geq 1} |N_k(t) - \mathbb{E}[N_k(t)]|$ exceeds

$C\sqrt{t\log t}$ is $o(1)$ for some $C > 0$ sufficiently large. We omit the details of this argument.

We conclude that Proposition 3.7 has been proved for all graphs $G(t)$ satisfying that $W_i \leq t^a$ for arbitrary $a \in (0, 1/2)$. Naturally, this assumption may not be true. When the initial degrees are bounded, the assumption is true, even with t^a replaced by m, but we are interested in graphs having initial degrees with finite $(1+\varepsilon)$-moments. We next extend the proof to this setting by a coupling argument.

Fix $a \in (0, \frac{1}{2})$, arbitrary, and define, for $i = 1, 2, \ldots, t$ and $1 \leq s \leq t$,

$$W_i' = W_i \wedge t^a, \qquad L_s' = \sum_{i=1}^{s} W_i', \tag{3.2.8}$$

where $x \wedge y = \min\{x, y\}$. Then, the above argument shows that the PARID-model with initial degrees $\{W_i'\}_{i=1}^{t}$ satisfies the claim in Proposition 3.7. Denote the graph process with initial degrees $\{W_i'\}_{i=1}^{t}$ by $\{G'(i)\}_{i=1}^{t}$ and its degrees by $d_i'(s)$, $i \leq s \leq t$. We now present a coupling between $\{G(i)\}_{i=1}^{t}$ and $\{G'(i)\}_{i=1}^{t}$.

Define the attachment probabilities in $\{G(i)\}_{i=1}^{t}$ and $\{G'(i)\}_{i=1}^{t}$ by

$$p_i(s) = \frac{d_i(s-1) + \delta}{2L_{s-1} + \delta s}, \qquad p_i'(s) = \frac{d_i'(s-1) + \delta}{2L_{s-1}' + \delta s}. \tag{3.2.9}$$

Observe that $p_i'(s)$ is properly defined since $d_i'(s-1) + \delta \geq W_i' + \delta = W_i \wedge t^a + \delta \geq 0$, for $t^a \geq \min\{x : x \in S_w\}$, which is true for t not too small.

We number the edges by saying that the edge (s, l) is the lth edge of vertex s, where $1 \leq l \leq W_s$. The aim is to couple all edges in such a way that most edges have the same starting and end vertex in G and G'. For this, we shall split the set of edges into two classes, the *successfully coupled edges*, and the *miscoupled edges*. The successfully coupled edges will have an identical starting and end vertex in both G and in G', while the miscoupled edges will either only exist in G (when $l > W_s'$ for edge (s, l)) or will have the same starting vertex, but a different end vertex in G and in G' (when $l \leq W_s'$ for edge (s, l)). We shall denote the set of miscoupled edges with number (s, l) with $s \leq t$ by $U(t)$. We now explain when an edge is miscoupled. For any $W_s' < l \leq W_s$, the edge with number (s, l) is miscoupled. In the graph $G(s)$, we attach the edge to a vertex i with probability $p_i(s)$, while in $G'(s)$, this edge is absent. For $1 \leq l \leq W_s'$, the edge with number (s, l) in *both* graphs is attached to i, where $i = 0, 1, \ldots, s-1$, with probability $m_i(s) = p_i(s) \wedge p_i'(s)$. Observe that $\sum_{i=0}^{s-1} p_i(s) = \sum_{i=0}^{s-1} p_i'(s) = 1$, but $\sum_{i=0}^{s-1} m_i(s) \leq 1$. For each edge with number (s, l) with $1 \leq l \leq W_s'$, we take *one* trial, independent of all randomness involved, with probability vector

$$(m_0(s), m_1(s), \ldots, m_{s-1}(s), \nu(s)), \tag{3.2.10}$$

where $\nu(s) = 1 - \sum_{i=0}^{s-1} m_i(s)$. If the trial ends in cell i, which happens with probability $m_i(s)$, then we attach the edge (s, l) to vertex i in both $G(s)$ and $G'(s)$, and the edge (s, l) is coupled successfully. If the trial ends in cell s, which happens with probability $\nu(s)$, then the edge (s, l) is miscoupled, so that $(s, l) \in U(s)$. Then, in the graphs $G(s)$ and $G'(s)$, respectively, we attach the edge (s, l) to

vertex $0, 1, \ldots, s-1$ according to two further, independent trials with probability vectors

$$\frac{1}{\nu(s)}(p_0(s) - m_0(s), \ldots, p_{s-1}(s) - m_{s-1}(s)), \tag{3.2.11}$$

and

$$\frac{1}{\nu(s)}(p'_0(s) - m_0(s), \ldots, p'_{s-1}(s) - m_{s-1}(s)),$$

respectively (note that since $m_i(s) = p_i(s) \wedge p'_i(s)$, these draws are indeed different a.s.). From this definition, we conclude that the probability of attaching any edge of vertex s to vertex i in the graph G has marginal probability

$$m_i(s) + \nu(s)\frac{p_i(s) - m_i(s)}{\nu(s)} = p_i(s), \tag{3.2.12}$$

as required. Similarly, this marginal probability equals $p'_i(s)$ in G', so that the graphs G and G' have the correct marginal distributions. We note that each miscoupled edge in $U(s)$ creates a difference in degrees of at most 2 in $G(s)$ and $G'(s)$, so that

$$\sum_{i=1}^{s} |d_i(s) - d'_i(s)| \leq 2|U(s)|. \tag{3.2.13}$$

Indeed, when $l > W'_s$, the edge (s, l) is absent in $G'(s)$ and present in $G(s)$, so that the sum of absolute difference in degrees is increased by at most 2, while if $l \leq W'_s$ and $(s, l) \in U(s)$, then only the end vertices of the edge (s, l) are different in $G(s)$ and $G'(s)$, so that the sum of absolute difference in degrees is again increased by at most 2.

From the above construction we get

$$\mathbb{E}[|U(s)|] = \mathbb{E}[|U(s-1)|] + 2\mathbb{E}[W_s - W'_s] + \mathbb{E}[R_s], \tag{3.2.14}$$

where R_s is the total number of miscoupled edges during the attachment of the edges with numbers (s, l) and $l \leq W'_s$. From (3.2.11), we obviously obtain

$$\mathbb{E}[R_s] = \mathbb{E}[\mathbb{E}[R_s|W_s]] = \mathbb{E}[W'_s \nu(s)] = \mathbb{E}[W'_s]\mathbb{E}[\nu(s)], \tag{3.2.15}$$

because W'_s is independent of $m_i(s)$, $i = 0, 1, \ldots, s-1$, and hence of $\nu(s)$.

In order to bound $\mathbb{E}[R_s]$, we observe that

$$\nu(s) = 1 - \sum_{i=0}^{s-1} m_i(s) = \sum_{i=0}^{s-1}[p_i(s) - (p_i(s) \wedge p'_i(s))] = \frac{1}{2}\sum_{i=0}^{s} |p_i(s) - p'_i(s)|.$$

We bound

$$|p_i(s) - p'_i(s)| = \left|\frac{d_i(s-1) + \delta}{2L_{s-1} + \delta s} - \frac{d'_i(s-1) + \delta}{2L'_{s-1} + \delta s}\right|$$

$$\leq \frac{|d_i(s-1) - d'_i(s-1)|}{2L_{s-1} + \delta s} + \frac{2(L_{s-1} - L'_{s-1})(d'_i(s-1) + \delta)}{(2L_{s-1} + \delta s)(2L'_{s-1} + \delta s)}, \tag{3.2.16}$$

because $L'_{s-1} \leq L_{s-1}$. From (3.2.16) we obtain the following upper bound for $\nu(s)$:

$$
\begin{aligned}
\nu(s) &= \frac{1}{2} \sum_{i=0}^{s-1} |p_i(s) - p'_i(s)| \leq \frac{1}{2} \sum_{i=0}^{s-1} \frac{|d_i(s-1) - d'_i(s-1)|}{2L_{s-1} + \delta s} \\
&\quad + \frac{1}{2} \sum_{i=0}^{s-1} \frac{2(L_{s-1} - L'_{s-1})(d'_i(s-1) + \delta)}{(2L_{s-1} + \delta s)(2L'_{s-1} + \delta s)} \\
&\leq \frac{|U(s-1)|}{2L_{s-1} + \delta s} + \frac{L_{s-1} - L'_{s-1}}{2L_{s-1} + \delta s}, \quad\quad\quad (3.2.17)
\end{aligned}
$$

by (3.2.13). The following lemma bounds the expected value of $|U(t)|$:

Lemma 3.9 *There exist constants $K > 0$ and $b \in (0,1)$ such that for all $t \in \mathbb{N}$,*

$$
\mathbb{E}[|U(t)|] \leq Kt^b. \quad\quad\quad (3.2.18)
$$

Proof: We prove Lemma 3.9 by induction. We start with some preparations for the induction step. Obviously, $\mathbb{E}[W'_s] \leq \mathbb{E}[W_s] = \mu$ and, from the existence of the $1 + \varepsilon$ moment of W_s, we obtain:

$$
\mathbb{E}[W_s - W'_s] = \mathbb{E}[(W_s - t^a)\mathbf{1}_{\{W_s > t^a\}}] \leq t^{-a\varepsilon} \mathbb{E}[W_s^{1+\varepsilon}] \leq Ct^{-a\varepsilon}. \quad\quad\quad (3.2.19)
$$

Secondly, from the strong law of large numbers $L_s/s \to \mu$, a.s. Using this in combination with (3.2.17), we find that, taking $\zeta > 0$ such that $2(1-\zeta)\mu + \delta = (1+\zeta)\mu > 1$, which is possible since $2\mu + \delta > \mu$,

$$
\begin{aligned}
\mathbb{E}[\nu(s)] &\leq \frac{\mathbb{E}[|U(s-1)|]}{(s-1)(1+\zeta)\mu} + \frac{2\mathbb{E}[L_{s-1} - L'_{s-1}]}{s-1} + \mathbb{P}\big(L_{s-1} \leq (1-\zeta)\mu(s-1)\big) \\
&= \frac{\mathbb{E}[|U(s-1)|]}{(s-1)(1+\zeta)\mu} + 2\mathbb{E}[W_{s-1} - W'_{s-1}] + \mathbb{P}\big(L_{s-1} \leq (1-\zeta)\mu(s-1)\big).
\end{aligned}
$$
$$\quad\quad\quad (3.2.20)$$

We are now ready to prove (3.2.18). Obviously, for any finite set of natural numbers t, the inequality (3.2.18) holds by making K sufficiently large. This initializes the induction hypothesis, and we may assume in the induction step that t is large. So assume (3.2.18) for $s-1 < t$, with s large and we will show that (3.2.18) holds for s. From (3.2.14), (3.2.15), (3.2.19), (3.2.20) and the induction hypothesis, it follows that

$$
\begin{aligned}
\mathbb{E}[|U(s)|] &\leq \mathbb{E}[|U(s-1)|] + 2\mathbb{E}[W_s - W'_s] + \mathbb{E}[R_s] \\
&\leq K(s-1)^b + 2C(1+\mu)t^{-a\varepsilon} + \frac{K(s-1)^b}{(1+\zeta)(s-1)} \\
&\quad + \mu\mathbb{P}(L_{s-1} \leq (1-\zeta)\mu(s-1)) \\
&= Ks^b \left\{ (1-1/s)^b + \frac{2C(1+\mu)}{Ks^{b+a\varepsilon}} + \frac{(1-1/s)^b}{(1+\zeta)(s-1)} \right\} \\
&\quad + \mu\mathbb{P}(L_{s-1} \leq (1-\zeta)\mu(s-1)).
\end{aligned}
$$

Standard large deviation techniques and the fact that L_t is a sum of t i.i.d. *non-negative* random variables show that $s \mapsto \mathbb{P}(L_{s-1} \leq (1-\zeta)\mu(s-1))$ converges to 0 exponentially fast for any $\zeta > 0$, so that we obtain the required bound Ks^b whenever s is sufficiently large and

$$(1-1/s)^b + \frac{2C(1+\mu)}{Ks^{b+a\varepsilon}} + \frac{(1-1/s)^b}{(1+\zeta)(s-1)} < 1.$$

This can be established when $b + a\varepsilon \geq 1$, by taking s and K sufficiently large. $\square$

We now complete the proof of Proposition 3.7. The Azuma-Hoeffding argument proves that $N_k'(t)$, the number of vertices with degree k in $G'(t)$, satisfies the bound in Proposition 3.7, i.e., that (recall (3.2.7))

$$\mathbb{P}\left(\max_{k \geq 1} |N_k'(t) - \mathbb{E}[N_k'(t)]| \geq t^\alpha\right) \leq 2t^\eta \exp\left\{-t^{2\alpha-1-2a}/8\right\} + \mathbb{P}(L_t' > t^{\eta+\alpha}),$$
(3.2.21)

for $\alpha \in (\frac{1}{2}, 1)$ and $\eta > 0$ such that $\alpha + \eta > 1$ and $a \in (0, \frac{1}{2})$. Moreover, we have for every $k \geq 1$, that

$$|N_k(t) - N_k'(t)| \leq |U(t)|,$$
(3.2.22)

since every miscoupling can change the degree of at most one vertex. By (3.2.22) and (3.2.18), there is a $b \in (0,1)$ such that

$$\left|\mathbb{E}[N_k(t)] - \mathbb{E}[N_k'(t)]\right| \leq \mathbb{E}[|U(t)|] \leq Kt^b.$$
(3.2.23)

Also, by the Markov inequality, (3.2.22) and (3.2.18), for every $\alpha \in (b, 1)$, we have that

$$\mathbb{P}\left(\max_{k \geq 1} |N_k(t) - N_k'(t)| > t^\alpha\right) \leq \mathbb{P}(|U(t)| > t^\alpha) \leq t^{-\alpha}\mathbb{E}[|U(t)|] = o(1).$$
(3.2.24)

Now fix $\alpha \in (b \vee (a + \frac{1}{2}), 1)$, where $x \vee y = \max\{x, y\}$, and decompose

$$\max_{k \geq 1} |N_k(t) - \mathbb{E}[N_k(t)]| \leq \max_{k \geq 1} |N_k'(t) - \mathbb{E}[N_k'(t)]| + \max_{k \geq 1} |\mathbb{E}[N_k(t)] - \mathbb{E}[N_k'(t)]|$$
$$+ \max_{k \geq 1} |N_k(t) - N_k'(t)|.$$
(3.2.25)

The first term on the right hand side is bounded by t^α with high probability by (3.2.21), the second term is, for t sufficiently large and with probability one, bounded by t^α by (3.2.23) while the third term is bounded by t^α with high probability by (3.2.24). This completes the proof. $\square$

3.2.4   Proof of Proposition 3.8

For $k \geq 1$, let

$$\bar{N}_k(t) = \mathbb{E}[N_k(t)|\{W_i\}_{i=1}^t]$$
(3.2.26)

denote the expected number of vertices with degree k at time t given the initial degrees $W_1, \ldots, W_t$, and define

$$\varepsilon_k(t) = \bar{N}_k(t) - (t+1)p_k, \qquad k \geq 1.$$
(3.2.27)

Also, for $Q = \{Q_k\}_{k \geq 1}$ a sequence of real numbers, define the supremum norm of Q as $\|Q\| = \sup_{k \geq 1} |Q_k|$. Using this notation, since $\mathbb{E}[\bar{N}_k(t)] = \mathbb{E}[N_k(t)]$, we have to show that there are constants $c > 0$ and $\beta \in [0, 1)$ such that

$$\|\mathbb{E}[\varepsilon(t)]\| = \sup_{k \geq 1} |\mathbb{E}[\bar{N}_k(t)] - (t+1)p_k| \leq ct^\beta, \qquad \text{for} \quad t = 0, 1, \ldots, \qquad (3.2.28)$$

where $\varepsilon(t) = \{\varepsilon_k(t)\}_{k=1}^\infty$. The plan to do this is to formulate a recursion for $\varepsilon(t)$, and then to use induction in t to establish (3.2.28). The recursion for $\varepsilon(t)$ is obtained by combining a recursion for $\bar{N}(t) = \{\bar{N}_k(t)\}_{k \geq 1}$, that will be derived below, and the recursion for p_k in (3.1.4). The hard work then is to bound the error terms in this recursion; see Lemma 3.10 below.

Let us start by deriving a recursion for $\bar{N}(t)$. To this end, for a real-valued sequence $Q = \{Q_k\}_{k \geq 0}$, with $Q_0 = 0$, introduce the operator T_t, defined as

$$(T_t Q)_k = \left(1 - \frac{k + \delta}{2L_{t-1} + t\delta}\right) Q_k + \frac{k - 1 + \delta}{2L_{t-1} + t\delta} Q_{k-1}, \qquad k \geq 1. \qquad (3.2.29)$$

When applied to $\bar{N}(t-1)$, the operator T_t describes the effect of the addition of a single edge emanating from the vertex v_t, the vertex v_t itself being excluded from the degree sequence. Indeed, there are on the average $\bar{N}_{k-1}(t-1)$ vertices with degree $k - 1$ at time $t - 1$ and a new edge is connected to such a vertex with probability $(k - 1 + \delta)/(2L_{t-1} + t\delta)$. After this connection is made, the vertex will have degree k. Similarly, there are on the average $\bar{N}_k(t-1)$ vertices with degree k at time $t - 1$. Such a vertex is hit by a new edge with probability $(k + \delta)/(2L_{t-1} + t\delta)$, and will then have degree $k + 1$. The expected number of vertices with degree k after the addition of one edge is hence given by the operator in (3.2.29) applied to $\bar{N}(t)$.

Write T_t^n for the n-fold application of T_t, and define $T_t' = T_t^{W_t}$. Then T_t' describes the change in the expected degree sequence $\bar{N}(t)$ when all the W_t edges emanating from vertex v_t have been connected, ignoring vertex v_t itself. Hence, $\bar{N}(t)$ satisfies

$$\bar{N}_k(t) = (T_t' \bar{N}(t-1))_k + \mathbf{1}_{\{W_t = k\}}, \qquad k \geq 1. \qquad (3.2.30)$$

Introduce a second operator S on sequences of real numbers $Q = \{Q_k\}_{k \geq 0}$, with $Q_0 = 0$, by (compare to (3.1.4))

$$(SQ)_k = \frac{k - 1 + \delta}{\theta} Q_{k-1} - \frac{k + \delta}{\theta} Q_k, \qquad k \geq 1, \qquad (3.2.31)$$

where $\theta = 2 + \delta/\mu$ and μ is the expectation of W_1.

The recursion (3.1.4) is given by $p_k = (Sp)_k + r_k$, with initial condition $p_0 = 0$. It is solved by $p = \{p_k\}_{k \geq 1}$, as defined in (3.1.5). Observe that

$$(t+1)p_k \;\; = \;\; tp_k + (Sp)_k + r_k = t(T_t'p)_k + r_k - \kappa_k(t), \qquad k \geq 1, \quad (3.2.32)$$

where

$$\kappa_k(t) = t(T_t'p)_k - (Sp)_k - tp_k. \qquad (3.2.33)$$

Combining (3.2.27), (3.2.30) and (3.2.32), and using the linearity of T'_t, it follows that $\varepsilon(t) = \{\varepsilon_k(t)\}_{k \geq 1}$ satisfies the recursion

$$\varepsilon_k(t) = (T'_t \varepsilon(t-1))_k + \mathbf{1}_{\{W_t = k\}} - r_k + \kappa_k(t), \qquad (3.2.34)$$

indeed,

$$\begin{aligned}
\varepsilon_k(t) &= \bar{N}_k(t) - (t+1)p_k \\
&= (T'_t \bar{N}(t-1))_k + \mathbf{1}_{\{W_t = k\}} - t(T'_t p)_k - r_k + \kappa_k(t) \\
&= (T'_t \varepsilon(t-1))_k + \mathbf{1}_{\{W_t = k\}} - r_k + \kappa_k(t).
\end{aligned}$$

Now we define $k_t = \eta t$, where $\eta \in (\mu, 2\mu + \delta)$. Since, by (3.1.2), $\delta > -\min\{x : x \in S_\mathrm{w}\} \geq -\mu$, the interval $(\mu, 2\mu + \delta) \neq \varnothing$. Also, by the law of large numbers, $L_t \leq k_t$, as $t \to \infty$, with high probability. Further, we define $\tilde{\varepsilon}_k(t) = \varepsilon_k(t)\mathbf{1}_{\{k \leq k_t\}}$ and note that, for $k \leq k_t$, the sequence $\{\tilde{\varepsilon}_k(t)\}_{k \geq 1}$ satisfies

$$\tilde{\varepsilon}_k(t) = \mathbf{1}_{\{k \leq k_t\}}(T'_t \varepsilon(t-1))_k + \mathbf{1}_{\{W_t = k\}} - r_k + \tilde{\kappa}_k(t), \qquad (3.2.35)$$

where $\tilde{\kappa}_k(t) = \kappa_k(t)\mathbf{1}_{\{k \leq k_t\}}$. It follows from $\mathbb{E}\big[\mathbf{1}_{\{W_t = k\}}\big] = r_k$ and the triangle inequality that

$$\begin{aligned}
\|\mathbb{E}[\varepsilon(t)]\| &\leq \|\mathbb{E}[\varepsilon(t) - \tilde{\varepsilon}(t)]\| + \|\mathbb{E}[\tilde{\varepsilon}(t)]\| \\
&\leq \|\mathbb{E}[\varepsilon(t) - \tilde{\varepsilon}(t)]\| + \|\mathbb{E}\big[\mathbf{1}_{(-\infty, k_t]}(\cdot)T'_t \varepsilon(t-1)\big]\| + \|\mathbb{E}[\tilde{\kappa}(t)]\|, \quad (3.2.36)
\end{aligned}$$

where $\mathbf{1}_{(-\infty, k_t]}(k) = \mathbf{1}_{\{k \leq k_t\}}$. Inequality (3.2.36) is the key ingredient in the proof of Proposition 3.8. We will derive the following bounds for the terms in (3.2.36).

Lemma 3.10 *There are constants $C_{\tilde{\varepsilon}}$, $C_\varepsilon^{(1)}$, $C_\varepsilon^{(2)}$ and $C_{\tilde{\kappa}}$, independent of t, such that for t sufficiently large and some $\beta \in [0, 1)$,*

  (a) $\|\mathbb{E}[\varepsilon(t) - \tilde{\varepsilon}(t)]\| \leq \frac{C_{\tilde{\varepsilon}}}{t^{1-\beta}}$,

  (b) $\|\mathbb{E}\big[\mathbf{1}_{(-\infty, k_t]}(\cdot)T'_t \varepsilon(t-1)\big]\| \leq \left(1 - \frac{C_\varepsilon^{(1)}}{t}\right)\|\mathbb{E}[\varepsilon(t-1)]\| + \frac{C_\varepsilon^{(2)}}{t^{1-\beta}}$,

  (c) $\|\mathbb{E}[\tilde{\kappa}(t)]\| \leq \frac{C_{\tilde{\kappa}}}{t^{1-\beta}}$.

When $r_m = 1$ for some integer $m \geq 1$, then the above bounds hold with $\beta = 0$.

Given these bounds, Proposition 3.8 is easily established.

Proof of Proposition 3.8: Recall that we want to establish (3.2.28). We shall prove this by induction on t. Fix $t_0 \in \mathbb{N}$. We start by verifying the induction hypothesis for $t \leq t_0$, thus initializing the induction hypothesis. For any $t \leq t_0$, we have

$$\|\mathbb{E}[\varepsilon(t)]\| \leq \sup_{k \geq 1} \mathbb{E}\big[\bar{N}_k(t)\big] + (t_0 + 1)\sup_{k \geq 1} p_k \leq 2(t_0 + 1), \qquad (3.2.37)$$

since there are precisely $t_0 + 1$ vertices at time t_0 and $p_k \leq 1$. This initializes the induction hypothesis, when c is so large that $2(t_0 + 1) \leq ct_0^\beta$. Next, we advance

the induction hypothesis. Assume that (3.2.28) holds at time $t-1$ and apply Lemma 3.10 to (3.2.36) to get that

$$
\begin{aligned}
\|\mathbb{E}[\varepsilon(t)]\,\| &\leq \|\mathbb{E}[\varepsilon(t)-\tilde{\varepsilon}(t)]\,\| + \|\mathbb{E}\big[\mathbf{1}_{(-\infty,k_t]}(\cdot)T_t'\varepsilon(t-1)\big]\,\| + \|\mathbb{E}[\tilde{\kappa}(t)]\,\| \\
&\leq \frac{C_{\tilde{\varepsilon}}}{t^{1-\beta}} + \left(1-\frac{C_{\varepsilon}^{(1)}}{t}\right)c(t-1)^{\beta} + \frac{C_{\varepsilon}^{(2)}}{t^{1-\beta}} + \frac{C_{\tilde{\kappa}}}{t^{1-\beta}} \\
&\leq ct^{\beta} - \frac{c\cdot C_{\varepsilon}^{(1)} - (C_{\varepsilon}^{(2)}+C_{\tilde{\varepsilon}}+C_{\tilde{\kappa}})}{t^{1-\beta}},
\end{aligned}
$$

as long as $1-\frac{C_{\varepsilon}^{(1)}}{t}\geq 0$, which is equivalent to $t\geq C_{\varepsilon}^{(1)}$. If we then choose c large so that $c\cdot C_{\varepsilon}^{(1)}\geq C_{\varepsilon}^{(2)}+C_{\varepsilon}+C_{\tilde{\kappa}}$ and $c\geq 2(t_0+1)t_0^{-\beta}$ (recall (3.2.37)) and $t_0\geq C_{\varepsilon}^{(1)}$, then we have that $\|\mathbb{E}[\varepsilon(t)]\,\|\leq ct^{\beta}$, and (3.2.28) follows by induction in t. $\qquad\square$

It remains to prove Lemma 3.10. We shall prove Lemma 3.10 (a)-(c) one by one, starting with (a).

Proof of Lemma 3.10(a): We have $\|\mathbb{E}[\varepsilon(t)-\tilde{\varepsilon}(t)]\,\|\leq \mathbb{E}[\|\varepsilon(t)-\tilde{\varepsilon}(t)\|]$, and, using the definition of $\tilde{\varepsilon}(t)$, we get that

$$
\|\varepsilon(t)-\tilde{\varepsilon}(t)\| = \sup_{k>k_t}\big|\bar{N}_k(t)-(t+1)p_k\big| \leq \sup_{k>k_t}\bar{N}_k(t) + (t+1)\sup_{k>k_t}p_k.
$$

The maximal possible degree of a vertex at time t is L_t, implying that

$$
\sup_{k>k_t}\bar{N}_k(t) = 0,\ \text{when}\ L_t\leq k_t.
$$

The latter is true almost surely when $r_m=1$ for some integer m, when t is sufficiently large, since for t large $L_t=mt\leq \eta t=k_t$, where $\eta\in(m,2m+\delta)$, by the fact that $\mu=m$ and $\delta>-m$. On the other hand, by (3.2.6), with $N_k(t)$ replaced by $\bar{N}_k(t)$ we find $\bar{N}_k(t)\leq\frac{L_t}{k_t}$ for $k\geq k_t$, and we obtain that

$$
\mathbb{E}[\sup_{k>k_t}\bar{N}_k(t)]\leq (k_t)^{-1}\mathbb{E}[L_t\mathbf{1}_{\{L_t>k_t\}}]. \tag{3.2.38}
$$

With $k_t=\eta t$ for some $\eta\in(\mu,2\mu+\delta)$, we have that

$$
\begin{aligned}
\mathbb{E}[L_t\mathbf{1}_{\{L_t>k_t\}}] &\leq k_t^{-\varepsilon}\mathbb{E}[L_t^{1+\varepsilon}\mathbf{1}_{\{L_t>k_t\}}] \\
&\leq k_t^{-\varepsilon}\mathbb{E}[|L_t-\mu t|^{1+\varepsilon}] + (\mu t)^{1+\varepsilon}k_t^{-\varepsilon}\mathbb{P}(L_t>k_t),
\end{aligned}
$$

and, by the Markov inequality

$$
\mathbb{P}(L_t>k_t)\leq\mathbb{P}\Big(|L_t-\mu t|^{1+\varepsilon}>(k_t-\mu t)^{1+\varepsilon}\Big)\leq (k_t-\mu t)^{-(1+\varepsilon)}\mathbb{E}[|L_t-\mu t|^{1+\varepsilon}].
$$

Combining the two latter results, we obtain

$$
\mathbb{E}[L_t\mathbf{1}_{\{L_t>k_t\}}]\leq k_t^{-\varepsilon}\Big(1+\Big(\frac{\mu}{\eta-\mu}\Big)^{1+\varepsilon}\Big)\mathbb{E}[|L_t-\mu t|^{1+\varepsilon}]. \tag{3.2.39}
$$

To bound the last expectation, we will use a consequence of the Marcinkiewicz-Zygmund inequality, see e.g [48, Corollary 8.2 in §3], which runs as follows. Let

$q \in [1,2]$, and suppose that $\{X_i\}_{i \geq 1}$ is an i.i.d. sequence with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[|X_1|^q] < \infty$. Then there exists a constant c_q depending only on q, such that

$$\mathbb{E}\Big[\Big|\sum_{i=1}^{t} X_i\Big|^q\Big] \leq c_q t \mathbb{E}[|X_1{}^q|]. \tag{3.2.40}$$

Applying (3.2.40) with $q = 1 + \varepsilon$, we obtain

$$\mathbb{E}[\sup_{k > k_t} \bar{N}_k(t)] \leq k_t^{-(1+\varepsilon)}\Big(1 + \Big(\frac{\mu}{\eta - \mu}\Big)^{1+\varepsilon}\Big)\mathbb{E}[|L_t - \mu t|^{1+\varepsilon}] \leq c_{1+\varepsilon} t^{-\varepsilon}. \tag{3.2.41}$$

Furthermore, since by Proposition 3.3, we have $p_k \leq ck^{-\gamma}$, for some $\gamma > 2$ (see also (3.1.6)), we have that $\sup_{k > k_t} p_k \leq ct^{-\gamma}$ for some constant c. It follows that

$$(t+1)\sup_{k > k_t} p_k \quad \leq \quad \frac{C_p}{t^{\gamma - 1}},$$

and, since $\gamma > 2$, part (a) is established with $C_{\tilde{\varepsilon}} = c_{1+\varepsilon} + C_p$, and $1 - \beta = (\varepsilon \wedge \gamma) - 1$. $\square$

Proof of Lemma 3.10(b):   Moving on to (b), we will start by showing that for t sufficiently large,

$$\|\mathbb{E}\big[\mathbf{1}_{(-\infty, k_t]}(\cdot) T_t \varepsilon(t-1)\big]\| \leq \Big(1 - \frac{C_\varepsilon^{(1)}}{t}\Big)\|\mathbb{E}\big[\mathbf{1}_{(-\infty, k_t]}(\cdot)\varepsilon(t-1))\big]\| + \frac{C_\varepsilon^{(3)}}{t^{1-\beta}}, \tag{3.2.42}$$

which is (b) when we condition on $W_t = 1$. We shall extend the proof to the case where $W_t \geq 1$ at a later stage. To prove (3.2.42), we shall prove a related bound, which also proves useful in the extension to $W_t \geq 1$. Indeed, we shall prove, for any real-valued sequence $Q = \{Q_k\}_{k \geq 0}$, satisfying (i) $Q_0 = 0$ and (ii)

$$\sup_{k \geq 1} |k + \delta||Q_k| \leq C_Q L_{t-1}, \tag{3.2.43}$$

that there exists a $\beta \in (0,1)$ (independent of Q) and a constant $c > 0$ such that for t sufficiently large,

$$\|\mathbb{E}\big[\mathbf{1}_{(-\infty, k_t]}(\cdot) T_t Q\big]\| \leq \Big(1 - \frac{C_\varepsilon^{(1)}}{t}\Big)\|\mathbb{E}\big[\mathbf{1}_{(-\infty, k_t]}(\cdot) Q\big]\| + \frac{cC_Q}{t^{1-\beta}}. \tag{3.2.44}$$

Here we stress that Q can be *random*, for example, we shall apply (3.2.44) to $\varepsilon(t-1)$ in order to derive (3.2.42).
    In order to prove (3.2.44), we recall that

$$\mathbb{E}[(T_t Q)_k] = \mathbb{E}\left[\Big(1 - \frac{k + \delta}{2L_{t-1} + t\delta}\Big) Q_k + \frac{k - 1 + \delta}{2L_{t-1} + t\delta} Q_{k-1}\right], \qquad k \geq 1. \tag{3.2.45}$$

In bounding this expectation we will encounter a problem in that Q_k, which is allowed to be random, and L_{t-1} are not independent (for example when $Q = \varepsilon(t-1)$). To get around this, we add and subtract the expression on the right

hand side but with the random quantities replaced by their expectations, that is, for $k \geq 1$, we write

$$\mathbb{E}[(T_t Q)_k] = \left(1 - \frac{k+\delta}{2\mu(t-1)+t\delta}\right)\mathbb{E}[Q_k] + \frac{k-1+\delta}{2\mu(t-1)+t\delta}\mathbb{E}[Q_{k-1}] \qquad (3.2.46)$$

$$+ \quad (k+\delta)\mathbb{E}\left[Q_k \frac{2L_{t-1} - 2\mu(t-1)}{(2L_{t-1}+t\delta)(2\mu(t-1)+t\delta)}\right] \qquad (3.2.47)$$

$$+ \quad (k+\delta-1)\mathbb{E}\left[Q_{k-1}\frac{2\mu(t-1) - 2L_{t-1}}{(2L_{t-1}+t\delta)(2\mu(t-1)+t\delta)}\right]. \qquad (3.2.48)$$

Note that, when $r_m = 1$ for some integer $m \geq 1$, then $L_t = \mu t = mt$. Hence the terms in (3.2.47, 3.2.48) are both equal to zero, and only (3.2.46) contributes. We first deal with (3.2.46). Observe that $k \leq k_t = \eta t$, with $\eta \in (\mu, 2\mu + \delta)$, implies that $k \leq (2\mu + \delta)(t-1)$ for t sufficiently large, and hence

$$1 - \frac{k+\delta}{2\mu(t-1)+t\delta} \geq 0. \qquad (3.2.49)$$

It follows that, for t sufficiently large,

$$\sup_{k \leq k_t}\left|\left(1 - \frac{k+\delta}{2\mu(t-1)+t\delta}\right)\mathbb{E}[Q_k] + \frac{k-1+\delta}{2\mu(t-1)+t\delta}\mathbb{E}[Q_{k-1}]\right| \qquad (3.2.50)$$

$$\leq \left(1 - \frac{1}{2\mu(t-1)+t\delta}\right)\|\mathbb{E}[\mathbf{1}_{(-\infty,k_t]}(\cdot)Q]\|$$

$$\leq \left(1 - \frac{C_\varepsilon^{(1)}}{t}\right)\|\mathbb{E}[\mathbf{1}_{(-\infty,k_t]}(\cdot)Q]\|,$$

for some constant $C_\varepsilon^{(1)}$. This proves (3.2.44) — with $C_Q = 0$ — when the number of edges is a.s. constant since (3.2.47, 3.2.48) are zero. It remains to bound the terms (3.2.47, 3.2.48) in the case where the number of edges is not a.s. constant. We will prove that the supremum over k of the absolute values of both these terms are bounded by constants divided by $t^{1-\beta}$ for some $\beta \in [0, 1)$. Starting with (3.2.47), by using the assumption (ii) in (3.2.43), as well as $2L_{t-1} + \delta t \geq L_{t-1}$ for t sufficiently large, it follows that

$$\sup_{k \geq 1}\left|(k+\delta)\mathbb{E}\left[Q_k \frac{2L_{t-1} - 2\mu(t-1)}{(2L_{t-1}+t\delta)(2\mu(t-1)+t\delta)}\right]\right| \leq \frac{cC_Q}{t}\mathbb{E}[|L_{t-1} - \mu(t-1)|].$$

To bound the latter expectation, we combine (3.2.40) for $q = 1 + \varepsilon$, with Hölders inequality, to obtain

$$\mathbb{E}[|L_t - \mu t|] \leq \left(\mathbb{E}\left[|L_t - \mu t|^{1+\varepsilon}\right]\right)^{1/(1+\varepsilon)}$$

$$\leq \left(c_{1+\varepsilon}t\mathbb{E}\left[|W_1 - \mu|^{1+\varepsilon}\right]\right)^{1/(1+\varepsilon)} \leq ct^{1/(1+\varepsilon)}, \qquad (3.2.51)$$

since W_i have finite moment of order $1 + \varepsilon$ by assumption, where, without loss of generality, we can assume that $\varepsilon \leq 1$. Hence, we have shown that the supremum over k of the absolute value of (3.2.47) is bounded from above by a constant

divided by $t^{1-\beta}$, where $\beta = 1/(1+\varepsilon)$. That the same is true for the term (3.2.48) can be seen analogously. This completes the proof of (3.2.44).

To prove (3.2.42), we note that, by convention, $\varepsilon_0(t-1) = 0$, so that we only need to prove that $\sup_{k\geq1} |k+\delta||\varepsilon_k(t-1)| \leq cL_{t-1}$. For this, note from (3.2.6), the bound $p_k \leq ck^{-\gamma}$, $\gamma > 2$, and from the lower bound $L_t \geq t$ that

$$
\begin{aligned}
\sup_{k\geq1}|k+\delta||\varepsilon_k(t-1)| &\leq \sum_{k\geq1}(k+|\delta|)|\varepsilon_k(t-1)| \\
&\leq \sum_{k\geq1}(k+|\delta|)\bar{N}_k(t-1) + t\sum_{k\geq1}(k+|\delta|)p_k \\
&\leq L_{t-1} + |\delta|(t-1) + t\sum_{k\geq1}(k+|\delta|)p_k \leq cL_{t-1}, \quad (3.2.52)
\end{aligned}
$$

for some constant c. This completes the proof of (3.2.42).

To complete the proof of Lemma 3.10(b), we first show that (3.2.44) implies, for every $1 \leq n \leq t$, and all $k \geq 1$,

$$
\mathbb{E}\big[\mathbf{1}_{\{k\leq k_t\}}\big(T_t^n\varepsilon(t-1)\big)_k\big] \leq \Big(1 - \frac{C_\varepsilon^{(1)}}{t}\Big)\|\mathbb{E}\big[\mathbf{1}_{(-\infty,k_t]}(\cdot)\varepsilon(t-1)\big]\| + \frac{nC_\varepsilon^{(3)}}{t^{1-\beta}}. \tag{3.2.53}
$$

To see (3.2.53), we use induction on n. We note that (3.2.53) for $n = 1$ is precisely equal to (3.2.42), and this initializes the induction hypothesis. To advance the induction hypothesis, we note that

$$
\mathbf{1}_{\{k\leq k_t\}}\big(T_t^n\varepsilon(t-1)\big)_k = \mathbf{1}_{\{k\leq k_t\}}T_t\big(Q(n-1)\big)_k, \tag{3.2.54}
$$

where $Q_k(n-1) = \mathbf{1}_{\{k\leq k_t\}}\big(T_t^{n-1}\varepsilon(t-1)\big)_k$. We wish to use (3.2.44), and we first check the assumptions (i-ii). By definition, $Q_0(n-1) = 0$, which establishes (i). For assumption (ii), we need to do some more work. According to (3.2.29), and using that $2L_{t-1} + t\delta > L_{t-1} \geq t - 1$, for t sufficiently large,

$$
\sum_{k=1}^{\infty}(k+|\delta|)(T_tQ)_k \leq \Big(1 + \frac{1}{t}\Big)\sum_{k=1}^{\infty}(k+|\delta|)Q_k,
$$

and hence, by induction,

$$
\sum_{k=1}^{\infty}(k+|\delta|)(T_t^{n-1}Q)_k \leq \Big(1 + \frac{1}{t}\Big)^{n-1}\sum_{k=1}^{\infty}(k+|\delta|)Q_k.
$$

Substituting $Q_k = \varepsilon_k(t-1)$ and using $|\varepsilon_k(t-1)| \leq N_k(t-1) + tp_k$, yields

$$
\begin{aligned}
\sum_{k\leq k_t}&(k+|\delta|)(T_t^{n-1}N(t-1))_k + t\sum_{k\leq k_t}(k+|\delta|)(T_t^{n-1}p)_k \\
&\leq \Big(1+\frac{1}{t}\Big)^{n-1}\sum_{k=1}^{\infty}(k+|\delta|)N_k(t-1) + \Big(1+\frac{1}{t}\Big)^{n-1}t\sum_{k=1}^{\infty}(k+|\delta|)p_k \\
&\leq \Big(1+\frac{1}{t}\Big)^{n-1}\cdot cL_{t-1}, \tag{3.2.55}
\end{aligned}
$$

according to (3.2.52). Using the inequality $1+x \le e^x$, $x \ge 0$, together with $n \le t$, this in turn yields,

$$\sup_{k \ge 1} |k + \delta||Q_k(n-1)| \le ecL_{t-1}, \tag{3.2.56}$$

which implies assumption (ii).

By the induction hypothesis, we have that, for $k \le k_t$,

$$\mathbb{E}[Q_k(n-1)] \le \left(1 - \frac{C_\varepsilon^{(1)}}{t}\right) \|\mathbb{E}[\mathbf{1}_{(-\infty,k_t]}(\cdot)\varepsilon(t-1)]\| + \frac{(n-1)C_\varepsilon^{(3)}}{t^{1-\beta}}, \tag{3.2.57}$$

so that we obtain, from (3.2.44), with $Q = \mathbf{1}_{(-\infty,k_t]}(\cdot)T_t\varepsilon(t-1)$,

$$\mathbb{E}\big[\mathbf{1}_{\{k \le k_t\}}\big(T_t^n\varepsilon(t-1)\big)_k\big] \le \left(1 - \frac{C_\varepsilon^{(1)}}{t}\right) \|\mathbb{E}[\mathbf{1}_{(-\infty,k_t]}(\cdot)\varepsilon(t-1)]\|$$
$$+ \frac{(n-1)C_\varepsilon^{(3)} + cC_Q}{t^{1-\beta}}, \tag{3.2.58}$$

which advances the induction hypothesis when $C_\varepsilon^{(3)} > cC_Q$.

By (3.2.58), we obtain that, for $W_t \le t$,

$$\mathbb{E}\big[\mathbf{1}_{\{k \le k_t\}}\big(T_t'\varepsilon(t-1)\big)_k\big|W_t\big] \le \left(1 - \frac{C_\varepsilon^{(1)}}{t}\right) \|\mathbb{E}[\varepsilon(t-1)|W_t]\| + \frac{W_tC_\varepsilon^{(3)}}{t^{1-\beta}}$$
$$= \left(1 - \frac{C_\varepsilon^{(1)}}{t}\right) \|\mathbb{E}[\varepsilon(t-1)]\| + \frac{W_tC_\varepsilon^{(3)}}{t^{1-\beta}},$$

where we use that $\varepsilon(t-1)$ is independent of W_t. In the case that $W_t > t$, we bound, similarly as in (3.2.52),

$$\sup_{k \le k_t} \big|\big(T_t'\varepsilon(t-1)\big)_k\big| \le cL_t, \tag{3.2.59}$$

so that

$$\mathbb{E}\big[\mathbf{1}_{\{k \le k_t\}}\big(T_t'\varepsilon(t-1)\big)_k\big|W_t\big] \le \left(1 - \frac{C_\varepsilon^{(1)}}{t}\right) \|\mathbb{E}[\varepsilon(t-1)]\| + \frac{W_tC_\varepsilon^{(3)}}{t^{1-\beta}}$$
$$+ c\mathbb{E}[L_t\mathbf{1}_{\{W_t>t\}}\big|W_t].$$

The bound in (b) follows from this by taking expectations on both sides, using

$$\mathbb{E}[L_t\mathbf{1}_{\{W_t>t\}}] = \mu(t-1)\mathbb{P}(W_t > t) + \mathbb{E}[W_t\mathbf{1}_{\{W_t>t\}}] \le \left(\frac{\mu}{t^\varepsilon} + \frac{1}{t^\varepsilon}\right)\mathbb{E}[W_t^{1+\varepsilon}], \tag{3.2.60}$$

after which we use that $\beta = 1/(1+\varepsilon) \ge 1-\varepsilon$ and choose the constants appropriately. This completes the proof of Lemma 3.10(b). □

Proof of Lemma 3.10(c): For part (c) of the lemma, recall that

$$\tilde{\kappa}_k(t) = \kappa_k(t)\mathbf{1}_{\{k \le k_t\}} \qquad \text{with} \qquad \kappa_k(t) = t((T_t' - I)p)_k - (Sp)_k, \tag{3.2.61}$$

where T_t is defined in (3.2.29), $T'_t = T_t^{W_t}$, S is defined in (3.2.31), and where I denotes the identity operator. In what follows, we will assume that $k \leq k_t$, so that $\tilde{\kappa}_k(t) = \kappa_k(t)$. We start by proving a trivial bound on $\kappa_k(t)$. By (3.2.34), we have that

$$\kappa_k(t) = \varepsilon_k(t) - (T'_t \varepsilon(t-1))_k - \mathbf{1}_{\{W_t = k\}} + r_k, \qquad (3.2.62)$$

where $\sup_{k \geq 1} |\varepsilon_k(t)| \leq cL_t$ by (3.2.52) and $\sup_{1 \leq k \leq k_t} |(T'_t \varepsilon(t-1))_k| \leq cL_t$ by (3.2.59), so that hence

$$\sup_{k \leq k_t} |\kappa_k(t)| \leq C_\eta L_t \qquad (3.2.63)$$

for some C_η (recall that $k_t = \eta t$ where $\eta \in (\mu, 2\mu + \delta)$). For $x \in [0, 1]$ and $w \in \mathbb{N}$, we denote

$$f_k(x; w) = \big((I + x(T_t - I))^w p\big)_k.$$

Then $\kappa_k(t) = \kappa_k(t; W_t)$, where

$$\kappa_k(t; w) = t[f_k(1; w) - f_k(0; w)] - (Sp)_k, \qquad (3.2.64)$$

and $x \mapsto f_k(x; w)$ is a polynomial in x of degree w. By a Taylor expansion around $x = 1$,

$$f_k(1; w) = p_k + w\big((T_t - I)p\big)_k + \frac{1}{2} f''_k(x_k; w), \qquad (3.2.65)$$

for some $x_k \in (0, 1)$, and, since $I + x(T_t - I)$ and $T_t - I$ commute,

$$f''_k(x; w) = w(w-1)\Big((I + x(T_t - I))^{w-2}(T_t - I)^2 p\Big)_k.$$

We next claim that, on the event $\{k_t \leq 2L_{t-1} + (t-1)\delta\}$,

$$\sup_{k \leq k_t} \Big|\big((I + x(T_t - I))Q\big)_k\Big| \leq \sup_{k \leq k_t} |Q_k|.$$

Indeed, $I + x(T_t - I) = (1 - x)I + xT_t$ and $x \in [0, 1]$, so that the claim follows when $\sup_{k \leq k_t} |(T_t Q)_k| \leq \sup_{k \leq k_t} |Q_k|$. The latter is the case, since, on the event that $k + \delta \leq 2L_{t-1} + t\delta$, and arguing as in (3.2.50), we have

$$
\begin{aligned}
\sup_{k \leq k_t} |(T_t Q)_k| &\leq \sup_{k \leq k_t} \Big[\Big(1 - \frac{k+\delta}{2L_{t-1} + t\delta}\Big)|Q_k| + \frac{k-1+\delta}{2L_{t-1} + t\delta}|Q_{k-1}|\Big] \\
&\leq \Big(1 - \frac{1}{2L_{t-1} + t\delta}\Big) \sup_{k \leq k_t} |Q_k|.
\end{aligned}
$$

Since $k \leq k_t$, the inequality $k + \delta \leq 2L_{t-1} + t\delta$ follows when $k_t \leq 2L_{t-1} + (t-1)\delta$.

As a result, on the event $\{k_t \leq 2L_{t-1} + (t-1)\delta\}$, we have that

$$\max_{x \in [0,1]} \sup_{k \leq k_t} |f''_k(x; w)| \leq w(w-1) \sup_{k \leq k_t} \big|((T_t - I)^2 p)_k\big|. \qquad (3.2.66)$$

Now recall the definition (3.2.31) of the operator S, and note that, for any sequence $Q = \{Q_k\}_{k=1}^\infty$, we can write

$$((T_t - I)Q)_k = \frac{\theta}{(2L_{t-1} + t\delta)}(SQ)_k = \frac{1}{t\mu}(SQ)_k + (R_t Q)_k, \qquad (3.2.67)$$

where the remainder operator R_t is defined as

$$(R_t Q)_k = \left(\frac{k+\delta}{2t\mu + t\delta} - \frac{k+\delta}{2L_{t-1} + t\delta} \right) Q_k + \left(\frac{k-1+\delta}{2L_{t-1} + t\delta} - \frac{k-1+\delta}{2t\mu + t\delta} \right) Q_{k-1}.$$
(3.2.68)

Combining (3.2.64), (3.2.65), (3.2.66) and (3.2.67), on the event $\{ k_t \leq 2L_{t-1} + (t-1)\delta \}$ and uniformly for $k \leq k_t$, we obtain that

$$\kappa_k(t; w) \leq \left(\frac{w}{\mu} - 1 \right) (Sp)_k + wt \sup_{k \leq k_t} |(R_t p)_k| + \frac{1}{2} w(w-1)t \sup_{k \leq k_t} \left| \left((T_t - I)^2 p \right)_k \right|,$$
(3.2.69)

together with a similar lower bound with minus signs in front of the last two terms. Indeed,

$$
\begin{aligned}
\kappa_k(t; w) &= t[f_k(1; w) - f_k(0; w)] - (Sp)_k \\
&= tw\big((T_t - I)p \big)_k + \frac{1}{2} t f_k''(x_k; w) - (Sp)_k \\
&= \frac{wt}{\mu t} (Sp)_k + wt(Rp)_k - (Sp)_k + \frac{1}{2} t f_k''(x_k; w),
\end{aligned}
$$

and (3.2.69) follows from this identity and (3.2.66).

With (3.2.69) at hand, we are now ready to complete the proof of (c). We start by treating the case where $r_m = 1$ for some integer $m \geq 1$. In this case, with $w = W_t = m = \mu$, we have that $(\frac{w}{\mu} - 1)(Sp)_k \equiv 0$. Furthermore, the inequality $k_t \leq 2L_{t-1} + (t-1)\delta$ is true almost surely when t is sufficiently large. Hence, we are done if we can bound the last two terms in (3.2.69) with $w = W_t$. To do this, note that, by the definition (3.2.29) of T_t and the fact that $2L_{t-1} + t\delta \geq k_t = \eta t$, with $\eta > \mu$,

$$\sup_{k \geq 1} \left| \left((T_t - I)Q \right)_k \right| \leq \frac{2}{\eta t} \sup_{k \geq 1} (k + |\delta|)|Q_k|.$$
(3.2.70)

Applying (3.2.70) twice yields that

$$\left| \left((T_t - I)^2 p \right)_k \right| \leq \frac{4}{\eta^2 t^2} \sup_{k \geq 1} (k + |\delta|)^2 p_k,$$

and, since by Proposition 3.3, $p_k \leq ck^{-\gamma}$ for some $\gamma > 2$, there is a constant $\tilde{C}_p$ such that

$$\sup_{k \geq 1} (k + |\delta|)^2 p_k \leq \tilde{C}_p.$$
(3.2.71)

Finally, since $L_t = mt$, we have that

$$\left| (R_t p)_k \right| \leq \frac{2}{m(t-1)t} \sup_{k \geq 1} (k + |\delta|) p_k \leq \frac{2\tilde{C}_p}{m(t-1)t}.$$

Summarizing, we arrive at the statement that there exists $c_{m,\delta}$ such that

$$\sup_{k \leq k_t} |\kappa_k(t; m)| \leq \frac{c_{m,\delta}}{t},$$

which proves the claim in (c) with $\beta = 0$ when $r_m = 1$.

We now move to random initial degrees. For any $a \in (0, 1)$, we can split

$$\kappa_k(t) = \kappa_k(t)\mathbf{1}_{\{W_t \leq t^a\}} + \kappa_k(t)\mathbf{1}_{\{W_t > t^a\}}. \tag{3.2.72}$$

On the event $\{k_t \leq 2L_{t-1} + (t-1)\delta\}$, the first term of (3.2.72) can be bounded by the right side of (3.2.69), i.e.,

$$\kappa_k(t)\mathbf{1}_{\{W_t \leq t^a\}} \leq \Big((W_t/\mu - 1)(Sp)_k + tW_t \sup_{k \leq k_t} |(R_t p)_k| $$
$$+ \frac{W_t(W_t - 1)}{2} t \sup_{k \leq k_t} \big|\big((T_t - I)^2 p\big)_k\big| \Big) \mathbf{1}_{\{W_t \leq t^a\}},$$

with a similar lower bound where the last two terms have a minus sign. From (3.2.63), we obtain the upper bound

$$\kappa_k(t)\mathbf{1}_{\{W_t > t^a\}} \leq C_\eta L_t \mathbf{1}_{\{W_t > t^a\}}.$$

Combining these two upper bounds with (3.2.72), and adding the term $(W_t/\mu - 1)(Sp)_k\mathbf{1}_{\{W_t > t^a\}}$ to the right side, yields that on the event that $\{k_t \leq 2L_{t-1} + (t-1)\delta\}$,

$$\kappa_k(t) \leq \left(\frac{W_t}{\mu} - 1\right)(Sp)_k + tW_t\mathbf{1}_{\{W_t \leq t^a\}} \sup_{k \leq k_t} |(R_t p)_k| \tag{3.2.73}$$
$$+ tW_t^2 \mathbf{1}_{\{W_t \leq t^a\}} \sup_{k \leq k_t} \big|\big((T_t - I)^2 p\big)_k\big| + \mathbf{1}_{\{W_t > t^a\}} C_\eta L_t,$$

and similarly we get as a lower bound, using $|W_t/\mu - 1| \leq W_t$,

$$\kappa_k(t) \geq \left(\frac{W_t}{\mu} - 1\right)(Sp)_k - tW_t\mathbf{1}_{\{W_t \leq t^a\}} \sup_{k \leq k_t} |(R_t p)_k| \tag{3.2.74}$$
$$- tW_t^2 \mathbf{1}_{\{W_t \leq t^a\}} \sup_{k \leq k_t} \big|\big((T_t - I)^2 p\big)_k\big| - \mathbf{1}_{\{W_t > t^a\}} \left(C_s W_t + C_\eta L_t\right),$$

where we used that $\sup_{k \geq 1} |(Sp)_k| \leq C_s$. We use (3.2.73) and (3.2.74) on $\{k_t \leq 2L_{t-1} + (t-1)\delta\}$, and (3.2.63) on the event $\{k_t > 2L_{t-1} + (t-1)\delta\}$ to arrive at

$$\kappa_k(t) \leq \left(\frac{W_t}{\mu} - 1\right)(Sp)_k + tW_t\mathbf{1}_{\{W_t \leq t^a\}} \sup_{k \leq k_t} |(R_t p)_k| \tag{3.2.75}$$
$$+ tW_t^2 \mathbf{1}_{\{W_t \leq t^a\}} \sup_{k \leq k_t} \big|\big((T_t - I)^2 p\big)_k\big|$$
$$+ \big(\mathbf{1}_{\{W_t > t^a\}} + \mathbf{1}_{\{k_t > 2L_{t-1} + (t-1)\delta\}}\big) \left((C_s + C_\eta)W_t + C_\eta L_{t-1}\right),$$

with a similar lower bound where the last three terms have a minus sign. We now take expectations on both sides of (3.2.75) and take advantage of the equality $\mathbb{E}[W_t/\mu] = 1$ and the property that $(Sp)_k$ is deterministic, so that the first term on the right side drops out. Moreover, using that W_t and L_{t-1} are independent,

as well as that $k_t > 2L_{t-1} + (t-1)\delta$ implies that $L_{t-1} \le k_t$, we arrive at

$$\left|\mathbb{E}[\kappa_k(t)]\right| \le \mathbb{E}\Big[\mathbf{1}_{\{W_t > t^a\}}\big((C_s + C_\eta)W_t + C_\eta\mu t\big)\Big] \tag{3.2.76}$$

$$+ \Big(C_\eta k_t + (C_s + C_\eta)\mu\Big)\mathbb{P}(k_t > 2L_{t-1} + (t-1)\delta) \tag{3.2.77}$$

$$+ t\mathbb{E}\Big[\sup_{k\le k_t}|(R_tp)_k|\Big]\mathbb{E}\Big[W_t\mathbf{1}_{\{W_t > t^a\}}\Big] \tag{3.2.78}$$

$$+ t\mathbb{E}[W_t^2\mathbf{1}_{\{W_t \le t^a\}}]\mathbb{E}\Big[\sup_{k\le k_t}\big|\big((T_t - I)^2p\big)_k\big|\Big]. \tag{3.2.79}$$

We now bound each of these four terms one by one. To bound (3.2.76), we use that W_t has finite $(1+\varepsilon)$-moment, to obtain that

$$\mathbb{E}\big[\mathbf{1}_{\{W_t > t^a\}}W_t\big] = \mathbb{E}\big[\mathbf{1}_{\{W_t > t^a\}}W_t^{-\varepsilon}W_t^{1+\varepsilon}\big] \le t^{-a\varepsilon}\mathbb{E}\big[W_t^{1+\varepsilon}\big] = O(t^{-a\varepsilon}),$$

and,

$$t\mathbb{E}\big[\mathbf{1}_{\{W_t > t^a\}}\big] = t\mathbb{P}\big(W_t^{1+\varepsilon} > t^{a(1+\varepsilon)}\big) \le t^{1-a(1+\varepsilon)}\mathbb{E}\big[W_t^{1+\varepsilon}\big] = O(t^{1-a(1+\varepsilon)}),$$

which bounds (3.2.76) as

$$\mathbb{E}\Big[\mathbf{1}_{\{W_t > t^a\}}\big((C_s + C_\eta)W_t + C_\eta\mu t\big)\Big] = O(t^b), \tag{3.2.80}$$

with $b = \max\{-a\varepsilon, 1 - a(1+\varepsilon)\}$.

To bound (3.2.77), we use that when $k_t > 2L_{t-1} + (t-1)\delta$, then $L_{t-1} < \frac{1}{2}(\eta t - \delta(t-1)) = \frac{1}{2}(\eta - \delta)(t-1) + \frac{1}{2}\eta$. Now, since $\eta \in (\mu, 2\mu + \delta)$, we have that $\frac{1}{2}(\eta - \delta) < \mu$. Standard large deviation theory and the fact that the initial degrees W_i are non-negative give that the probability that $L_{t-1} < \sigma(t-1)$, with $\sigma < \mu$, is exponentially small in t. As a result, we obtain that

$$\Big(C_\eta k_t + (C_s + C_\eta)\mu\Big)\mathbb{P}\Big(k_t > 2L_{t-1} + (t-1)\delta\Big) = O(t^{-1}). \tag{3.2.81}$$

To bound (3.2.78), we use that $2L_{t-1} + t\delta \ge L_{t-1} \ge t - 1 \ge t/2$, and also use (3.2.71), to obtain that

$$\mathbb{E}\Big[\sup_{k\le k_t}|(R_tp)_k|\Big] \le \frac{c}{t^2}\mathbb{E}|L_{t-1} - t\mu|\sup_{k\ge 1}(k + |\delta|)p_k \le \frac{c}{t^2}\mathbb{E}|L_{t-1} - t\mu|.$$

Thus,

$$t\mathbb{E}\Big[\sup_{k\le k_t}|(R_tp)_k|\Big]\mathbb{E}\Big[W_t\mathbf{1}_{\{W_t > t^a\}}\Big] \le \frac{c}{t}\mathbb{E}|L_{t-1} - t\mu| \cdot t^{-a\varepsilon} \le O\left(t^{-a\varepsilon - \varepsilon/(1+\varepsilon)}\right), \tag{3.2.82}$$

where the final bound follows from (3.2.51).

Finally, to bound (3.2.79), note that

$$\mathbb{E}[W_t^2\mathbf{1}_{\{W_t \le t^a\}}] = \mathbb{E}[W_t^{1-\varepsilon}W_t^{1+\varepsilon}\mathbf{1}_{\{W_t \le t^a\}}] \le t^{a(1-\varepsilon)}\mathbb{E}[W_t^{1+\varepsilon}] = O\left(t^{a(1-\varepsilon)}\right),$$

and, by (3.2.29) and the fact that $2L_{t-1} + t\delta \ge \eta t$ for some $\eta > 0$, we have

$$\mathbb{E}\Big[\sup_{k\le k_t}\big|\big((T_t - I)^2p\big)_k\big|\Big] \le \frac{c}{t^2}\sup_{k\ge 1}(k + |\delta|)^2p_k. \tag{3.2.83}$$

This leads to the bound that

$$t\mathbb{E}[W_t^2 \mathbf{1}_{\{W_t \le t^a\}}]\mathbb{E}\left[\sup_{k \le k_t} \left|((T_t - I)^2 p)_k\right|\right] \le O\left(t^{a(1-\varepsilon)-1}\right). \tag{3.2.84}$$

Combining the bounds in (3.2.80), (3.2.81), (3.2.82) and (3.2.84) completes the proof of part (c) of Lemma 3.10, for any a such that $1/(\varepsilon + 1) < a < 1$. $\qquad\square$

3.3   Proof of Theorem 3.5

In this section, we write $F(x) = \mathbb{P}(W_1 \le x)$, and assume that $1 - F(x) = x^{1-\tau}L(x)$ for some slowly varying function $x \mapsto L(x)$. Throughout this section, we write $\tau = \tau_{\mathrm{w}}$.

From (3.1.1) it is immediate that

$$d_i(t) = d_i(t-1) + X_{i,t}, \qquad \text{for } i = 0, 1, 2, \dots, t-1, \tag{3.3.1}$$

where, conditionally on $d_i(t-1)$ and $\{W_j\}_{j=1}^t$, the distribution of $X_{i,t}$ is binomial with parameters W_t and success probability

$$q_i(t) = \frac{d_i(t-1) + \delta}{2L_{t-1} + t\delta}. \tag{3.3.2}$$

Hence, for $t > i$,

$$\begin{aligned} \mathbb{E}\big[(d_i(t) + \delta)^s \,|\{W_j\}_{j=1}^t\big] \\ &= \mathbb{E}[\mathbb{E}\big[(d_i(t-1) + \delta + X_{i,t})^s \,|d_i(t-1), \{W_j\}_{j=1}^t\big]|\{W_j\}_{j=1}^t] \\ &\le \mathbb{E}\big[(d_i(t-1) + \delta + \mathbb{E}[X_{i,t}|d_i(t-1), \{W_j\}_{j=1}^t])^s\big], \end{aligned} \tag{3.3.3}$$

where we have used the Jensen inequality $\mathbb{E}[(a + X)^s] \le (a + \mathbb{E}[X])^s$, which follows from concavity of $t \mapsto (a + t)^s$ for $0 < s < 1$. Next, we substitute $\mathbb{E}\big[X_{i,t}|d_i(t-1), \{W_j\}_{j=1}^t\big] = W_t q_i(t)$ and use the inequality $2L_{t-1}+t\delta \ge L_{t-1}+\delta$, to obtain that

$$\begin{aligned} \mathbb{E}\big[(d_i(t) + \delta)^s \,|\{W_j\}_{j=1}^t\big] &\le \mathbb{E}\big[(d_i(t-1) + \delta)^s |\{W_j\}_{j=1}^t\big] \left(1 + \frac{W_t}{2L_{t-1} + t\delta}\right)^s \\ &\le \mathbb{E}\big[(d_i(t-1) + \delta)^s |\{W_j\}_{j=1}^t\big] \left(\frac{L_t + \delta}{L_{t-1} + \delta}\right)^s. \end{aligned}$$

Thus, by induction, and because $d_i(i) = W_i$, we get that, for all $t > i \ge 1$,

$$\begin{aligned} \mathbb{E}\big[(d_i(t) + \delta)^s \,|\{W_j\}_{j=1}^t\big] &\le (W_i + \delta)^s \prod_{n=i+1}^t \left(\frac{L_n + \delta}{L_{n-1} + \delta}\right)^s \\ &= (W_i + \delta)^s \left(\frac{L_t + \delta}{L_i + \delta}\right)^s. \end{aligned} \tag{3.3.4}$$

The case $i = 0$ can be treated by $(d_0(t) + \delta)^s = (d_1(t) + \delta)^s$, which is immediate from the definition of $G(1)$. Thus,

$$\mathbb{E}[(d_i(t) + \delta)^s] \le \mathbb{E}\left[(W_i + \delta)^s \left(\frac{L_t + \delta}{L_i + \delta}\right)^s\right]. \tag{3.3.5}$$

Define $f(W_i) = (W_i + \delta)^s$ and

$$g(W_i) = \left(\frac{L_t + \delta}{L_i + \delta}\right)^s = \left(1 + \frac{W_{i+1} + W_{i+2} + \cdots + W_t}{W_1 + W_2 + \ldots + W_i + \delta}\right)^s,$$

and notice that when we condition on all W_j, $1 \le j \le t$, except W_i, then the map $W_i \mapsto f(W_i)$ is increasing in its argument, whereas $W_i \mapsto g(W_i)$ is decreasing. This implies that,

$$\mathbb{E}[f(W_i)g(W_i)] \le \mathbb{E}[f(W_i)]\mathbb{E}[g(W_i)]. \qquad (3.3.6)$$

Hence,

$$\begin{aligned}
\mathbb{E}[(d_i(t) + \delta)^s] &\le \mathbb{E}[(W_i + \delta)^s]\mathbb{E}\left[\left(\frac{L_t + \delta}{L_i + \delta}\right)^s\right] \\
&\le \mathbb{E}[(W_i + \delta)^s]\mathbb{E}\left[(L_t + \delta)^s\right]\mathbb{E}\left[(L_i + \delta)^{-s}\right], \qquad (3.3.7)
\end{aligned}$$

where in the final step we have applied the inequality (3.3.6) once more.

For $i, t \to \infty$,

$$\mathbb{E}\left[(L_i + \delta)^{-s}\right] = (1 + o(1))\mathbb{E}\left[L_i^{-s}\right], \quad \mathbb{E}\left[(L_t + \delta)^s\right] = (1 + o(1))\mathbb{E}\left[L_t^s\right]. \qquad (3.3.8)$$

The moment of order s of $W_i + \delta$ can be bounded by

$$\mathbb{E}[(W_i + \delta)^s] \le \mathbb{E}\left[W_i^s\left(1 + \frac{|\delta|}{W_i}\right)^s\right] \le (1 + |\delta|)^s\mathbb{E}[W_i^s] = (1 + |\delta|)^s\mathbb{E}[W_1^s], \qquad (3.3.9)$$

since $W_i \ge 1$. Combining (3.3.7), (3.3.8) and (3.3.9) gives for i sufficiently large and $t > i$,

$$\mathbb{E}[(d_i(t) + \delta)^s] \le (1 + |\delta|)^s\mathbb{E}[W_1^s]\mathbb{E}\left[L_i^{-s}\right]\mathbb{E}\left[L_t^s\right](1 + o(1)). \qquad (3.3.10)$$

We will bound each of the terms $\mathbb{E}[W_1^s]$, $\mathbb{E}\left[L_t^s\right]$ and $\mathbb{E}\left[L_i^{-s}\right]$ separately.

Evidently, $\mathbb{E}[W_1^s]$ can be bounded by some constant, since all moments smaller than $\tau - 1$ are finite. We will show that, for some constant C_s,

$$\mathbb{E}[L_t^s] \le C_s t^{s/(\tau-1)}l(t)^s \qquad (3.3.11)$$

and, that, for i sufficiently large,

$$\mathbb{E}\left[L_i^{-s}\right] \le C_s i^{-s/(\tau-1)}l(i)^{-s}. \qquad (3.3.12)$$

We will first show claim (3.3.12) and then (3.3.11). For claim (3.3.12), we define the norming sequence $\{a_n\}_{n \ge 1}$ by

$$a_n = \sup\left\{x : 1 - F(x) \ge n^{-1}\right\}, \qquad (3.3.13)$$

so that it is immediate that $a_n = n^{1/(\tau-1)}l(n)$, where $n \mapsto l(n)$ is slowly varying. We use that $L_i \ge W_{(i)} = \max_{1 \le j \le i} W_j$, so that

$$\mathbb{E}\left[L_i^{-s}\right] \le \mathbb{E}\left[W_{(i)}^{-s}\right] = -\mathbb{E}\left[(-Y_{(i)})^s\right], \qquad (3.3.14)$$

where $Y_j = -W_j^{-1}$ and $Y_{(i)} = \max_{1 \le j \le i} Y_j$. Clearly, $Y_j \in [-1,0]$, so that $\mathbb{E}[(-Y_1)^s] < \infty$. Also, $a_i Y_{(i)} = -a_i / W_{(i)}$ converges in distribution to the random variable $-E^{-1/(\tau_{\mathrm{w}}-1)}$, where E is exponential with mean 1, so it follows from [73, Theorem 2.1] that as $i \to \infty$,

$$\mathbb{E}\left[(a_i/L_i)^s\right] \le -\mathbb{E}\left[(-a_i Y_{(i)})^s\right] \to \mathbb{E}[E^{-1/(\tau-1)}] < \infty, \tag{3.3.15}$$

which proves the claim (3.3.12).

We now turn to claim (3.3.11). The discussion on page 565 and Corollary 1 of [50] yields, for $s < \tau - 1$, $\mathbb{E}[L_t^s] = \mathbb{E}[|L_t|^s] \le 2^{s/2} \lambda_s(t)$, for some function $\lambda_s(t)$ depending on s, t and F. Using the discussion on page 564 of [50], we have that $\lambda_s(t) \le C_s t^{s/(\tau-1)} M^*(t^{1/(\tau-1)})^s$, where $M^*(\cdot)$ is a slowly varying function. With some more effort, it can be shown that we can replace $M^*(t^{1/(\tau-1)})$ by $l(t)$, which gives (3.3.11).

Combining (3.3.10), (3.3.11) and (3.3.12), we obtain

$$\mathbb{E}[(d_i(t) + \delta)^s] \le C\Big(\frac{t}{i \vee 1}\Big)^{s/(\tau-1)} \Big(\frac{l(t)}{l(i)}\Big)^s. \tag{3.3.16}$$

Finally, we note that, since $d_i(t) \ge \min\{x : x \in S_{\mathrm{w}}\} \equiv \delta + \nu$ where $\nu > 0$, and using (3.1.2), we can bound $\mathbb{E}[d_i(t)^s] \le (1 \vee \nu^{-1})^s \mathbb{E}[(d_i(t) + \delta)^s]$, which together with (3.3.16) establishes the proof of Theorem 3.5. $\qquad\square$

Chapter 4

Universality for the distance in finite variance random graphs

Joint work with R. van der Hofstad and G. Hooghiemstra

Abstract

The asymptotic behavior of the graph distance between two uniformly chosen nodes in the configuration model is generalized to a wide class of random graphs, where the degrees have finite variance. Among others, this class contains the Poissonian random graph and the generalized random graph (including the classical Erdős-Rényi graph).

We prove that the graph distance grows like $\log_\nu N$, when the base of the logarithm equals $\nu = \mathbb{E}\left[\Lambda^2\right]/\mathbb{E}[\Lambda]$, where Λ is a positive random variable with $\mathbb{P}(\Lambda > x) \leq cx^{1-\tau}$, for some constant c and $\tau > 3$. In addition, the random fluctuations around this asymptotic mean $\log_\nu N$ are characterized and shown to be uniformly bounded.

The proof of this result uses that the graph distance of all members of the class can be coupled successfully to the graph distance in the Poissonian random graph.

4.1   Introduction

Various papers (see e.g., [17, 25, 54, 68, 71]) study properties of random graphs with a given degree sequence. Among such properties as connectivity, cluster size and diameter, the graph distance between two uniformly chosen nodes is an important one. For two connected nodes the graph distance is defined as the minimum number of edges of a path that connects these nodes. If the nodes are not connected, then the graph distance is put equal to infinity.

For the configuration model (see Section 1.4 for a definition) a distance result appeared in [54], when the distribution of the i.i.d. degrees $D^{(C)}$ satisfies

$$\mathbb{P}(D^{(C)} > x) \leq cx^{1-\tau}, \tag{4.1.1}$$

for some constant c, all $x \geq 0$, and with $\tau > 3$. We use the superscript (C) to differentiate between models. The result in [54] states that with probability converging to 1 (**whp**), the typical distance between nodes in the giant component has, for

$$\tilde{\nu} = \frac{\mathbb{E}[D^{(C)}(D^{(C)} - 1)]}{\mathbb{E}[D^{(C)}]} > 1, \qquad (4.1.2)$$

bounded fluctuations around $\log_{\tilde{\nu}} N$. The condition $\tilde{\nu} > 1$ corresponds to the supercritical case of an associated branching process.

In this chapter we extend the above distance result to a wide class of random graph models. Models which fall in this class are the generalized random graph (GRG), the expected degree random graph (EDRG) and the Poissonian random graph (PRG). All three models will be introduced in more detail below.

The method of proof is coupling. It is shown that the distance result holds for all models in the general class if and only if the result holds for the PRG (Section 2). In Section 4 we prove the distance result for the Poissonian random graph. This proof is parallel to that in [54] for the configuration model. In this chapter we included full proofs of the auxiliary lemmas contained in Section 4.3, since details of these proofs are different from those in [54].

4.1.1   Model definition

The graph models considered here are *static* models, meaning that the number of nodes is fixed. The graph G_N has N nodes, numbered $1, 2, \ldots, N$. Associated with the nodes is a sequence $\{\Lambda_i\}_{i=1}^N$ of positive i.i.d. random variables, with distribution $F_\Lambda(x) = \mathbb{P}(\Lambda \leq x)$. We call Λ_i the *capacity* of node i. In all graphs below nodes with a large capacity will obtain a high degree, whereas nodes with small capacity have only a limited number of edges. Furthermore, we define

$$L_N = \Lambda_1 + \Lambda_2 + \cdots + \Lambda_N, \qquad (4.1.3)$$

i.e., L_N is the total capacity of all nodes of the graph G_N.

The binary random variables $\{X_{ij}\}_{1 \leq i \leq j \leq N}$, are defined by setting $X_{ij} = 1$, if there is a connection, i.e., one or more edges, between node i and node j in the graph G_N, otherwise we set $X_{ij} = 0$. If $i > j$, then by convention $X_{ji} = X_{ij}$. We call X_{ij} the *connection variable* and $p_{ij} = \mathbb{P}_N(X_{ij} = 1)$ the *connection probability*, where $\mathbb{P}_N(\cdot)$ is the conditional distribution given the capacities $\{\Lambda_i\}_{i=1}^N$. The graph G_N obeys the following two assumptions:

A1: Conditionally on the capacities, the connection variables $\{X_{ij}\}_{1 \leq i < j \leq N}$, are independent.

A2: The connection probability p_{ij}, for $1 \leq i < j \leq N$, can be written as $p_{ij} = h(\Lambda_i \Lambda_j / L_N)$, for some function $h : [0, 1] \to [0, 1]$, satisfying

$$h(x) - x = \mathcal{O}(x^2), \text{ for } x \downarrow 0. \qquad (4.1.4)$$

The current chapter presents a derivation for the fluctuations of the graph distance in the graph G_N with finite variance degrees, that is, we assume that

for some fixed $\tau > 3$ and some positive constant c, the capacity distribution $F_\Lambda(x) = \mathbb{P}(\Lambda \leq x)$ satisfies:

$$1 - F_\Lambda(x) \leq cx^{1-\tau}, \text{ for all } x \geq 0. \tag{4.1.5}$$

The often used condition that $1 - F_\Lambda(x) = x^{1-\gamma}L(x)$, for some $\gamma > 3$, with $L(x)$ a slowly varying function is covered by (4.1.5), because by Potter's Theorem [41, Lemma 2, p. 277], any slowly varying function $L(x)$ can be bounded from above and below by an arbitrary small power of x, so that (4.1.5) holds for any τ, with $3 < \tau < \gamma$.

4.1.2   Three special cases

We give three examples of random graph models, which satisfy assumptions **A1** and **A2**, and hence fall in the class of models considered here.

The first example is the Poissonian random graph (PRG), which was introduced by Norros and Reittu in [71]. The second and third example are variants of random graph models found in the literature. The second random graph model, which we call the expected degree random graph (EDRG), is a variant of a random graph model introduced by Chung and Lu in [25, 26]. Instead of fixed weights we consider the model with i.i.d. weights $\{\Lambda_i\}_{i=1}^N$. The third and last example is the generalized random graph (GRG), which was introduced by Britton, Deijfen and Martin-Löf [23].

We now define the three models and verify that they satisfy the conditions **A1** and **A2** .

- **The Poissonian pandom graph:** In [71] the Poissonian random graph is introduced. The main feature of such a graph $G_N^{(P)}$ is that, conditionally on the capacities, the number of edges between any pair of nodes i and j is a Poisson random variable. The model in [71] is introduced as a growth model, but as a consequence of [71, Proposition 2.1], it can be formulated as a static model, and we will do so. Start with the graph $G_N^{(P)}$ consisting of N nodes and capacities $\{\Lambda_i\}_{i=1}^N$. The number of edges between two different nodes i and j is given by an independent Poisson random variable $E_{ij}^{(P)}$ with *random* parameter

$$\Lambda_i\Lambda_j/L_N. \tag{4.1.6}$$

  The connection variables are then $X_{ij}^{(P)} = \mathbf{1}_{\{E_{ij}^{(P)}>0\}}$, so that, for $1 \leq i \leq j \leq N$, the connection probabilities are given by

$$p_{ij}^{(P)} = \mathbb{P}_N\big(X_{ij}^{(P)} = 1\big) = \mathbb{P}_N\big(E_{ij}^{(P)} > 0\big) = 1 - \exp\left(-\frac{\Lambda_i\Lambda_j}{L_N}\right) = h^{(P)}(\Lambda_i\Lambda_j/L_N),$$

  where $h^{(P)}(x) = 1 - e^{-x}$. Obviously, $h^{(P)}(x) - x = \mathcal{O}(x^2)$ for $x \downarrow 0$. Since, by definition, the random variables $\{X_{ij}^{(P)}\}_{1 \leq i < j \leq N}$ are independent given the capacities, we conclude that the assumptions **A1** and **A2** are satisfied.

- **The expected degree random graph:** In [25, 26] a random graph model is introduced starting from a sequence of deterministic weights $\{w_i\}_{i=1}^N$. We

give a variant of this random graph model, where we replace the deterministic weights by the sequence $\{\Lambda_i\}_{i=1}^N$. We construct the EDRG $G_N^{(E)}$ as follows. Let $\{X_{ij}^{(E)}\}_{1 \le i \le j \le N}$ be a sequence of independent Bernoulli random variables with success probability

$$p_{ij}^{(E)} = \mathbb{P}_N\big(X_{ij}^{(E)} = 1\big) = (\Lambda_i \Lambda_j / L_N) \wedge 1, \quad \text{for } 1 \le i \le j \le N,$$

where $x \wedge y$ denotes the minimum of x and y. This minimum is to ensure that the result is a probability.

Assumption **A1** is satisfied, as by definition the connection variables conditionally on the capacities are independent Bernoulli variables, and assumption **A2** is also satisfied if we pick $h^{(E)}(x) = x \wedge 1$.

If we assume that **whp** $\Lambda_i \Lambda_j / L_N < 1$ for all $1 \le i \le j \le N$, which is the case for $\tau > 3$, then the expected degree of a node i is given by Λ_i, as

$$\mathbb{E}_N\left[\sum_{j=1}^N X_{ij}^{(E)}\right] = \sum_{j=1}^N \Lambda_i \Lambda_j / L_N = \Lambda_i,$$

where we used the notation $\mathbb{E}_N[\,\cdot\,]$ as the conditional expectation under the probability measure $\mathbb{P}_N(\,\cdot\,)$.

The Erdős-Rényi random graph, usually denoted by $G(N, p)$, is a special case of the EDRG. In the graph $G(N, p)$, an edge between a pair of nodes is present with probability $p \in [0, 1]$, independently of the other edges. When $p = \lambda/N$ for some constant $\lambda > 0$, then we obtain the graph $G(N, \lambda/N)$ from the EDRG by picking $\Lambda_i = \lambda$ for all i, since then $p_{ij}^{(E)} = \Lambda_i \Lambda_j / L_N = \lambda/N = p$, for all $1 \le i < j \le N$.

- **The generalized random graph:** The GRG model is an adapted version of the EDRG model, see the previous example. We define $G_N^{(G)}$ with N nodes as follows. The sequence of connection variables, is, conditionally on the capacities, again given by a sequence of independent Bernoulli random variables $\{X_{ij}^{(G)}\}_{1 \le i < j \le N}$ with

$$\mathbb{P}_N\big(X_{ij}^{(G)} = 1\big) = p_{ij}^{(G)} = \frac{\Lambda_i \Lambda_j / L_N}{1 + \Lambda_i \Lambda_j / L_N}.$$

In [23] the edge probabilities are given by $p_{ij}^{(G)} = (\Lambda_i \Lambda_j / N)/(1 + \Lambda_i \Lambda_j / N)$, so that we have replaced $\Lambda_i / N^{1/2}$ and $\Lambda_j / N^{1/2}$ by $\Lambda_i / L_N^{1/2}$ and $\Lambda_j / L_N^{1/2}$, respectively. This makes hardly any difference since by the strong law of large numbers $L_N / N \to \mathbb{E}[\Lambda]$.

Again, the assumptions **A1** and **A2** are satisfied. To satisfy assumption **A2** we pick $h^{(G)}(x) = x/(1+x) = x + \mathcal{O}(x^2)$.

4.1.3   Main results

Before we can state the main result, we introduce a specific delayed branching process (BP), which we need in the formulation of the main theorems. We define

the process $\{\mathcal{Z}_l\}_{l \geq 0}$ as a BP starting from $\mathcal{Z}_0 = 1$, where in the first generation the offspring distribution is equal to

$$f_n = \int_0^\infty e^{-x} \frac{x^n}{n!} dF_\Lambda(x) = \mathbb{E}\left[e^{-\Lambda} \frac{\Lambda^n}{n!} \right], \, n \geq 0, \tag{4.1.7}$$

whereas in the second and further generations the offspring is chosen in accordance to

$$g_n = \frac{(n+1)f_{n+1}}{\mu} = \frac{1}{\mu} \mathbb{E}\left[e^{-\Lambda} \frac{\Lambda^{n+1}}{n!} \right], \, n \geq 0. \tag{4.1.8}$$

Here μ is the expected number of offspring in the first generation:

$$\mu = \sum_{n=1}^\infty n f_n = \mathbb{E}\left[\sum_{n=1}^\infty n e^{-\Lambda} \frac{\Lambda^n}{n!} \right] = \mathbb{E}\left[\Lambda \sum_{n=0}^\infty \mathbb{P}(Poi(\Lambda) = n) \right] = \mathbb{E}[\Lambda], \tag{4.1.9}$$

and we have used the notation $Poi(\lambda)$, for $\lambda > 0$, to denote a Poisson random variable with parameter λ. Similarly, the expected number of offspring in the second and further generations is given by $\nu = \sum_{n=1}^\infty n g_n$, wich expands to

$$\nu = \frac{1}{\mu} \mathbb{E}\left[\sum_{n=1}^\infty n(n+1) e^{-\Lambda} \frac{\Lambda^{n+1}}{(n+1)!} \right] = \frac{1}{\mu} \mathbb{E}\left[\Lambda^2 \sum_{n=0}^\infty \mathbb{P}(Poi(\Lambda) = n) \right] = \frac{\mathbb{E}[\Lambda^2]}{\mu}. \tag{4.1.10}$$

We define the graph distance or hopcount H_N between two different randomly chosen nodes A_1 and A_2 in the graph G_N as the minimum number of edges that form a path from the node A_1 to node A_2 where, by convention, the distance equals ∞ if the nodes A_1 and A_2 are not connected.

Theorem 4.1 (Fluctuations of the graph distance)
*Fix $\tau > 3$ in (4.1.5), assume that $\nu > 1$ and that assumptions **A1** and **A2** are satisfied. For $k \geq 1$, let $\sigma_k = \lfloor \log_\nu k \rfloor$ and $a_k = \sigma_k - \log_\nu k$. There exists random variables $(R_a)_{a \in (-1,0]}$ such that, as $N \to \infty$,*

$$\mathbb{P}(H_N = \sigma_N + l \,|\, H_N < \infty) = \mathbb{P}(R_{a_N} = l) + o(1). \tag{4.1.11}$$

We identify the random variables $(R_a)_{a \in (-1,0]}$ in Theorem 4.3 below. Before doing so, we state a consequence of Theorem 4.1:

Corollary 4.2 (Concentration of the graph distance) *Under the given assumptions of Theorem 4.1,*

- *with probability $1 - o(1)$ and conditionally on $H_N < \infty$, the random variable H_N is in between $(1 \pm \varepsilon) \log_\nu N$ for any $\varepsilon > 0$;*

- *conditionally on $H_N < \infty$, the sequence of random variables $H_N - \log_\nu N$ forms a tight sequence, i.e.,*

$$\lim_{K \to \infty} \limsup_{N \to \infty} \mathbb{P}\big(|H_N - \log_\nu N| \leq K \big| H_N < \infty \big) = 1. \tag{4.1.12}$$

We use a limit result from branching process theory to identify the limiting random variables $(R_a)_{a\in(-1,0]}$. It is well known, see [41, p. 244], that the process $\{\mathcal{Z}_l/\mu\nu^{l-1}\}_{l\geq 1}$ is a martingale with uniformly bounded expectation and consequently converges almost surely to a limit $\mathcal{W}$:

$$\lim_{l\to\infty}\frac{\mathcal{Z}_l}{\mu\nu^{l-1}}=\mathcal{W},\ \text{a.s.} \tag{4.1.13}$$

Let $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ be two independent copies of $\mathcal{W}$ in (4.1.13), then we can identify the limit random variables $(R_a)_{a\in(-1,0]}$ as follows:

Theorem 4.3 *Under the assumptions of Theorem 4.1 and for $a\in(-1,0]$,*

$$\mathbb{P}(R_a>j)=\mathbb{E}\big[\exp\big\{-\kappa\nu^{a+j}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\big\}\,|\mathcal{W}^{(1)}\mathcal{W}^{(2)}>0\big],$$

where $\kappa=\mu(\nu-1)^{-1}$.

4.1.4   Relations with the configuration model

The configuration model (CM) appeared in the context of random regular graphs as early as 1978 (see [9, 58]). Here we consider the CM as introduced in [54]. Start with an i.i.d. sequence $\{D_i^{(C)}\}_{i=1}^N$ of positive integer valued random variables , where $D_i^{(C)}$ will denote the degree of node i. To built a graph it is mandatory that $D_1^{(C)}+D_2^{(C)}+\ldots+D_N^{(C)}$ is even, so if $D_1^{(C)}+D_2^{(C)}+\ldots+D_N^{(C)}$ is odd we increase $D_N^{(C)}$ by one, which will have little effect. We build the graph model by attaching $D_i^{(C)}$ stubs or half edges to node i and pair the stubs at random, so that two half edges will form one edge.

In [54], the authors prove a version of Theorem 4.1-4.3 for the configuration model. The Theorems 4.1-4.3 hold verbatim for the configuration model with only two changes:

1. Replace the condition $\nu>1$ in Theorem 4.1 by the condition $\tilde{\nu}>1$, defined in (4.1.2).

2. Replace the offspring distributions of the BP $\{\mathcal{Z}_l\}_{l\geq 0}$, by

   (a)
   $$\tilde{f}_n=\mathbb{P}(D^{(C)}=n),\quad n\geq 1,$$

   (b)
   $$\tilde{g}_n=\frac{(n+1)\tilde{f}_{n+1}}{\sum_{m=1}^{\infty}m\tilde{f}_m},\quad n\geq 0.$$

One wonders why a result like the Theorems 1.1-1.3, holds true for the class of models introduced in Section 1.1, especially if one realizes that in the CM the degrees are *independent*, and the edges are not, whereas for instance in the GRG (and in the other two examples) precisely the opposite is true, i.e., in the GRG the edges are independent and the degrees are not. To understand at least at an intuitive level why the distance result holds true, we compare the configuration model with the generalized random graph.

By construction the degree sequence $D_1^{(C)}, D_2^{(C)}, \ldots, D_N^{(C)}$ of the CM is an i.i.d. sequence, and conditionally on $\mathcal{D} = \{D_1^{(C)} = d_1, D_2^{(C)} = d_2, \ldots, D_N^{(C)} = d_N\}$, the graph configuration is uniform over all configurations satisfying $\mathcal{D}$, because the pairing is at random. Hence if we condition on both the event $\mathcal{D}$ and the event $\mathcal{S} = \{$the resulting graph has no self-loops and no multiple edges$\}$, then the CM renders a simple graph, which is picked uniformly from all possible *simple* configurations with degree sequence satisfying $\mathcal{D}$. Since for $N \to \infty$ the probability of the event $\mathcal{S}$ converges to $\exp(-\nu/2 - \nu^2/4) > 0$ (see [16, p. 51]), it follows from [58, Theorem 9.9] that properties that hold **whp** in the CM also hold **whp** in the conditioned simple graph. Hence a property as tightness of the graph distance $H_N^{(C)}$ in the CM is inherited by the conditioned simple graph, with the same degree sequence. This suggests that also the limiting distribution of the fluctuations of the graph distance in the CM conditioned on $\mathcal{S}$ is the same as the one in the CM as identified in [54]. A direct proof of this claim is missing.

On the other hand the GRG with given degree sequence $d_1, d_2, \ldots, d_N$ is also uniform over all possible (simple) configurations. Moreover [23, Theorem 3.1] shows that the degree sequence $D_1^{(G)}, D_2^{(G)}, \ldots, D_N^{(G)}$ of the GRG is asymptotically independent with marginal distribution a mixed Poisson distribution:

$$\mathbb{P}(D^{(G)} = k) = \int_0^\infty e^{-x} \frac{x^k}{k!} \, dF_\Lambda(x), \quad k = 0, 1, 2 \ldots, \tag{4.1.14}$$

where F_Λ is the capacity distribution. Hence starting from $D_1^{(G)}, D_2^{(G)}, \ldots, D_N^{(G)}$ as an i.i.d. sequence with common distribution given by (4.1.14), the (conditioned) CM with these degrees is close to the GRG, at least in an asymptotic sense, so that one expects that the asymptotic fluctuations of the graph distance of the CM also hold for the generalized random graph. Also note from the mixed Poisson distribution (4.1.14), that

$$\tilde{\nu} = \frac{\mathbb{E}[D^{(C)}(D^{(C)} - 1)]}{\mathbb{E}[D^{(C)}]} = \frac{\mathbb{E}[\Lambda^2]}{\mathbb{E}[\Lambda]},$$

which is equal to ν, according to (4.1.9) and (4.1.10). As said earlier, a proof of this intuitive reasoning is missing, and our method of proof is by coupling each random graph satisfying **A1** and **A2** to the Poisson random graph (PRG), and by giving a separate proof of Theorem 1.1-1.3 for the PRG.

We finish this section by giving an overview of different distance results in random graphs. Let τ denote the exponent of the probability mass function of the degree distribution. In this chapter and in [25, 54] the case $\tau > 3$ is studied. Results for $2 < \tau < 3$ for various models appeared in [25, 56, 71, 70]. Typically in that case, the distance fluctuates around a constant times $2 \log \log N / |\log(\tau - 2)|$. For $1 < \tau < 2$, there exists a subset of nodes with a high degree, called the *core* (see [38]). The core forms a complete graph and almost every node is attached to the core and, thus, the graph distance is **whp** at most 3.

4.1.5 Organization of the chapter

The coupling argument that ties the fluctuations of the graph distance $H_N^{(P)}$ in the PRG to the fluctuations of the graph distance in random graphs satisfying

assumptions **A1** and **A2** is treated in Section 4.2. In Section 4 we show that the fluctuations of the graph distance $H_N^{(P)}$ is given by Theorem 4.1. The derivation of the fluctuations of the graph distance $H_N^{(P)}$ is similar to the derivation of the fluctuations of the graph distance $H_N^{(C)}$ in the configuration model, see [54]. The proof in [54] is more complicated than the proof presented here for the PRG model, mainly because in the latter the expansion of a given node (e.g. the nodes on a given distance) can be described by means of the so called Reittu-Norros process, a marked branching process. This branching process will be introduced in Section 4.3.

In this paper full proofs of the auxiliary propositions and lemmas introduced in Sections 4.3 and 4.4 are presented in the appendix. These proofs were omitted in [37].

4.2   Coupling

In this section we denote by G_N the PRG and by G'_N some other random graph satisfying the assumptions **A1** and **A2**, given in Section 4.1.1. We number the nodes of both G_N and G'_N from 1 to N and we assign the capacity Λ_i, for $1 \leq i \leq N$, to node i in each graph. We denote by H_N and H'_N the graph distance between two randomly chosen nodes A_1 and A_2, such that $A_1 \neq A_2$, in G_N and G'_N, respectively. We will show that for $N \to \infty$,

$$\mathbb{P}(H_N \neq H'_N) = o(1). \tag{4.2.1}$$

The above implies that **whp** the coupling of the graph distances is successful. Therefore, given the succesful coupling (4.2.1), it is sufficient to show Theorem 4.1 for the PRG.

4.2.1   Coupling of G_N and G'_N

We next describe the coupling of the connection variables of the graphs G_N and G'_N. A classical coupling is used, see e.g. [80]. Denote by $\{X_{ij}\}_{1 \leq i < j \leq N}$ and $\{X'_{ij}\}_{1 \leq i < j \leq N}$ the connection variables of the graphs G_N and G'_N, and, similarly, denote the connection probabilities by $\{p_{ij}\}_{1 \leq i < j \leq N}$ and $\{p'_{ij}\}_{1 \leq i < j \leq N}$. For the coupling we introduce independent random variables $\{K_{ij}\}_{1 \leq i < j \leq N}$. Set $\underline{p}_{ij} = \min\{p_{ij}, p'_{ij}\}$ and $\overline{p}_{ij} = \max\{p_{ij}, p'_{ij}\}$, and define random variables $\hat{X}_{ij}$ and $\hat{X}'_{ij}$ with

$$\mathbb{P}_N\left(\hat{X}_{ij} = 1, \hat{X}'_{ij} = 1, K_{ij} = 0\right) = \underline{p}_{ij},$$

$$\mathbb{P}_N\left(\hat{X}_{ij} = 1, \hat{X}'_{ij} = 0, K_{ij} = 1\right) = p_{ij} - \underline{p}_{ij},$$

$$\mathbb{P}_N\left(\hat{X}_{ij} = 0, \hat{X}'_{ij} = 1, K_{ij} = 1\right) = \overline{p}_{ij} - p_{ij},$$

$$\mathbb{P}_N\left(\hat{X}_{ij} = 0, \hat{X}'_{ij} = 0, K_{ij} = 0\right) = 1 - \overline{p}_{ij},$$

whereas all other combinations have probability 0. Then the laws of $\hat{X}_{ij}$ and $\hat{X}'_{ij}$ are the same as the laws of X_{ij} and X'_{ij}, respectively. Furthermore, K_{ij} assumes

the value 1 with probability $|p_{ij} - p'_{ij}|$, and is 0 otherwise. Note that we do abuse the notation in the above display. We should replace the probability measure $\mathbb{P}_N$ in the above display by some other probability measure $\mathbb{Q}_N$, because the probability space is defined by the graphs G_N and G'_N, instead of only the graph G_N. Since the graphs, conditioned on the capacities, are constructed independently from each other, this abuse of notation is not a problem.

Consider the nodes i and j, $1 \leq i < j \leq N$, in the graphs G_N and G'_N simultaneously. Then the event $\{K_{ij} = 0\} = \{\hat{X}_{ij} = \hat{X}'_{ij}\}$ corresponds to the event that in both graphs there exists a connection between nodes i and j, or that in both graphs there is no connection between nodes i and j. The event $\{K_{ij} = 1\} = \{\hat{X}_{ij} \neq \hat{X}'_{ij}\}$ corresponds with the event that there exists a connection in one of the graphs, but not in the other one. We call the event $\{K_{ij} = 1\}$ a *mismatch* between the nodes i and j.

Assumption **A2** implies that for some constant $C' > 0$,

$$\mathbb{P}_N(K_{ij} = 1) = \left|p_{ij} - p'_{ij}\right| \leq \left|p_{ij} - \Lambda_i \Lambda_j / L_N\right| + \left|p'_{ij} - \Lambda_i \Lambda_j / L_N\right| \leq C' \frac{\Lambda_i^2 \Lambda_j^2}{L_N^2},$$

(4.2.2)

for all $1 \leq i < j \leq N$. The number of mismatches due to all the nodes incident to node i, $1 \leq i \leq N$, is given by

$$K_i = \sum_{j \neq i} K_{ij}.$$

(4.2.3)

Obviously, we cannot couple all the connections in the graphs G_N and G'_N successfully, but the total number of mismatches due to all the nodes can be bounded from above by any positive power of N. Before proving this, we introduce a lemma. In its statement we use the event

$$\mathcal{S}_{q,\alpha_0} = \left\{ |S_{N,q} - \mathbb{E}[\Lambda^q]| \leq N^{-\alpha_0} \right\},$$

(4.2.4)

where $\varepsilon, q > 0$, and where the random variable $S_{N,q}$ is defined as

$$S_{N,q} = \frac{1}{N} \sum_{i=1}^{N} \Lambda_i^q.$$

(4.2.5)

Lemma 4.4 *For each fixed $q \in (0, \tau - 1)$, there exist constants $\alpha_0, \beta_0 > 0$ such that*

$$\mathbb{P}\big(\mathcal{S}_{q,\alpha_0}^c\big) \leq N^{-\beta_0}.$$

(4.2.6)

Proof. The proof is deferred to Section 4.A.1. $\qquad\qquad\square$

As a consequence of Lemma 4.4, and since $\tau > 3$, there exist positive constants $\underline{S}$ and $\overline{S}$ such that **whp**

$$\underline{S} \leq S_{N,q} \leq \overline{S},$$

(4.2.7)

for $q = 1, 2$ and N sufficiently large.

Pick $\alpha_0, \beta_0 > 0$, using Lemma 4.4, such that $\mathbb{P}(\mathcal{S}^c_{q,\alpha_0}) \leq N^{-\beta_0}$ for $q = 1, 2$. On the event $\mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0}$, using (4.2.2), (4.2.3), (4.2.5) and (4.2.7), we bound **whp** the total expected number of mismatches due to a single node by

$$\mathbb{E}_N[K_i] = \sum_{j \neq i} \mathbb{E}_N[K_{ij}] = \sum_{j \neq i} \mathbb{P}_N(K_{ij} = 1) \leq \frac{C\Lambda_i^2}{L_N^2} \sum_{j=1}^N \Lambda_j^2$$

$$= \frac{C\Lambda_i^2}{N} \frac{S_{N,2}}{S_{N,1}^2} \leq C\overline{S}\underline{S}^{-2}\Lambda_i^2 N^{-1} \leq C\Lambda_i^2 N^{-1}, \qquad (4.2.8)$$

where C is a constant that may change from line to line. Thus, **whp** the total number of mismatches is bounded from above by N^γ, for any $\gamma > 0$, since

$$\mathbb{P}\left(\sum_{i=1}^N K_i > N^\gamma\right) \leq N^{-\gamma} \mathbb{E}\left[\sum_{i=1}^N \mathbb{E}_N[K_i] \, \mathbf{1}_{\mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0}}\right] + \mathbb{P}(\mathcal{S}^c_{1,\alpha_0} \cup \mathcal{S}^c_{2,\alpha_0})$$

$$\leq CN^{-\gamma} + \mathcal{O}(N^{-\beta_0}).$$

We see that the right hand goes to zero if N goes to infinity, which implies that **whp** the total number of mismatches is bounded from above by N^γ for any $\gamma > 0$.

Define the event $\mathcal{A}_N$ as

$$\mathcal{A}_N = \bigcap_{i=1}^N \left\{K_i \mathbf{1}_{\{\Lambda_i > c_N\}} = 0\right\} = \left\{\sum_{i=1}^N K_i \mathbf{1}_{\{\Lambda_i > c_N\}} = 0\right\}, \qquad (4.2.9)$$

where $c_N = N^\xi$ for each $\xi > 0$. Then, on the event $\mathcal{A}_N$, all nodes with capacity greater than c_N are successfully coupled.

Lemma 4.5 *For each $\xi > 0$ there exists a constant $\theta > 0$ such that*

$$\mathbb{P}_N(\mathcal{A}_N^c) = \mathcal{O}(N^{-\theta}). \qquad (4.2.10)$$

Proof.   Fix $\alpha_0 > 0$ as in Lemma 4.4. On the event $\mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0}$, we bound $\mathbb{P}_N(\mathcal{A}_N^c)$ using Boole's inequality, the Markov inequality and (4.2.8), which yields

$$\mathbb{P}_N(\mathcal{A}_N^c) \leq \sum_{i=1}^N \mathbb{P}_N(K_i \mathbf{1}_{\{\Lambda_i > c_N\}} > 0) \leq \sum_{i=1}^N \mathbb{E}_N[K_i] \, \mathbf{1}_{\{\Lambda_i > c_N\}} \leq \frac{C}{N} \sum_{i=1}^N \Lambda_i^2 \mathbf{1}_{\{\Lambda_i > c_N\}}.$$

Therefore,

$$\mathbb{P}(\mathcal{A}_N^c \cap \mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0}) = \mathbb{E}\left[\mathbb{P}_N(\mathcal{A}_N^c) \, \mathbf{1}_{\mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0}}\right]$$

$$\leq \frac{C}{N} \sum_{i=1}^N \mathbb{E}\left[\Lambda_i^2 \mathbf{1}_{\{\Lambda_i > c_N\}}\right] = C\mathbb{E}\left[\Lambda^2 \mathbf{1}_{\{\Lambda > c_N\}}\right],$$

for some constant $C > 0$. Using integration by parts and $1 - F_\Lambda(x) \leq cx^{1-\tau}$ we have that $\mathbb{E}\left[\Lambda^2 \mathbf{1}_{\{\Lambda > c_N\}}\right]$ is equal to

$$= -w^2[1 - F_\Lambda(w)]\big|_{w=c_N}^\infty + 2\int_{c_N}^\infty w[1 - F_\Lambda(w)]dw = \mathcal{O}\left(N^{-(\tau-3)\xi}\right), \qquad (4.2.11)$$

since

$$w^2[1 - F_{\Lambda}(w)]\big|_{w=\infty}^{c_N} \leq c_N^2[1 - F_{\Lambda}(c_N)] = \mathcal{O}\Big(N^{-(\tau-3)\xi}\Big)$$

and

$$\int_{c_N}^{\infty} w[1 - F_{\Lambda}(w)]dw \leq c \int_{c_N}^{\infty} w^{2-\tau}dw = \mathcal{O}\Big(N^{-(\tau-3)\xi}\Big).$$

This yields

$$\mathbb{P}(\mathcal{A}_N^c) \leq \mathbb{P}(\mathcal{A}_N^c \cap \mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0}) + \mathbb{P}\big(\mathcal{S}_{1,\alpha_0}^c\big) + \mathbb{P}\big(\mathcal{S}_{2,\alpha_0}^c\big) = \mathcal{O}\big(N^{-\theta}\big)$$

for some $\theta = \min\{\beta_0, (\tau-3)\xi\} > 0$. $\qquad\square$

4.2.2   Couple the graph distances of G_N and G_N'

In this subsection we couple the graph distance of the PRG with any random graph satisfying assumptions **A1** and **A2**.

Theorem 4.6 *Let G_N be a PRG and let and G_N' be a random graph satisfying assumption **A1** and **A2**. Let H_N and H_N' be the graph distances between two different uniformly chosen nodes A_1 and A_2 in, respectively, the graphs G_N and G_N'. Then*

$$\mathbb{P}(H_N \neq H_N') = o(1). \tag{4.2.12}$$

The above theorem implies that, **whp** , the coupling of the graph distances H_N and H_N' is successful.

In order to prove Theorem 4.6, we use the following proposition. In its statement, we consider the neighborhood shells of a uniformly chosen node $A \in \{1, 2, \ldots, N\}$, i.e., all nodes on a fixed graph distance of node A. More precisely,

$$\partial\mathcal{N}_0 = \{A\} \qquad \text{and} \qquad \partial\mathcal{N}_l = \{1 \leq j \leq N : d(A, j) = l\}, \tag{4.2.13}$$

where $d(i, j)$ denotes the graph distance between nodes i and j, i.e., the minimum number of edges in a path between the nodes i and j. Furthermore, define the set of nodes reachable in at most j steps from root A by

$$\mathcal{N}_l = \{1 \leq j \leq N : d(A, j) \leq l\} = \bigcup_{k=0}^{l} \partial\mathcal{N}_k. \tag{4.2.14}$$

Proposition 4.7 *For N sufficiently large, $l \in \mathbb{N}$, some constant $C' > 0$, and every $b \in (0, 1)$,*

$$\mathbb{P}(H_N \neq H_N') \leq \mathbb{P}(\mathcal{A}_N^c) + \mathbb{P}(H_N > 2l) + 2l\mathbb{P}\big(|\mathcal{N}_{l-1}| > N^b\big) + 2C'\underline{S}^{-2}lN^{-(1-b)}c_N^4. \tag{4.2.15}$$

Before giving a proof, we show that Theorem 4.6 is a consequence of Proposition 4.7.

Proof of Theorem 4.6. By Lemma 4.5, we have that, for $\tau > 3$, $\mathbb{P}(\mathcal{A}_N^c) \leq N^{-\theta}$. We fix $l = l_N = \lfloor (\frac{1}{2} + \eta) \log_\nu N \rfloor$. Then, from Corollary 4.2, applied to the PRG model, we obtain that $\mathbb{P}(H_N > 2l) = o(1)$. The third term in the right hand side of (4.2.15) can be bounded using the following lemma:

Lemma 4.8 *Let $\{\mathcal{N}_l\}_{l \geq 0}$ be the reachable sets of a uniformly chosen node A in the PRG G_N. Then for $\eta, \delta \in (-1/2, 1/2)$ and all $l \leq (1/2 + \eta) \log_\nu N$, there exists a constant $\beta_1 > 0$ such that*

$$\mathbb{P}\left(|\mathcal{N}_l| > N^{1/2+\delta} \right) = \mathcal{O}\left((\log_\nu N) N^{-\min\{\delta - \eta, \beta_1\}} \right). \tag{4.2.16}$$

Proof.   See the proof of Lemma 4.18. □

We now prove that all terms in the right hand of (4.2.15) are $o(1)$ for an appropriate choice of b. Lemma 4.8 implies that $2l\mathbb{P}(|\mathcal{N}_l| > N^b) = o(1)$ for some appropriately chosen $b > \frac{1}{2}$. Then, provided that $b < 1$, we see that $2C'\underline{S}^{-2}lN^{b-1}c_N^4 = 2C'\underline{S}^{-2}lN^{4\xi+b-1} = o(1)$, where we substitute $c_N = N^\xi$, and picking $\xi \in (0, (1-b)/4)$. Hence, by Proposition 4.7, $\mathbb{P}(H_N \neq H_N') = o(1)$, which is precisely the content of Theorem 4.6. □

Proof of Proposition 4.7. We use that

$$\mathbb{P}(H_N \neq H_N') \leq \mathbb{P}(\mathcal{A}_N^c) + \mathbb{P}(H_N > 2l) + \mathbb{P}(\{H_N \leq 2l\} \cap \mathcal{A}_N \cap \{H_N \neq H_N'\}). \tag{4.2.17}$$

Let $\mathcal{N}_l^{(i)}$ and $\mathcal{N}_l'^{(i)}$, for $i = 1, 2$, be the union of neighborhood shells of the nodes A_i in G_N and G_N', respectively. Now, we use the fact that if $H_N \leq 2l$ and if $H_N \neq H_N'$, then $\mathcal{N}_l^{(1)} \neq \mathcal{N}_l'^{(1)}$ and/or $\mathcal{N}_l^{(2)} \neq \mathcal{N}_l'^{(2)}$. By the exchangeability of the nodes, we have

$$\mathbb{P}(\{H_N \leq 2l\} \cap \mathcal{A}_N \cap \{H_N \neq H_N'\}) \leq 2\mathbb{P}(\{\mathcal{N}_l \neq \mathcal{N}_l'\} \cap \mathcal{A}_N), \tag{4.2.18}$$

where $\mathcal{N}_l$ and $\mathcal{N}_l'$ are the neighborhood shells of a uniformly chosen node A in respectively G_N and G_N'. If $\mathcal{N}_l \neq \mathcal{N}_l'$, then there must be a $k \in \{1, \ldots, l\}$ for which $\mathcal{N}_k \neq \mathcal{N}_k'$, but $\mathcal{N}_{k-1} = \mathcal{N}_{k-1}'$. Thus,

$$\mathbb{P}(\{H_N \leq 2l\} \cap \mathcal{A}_N \cap \{H_N \neq H_N'\})$$
$$\leq 2\sum_{k=1}^{l} \mathbb{P}\left(\{\mathcal{N}_k \neq \mathcal{N}_k'\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}_{k-1}'\} \cap \mathcal{A}_N\right). \tag{4.2.19}$$

In turn, the event $\{\mathcal{N}_k \neq \mathcal{N}_k'\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}_{k-1}'\}$ implies that one of the edges from $\partial\mathcal{N}_{k-1}$ must be miscoupled, thus $K_{ij} = 1$ for some $i \in \partial\mathcal{N}_{k-1}$ and $j \in \mathcal{N}_{k-1}^c$, where $\mathcal{N}_{k-1}^c = \{1, 2, \ldots, N\} \backslash \mathcal{N}_{k-1}$. The event $\mathcal{A}_N$ implies that $\Lambda_i, \Lambda_j \leq c_N$. Therefore, we bound

$$\mathbb{P}_N\left(\{\mathcal{N}_k \neq \mathcal{N}_k'\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}_{k-1}'\} \cap \mathcal{A}_N\right) \leq \mathbb{P}_N\left(|\mathcal{N}_{k-1}| > N^b\right)$$
$$+ \sum_{i,j} \mathbb{P}_N\left(\{i \in \partial\mathcal{N}_{k-1}, j \in \mathcal{N}_{k-1}^c, K_{ij} = 1\} \cap \{|\mathcal{N}_{k-1}| \leq N^b\}\right) \mathbf{1}_{\{\Lambda_i, \Lambda_j \leq c_N\}}.$$

$$\tag{4.2.20}$$

Also, under $\mathbb{P}_N$, since $i \in \mathcal{N}_{k-1}^c$ and $j \in \partial\mathcal{N}_{k-1}$, the event $\{K_{ij} = 1\}$ is independent of $\mathcal{N}_{k-1}$ and, therefore, from $\partial\mathcal{N}_{k-1}$ as $\partial\mathcal{N}_{k-1} \subset \mathcal{N}_{k-1}$. The edge between the nodes i and j points out of $\mathcal{N}_{k-1}$, while $\mathcal{N}_{k-1}$ is determined by the occupation status of edges that are between nodes in $\mathcal{N}_{k-2}$ or pointing out of $\partial\mathcal{N}_{k-2}$. Thus, we can replace each term in the sum of (4.2.20) by

$$\mathbb{P}_N(K_{ij} = 1)\,\mathbb{P}_N\big(\{i \in \partial\mathcal{N}_{k-1}, j \in \mathcal{N}_{k-1}^c\} \cap \{|\mathcal{N}_{k-1}| \le N^b\}\big)\,\mathbf{1}_{\{\Lambda_i,\Lambda_j \le c_N\}}. \quad (4.2.21)$$

Since by (4.2.2) and (4.2.7), we have

$$\mathbb{P}_N(K_{ij} = 1)\,\mathbf{1}_{\{\Lambda_i,\Lambda_j \le c_N\}} = \mathbb{E}_N[K_{ij}]\,\mathbf{1}_{\{\Lambda_i,\Lambda_j \le c_N\}}$$
$$\le C'\frac{\Lambda_i^2\Lambda_j^2}{L_N^2}\mathbf{1}_{\{\Lambda_i,\Lambda_j \le c_N\}} \le C'\underline{S}^{-2}c_N^4 N^{-2}, \quad (4.2.22)$$

we can bound the right hand side of (4.2.20) from above by

$$\mathbb{P}_N\big(|\mathcal{N}_{k-1}| > N^b\big)$$
$$+ C'\underline{S}^{-2}c_N^4 N^{-2}\sum_{i,j}\mathbb{P}_N\big(\{i \in \partial\mathcal{N}_{k-1}, j \in \mathcal{N}_{k-1}^c\} \cap \{|\mathcal{N}_{k-1}| \le N^b\}\big).$$

Finally, we bound the sum on the right side by

$$N\mathbb{E}_N\big[|\partial\mathcal{N}_{k-1}|\mathbf{1}_{\{|\mathcal{N}_{k-1}| \le N^b\}}\big] \le N^{1+b}.$$

Therefore, we can bound each term in the sum of (4.2.19), with $\mathbb{P}$ replaced by $\mathbb{P}_N$,

$$\mathbb{P}_N\big(\{\mathcal{N}_k \ne \mathcal{N}_k'\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}_{k-1}'\} \cap \mathcal{A}_N\big) \le \mathbb{P}_N\big(|\mathcal{N}_{k-1}| > N^b\big) + C'\underline{S}^{-2}c_N^4 N^{-1+b}.$$

Since, for $k \le l$, we have that $\mathbb{P}_N\big(|\mathcal{N}_{k-1}| > N^b\big) \le \mathbb{P}_N\big(|\mathcal{N}_{l-1}| > N^b\big)$, by summing over $k = 1, \ldots, l$ in (4.2.19), we arrive at

$$\mathbb{P}_N(\{H_N \le 2l\} \cap \mathcal{A}_N \cap \{H_N \ne H_N'\}) \le 2l\mathbb{P}_N\big(|\mathcal{N}_{l-1}| > N^b\big) + C'\underline{S}^{-2}lc_N^4 N^{-(1-b)},$$

which in turn implies

$$\mathbb{P}(\{H_N \le 2l\} \cap \mathcal{A}_N \cap \{H_N \ne H_N'\}) \le 2l\mathbb{P}\big(|\mathcal{N}_{l-1}| > N^b\big) + C'\underline{S}^{-2}lc_N^4 N^{-(1-b)}.$$

Therefore, we can bound $\mathbb{P}(H_N \ne H_N')$, see (4.2.17) by

$$\mathbb{P}(\mathcal{A}_N^c) + \mathbb{P}(H_N > 2l) + 2l\mathbb{P}\big(|\mathcal{N}_{l-1}| > N^b\big) + C'\underline{S}^{-2}lc_N^4 N^{-(1-b)},$$

which is precisely the claim (4.2.15). $\qquad\qquad\qquad\qquad\qquad\qquad\Box$

4.3   The Poissonian random graph model

The proof of the fluctuations of the graph distance in the configuration model in [54] is done in a number of steps. One of the most important steps is the coupling of the expansion of the neighborhood shells of a node to a BP. For the PRG, we follow the same strategy as in [54], although the details differ substantially.

The first step is to introduce the $\overline{\text{NR}}$-process, which is a marked BP. The $\overline{\text{NR}}$-process was introduced by Norros and Reittu in [71]. We can thin the $\overline{\text{NR}}$-process in such a way that the resulting process, the $\underline{\text{NR}}$-process, can be coupled to the expansion of the neighborhood shells of a randomly chosen node in the PRG. Finally, we introduce capacities for the $\overline{\text{NR}}$-process and the $\underline{\text{NR}}$-process.

4.3.1 The NR-process and its thinned version

The $\overline{\text{NR}}$-process is a marked delayed BP denoted by $\{\overline{Z}_l, \overline{\boldsymbol{M}}_l\}_{l \geq 0}$, where $\overline{Z}_l$ denotes the number of individuals of generation l, and where the vector

$$\overline{\boldsymbol{M}}_l = (\overline{M}_{l,1}, \overline{M}_{l,2}, \ldots, \overline{M}_{l,\overline{Z}_l}) \in \{1, 2, \ldots, N\}^{\overline{Z}_l},$$

denotes the marks of the individuals in generation l. We now give a more precise definition of the $\overline{\text{NR}}$-process and describe its connection with G_N, the PRG. We define $\overline{Z}_0 = 1$ and take $\overline{M}_{0,1}$ uniformly from the set $\{1, 2, \ldots, N\}$, corresponding to the choice of A_1, which is uniformly over all the nodes. The offspring of an individual with mark $m \in \{1, 2, \ldots, N\}$ is as follows: the total number of children has a Poisson distribution with parameter Λ_m, of which, for each $i \in \{1, 2, \ldots, N\}$, a Poisson distributed number with parameter

$$\frac{\Lambda_i \Lambda_m}{L_N}, \tag{4.3.1}$$

bears mark i, independently of the other individuals. Since

$$\sum_{i=1}^{N} \frac{\Lambda_i \Lambda_m}{L_N} = \frac{\Lambda_m}{L_N} \sum_{i=1}^{N} \Lambda_i = \Lambda_m,$$

and sums of independent Poisson random variables are again Poissonian, we may take the number of children with different marks mutually independent. As a result of this definition, the marks of the children of an individual in $\{\overline{Z}_l, \overline{\boldsymbol{M}}_l\}_{l \geq 0}$ can be seen as independent realizations of a random variable M, with distribution

$$\mathbb{P}_N(M = m) = \frac{\Lambda_m}{L_N}, \quad 1 \leq m \leq N, \tag{4.3.2}$$

and, consequently,

$$\mathbb{E}_N[\Lambda_M] = \sum_{m=1}^{N} \Lambda_m \mathbb{P}_N(\Lambda_M = \Lambda_m) = \sum_{m=1}^{N} \Lambda_m \mathbb{P}_N(M = m) = \frac{1}{L_N} \sum_{m=1}^{N} \Lambda_m^2. \tag{4.3.3}$$

For the definition of the $\underline{\text{NR}}$-process we start with a copy of the $\overline{\text{NR}}$-process $\{\overline{Z}_l, \overline{\boldsymbol{M}}_l\}_{l \geq 0}$, and reduce this process generation by generation, i.e., in the order

$$\overline{M}_{0,1}, \overline{M}_{1,1}, \ldots \overline{M}_{1,\overline{Z}_1}, \overline{M}_{2,1}, \ldots \tag{4.3.4}$$

by discarding each individual and all its descendants whose mark has appeared before. The process obtained in this way is called the $\underline{\text{NR}}$-process and is denoted by the sequence $\{\underline{Z}_l, \underline{\boldsymbol{M}}_l\}_{l \geq 0}$. One of the main results of [71] is Proposition 3.1, but for completeness we also add the proof.:

Proposition 4.9 *Let $\{\underline{Z}_l, \underline{\boldsymbol{M}}_l\}_{l \geq 0}$ be the $\underline{\text{NR}}$-process and let $\underline{\boldsymbol{M}}_l$ be the set of marks in the l−th generation, then the sequence of sets $\{\underline{\boldsymbol{M}}_l\}_{l \geq 0}$ has the same distribution as the sequence $\{\partial \mathcal{N}_l\}_{l \geq 0}$ given by (4.2.13).*

Proof.   We proceed by induction. By definition A_1 and $\underline{M}_{0,1} = \overline{M}_{0,1}$ have the same distribution. It is clear that the $\underline{\text{NR}}$-process $\{\underline{Z}_l, \underline{M}_l\}_{l \geq 0}$ can be generated simultaneously with $\{\overline{Z}_l, \overline{M}_l\}_{l \geq 0}$, by deleting each new individual whose mark has already been seen without letting it reproduce. Let us do this, and assume that generation k has been fully generated and is identical in both processes, and that the same holds for the children of the first $i+1$ members of the $k^{-\text{th}}$ generation. It now suffices to show that we obtain a similarly distributed result for the children of the $(i+1)^{-th}$ member of the $k^{-\text{th}}$ generation.

First we consider the neighborhood expansion process $\{\partial \mathcal{N}_l\}_{l \geq 0}$. Here we need only to consider vertices that have not yet been reached. With probability

$$1 - \exp\left(-\frac{\Lambda_{\underline{M}_{k,i+1}} \Lambda_j}{L_N}\right)$$

there is at least one edge between such vertex v_j and the present vertex. Moreover, the numbers of edges to different vertices are independently, by assumption. In the branching process, on the other hand, for each $j = 1, 2, \dots, N$ the corresponding individual produces a Poisson number of offspring each bearing mark j, where the parameter is given by $\Lambda_{\underline{M}_{k,i+1}} \Lambda_j / L_N$. For different values of j, these numbers are independent. The pruning then removes all vertices whose mark were generated in previously steps, and also all newly generated duplicates. Clearly the set of remaining marks is distributed in the same way as the above-described set of new neighbors in the neighborhood expansion. $\square$

As a consequence of the previous proposition, we can couple the $\underline{\text{NR}}$-process to the neighborhood shells of a uniformly chosen node $A \in \{1, 2, \dots, N\}$, i.e., all nodes on a fixed graph distance of A, see (4.2.13) and note that $A \sim \overline{M}_{0,1}$. Thus, using the above proposition, we can couple the expansion of the neighborhood shells and the $\underline{\text{NR}}$-process in such a way that

$$\underline{M}_l = \partial \mathcal{N}_l \qquad \text{and} \qquad \underline{Z}_l = |\partial \mathcal{N}_l|, \; l \geq 0. \qquad (4.3.5)$$

Furthermore, we see that an individual with mark m, $1 \leq m \leq N$, in the $\underline{\text{NR}}$-process is identified with node m in the graph G_N, whose capacity is given by Λ_m.

We will now show that the offspring distribution of the BP $\{\overline{Z}_l\}_{l \geq 0}$ converges as $N \to \infty$ to the offspring distribution of $\{\mathcal{Z}_l\}_{l \geq 0}$, introduced in Section 4.1.3. The offspring distribution $f^{(N)}$ of $\overline{Z}_1$, i.e., the first generation of $\{\overline{Z}_l\}_{l \geq 0}$, is given by

$$f_n^{(N)} = \mathbb{P}_N(Poi(\Lambda_A) = n) = \sum_{m=1}^{N} \mathbb{P}_N(Poi(\Lambda_A) = n | A = m) \, \mathbb{P}_N(A = m)$$

$$= \frac{1}{N} \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\Lambda_m^n}{n!}, \qquad (4.3.6)$$

for $n \geq 0$. Recall that individuals in the second and further generations have a random mark distributed as M, given by (4.3.2). Hence, if we denote the offspring

distribution of the second and further generations by $g_n^{(N)}$, then we obtain

$$g_n^{(N)} = \mathbb{P}_N\big(Poi(\Lambda_M) = n\big) = \sum_{m=1}^{N} \mathbb{P}_N\big(Poi(\Lambda_M) = n \big| M = m\big)\, \mathbb{P}_N(M = m) \quad (4.3.7)$$

$$= \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\Lambda_m^n}{n!} \frac{\Lambda_m}{L_N} = \frac{1}{L_N} \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\Lambda_m^{n+1}}{n!}, \quad (4.3.8)$$

for $n \geq 0$. Furthermore, we can relate $g_n^{(N)}$ and $f_n^{(N)}$ by

$$g_n^{(N)} = \frac{1}{L_N} \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\Lambda_m^{n+1}}{n!} = \frac{(n+1)}{L_N/N} \frac{1}{N} \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\Lambda_m^{n+1}}{(n+1)!} = \frac{(n+1)f_{n+1}^{(N)}}{L_N/N}.$$
$$(4.3.9)$$

Observe that $\{f_n^{(N)}\}_{n\geq 0}$ is the average over Poisson probabilities, whereas (4.3.9) shows that $\{g_n^{(N)}\}_{n\geq 0}$ comes from size biased sampling of $\{f_{n+1}^{(N)}\}_{n\geq 0}$. Since the sequence $\{\Lambda_m\}_{m=1}^{N}$ is i.i.d. the strong law of large number states that for $N \to \infty$ the limit distributions of $f^{(N)}$ and $g^{(N)}$ are given by

$$f_n^{(N)} = \frac{1}{N} \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\Lambda_m^n}{n!} \to \mathbb{E}\Big[e^{-\Lambda} \frac{\Lambda^n}{n!}\Big] = f_n, \; n \geq 0, \quad (4.3.10)$$

and, as a consequence,

$$g_n^{(N)} = \frac{(n+1)f_{n+1}^{(N)}}{L_N/N} \to \frac{(n+1)f_{n+1}}{\mu} = g_n, \; n \geq 0. \quad (4.3.11)$$

Indeed, according to (4.1.7) and (4.1.8) the limit distributions are equal to the offspring distributions of the delayed BP $\{\mathcal{Z}_l\}_{l\geq 0}$ introduced in Section 4.1.3.

4.3.2   Coupling the thinned NR-process with a delayed BP

In this subsection we will introduce a coupling between the $\underline{\text{NR}}$-process with the delayed BP $\{\mathcal{Z}_l\}_{l\geq 0}$, which is defined by (4.1.7) and (4.1.8) in Section 4.1.3. This coupling is used in the proof of Theorem 4.1 and 4.3 for the PRG, to express the probability distribution of H_N in terms of the BP $\{\mathcal{Z}_l\}_{l\geq 0}$.

Introduce the capacity of the l-th generation of the $\overline{\text{NR}}$-process $\{\overline{Z}_l, \overline{M}_l\}_{l\geq 0}$ and the $\underline{\text{NR}}$-process $\{\underline{Z}_l, \underline{M}_l\}_{l\geq 0}$ as, respectively,

$$\overline{C}_{l+1} = \sum_{i=1}^{\overline{Z}_l} \Lambda\big(\overline{M}_{l,i}\big) \qquad \text{and} \qquad \underline{C}_{l+1} = \sum_{i=1}^{\underline{Z}_l} \Lambda\big(\underline{M}_{l,i}\big), \; l \geq 0. \quad (4.3.12)$$

Using the coupling given by (4.3.5), we can rewrite the capacity $\underline{C}_{l+1}$ as

$$\underline{C}_{l+1} = \sum_{i\in\partial\mathcal{N}_l} \Lambda_i. \quad (4.3.13)$$

For the proof of Theorem 4.1 and 4.3, in the case of the PRG, we need to control the difference between $\overline{C}_l$ and $\underline{C}_l$ for fixed l. For this we will use the following proposition:

Proposition 4.10 *There exist constants $\alpha_2, \beta_2 > 0$, such that for all $0 < \eta < \alpha_2$ and all $l \leq (1/2 + \eta) \log_\nu N$,*

$$\mathbb{P}\left(\sum_{k=1}^{l}(\overline{C}_k - \underline{C}_k) > N^{1/2-\alpha_2}\right) \leq N^{-\beta_2}. \quad (4.3.14)$$

Proof.   The proof is deferred to Section 4.A.3.   □

In order to prove Theorem 4.1 and Theorem 4.3 we will grow two $\underline{\text{NR}}$-processes $\{\underline{Z}_l^{(i)}, \underline{M}_l^{(i)}\}_{l\geq0}$, for $i = 1, 2$. The root of $\{\underline{Z}_l^{(i)}, \underline{M}_l^{(i)}\}_{l\geq0}$ starts from a uniformly chosen node or mark $A_i \in \{1, 2, \ldots, N\}$. These two nodes are different **whp** , because

$$\mathbb{P}(A_1 = A_2) = \frac{1}{N}.$$

By (4.3.5) the $\underline{\text{NR}}$-process can be coupled to the neighborhood expansion shells $\{\mathcal{N}_l^{(1)}\}_{l\geq0}$ and $\{\mathcal{N}_l^{(2)}\}_{l\geq0}$. In the following lemma we compute the distribution of the number of edges between two shells with different subindeces, i.e., $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_l^{(2)}$.

Lemma 4.11 *Consider the neighborhood expansion shells $\{\mathcal{N}_l^{(1)}\}$ and $\{\mathcal{N}_l^{(2)}\}$. Then conditionally on $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_l^{(2)}$ and given that $\mathcal{N}_k^{(1)} \cap \mathcal{N}_l^{(2)} = \emptyset$ the number of edges between the nodes in $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_l^{(2)}$, for fixed positive integers k and l, is distributed as a Poisson random variable with mean*

$$\frac{C_{k+1}^{(1)}C_{l+1}^{(2)}}{L_N}. \quad (4.3.15)$$

Proof.   Conditioned on $\mathcal{N}_k^{(1)}$, $\mathcal{N}_l^{(2)}$ and $\mathcal{N}_k^{(1)} \cap \mathcal{N}_l^{(2)} = \emptyset$, the number of edges between $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_l^{(2)}$ is given by

$$\sum_{i\in\partial\mathcal{N}_k^{(1)}}\sum_{j\in\partial\mathcal{N}_l^{(2)}} E_{ij}^{(P)}, \quad (4.3.16)$$

where $E_{ij}^{(P)}$ are independent Poisson random variables with mean $\Lambda_i\Lambda_j/L_N$, see (4.1.6). Sums of independent Poisson random variables are again Poissonian, thus (4.3.16) is a Poisson random variable with mean the expected value of (4.3.16):

$$\sum_{i\in\partial\mathcal{N}_k^{(1)}}\sum_{j\in\partial\mathcal{N}_l^{(2)}} \frac{\Lambda_i\Lambda_j}{L_N} = \frac{1}{L_N}\left(\sum_{i\in\partial\mathcal{N}_k^{(1)}}\Lambda_i\right)\left(\sum_{i\in\partial\mathcal{N}_l^{(2)}}\Lambda_j\right) = \frac{C_{k+1}^{(1)}C_{l+1}^{(2)}}{L_N},$$

where we have used (4.3.13) in the last step.   □

The further proof of Theorems 4.1-4.3 crucially relies on the following technical claim:

Proposition 4.12 *There exist constants $\alpha_3, \beta_3, \eta > 0$ such that for all $l \leq (1 + 2\eta) \log_\nu N$, as $N \to \infty$,*

$$\mathbb{P}\left(\frac{1}{N}\left|\sum_{k=2}^{l+1}\mathcal{Z}_{\lceil k/2\rceil}^{(1)}\mathcal{Z}_{\lfloor k/2\rfloor}^{(2)} - \sum_{k=2}^{l+1}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}\right| > N^{-\alpha_3}\right) = \mathcal{O}\left(N^{-\beta_3}\right). \quad (4.3.17)$$

Proof.   The proof is deferred to Section 4.A.4.   □

In the next section we will use this proposition in combination with Lemma 4.11 to replace sums over capacities, which do depend on N, by sums over sizes of a BP, which do not depend on N anymore.

4.4   Proof of Theorem 4.1 and 4.3 for the PRG

In this section, we give the proof of Theorem 4.1 and 4.3 for the PRG model. Using the coupling result in Proposition 4.7 we obtain Theorem 4.1 and 4.3 for all random graphs satisfying the assumptions **A1** and **A2**. As in the previous section, we denote by G_N a PRG.

We grow two $\underline{\text{NR}}$-processes. Each $\underline{\text{NR}}$-process starts from a uniformly chosen node $A_i \in \{1, 2, \ldots, N\}$, $i = 1, 2$, such that $A_1 \neq A_2$, **whp** .

Step 1: Expressing $\mathbb{P}(H_N > l)$ in capacities. We have $H_N > 1$ iff (if and only if) there are no edges between the nodes A_1 and A_2, which is equivalent to the fact that $X_{A_1 A_2} = 0$. Given the capacities $\underline{C}_1^{(1)}$ and $\underline{C}_1^{(2)}$, the number of edges between the nodes A_1 and A_2 has, according to Lemma 4.11, a Poisson distribution with mean

$$\frac{\underline{C}_1^{(1)} \underline{C}_1^{(2)}}{L_N}. \tag{4.4.1}$$

By taking expectations

$$\mathbb{P}(H_N > 1) = \mathbb{E}[\mathbb{P}_N(X_{A_1 A_2} = 0)] = \mathbb{E}\left[\exp\left\{-\frac{C_1^{(1)} C_1^{(2)}}{L_N}\right\}\right]. \tag{4.4.2}$$

We next inspect the capacity of the first generation of $\underline{Z}_1^{(1)}$ given by $\underline{C}_2^{(1)}$. Given $\underline{C}_2^{(1)}$ and $\underline{C}_1^{(2)}$, that is the total capacity of the nodes in $\underline{Z}_1^{(1)}$ and the capacity of node A_2, we again have a Poisson number of edges between node A_2 and the nodes in $\underline{Z}_1^{(1)}$, however, this time with parameter

$$\frac{\underline{C}_2^{(1)} \underline{C}_1^{(2)}}{L_N}. \tag{4.4.3}$$

In order to compute the survival probability $\mathbb{P}(H_N > l)$ we need more notation. We write $\mathbb{Q}_C^{(l_1, l_2)}$ for the conditional probabilities given $\{\underline{C}_k^{(1)}\}_{k=1}^{l_1}$ and $\{\underline{C}_k^{(2)}\}_{k=1}^{l_2}$. We further write $\mathbb{E}_C^{(l_1, l_2)}$ for the expectation with respect to $\mathbb{Q}_C^{(l_1, l_2)}$. For $l_2 = 0$, we only condition on $\{\underline{C}_k^{(1)}\}_{k=1}^{l_1}$. Lemma 4.11 implies that

$$\mathbb{Q}_C^{(k,l)}(H_N > k + l - 1 | H_N > k + l - 2) = \exp\left\{-\frac{\underline{C}_k^{(1)} \underline{C}_l^{(2)}}{L_N}\right\}.$$

Indeed, the event $\{H_N > k+l-2\}$ implies that $\mathcal{N}_k^{(1)} \cap \mathcal{N}_l^{(2)} = \emptyset$. Then from (4.4.1)

and the above statement,

$$
\begin{aligned}
\mathbb{P}(H_N > 2) &= \mathbb{E}\big[\mathbb{Q}_C^{(1,1)}(H_N > 1)\,\mathbb{Q}_C^{(1,1)}(H_N > 2|H_N > 1)\big] \\
&= \mathbb{E}\big[\mathbb{Q}_C^{(1,1)}(H_N > 1)\,\mathbb{E}_C^{(1,1)}\big[Q_C^{(2,1)}(H_N > 2|H_N > 1)\big]\big] \\
&= \mathbb{E}\big[\mathbb{E}_C^{(1,1)}\big[\mathbb{Q}_C^{(1,1)}(H_N > 1)\,\mathbb{Q}_C^{(2,1)}(H_N > 2|H_N > 1)\big]\big] \\
&= \mathbb{E}\left[\exp\left\{-\frac{\underline{C}_1^{(1)}\underline{C}_1^{(2)}}{L_N}\right\}\cdot\exp\left\{-\frac{\underline{C}_2^{(1)}\underline{C}_1^{(2)}}{L_N}\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{-\frac{\sum_{k=2}^3 \underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}}{L_N}\right\}\right].
\end{aligned}
$$

By induction we obtain as in [54, Lemma 4.1],

$$
\mathbb{P}(H_N > l) = \mathbb{E}\left[\exp\left\{-\frac{\sum_{k=2}^{l+1}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}}{L_N}\right\}\right]. \tag{4.4.4}
$$

Note that (4.4.4) is an equality, while in [54] an error needed to be taken along.

Step 2: Coupling with the BP with offspring distribution $\{g_n\}$. In this step we replace $\underline{C}_l^{(1)}$ and $\underline{C}_l^{(2)}$ by $\mathcal{Z}_l^{(1)}$ and $\mathcal{Z}_l^{(2)}$.

For each event $\mathcal{B}$, and any two nonnegative random variables V and W,

$$
\begin{aligned}
\big|\mathbb{E}\big[e^{-V}\big] - \mathbb{E}\big[e^{-W}\big]\big| &\leq \big|\mathbb{E}\big[(e^{-V} - e^{-W})\mathbf{1}_{\mathcal{B}}\big]\big| + \big|\mathbb{E}\big[(e^{-V} - e^{-W})\mathbf{1}_{\mathcal{B}^c}\big]\big| \\
&\leq \big|\mathbb{E}\big[(e^{-V} - e^{-W})\mathbf{1}_{\mathcal{B}}\big]\big| + \mathbb{P}(\mathcal{B}^c).
\end{aligned}
$$

Now take

$$
\mathcal{B} = \left\{\frac{1}{N}\left|\sum_{k=2}^{l+1}\mathcal{Z}_{\lceil k/2\rceil}^{(1)}\mathcal{Z}_{\lfloor k/2\rfloor}^{(2)} - \sum_{k=2}^{l+1}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}\right| \leq N^{-\alpha_3}\right\},
$$

and the random variables V and W as

$$
V = \frac{1}{N}\sum_{k=2}^{l+1}\mathcal{Z}_{\lceil k/2\rceil}^{(1)}\mathcal{Z}_{\lfloor k/2\rfloor}^{(2)}, \qquad W = \frac{1}{N}\sum_{k=2}^{l+1}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}.
$$

Then for $l \leq (1+2\eta)\log_\nu N$ the Proposition 4.12 implies that $\mathbb{P}(\mathcal{B}^c) = \mathcal{O}(N^{-\beta_3})$, whereas on the event $\mathcal{B}$ we have $|V - W| \leq N^{-\alpha_3}$. Hence, using that $e^{-v} - e^{-w} = \mathcal{O}(v - w)$ when v, w are small, and that $e^{-v} \leq 1$, $v \geq 0$, we obtain

$$
\big|\mathbb{E}\big[e^{-V}\big] - \mathbb{E}\big[e^{-W}\big]\big| \leq \mathcal{O}\big(N^{-\alpha_3}\big)\mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{B}^c) = \mathcal{O}\big(N^{-\alpha_3}\big) + \mathcal{O}\big(N^{-\beta_3}\big). \tag{4.4.5}
$$

It is now clear from *step* 1, the above result and Lemma 4.4, where we take $q = 1$, that for some $\beta > 0$ and all $l \leq (1+2\eta)\log_\nu N$,

$$
\mathbb{P}(H_N > l) = \mathbb{E}\left[\exp\left\{-\frac{\sum_{k=2}^{l+1}\mathcal{Z}_{\lceil k/2\rceil}^{(1)}\mathcal{Z}_{\lfloor k/2\rfloor}^{(2)}}{\mu N}\right\}\right] + \mathcal{O}\big(N^{-\beta}\big). \tag{4.4.6}
$$

Step 3: Evaluation of the limit points. From this step on, the remainder of the proof of our main theorem is identically to the proof of Theorem 1.1 in [54]. To let the proof be self-contained, we finish the main line of the argument. Starting from (4.4.6) we replace l by $\sigma_N + l$ and assume that $\sigma_N + l \leq (1 + 2\eta) \log_\nu N$, where, as before, $\sigma_N = \lfloor \log_\nu N \rfloor$, to obtain

$$\mathbb{P}(H_N > \sigma_N + l) = \mathbb{E}\left[\exp\left\{ -\frac{\sum_{k=2}^{\sigma_N + l + 1} \mathcal{Z}_{\lceil k/2 \rceil}^{(1)} \mathcal{Z}_{\lfloor k/2 \rfloor}^{(2)}}{\mu N} \right\}\right] + \mathcal{O}(N^{-\beta}). \qquad (4.4.7)$$

We write $N = \nu^{\log_\nu N} = \nu^{\sigma_N - a_N}$, where we recall that $a_N = \lfloor \log_\nu N \rfloor - \log_\nu N$. Then

$$\frac{\sum_{k=2}^{\sigma_N + l + 1} \mathcal{Z}_{\lceil k/2 \rceil}^{(1)} \mathcal{Z}_{\lfloor k/2 \rfloor}^{(2)}}{\mu N} = \mu \nu^{a_N + l} \frac{\sum_{k=2}^{\sigma_N + l + 1} \mathcal{Z}_{\lceil k/2 \rceil}^{(1)} \mathcal{Z}_{\lfloor k/2 \rfloor}^{(2)}}{\mu^2 \nu^{\sigma_N + l}}.$$

In the above expression, the factor ν^{a_N} prevents proper convergence.

Without the factor $\mu \nu^{a_N + l}$, we obtain from Appendix A4 of [54] that, with probability 1,

$$\lim_{N \to \infty} \frac{\sum_{k=2}^{\sigma_N + l + 1} \mathcal{Z}_{\lceil k/2 \rceil}^{(1)} \mathcal{Z}_{\lfloor k/2 \rfloor}^{(2)}}{\mu^2 \nu^{\sigma_N + l}} = \frac{\mathcal{W}^{(1)} \mathcal{W}^{(2)}}{\nu - 1}, \qquad (4.4.8)$$

We now use the speed of convergence result of [6], which was further developed in Section 2 of [54] and which reads that there exists a positive β such that:

$$\mathbb{P}(|\mathcal{W} - \mathcal{W}_k| > (\log N)^{-\alpha}) = \mathcal{O}(N^{-\beta}), \; k \leq \lfloor \tfrac{1}{2} \log_\nu N \rfloor. \qquad (4.4.9)$$

for each positive α. Combining (4.4.7) and (4.4.9) we obtain that for each $\alpha > 0$ and for $l \leq 2\eta \log_\nu N$,

$$\mathbb{P}(H_N > \sigma_N + l) = \mathbb{E}\big[\exp\{-\kappa \nu^{a_N + l} \mathcal{W}_1 \mathcal{W}_2\}\big] + \mathcal{O}((\log N)^{-\alpha}). \qquad (4.4.10)$$

From (4.4.10) one can finally derive as in [54], that, asymptotically as $N \to \infty$, the probability $\mathbb{P}(H_N < \infty)$ is equivalent to the probability $q^2 = \mathbb{P}(\mathcal{W}_1 \mathcal{W}_2 > 0)$, where q is the survival probability of the branching process $\{\mathcal{Z}_l\}_{l \geq 0}$, so that (4.4.10) induces for $l \leq 2\eta \log_\nu N$,

$$\mathbb{P}(H_N \leq \sigma_N + l \,|\, H_N < \infty) = \mathbb{E}\big[1 - \exp\{-\kappa \nu^{a_N + l} \mathcal{W}_1 \mathcal{W}_2\}\,|\,\mathcal{W}_1 \mathcal{W}_2 > 0\big] + o(1). \qquad (4.4.11)$$

4.A   Appendix

4.A.1   Proof of Lemma 4.4

In this section we prove Lemma 4.4. For convenience we restate the lemma here as Lemma 4.13. Recall the definition of the event $\mathcal{S}_{q,\alpha_0}$ given in (4.2.4).

Lemma 4.13 *For each fixed $q \in (0, \tau - 1)$, there exist constants $\alpha_0, \beta_0 > 0$ such that*

$$\mathbb{P}\big(\mathcal{S}_{q,\alpha_0}^c\big) \leq N^{-\beta_0}. \qquad (4.A.1)$$

Proof.   For $q < (\tau - 1)/2$, in which case $\mathbb{E}[\Lambda^{2q}] < \infty$, we simply apply Chebychev's inequality:

$$\mathbb{P}\big(\mathcal{S}_{q,\alpha_0}^c\big) \leq N^{2\alpha_0}\mathrm{Var}(S_{N,q}) = N^{2\alpha_0-1}\mathrm{Var}(\Lambda^q) = O(N^{2\alpha_0-1}), \qquad (4.A.2)$$

so each $0 < \alpha_0 < \frac{1}{2}$ and $\beta_0 = -1 + 2\alpha_0$ will do the job.

For $q \in [(\tau - 1)/2, \tau - 1)$, the proof is more involved. Since $\mathbb{E}[\Lambda_1^{2q}] < \infty$ no longer holds, we have to cut off the large values of $S_{N,q}$. Take δ such that $\frac{1}{2} < \delta < (\tau - 1)/(2q) \leq 1$, and define for $r = 2q\delta$, the event

$$\mathcal{G}_r = \{S_{N,r} \leq N^{\alpha_0}\}.$$

Since $r = 2q\delta < \tau - 1$ implies $\mathbb{E}[\Lambda^r] < \infty$, the Markov inequality yields

$$\mathbb{P}(\mathcal{G}_r^c) \leq N^{-\alpha_0}\mathbb{E}[S_{N,r}] = N^{-\alpha_0}\mathbb{E}[\Lambda^r] = \mathcal{O}\big(N^{-\alpha_0}\big). \qquad (4.A.3)$$

We also use Minkowski's Inequality, [51, (2.11.24), page 30], implying that for each $s \in [0, 1]$,

$$\left(\sum_{i=1}^{N}\Lambda_i\right)^s \leq \sum_{i=1}^{N}\Lambda_i^s. \qquad (4.A.4)$$

Applying both (4.A.3) and (4.A.4), we arrive at

$$\mathbb{P}\big(\mathcal{S}_{q,\alpha_0}^c\big) \leq \mathbb{P}\Big(\{(S_{N,q} - \mathbb{E}[\Lambda^q])^2 > N^{-2\alpha_0}\} \cap \mathcal{G}_r\Big) + \mathbb{P}(\mathcal{G}_r^c)$$

$$\leq N^{2\alpha_0-2}\mathbb{E}\left[\sum_{i=1}^{N}\Lambda_i^{2q}\mathbf{1}_{\mathcal{G}_r}\right] + \mathbb{P}(\mathcal{G}_r^c)$$

$$\leq N^{2\alpha_0-2}\mathbb{E}\left[\left(\sum_{i=1}^{N}\Lambda_i^{2q\delta}\right)^{1/\delta}\mathbf{1}_{\mathcal{G}_r}\right] + \mathbb{P}(\mathcal{G}_r^c)$$

$$= N^{2\alpha_0-2}N^{(1+\alpha_0)/\delta}\mathbb{E}\left[\big(N^{-\alpha_0}S_{N,2q\delta}\big)^{1/\delta}\mathbf{1}_{\mathcal{G}_{2q\delta}}\right] + \mathbb{P}(\mathcal{G}_r^c)$$

$$= \mathcal{O}\Big(N^{-\min\{-2\alpha_0+2-(1+\alpha_0)/\delta,\alpha_0\}}\Big).$$

Consequently, for $q \in [(\tau - 1)/2, \tau - 1)$ we can take $\beta_0 = \min\{\alpha_0, 2 - 2\alpha_0 - (1 + \alpha_0)/\delta\}$, and $\beta_0 > 0$ provided that we choose $0 < \alpha_0 < (2\delta - 1)/(2\delta + 1)$.

4.A.2   Coupling of $\{\overline{\mathcal{Z}}_i\}_{i\geq 0}$ and $\{\mathcal{Z}_i\}_{i\geq 0}$

We will couple the delayed BP $\{\overline{\mathcal{Z}}_i\}_{i\geq 0}$ to the delayed BP $\{\mathcal{Z}_i\}_{i\geq 0}$ with law f in the first generation and law g in the second and further generations, using a classical coupling argument, see [80].

We give each individual of $\{\overline{\mathcal{Z}}_i\}_{i\geq 0}$ an independent indicator, which is 1 with probability

$$q_{0,N} = \frac{1}{2}\sum_{n=0}^{\infty}|f_n^{(N)} - f_n|,$$

in the first generation and with probability

$$q_{1,N} = \frac{1}{2} \sum_{n=0}^{\infty} |g_n^{(N)} - g_n|,$$

in the second and further generations. When this indicator is 0 for a certain individual, then the offspring of this individual is successfully coupled. When, on the other hand, the indicator is 1, then an error has occurred, and the coupling is not successful. In this case, the law of the offspring of $\{\overline{Z}_i\}_{i\geq 0}$ is different from the one in $\{\mathcal{Z}_i\}_{i\geq 0}$, and we record an error. The following proposition gives bounds on the probability that the *random* variables $q_{0,N}$ and $q_{1,N}$ are larger than N to a certain negative power:

Proposition 4.14 *For each $\tau > 3$, there exist constants $\alpha_1 > 0$ and $\beta_1 > 0$ such that*

$$\mathbb{P}\left(\sum_{n=0}^{\infty} (n+1) |g_n^{(N)} - g_n| \geq N^{-\alpha_1} \right) \leq N^{-\beta_1}. \tag{4.A.5}$$

Consequently,

$$\mathbb{P}\left(\max\{q_{0,N}, q_{1,N}\} \geq N^{-\alpha_1} \right) \leq N^{-\beta_1}. \tag{4.A.6}$$

Using Proposition 4.14 we will render a coupling between the sums $\sum_{k=1}^{l} \overline{Z}_k$ and $\sum_{k=1}^{l} \mathcal{Z}_k$.

In the proof of this proposition we need an additional lemma from analysis. Recall the definition of f_n in (4.1.7).

Lemma 4.15 *Fix $\tau > 3$, then for $s = 0, 1, 2$ and $\delta = \min\{(\tau-3)/4, 1/4\}$, there exists a constant $m_0 > 0$ such that for all $m \geq m_0$,*

$$\sum_{n=m}^{\infty} (n+1)^s f_{n+1} \leq m^{-\delta}. \tag{4.A.7}$$

We defer the proof of Lemma 4.15 to the end of this section.

Before we give the proof of Proposition 4.14, we introduce the function

$$t_n(x) = \frac{x^n}{n!} e^{-x}, \; x > 0. \tag{4.A.8}$$

Using this function we can rewrite (4.3.6) and (4.3.8) as:

$$f_n^{(N)} = \frac{1}{N} \sum_{i=1}^{N} t_n(\Lambda_i) \qquad \text{and} \qquad g_n^{(N)} = \frac{n+1}{L_N} \sum_{i=1}^{N} t_{n+1}(\Lambda_i). \tag{4.A.9}$$

Proof of Proposition 4.14. The proof is based on the proof of [54, Proposition 3.4]. For $0 < u < 1$, $v > 0$ and $\alpha_0 \in (0, 1/2)$ we define $\mathcal{B} = \mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{B}_3 \cap \mathcal{B}_4$,

where

$$\mathcal{B}_1 = \mathcal{B}_1(\alpha_0) = \left\{ \left| \frac{L_N}{\mu N} - 1 \right| \leq N^{-\alpha_0} \right\},$$

$$\mathcal{B}_2 = \mathcal{B}_2(u, v) = \left\{ \frac{1}{N} \sum_{i=1}^{N} \sum_{n=N^u}^{\infty} (n+1)^2 t_{n+1}(\Lambda_i) \leq N^{-v} \right\},$$

$$\mathcal{B}_3 = \mathcal{B}_3(u, v) = \left\{ \frac{1}{N} \sum_{n=0}^{N^u} (n+1)^2 \left| \sum_{i=1}^{N} \left(t_{n+1}(\Lambda_i) - f_{n+1} \right) \right| \leq N^{-v} \right\},$$

$$\mathcal{B}_4 = \mathcal{B}_4(\alpha_0) = \left\{ \frac{1}{N} \left| \sum_{i=1}^{N} (t_0(\Lambda_i) - f_0) \right| \leq N^{-\alpha_0} \right\}.$$

The precise values of α_0, u and v will be chosen later in the proof. The strategy of the proof is as follows. We will show that

$$\mathbb{P}(\mathcal{B}^c) \leq N^{-\beta_1}, \tag{4.A.10}$$

for some $\beta_1 > 0$, and that on the event $\mathcal{B}$,

$$\sum_{n=0}^{\infty} (n+1) |g_n^{(N)} - g_n| = \mathcal{O}(N^{-\alpha_1}), \tag{4.A.11}$$

for some $\alpha_1 > 0$.

We start by proving (4.A.11). We bound

$$q_{1,N} = \sum_{n=0}^{\infty} (n+1)|g_n^{(N)} - g_n| \leq \sum_{n=0}^{\infty} (n+1)|g_n^{(N)} - \frac{N\mu}{L_N} g_n| + (\nu+1)\left| \frac{N\mu}{L_N} - 1 \right|, \tag{4.A.12}$$

using $\nu = \sum_{n=0}^{\infty} n g_n$. On $\mathcal{B}_1$, the second term on the right-hand side of (4.A.12) is bounded by $\mathcal{O}(N^{-\alpha_0})$. The first term on the right-hand side of (4.A.12) can be rewritten by (4.3.9), (4.A.8), (4.A.9) and (4.1.8), respectively, as

$$\sum_{n=0}^{\infty} (n+1)|g_n^{(N)} - \frac{N\mu}{L_N} g_n| = \frac{1}{L_N} \sum_{n=0}^{\infty} (n+1)^2 \left| \sum_{i=1}^{N} \left(t_{n+1}(\Lambda_i) - f_{n+1} \right) \right|.$$

Consequently, the first term on the right-hand side of (4.A.12) can be bounded on $\mathcal{B}_1$, for N sufficiently large, by

$$\sum_{n=0}^{\infty} (n+1)|g_n^{(N)} - \frac{N\mu}{L_N} g_n| \leq \frac{2}{\mu N} \sum_{n=0}^{\infty} (n+1)^2 \left| \sum_{i=1}^{N} \left(t_{n+1}(\Lambda_i) - f_{n+1} \right) \right|.$$

We next split the sum over n into $n < N^u$ and $n \geq N^u$. On $\mathcal{B}_3$, the contribution from $n < N^u$ is at most $\frac{2}{\mu} N^{-v}$, whereas we can bound the contribution from $n \geq N^u$ on $\mathcal{B}_2$ by

$$\frac{2}{\mu N} \sum_{n=N^u}^{\infty} (n+1)^2 \sum_{i=1}^{N} (t_{n+1}(\Lambda_i) + f_{n+1}) \leq \frac{2}{\mu} N^{-v} + \frac{2}{\mu} \sum_{n=N^u}^{\infty} (n+1)^2 f_{n+1}.$$

For $\tau > 3$, the latter term is bounded by $N^{-u\delta}$ by Lemma 4.15. Thus, we obtain the claim in (4.A.11) for each

$$0 < \alpha_1 < \min\{\alpha_0, u\delta, v\}. \tag{4.A.13}$$

To show (4.A.10), it is sufficient to show that

$$\mathbb{P}(\mathcal{B}_1^c) + \mathbb{P}(\mathcal{B}_2^c) + \mathbb{P}(\mathcal{B}_3^c) + \mathbb{P}(\mathcal{B}_4^c) \leq N^{-\beta_1}, \tag{4.A.14}$$

for some $\beta_1 > 0$. We will prove that $\mathbb{P}(\mathcal{B}_i^c) \leq N^{-a_i}$, where $a_1, a_2, \ldots, a_4$ depend on α_0, δ, u and v. Later we will show that we can take α_0, δ, u and v such that (4.A.13) is satisfied and $a_1, a_2, a_3, a_4 > 0$.

Lemma 4.13 states that $\mathbb{P}(\mathcal{B}_1^c) = \mathcal{O}(N^{-\beta_0})$, thus $a_1 = \beta_0$. We bound $\mathbb{P}(\mathcal{B}_2^c)$ by using the Markov inequality and Lemma 4.15:

$$\mathbb{P}(\mathcal{B}_2^c) \leq N^v \sum_{n=N^u}^{\infty} (n+1)^2 \mathbb{E}[t_{n+1}(\Lambda)] = N^v \sum_{n=N^u}^{\infty} (n+1)^2 f_{n+1} \leq N^{v-u\delta} = N^{-a_2},$$

for some constant $\delta > 0$ and we take $a_2 = u\delta - v$ and N sufficiently large.

Before we bound $\mathbb{P}(\mathcal{B}_3^c)$ from above, we first note that

$$\mathbb{E}\left[\left(\sum_{i=1}^{N}[t_{n+1}(\Lambda_i) - f_{n+1}]\right)^2\right] = \mathrm{Var}\left(\sum_{i=1}^{N} t_{n+1}(\Lambda_i)\right) \leq N\mathbb{E}\left[t_{n+1}(\Lambda)^2\right] \leq N, \tag{4.A.15}$$

where we used in the last step that $t_{n+1}(x)$ is a probability. In the sequence of inequalities below, which gives the bound for $\mathbb{P}(\mathcal{B}_3^c)$, we use the Markov inequality, Cauchy-Schwarz in the form $\sum_{n=0}^{N^u} b_n \leq (\sum_{n=0}^{N^u} 1^2 \sum_{n=0}^{N^u} b_n^2)^{\frac{1}{2}}$, the Jensen inequality applied to the concave function $x \mapsto \sqrt{x}$ and (4.A.15), respectively, to obtain

$$\mathbb{P}(\mathcal{B}_3^c) \leq N^{v-1}\mathbb{E}\Big[\sum_{n=0}^{N^u}(n+1)^2 \big|\sum_{i=1}^{N}\big(t_{n+1}(\Lambda_i) - f_{n+1}\big)\big|\Big]$$

$$\leq N^{v-1}(N^u+1)^{\frac{1}{2}}\mathbb{E}\left[\left(\sum_{n=0}^{N^u}(n+1)^4\left(\sum_{i=1}^{N}\big(t_{n+1}(\Lambda_i) - f_{n+1}\big)\right)^2\right)^{1/2}\right]$$

$$\leq 2N^{v+u/2-1}\left(\mathbb{E}\left[\sum_{n=0}^{N^u}(n+1)^4\left(\sum_{i=1}^{N}\big(t_{n+1}(\Lambda_i) - f_{n+1}\big)\right)^2\right]\right)^{1/2}$$

$$\leq 2N^{v+u/2-1}\Big(\sum_{n=0}^{N^u}(n+1)^4 N\Big)^{1/2} \leq 2N^{v+3u-1/2} = \mathcal{O}\left(N^{-a_3}\right),$$

for $a_3 = 1/2 - v - 3u$.

Finally, we bound the fourth term of (4.A.14) using the Chebychev inequality

$$\mathbb{P}(\mathcal{B}_4^c) \leq N^{2u-2}\mathrm{Var}\left(\sum_{i=1}^{N} t_0(\Lambda_i)\right) = N^{2u-1}\mathrm{Var}(t_0(\Lambda)) \leq N^{2u-1} = N^{-u_4},$$

where $a_4 = 2u - 1$, because $t_0(\Lambda) \leq 1$, as it is a probability.

We now verify that we can choose α_0, δ, u and v in such a way that both α_1 and β_1 can be taken positive. Recall that

$$a_1 = \beta_0, \quad a_2 = u\delta - v, \quad a_3 = 1/2 - 3u - v, \quad \text{and} \quad a_4 = 2u - 1. \quad (4.A.16)$$

The constant $\delta > 0$ follows from Lemma 4.15, we pick $0 < \alpha_0 < 1/2$ and use Lemma 4.13 to find a positive β_0, then finally we choose $u = 1/12$ and v such that $0 < v < \min\{1/4, \delta/12\}$. The reader easily verifies that these conditions imply that the constants a_1, a_2, a_3 and a_4 are positive, so that β_1 equal to the minimum of these four quantities does the job. For α_1 we can take any value that satisfies (4.A.13). This completes the proof of (4.A.13).

In order to show (4.A.6) it is sufficient to show that $\mathbb{P}(q_{0,N} \geq N^{-\alpha_1}) \leq N^{-\beta_1}$, i.e.,

$$\mathbb{P}\left(\sum_{n=0}^{\infty} (n+1) |f_n^{(N)} - f_n| \geq N^{-\alpha_1}\right) \leq N^{-\beta_1}, \quad (4.A.17)$$

because we already have shown that $\mathbb{P}(q_{1,N} \geq N^{-\alpha_1}) \leq N^{-\beta_1}$.

On the event $\mathcal{B}_4$ we have that

$$\left|f_0^{(N)} - f_0\right| = \frac{1}{N}\left|\sum_{i=1}^{N}(t_0(\Lambda_i) - f_0)\right| \leq N^{-\alpha_0}.$$

Thus, using the event $\mathcal{B}$, the, now proven, bound (4.A.11), $L_N/N < 2\mu$ for N sufficiently large, and the fact $\alpha_1 \leq \alpha_0$ we have that

$$q_{0,N} = \sum_{n=0}^{\infty}(n+1)|f_n^{(N)} - f_n| \leq \left|f_0^{(N)} - f_0\right| + 2\sum_{n=0}^{\infty}(n+1)\left|f_{n+1}^{(N)} - f_{n+1}\right|$$

$$\leq \left|f_0^{(N)} - f_0\right| + 2\frac{L_N}{N}\sum_{n=0}^{\infty}\left|g_{n+1}^{(N)} - g_{n+1}\right| + 2\frac{L_N}{N}\sum_{n=0}^{\infty}g_n\left|\frac{L_N}{N} - \mu\right|$$

$$\leq \left|f_0^{(N)} - f_0\right| + 4\mu\sum_{n=0}^{\infty}\left|g_{n+1}^{(N)} - g_{n+1}\right| + 4\mu\left|\frac{L_N}{N} - \mu\right| = \mathcal{O}\left(N^{-\min\{\alpha_1,\alpha_0\}}\right),$$

and remark that $\min\{\alpha_1, \alpha_0\} = \alpha_1$ □

We close this section with the proof of Lemma 4.15.

Proof of Lemma 4.15. It is sufficient to consider only the case $s = 2$, because

$$\sum_{n=m}^{\infty}(n+1)^s f_{n+1} \leq \sum_{n=m}^{\infty}(n+1)^2 f_{n+1}, \text{ for } s = 0, 1, 2.$$

Recall that we denote by $Poi(\lambda)$ a Poisson random variable with mean λ. Then,

using (4.1.7) we can bound the right side of the above equation by

$$\mathbb{E}\left[\sum_{n=m}^{\infty}(n+1)^2 e^{-\Lambda}\frac{\Lambda^{n+1}}{(n+1)!}\right]$$

$$= \mathbb{E}\left[\Lambda^2 \sum_{n=m}^{\infty} e^{-\Lambda}\frac{\Lambda^{n-1}}{(n-1)!}(1+1/n)\right] \le 2\mathbb{E}\left[\Lambda^2\mathbb{P}(Poi(\Lambda)\ge m|\Lambda)\right]$$

$$\le 2\mathbb{E}\left[\Lambda^2\mathbf{1}_{\{\Lambda\ge m^w\}}\right] + 2m^{2w}\mathbb{E}\left[\mathbb{P}(Poi(\Lambda)\ge m|\Lambda)\,\mathbf{1}_{\{\Lambda<m^w\}}\right], \quad (4.A.18)$$

for each $w \in (0,1)$. Using (4.2.11), where we replace c_N by m^w, we bound the first term on the right hand side by $Cm^{w(3-\tau)}$, for some constant C. To bound the second term on the right hand side of (4.A.18) we use that conditioned on the event $\{\Lambda < m^w\}$

$$\mathbb{P}(Poi(\Lambda)\ge m|\Lambda) \le m^{-1}\mathbb{E}[Poi(\Lambda)\,|\Lambda] = m^{-1}\Lambda \le m^{w-1}.$$

Therefore, we bound the second term on the right hand side of (4.A.18) by $2m^{3w-1}$.
  Thus, pick $w=1/4$ and $0 < \delta < \min\{(\tau-3)/4, 1/4\}$, then

$$\sum_{n=m}^{\infty}(n+1)^2 f_{n+1} \le Cm^{w(3-\tau)} + 2m^{3w-1} \le m^{-\delta}.$$

$\square$

4.A.3   Proof of Proposition 4.10.

In this part of the appendix we give the main result on the coupling between the capacities of the $\overline{\mathrm{NR}}$-process and those of the $\underline{\mathrm{NR}}$-process. For convenience we restate Proposition 4.10 as:

Proposition 4.16 *There exist constants $\alpha_2, \beta_2 > 0$, such that for all $0 < \eta < \alpha_2$ and all $l \le (1/2+\eta)\log_\nu N$,*

$$\mathbb{P}\left(\sum_{k=1}^{l}(\overline{C}_k - \underline{C}_k) > N^{1/2-\alpha_2}\right) \le N^{-\beta_2}. \qquad (4.A.19)$$

Before we go to the actual proof of Proposition 4.A.19, we explain where the difference between $\overline{C}_k$ and $\underline{C}_k$ stems from, and we give a outline of the proof.

  The difference between $\overline{C}_k$ and $\underline{C}_k$ stems from individuals, whose mark has appeared previously and which is consequently discarded together with all its descendants. Call such an individual a *duplicate*. To show the claim (4.A.19) we need some more details of the thinning procedure of the $\overline{\mathrm{NR}}$-process $\{\overline{Z}_l, \overline{M}_l\}_{l\ge 0}$. We relabel the marks of the $\overline{\mathrm{NR}}$-process

$$\overline{M}_{0,1}, \overline{M}_{1,1}, \ldots, \overline{M}_{1,\overline{Z}_1}, \overline{M}_{2,1}, \ldots,$$

given in Section 4.3.1, as

$$M_0, M_1, M_2, \ldots, M_{\overline{Z}_1}, M_{\overline{Z}_1+1}, \ldots \qquad (4.A.20)$$

By definition, M_0 is a random variable uniformly chosen from $\{1, 2, \ldots, N\}$ and the marks M_v, for $v > 0$, are independent copies of the random variable M given by (4.3.2). Introduce the random variable Y_v, such that $Y_v = 1$ if M_v is a duplicate and $Y_v = 0$ otherwise, so that

$$Y_v = \mathbf{1}_{\cup_{w=0}^{v-1}\{M_w = M_v\}}, \tag{4.A.21}$$

and denote by $s(v)$, $v \geq 0$, the generation of individual v in the $\overline{\mathrm{NR}}$-process. Let Dup_l be the set of all the duplicates in the first l generations of the $\overline{\mathrm{NR}}$-process, so that

$$\mathrm{Dup}_l = \{v \geq 1 : Y_v = 1 \text{ and } s(v) \leq l\}, \ l \geq 0.$$

Consider the subtree with root the duplicate d, $d \in \mathrm{Dup}_l$. The $\overline{\mathrm{NR}}$-process is a marked BP, therefore the subtree with root d is also a marked BP and we denote this subtree by $\{Z_l^{(d)}, \boldsymbol{M}_l^{(d)}\}_{l \geq 0}$. The offspring distribution of $\{Z_l^{(d)}\}_{l \geq 0}$ is in the first generation given by

$$\hat{f}_n^{(d)} = \mathbb{P}_N(Poi(\Lambda(M_d)) = n | Y_d = 1), \ n \geq 0, \tag{4.A.22}$$

and in the second and further generations by $\{g_n^{(N)}\}_{n \geq 0}$, defined in (4.3.8). Next we describe $\{\boldsymbol{M}_l^{(d)}\}_{l \geq 0}$. The capacity of a duplicate depends on the previous seen marks. The distribution of the mark of a duplicate d is given by

$$\mathbb{P}_N(M = m | Y_d = 1), \text{ for } m = 1, 2, \ldots, N. \tag{4.A.23}$$

Furthermore, the marks of each individual of the progeny of d are, by construction, independent copies of the random variable M, see (4.3.2). The vector

$$\boldsymbol{M}_l^{(d)} = (M_{l,1}^{(d)}, M_{l,2}^{(d)}, \ldots, M_{l,Z_l^{(d)}}^{(d)})$$

denotes the marks of the individuals in generation k of the subtree with root d. Using the above definitions we set $\boldsymbol{M}_0^{(d)} = (M_{0,1}^{(d)})$, where $M_{0,1}^{(d)}$ is an independent copy of the mark given by (4.A.23), and all the other marks are independent copies of the random variable M, see (4.3.2).

By construction,

$$\sum_{k=0}^{l} (\overline{Z}_k - \underline{Z}_k) \leq \sum_{d \in \mathrm{Dup}_l} \sum_{k=0}^{l-s(d)} Z_k^{(d)} = |\mathrm{Dup}_l| + \sum_{d \in \mathrm{Dup}_l} \sum_{i=1}^{l-s(d)} Z_k^{(d)}, \tag{4.A.24}$$

and, similarly,

$$\sum_{k=0}^{l} (\overline{C}_k - \underline{C}_k) \leq \sum_{d \in \mathrm{Dup}_l} \Lambda(M_{0,1}^{(d)}) + \sum_{d \in \mathrm{Dup}_l} \sum_{k=1}^{l-s(d)} \sum_{m=1}^{Z_k^{(d)}} \Lambda(M_{k,m}^{(d)}). \tag{4.A.25}$$

Suppose we can find an event $\mathcal{D}$, with $\mathbb{P}(\mathcal{D}^c) \leq N^{-\beta_2}$, such that the event $\mathcal{D}$, which we will introduce later on, ensures that there are few mismatches between the $\overline{\mathrm{NR}}$-process and $\underline{\mathrm{NR}}$-process. Then the Markov inequality, yields:

$$\mathbb{P}\left(\sum_{k=1}^{l} (\overline{C}_k - \underline{C}_k) > N^{1/2-\alpha_2}\right) \leq N^{-1/2+\alpha_2} \mathbb{E}\left[\sum_{k=1}^{l} (\overline{C}_k - \underline{C}_k)\mathbf{1}_{\mathcal{D}}\right] + N^{-\beta_2}. \tag{4.A.26}$$

According to (4.A.26) it is sufficient to show that there exists a event $\mathcal{D}$ with $\mathbb{P}(\mathcal{D}^c) < N^{-\beta_2}$ such that

$$\mathbb{E}\left[\sum_{k=1}^{l}(\overline{C}_k - \underline{C}_k)\mathbf{1}_\mathcal{D}\right] = \mathcal{O}\left(N^{1/2-\alpha_2-\beta_2}\right). \tag{4.A.27}$$

We will show below that we can bound the capacity of each duplicate by N^γ for some $0 < \gamma < 1/2$ on the event $\mathcal{D}$. This is a very crude bound, but it is sufficient. Furthermore, as the marks in the second and further generations of a duplicate are independent copies of the random variable M, we can bound $\mathbb{E}[\Lambda_M\mathbf{1}_\mathcal{D}]$ by $\overline{S}\underline{S}^{-1}$, where we recall (4.2.7) for the definition of $\underline{S}$ and $\overline{S}$. Then the expected value of (4.A.25) becomes

$$\mathbb{E}\left[\sum_{k=1}^{l}(\overline{C}_k - \underline{C}_k)\mathbf{1}_\mathcal{D}\right] \leq N^\gamma\mathbb{E}[|\mathrm{Dup}_l|\mathbf{1}_\mathcal{D}] + \overline{S}\underline{S}^{-1}\mathbb{E}\left[\sum_{d\in\mathrm{Dup}_l}\sum_{k=1}^{l-s(d)}Z_k^{(d)}\mathbf{1}_\mathcal{D}\right], \tag{4.A.28}$$

Using auxiliary lemmas we will show that

$$\mathbb{E}[|\mathrm{Dup}_l|\mathbf{1}_\mathcal{D}] = \mathcal{O}\left(N^{1/2-\varepsilon-\gamma}\right) \quad\text{and}\quad \mathbb{E}\left[\sum_{d\in\mathrm{Dup}_l}\sum_{k=1}^{l-s(d)}Z_i^{(d)}\mathbf{1}_\mathcal{D}\right] = \mathcal{O}\left(N^{1/2-\varepsilon}\right) \tag{4.A.29}$$

for some $\varepsilon > 0$. Combining (4.A.28) and (4.A.29) yields (4.A.27). We end the outline with a list of all statements that we will prove, which together imply the statement in Proposition 4.A.19. The list consists of the following steps:

1. Define $\mathcal{D}$ and show that $\mathbb{P}(\mathcal{D}^c) \leq N^{-\beta_2}$,

2. Show that $\max_{1\leq m\leq N}\Lambda_m < N^\gamma$, **whp**, and that $\mathbb{E}[\Lambda_M\mathbf{1}_\mathcal{D}] \leq \overline{S}\underline{S}^{-1}$,

3. Show the two statements in (4.A.29).

We start with Step 2. Define $\Lambda_N^{(N)}$ as

$$\Lambda_N^{(N)} = \max_{1\leq i\leq N}\Lambda_i.$$

Then,

Lemma 4.17 *For every $\gamma > 0$,*

$$\mathbb{P}(\Lambda_N^{(N)} \geq N^\gamma) = \mathcal{O}\left(N^{1-(\tau-1)\gamma}\right).$$

Proof. Boole's inequality yields:

$$\mathbb{P}(\Lambda_N^{(N)} \geq N^\gamma) \leq \sum_{i=1}^{N}\mathbb{P}(\Lambda_i \geq N^\gamma) \leq cN^{1-(\tau-1)\gamma}.$$

□

Now, define the event $\mathcal{C}$ as

$$\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3, \tag{4.A.30}$$

where

$$\mathcal{C}_1 = \mathcal{C}_1(\alpha_0) = \left\{ |S_{N,1} - \mu| \leq N^{-\alpha_0} \right\} \cap \left\{ |S_{N,2} - \nu| \leq N^{-\alpha_0} \right\}, \tag{4.A.31}$$

$$\mathcal{C}_2 = \mathcal{C}_2(\gamma) = \left\{ \Lambda_N^{(N)} \leq N^\gamma \right\}, \tag{4.A.32}$$

$$\mathcal{C}_3 = \mathcal{C}_3(\alpha_1) = \left\{ \sum_{n=1}^{\infty} n \, |f_n - f_n^{(N)}| \leq N^{-\alpha_1} \right\} \cap \left\{ \sum_{n=1}^{\infty} n \, |g_n - g_n^{(N)}| \right\}, \tag{4.A.33}$$

where $\gamma \in (1/(\tau - 1), 1/2)$, α_0 and α_1 are taken such that Proposition 4.4 and Proposition 4.14 hold, respectively. Then, according to Lemma 4.17 and the afore mentioned propositions, we have

$$\mathbb{P}(\mathcal{C}^c) \leq \mathbb{P}(\mathcal{C}_1^c) + \mathbb{P}(\mathcal{C}_2^c) + \mathbb{P}(\mathcal{C}_3^c) = \mathcal{O}(N^{-\beta_0}) + \mathcal{O}(N^{1-(\tau-1)\gamma}) + \mathcal{O}(N^{-\beta_1}), \tag{4.A.34}$$

where we can replace the right hand side by $\mathcal{O}(N^{-\varepsilon})$ with $\varepsilon = \min\{\beta_0, (\tau - 1)\gamma - 1, \beta_1\} > 0$. On the event $\mathcal{C}$ it follows from (4.2.7) that **whp**

$$\mathbb{E}_N[\Lambda_M \mathbf{1}_\mathcal{C}] = \frac{1}{L_N} \sum_{m=1}^{N} \Lambda_m^2 \mathbf{1}_\mathcal{C} = \frac{S_{N,2}}{S_{N,1}} \mathbf{1}_\mathcal{C} \leq \overline{S} \underline{S}^{-1}. \tag{4.A.35}$$

Hence, if the event $\mathcal{D}$, which will be defined later, contains $\mathcal{C}$, then Step 2 holds.

To control the size of the set Dup_l we will use the following two lemmas:

Lemma 4.18 *For $\eta, \delta \in (-1/2, 1/2)$ and all $l \leq (1/2 + \eta) \log_\nu N$, and with β_1 given in (A.2.1),*

$$\mathbb{P}\left(\sum_{k=0}^{l} \overline{Z}_k > N^{1/2+\delta} \right) = \mathcal{O}\left((\log_\nu N) N^{-\min\{\delta - \eta, \beta_1\}} \right),$$

and, consequently, if $\{\mathcal{N}_l\}_{l \geq 0}$ are the reachable sets of an uniformly chosen node A in the graph G_N, then

$$\mathbb{P}\left(|\mathcal{N}_l| > N^{1/2+\delta} \right) = \mathcal{O}\left((\log_\nu N) N^{-\min\{\delta - \eta, \beta_1\}} \right).$$

Proof. We have for $l \leq (1/2 + \eta) \log_\nu N$,

$$\mathbb{P}\left(\left\{ \sum_{k=0}^{l} \overline{Z}_k > N^{1/2+\delta} \right\} \cap \mathcal{C}_3 \right) \leq N^{-1/2-\delta} \sum_{k=0}^{k} \mathbb{E}\left[\overline{Z}_k \mathbf{1}_{\mathcal{C}_3} \right]$$

$$\leq (l+1) N^{-1/2-\delta} \max_{1 \leq k \leq l} \mathbb{E}\left[\overline{Z}_k \mathbf{1}_{\mathcal{C}_3} \right]. \tag{4.A.36}$$

Let

$$\mu_N = \sum_{n=1}^{\infty} n f_n^{(N)} \qquad \text{and} \qquad \nu_N = \sum_{n=1}^{\infty} n g_n^{(N)}, \tag{4.A.37}$$

then, by definition, we have $\mathbb{E}_N\big[\overline{Z}_k \mathbf{1}_{\mathcal{C}_3}\big] = \mu_N \nu_N^{k-1} \mathbf{1}_{\mathcal{C}_3}$ for each $k \geq 1$. Furthermore, on the event $\mathcal{C}_3$ we have

$$\nu^{-1} \max\{\nu, \nu_N\} = 1 + \nu^{-1} \max\left\{0, \sum_{n=0}^{\infty} n\left(g_n - g_n^{(N)}\right)\right\} = 1 + \mathcal{O}\big(N^{-\alpha_1}\big) \tag{4.A.38}$$

and, similarly,

$$\mu^{-1} \max\{\mu, \mu_N\} = 1 + \mathcal{O}\big(N^{-\alpha_1}\big). \tag{4.A.39}$$

Thus, on $\mathcal{C}_3$, and for $k \leq l \leq (1/2 + \eta)\log_\nu N$,

$$\mathbb{E}_N\big[\overline{Z}_k\big] = \mu_N \nu_N^{k-1} \leq \mu\nu^{l-1}\left(1 + \mathcal{O}\big(N^{-\alpha_1}\big)\right)^l = \mathcal{O}\big(N^{1/2+\eta}\big), \tag{4.A.40}$$

The above yields,

$$\max_{1 \leq k \leq l} \mathbb{E}\big[\overline{Z}_k \mathbf{1}_{\mathcal{C}_3}\big] = \max_{1 \leq k \leq l} \mathbb{E}\big[\mathbb{E}_N\big[\overline{Z}_k \mathbf{1}_{\mathcal{C}_3}\big]\big] = \mathcal{O}\big(N^{1/2+\eta}\big). \tag{4.A.41}$$

Therefore, if we substitute $(4.A.41)$ in $(4.A.36)$, then we end up with

$$\mathbb{P}\left(\left\{\sum_{k=0}^{l} \overline{Z}_k > N^{1/2+\delta}\right\} \cap \mathcal{C}_3\right) \leq (l+1)N^{-1/2-\delta}\mathcal{O}\big(N^{1/2+\eta}\big),$$

ans observe that the order of the right hand side is $\mathcal{O}\big((\log_\nu N)N^{\eta-\delta}\big)$. Combining this with Proposition 4.14 yields the main statement of the lemma. Finally, using the coupling (4.3.5),

$$\mathbb{P}\big(|\mathcal{N}_l| > N^{1/2+\delta}\big) = \mathbb{P}\left(\sum_{k=0}^{l} \underline{Z}_k > N^{1/2+\delta}\right)$$

$$\leq \mathbb{P}\left(\sum_{k=0}^{l} \overline{Z}_k > N^{1/2+\delta}\right) = \mathcal{O}\big((\log_\nu N)N^{\eta-\delta}\big).$$

$\square$

Lemma 4.19 *There exists a constant $\varepsilon > 0$, such that for each $\delta \in (-1/2, 1/2)$, $u \geq 0$ and $v \leq N^{1/2+\delta}$,*

$$\mathbb{P}\left(\sum_{w=1}^{v} Y_w \geq N^u\right) = \mathcal{O}\big(N^{-\min\{u-2\delta,\varepsilon\}}\big). \tag{4.A.42}$$

Proof. From the definition of Y_v, see (4.A.21), and Boole's inequality,

$$\mathbb{P}_N(Y_v = 1 | M_v = m) = \mathbb{P}_N\left(\cup_{w=0}^{v-1} \{M_v = m\}\right)$$

$$\leq \sum_{w=0}^{v-1} \mathbb{P}_N(M_v = m) = \frac{1}{N} + (v-1)\frac{\Lambda_m}{L_N}.$$

Therefore, using (4.2.7) and $v \leq N^{1/2+\delta}$, we have, on the event $\mathcal{C}$, that

$$\mathbb{P}_N(Y_v = 1) = \frac{1}{N}\sum_{m=1}^{N}\frac{\Lambda_m}{L_N} + (v-1)\sum_{m=1}^{N}\frac{\Lambda_m^2}{L_N^2} = \frac{1}{N} + \frac{v-1}{N}\frac{S_{N,2}}{S_{N,1}^2}$$

$$\leq \frac{1}{N} + \overline{S}\,\underline{S}^{-2}N^{\delta-1/2} = \mathcal{O}\left(N^{\delta-1/2}\right).$$

From the Markov inequality, we hence deduce, on the event $\mathcal{C}$, that

$$\mathbb{P}_N\left(\sum_{w=1}^{v}Y_w \geq N^u\right) \leq N^{-u}\sum_{w=1}^{v}\mathbb{P}_N(Y_w = 1) \leq v\mathcal{O}\left(N^{\delta-1/2-u}\right) = \mathcal{O}(N^{2\delta-u}).$$

Thus,

$$\mathbb{P}\left(\left\{\sum_{w=1}^{v}Y_w \geq N^u\right\}\cap\mathcal{C}\right) = \mathbb{E}\left[\mathbb{P}_N\left(\sum_{w=1}^{v}Y_w \geq N^u\right)\mathbf{1}_\mathcal{C}\right] + \mathbb{P}(\mathcal{C}^c)$$

$$= \mathcal{O}\left(N^{-\min\{u-2\delta,\varepsilon\}}\right),$$

which implies the claim (4.A.42). $\qquad\square$

Proof of Proposition 4.16. Define the event $\mathcal{D} = \mathcal{D}_1 \cap \mathcal{D}_2 \cap \mathcal{D}_3 \cap \mathcal{C}$, where

$$\mathcal{D}_1 = \left\{\mathrm{Dup}_{\lfloor(1/2-2\eta)\log_\nu N\rfloor} = \emptyset\right\}, \tag{4.A.43}$$

$$\mathcal{D}_2 = \left\{\left|\mathrm{Dup}_{\lfloor(1/2+\eta)\log_\nu N\rfloor}\right| \leq N^{5\eta}\right\}, \tag{4.A.44}$$

$$\mathcal{D}_3 = \bigcap_{d\in\mathrm{Dup}_{\lfloor(1/2+\eta)\log_\nu N\rfloor}}\left\{Z_1^{(d)} < N^\theta\right\} \tag{4.A.45}$$

and $\theta = 6\eta + \gamma$.

As Step 2 has been shown above, it remains to prove $\mathbb{P}(\mathcal{D}^c) \leq N^{-\beta_2}$, with $\mathcal{D}$ as defined above (Step 1), and the two statements in (4.A.29) (Step 3).

We start with the first statement in (4.A.29). We have,

$$\mathbb{E}[|\mathrm{Dup}_l|\mathbf{1}_\mathcal{D}] \leq \mathbb{E}\left[|\mathrm{Dup}_{\lfloor(1/2+\eta)\log_\nu N\rfloor}|\mathbf{1}_\mathcal{D}\right] \leq N^{5\eta} \leq N^{1/2-\varepsilon-\gamma}, \tag{4.A.46}$$

by choosing η and ε small and $\gamma \in (1/(\tau-1), 1/2)$ appropriately.

The second statement in (4.A.29) proceeds in the following way. On the event $\mathcal{D}_1 \cap \mathcal{D}_2$ duplicates do not appear in the first $(1/2-2\eta)\log_\nu N$ generations, implying that,

$$\max_{d\in\mathrm{Dup}_l}(l - s(d)) \leq l - (1/2 - 2\eta)\log_\nu N \leq 3\eta\log_\nu N. \tag{4.A.47}$$

Thus,

$$\mathbb{E}\left[\sum_{d\in\mathrm{Dup}_l}\sum_{k=1}^{l-s(d)}Z_k^{(d)}\mathbf{1}_{\mathcal{D}}\right]\leq\mathbb{E}\left[\sum_{d\in\mathrm{Dup}_l}\sum_{k=1}^{\lfloor 3\eta\log_\nu N\rfloor}Z_k^{(d)}\mathbf{1}_{\mathcal{D}}\right]. \qquad (4.A.48)$$

The right hand side is the expected size of the progeny of the duplicates and its offspring. The total children of all the duplicates on $\mathcal{D}_3$ is bounded from above by the the total number of duplicates times N^θ. Since on $\mathcal{D}_2$ the total number of duplicates is bounded from above by $N^{5\eta}$, therefore the total number of children of all the duplicates is bounded from above by $N^{5\eta+\theta}$ on $\mathcal{D}_2\cap\mathcal{D}_3$. Furthermore, the offspring distribution of each child of a duplicate is an independent copy of $\{\hat{Z}_l\}_{l\geq 0}$, where $\{\hat{Z}_l\}_{l\geq 0}$ is a BP with $\hat{Z}_0=1$ and where each individual has offspring distribution $\{g_n^{(N)}\}_{n\geq 0}$. Since, the mark of a child of a duplicate is by definition an independent copy of the random variable M.

Therefore, (4.A.48) can be bounded by

$$\mathbb{E}\left[\sum_{d\in\mathrm{Dup}_l}\sum_{k=1}^{l-s(d)}Z_k^{(d)}\mathbf{1}_{\mathcal{D}}\right]\leq N^{5\eta+\theta}+N^{5\eta+\theta}\mathbb{E}\left[\sum_{k=1}^{\lfloor 3\eta\log_\nu N\rfloor}\hat{Z}_k\mathbf{1}_{\mathcal{C}}\right]$$

On $\mathcal{C}$ we bound $\sum_{k=1}^{\lfloor 3\eta\log_\nu N\rfloor}\hat{Z}_k$ by

$$\sum_{k=1}^{\lfloor 3\eta(\log_\nu N)\rfloor}\hat{Z}_k\leq 3\eta(\log_\nu N)\nu_N^{3\eta\log_\nu N}=\mathcal{O}\big((\log_\nu N)N^{3\eta}\big)=\mathcal{O}\big(N^{4\eta}\big),$$

compare with (4.A.40) and (4.A.41). Thus, we bound (4.A.48) by

$$\mathbb{E}\left[\sum_{d\in\mathrm{Dup}_l}\sum_{k=1}^{l-s(d)}Z_k^{(d)}\mathbf{1}_{\mathcal{D}}\right]\leq N^{5\eta+\theta}+N^{5\eta+\theta}\mathcal{O}\big(N^{4\eta}\big)=\mathcal{O}\big(N^{\theta+9\eta}\big)=\mathcal{O}\big(N^{15\eta+\gamma}\big),$$

which is more than sufficient for the second statement of (4.A.29), since $\gamma<1/2$ and we can pick η arbitrarily small. This completes the proof of Step 3.

We continue with the proof of Step 1. For this, we bound

$$\mathbb{P}(\mathcal{D}^c)\leq\mathbb{P}(\mathcal{D}_1^c)+\mathbb{P}(\mathcal{D}_2^c)+\mathbb{P}(\mathcal{D}_3^c\cap\mathcal{D}_2\cap\mathcal{C}). \qquad (4.A.49)$$

We now bound these terms one by one. For $\mathbb{P}(\mathcal{D}_1^c)$, we use Lemmas 4.18 and 4.19 to obtain

$$\mathbb{P}(\mathcal{D}_1^c)\leq\mathbb{P}\left(\mathcal{D}_1^c\cap\left\{\sum_{k=0}^{\lfloor(1/2-2\eta)\log_\nu N\rfloor}\overline{Z}_k<N^{1/2-\eta}\right\}\right)+\mathcal{O}\Big((\log_\nu N)N^{-\min\{\eta,\beta_1\}}\Big)$$

$$\leq\mathbb{P}\left(\sum_{w=1}^{\lfloor N^{1/2-\eta}\rfloor}Y_w>0\right)+\mathcal{O}\Big(N^{-\min\{\eta/2,\beta_1/2\}}\Big)=\mathcal{O}\Big(N^{-\min\{\eta/2,\beta_1/2,\varepsilon\}}\Big)$$

$$(4.A.50)$$

and, similarly,

$$
\begin{aligned}
\mathbb{P}(\mathcal{D}_2^c) &\leq \mathbb{P}\left(\mathcal{D}_2^c \cap \left\{\sum_{k=0}^{\lfloor(1/2+\eta)\log_\nu N\rfloor} \overline{Z}_k < N^{1/2+2\eta}\right\}\right) + \mathcal{O}\Big((\log_\nu N)N^{-\min\{\eta,\beta_1\}}\Big) \\
&\leq \mathbb{P}\left(\sum_{w=1}^{\lfloor N^{1/2+2\eta}\rfloor} Y_w > N^{5\eta}\right) + \mathcal{O}\Big(N^{-\min\{\eta/2,\beta_1/2\}}\Big) \\
&= \mathcal{O}\Big(N^{-\min\{\eta/2,\beta_1/2,\varepsilon\}}\Big).
\end{aligned} \tag{4.A.51}
$$

To bound $\mathbb{P}(\mathcal{D}_3^c \cap \mathcal{D}_2 \cap \mathcal{C})$, we first bound $\mathbb{P}_N(\mathcal{D}_3^c \cap \mathcal{D}_2)$ on the event $\mathcal{C}$. Using the Markov inequality and Boole's inequality we have, on $\mathcal{C}$, that

$$
\mathbb{P}_N(\mathcal{D}_3^c \cap \mathcal{D}_2) \leq N^{-\theta}\mathbb{E}_N\left[\sum_{d\in\mathrm{Dup}_{\lfloor(1/2+\eta)\log_\nu N\rfloor}} Z_1^{(d)}\mathbf{1}_{\mathcal{D}_2}\right] = N^{-\theta}\mathbb{E}_N\left[\sum_{d\in\mathrm{Dup}_k} Z_1^{(d)}\mathbf{1}_{\mathcal{D}_2}\right], \tag{4.A.52}
$$

where we, for convenience, set $k = \lfloor(1/2+\eta)\log_\nu N\rfloor$. If we condition on the sequences $\{\Lambda_i\}_{i=1}^N$ and $\{M_v\}_{v\geq 0}$, recall (4.A.20), then $Z_1^{(d)}$, for $d\in\mathrm{Dup}_k$, is an independent Poisson random variable with mean $\Lambda(M_d)$. Thus, conditioned on the sequences $\{\Lambda_i\}_{i=1}^N$ and $\{M_v\}_{v\geq 0}$, the sum $\sum_{d\in\mathrm{Dup}_k} Z_1^{(d)}$ is distributed as a Poisson random variable with mean $\sum_{d\in\mathrm{Dup}_k} M_d$. Which in turn can be stochastically bounded from above by a Poisson random variable with mean $|\mathrm{Dup}_k|N^\gamma$ on the event $\mathcal{C}$, because $\Lambda(M_d) \leq \Lambda_N^{(N)} \leq N^\gamma$ for all $d\in\mathrm{Dup}_k$. Since $\mathcal{D}_2$ and Dup_k are measurable with respect to $\{M_v\}_{v\geq 0}$, therefore the above yields

$$
\begin{aligned}
\mathbb{E}_N\left[\sum_{d\in\mathrm{Dup}_k} Z_1^{(d)}\mathbf{1}_{\mathcal{D}_2}\right] &= \mathbb{E}_N\left[\sum_{d\in\mathrm{Dup}_k} \mathbb{E}_N\big[Poi\big(Z_1^{(d)}\big)\big|\{M_v\}_{v\geq 0}\big]\mathbf{1}_{\mathcal{D}_2}\right] \\
&= \mathbb{E}_N\left[\mathbb{E}_N\left[Poi\left(\sum_{d\in\mathrm{Dup}_k} Z_1^{(d)}\right)\bigg|\{M_v\}_{v\geq 0}\right]\mathbf{1}_{\mathcal{D}_2}\right] \\
&\leq \mathbb{E}_N[\mathbb{E}_N[Poi(|\mathrm{Dup}_k|N^\gamma)|\{M_v\}_{v\geq 0}]\mathbf{1}_{\mathcal{D}_2}] \\
&\leq \mathbb{E}_N\big[\mathbb{E}_N\big[Poi\big(N^{5\eta+\gamma}\big)\big|\{M_v\}_{v\geq 0}\big]\big] \\
&= \mathbb{E}_N\big[Poi\big(N^{5\eta+\gamma}\big)\big] = N^{5\eta+\gamma}.
\end{aligned} \tag{4.A.53}
$$

Thus, we bound (4.A.52) by

$$
\mathbb{P}_N(\mathcal{D}_3^c \cap \mathcal{D}_2) \leq N^{-\theta}\mathbb{E}_N\left[\sum_{d\in\mathrm{Dup}_{\lfloor(1/2+\eta)\log_\nu N\rfloor}} Z_1^{(d)}\mathbf{1}_{\mathcal{D}_2}\right] \leq N^{-\theta+5\eta+\gamma} = N^{-\eta}. \tag{4.A.54}
$$

Hence,

$$
\mathbb{P}(\mathcal{D}_3^2 \cap \mathcal{D}_2 \cap \mathcal{C}) = \mathbb{E}[\mathbf{1}_{\mathcal{C}}\mathbb{P}_N(\mathcal{D}_3^c \cap \mathcal{D}_2)] = \mathcal{O}\Big(N^{-\min\{\eta/2,\beta_1/2,\varepsilon\}}\Big) \tag{4.A.55}
$$

and, finally, by (4.A.50), (4.A.51), (4.A.55) and (4.A.49),

$$\mathbb{P}(\mathcal{D}^c) = \mathcal{O}\Big(N^{-\min\{\eta/2,\beta_1/2,\varepsilon\}}\Big),$$

which proves Step 1 and therefore the proposition. $\qquad\square$

4.A.4   Proof of Proposition 4.12

In this section we prove Proposition 4.12, which we restate here for convenience as Proposition 4.20.

Proposition 4.20 *There exist constants $\alpha_3, \beta_3, \eta > 0$ such that for all $l \le (1 + 2\eta)\log_\nu N$, as $N \to \infty$,*

$$\mathbb{P}\Big(\frac{1}{N}\Big|\sum_{k=2}^{l+1}\mathcal{Z}^{(1)}_{\lceil k/2\rceil}\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor} - \sum_{k=2}^{l+1}\underline{C}^{(1)}_{\lceil k/2\rceil}\underline{C}^{(2)}_{\lfloor k/2\rfloor}\Big| > N^{-\alpha_3}\Big) = \mathcal{O}\big(N^{-\beta_3}\big). \quad (4.A.56)$$

We start with an outline of the proof. Define

$$T_1 = \frac{1}{N}\sum_{k=2}^{l+1}\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor}\big|\mathcal{Z}^{(1)}_{\lceil k/2\rceil} - \underline{C}^{(1)}_{\lceil k/2\rceil}\big| \quad \text{and} \quad T_2 = \frac{1}{N}\sum_{k=2}^{l+1}\underline{C}^{(1)}_{\lceil k/2\rceil}\big|\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor} - \underline{C}^{(2)}_{\lfloor k/2\rfloor}\big|. \tag{4.A.57}$$

To show (4.A.56), it suffices to prove that for an appropriate chosen event $\mathcal{H}$,

$$\mathbb{P}\Big(T_1\mathbf{1}_\mathcal{H} > \frac{1}{2}N^{-\alpha_3}\Big) = \mathcal{O}\big(N^{-\beta_3}\big) \quad \text{and} \quad \mathbb{P}\Big(T_2\mathbf{1}_\mathcal{H} > \frac{1}{2}N^{-\alpha_3}\Big) = \mathcal{O}\big(N^{-\beta_3}\big), \tag{4.A.58}$$

and that $\mathbb{P}(\mathcal{H}^c) = \mathcal{O}\big(N^{-\xi}\big)$ for some $\xi > 0$. We choose the event $\mathcal{H}$ such that on this event, for each $k \le l \le (1 + 2\eta)\log_\nu N$,

$$\mathbb{E}\big[\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor}\mathbf{1}_\mathcal{H}\big] = \mathcal{O}\Big(N^{1/2+2\eta}\Big), \qquad \mathbb{E}\big[\big|\mathcal{Z}^{(1)}_{\lceil k/2\rceil} - \underline{C}^{(1)}_{\lceil k/2\rceil}\big|\mathbf{1}_\mathcal{H}\big] = \mathcal{O}\Big(N^{1/2-u}\Big), \tag{4.A.59}$$

$$\mathbb{E}\big[\underline{C}^{(1)}_{\lceil k/2\rceil}\mathbf{1}_\mathcal{H}\big] = \mathcal{O}\Big(N^{1/2+2\eta}\Big) \quad \text{and} \quad \mathbb{E}\big[\big|\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor} - \underline{C}^{(2)}_{\lfloor k/2\rfloor}\big|\mathbf{1}_\mathcal{H}\big] = \mathcal{O}\Big(N^{1/2-u}\Big), \tag{4.A.60}$$

for some $u, \alpha_3, \eta > 0$ such that

$$\alpha_3 < u - 2\eta. \tag{4.A.61}$$

The claims (4.A.59), (4.A.60) and (4.A.61) implies (4.A.58). This is straightforward from the Markov inequality, the independence of $\mathcal{Z}^{(1)}$ between $\mathcal{Z}^{(2)}$ and the associated capacities and the inequality $l \le (1 + 2\eta)\log_\nu N$. We leave it to the reader to verify this.

Instead of showing (4.A.59) and (4.A.60), it is sufficient to show that there exists constants $\alpha_3, \eta, u > 0$ satisfying (4.A.61) and such that for each $k \leq l \leq (1/2 + \eta) \log_\nu N$,

$$\mathbb{E}[|\mathcal{Z}_k - \underline{C}_k|\, \mathbf{1}_\mathcal{H}] = \mathcal{O}\left(N^{1/2-u}\right), \qquad \mathbb{E}[\mathcal{Z}_k \mathbf{1}_\mathcal{H}] = \mathcal{O}\left(N^{1/2+2\eta}\right)$$

and

$$\mathbb{E}[\underline{C}_k \mathbf{1}_\mathcal{H}] = \mathcal{O}\left(N^{1/2+2\eta}\right), \tag{4.A.62}$$

which simplifies the notation considerably.

We choose the event $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{H}_3 \cap \mathcal{H}_4 \cap \mathcal{C}$, where

$$\mathcal{H}_1 = \left\{ \sum_{k=1}^l (\overline{C}_k - \underline{C}_k) \leq N^{1/2-\alpha_2} \right\}, \tag{4.A.63}$$

$$\mathcal{H}_{2,k} = \left\{ \left| \overline{C}_k - \mathbb{E}_N[\Lambda_M]\, \overline{Z}_{k-1} \right| < N^{1/2-\delta} \right\}, \qquad \mathcal{H}_2 = \cap_{k=1}^l \mathcal{H}_{2,k}, \tag{4.A.64}$$

$$\mathcal{H}_{3,k} = \left\{ \left| \mathcal{Z}_k - \nu \mathcal{Z}_{k-1} \right| < N^{1/2-\delta} \right\}, \qquad \mathcal{H}_3 = \cap_{k=1}^l \mathcal{H}_{3,k}, \tag{4.A.65}$$

$$\mathcal{H}_4 = \left\{ \left| \mathbb{E}_N[\Lambda_M] - \nu \right| \leq N^{-\delta} \right\}, \tag{4.A.66}$$

and the event $\mathcal{C}$ is given by (4.A.30). Hence, for the proof of Proposition 4.20 it suffices to show the three claims in (4.A.62) with $u - 2\eta > 0$, and that for some $\xi > 0$ we have that $\mathbb{P}(\mathcal{H}^c) = \mathcal{O}(N^{-\xi})$. For the latter statement, we use the following lemma. In this lemma we denote by d_N a random variable with distribution $\{g_n^{(N)}\}_{n \geq 0}$ given by (4.3.8).

Lemma 4.21 *For every $\gamma > 0$, for $\{\Lambda_i\}_{i=1}^N$, such that $\mathcal{C}$ occurs,*

$$Var_N(\Lambda_M) = \mathcal{O}(N^\gamma) \qquad \text{and} \qquad Var_N(d_N) = \mathcal{O}(N^\gamma), \tag{4.A.67}$$

where $Var_N(\cdot)$ is the variance under $\mathbb{P}_N(\cdot)$.

The proof of this lemma is deferred to the end of this section. Although it seems that we can take any $\gamma > 0$, this will not be the case in the proof of Proposition 4.20. We need that the event $\mathcal{C}$ happens **whp** , which is the case when $\gamma > 1/(\tau - 1)$.

Proof of Proposition 4.20 Consider the first claim given by (4.A.62). Using the triangle inequality we arrive at

$$\mathbb{E}[|\underline{C}_k - \mathcal{Z}_k|\, \mathbf{1}_\mathcal{H}] \leq \mathbb{E}\left[\left| \underline{C}_k - \overline{C}_k \right| \mathbf{1}_\mathcal{H} \right] + \mathbb{E}\left[\left| \overline{C}_k - \mathbb{E}_N[\Lambda_M]\, \overline{Z}_{k-1} \right| \mathbf{1}_\mathcal{H} \right]$$
$$+ \mathbb{E}\left[\left| \overline{Z}_{k-1}(\mathbb{E}_N[\Lambda_M] - \nu) \right| \mathbf{1}_\mathcal{H} \right] + \nu \mathbb{E}\left[\left| \overline{Z}_{k-1} - \mathcal{Z}_{k-1} \right| \mathbf{1}_\mathcal{H} \right] + \mathbb{E}[|\nu \mathcal{Z}_{k-1} - \mathcal{Z}_k|\, \mathbf{1}_\mathcal{H}], \tag{4.A.68}$$

The first, second and the last term on the right hand of (4.A.68), we bound by $N^{1/2-\min\{\delta,\alpha_2\}}$ using the events $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$, respectively. We bound the third term of (4.A.68) using the event $\mathcal{H}_4$ and (4.A.40), which gives

$$\mathbb{E}\left[\left| \overline{Z}_{k-1}(\mathbb{E}_N[\Lambda_M] - \nu) \right| \mathbf{1}_{\mathcal{H}_4 \cap \mathcal{C}} \right] \leq N^{-\delta} \mathbb{E}\left[\left| \overline{Z}_{k-1} \mathbf{1}_\mathcal{C} \right| \right] = \mathcal{O}\left(N^{1/2+2\eta-\delta}\right). \tag{4.A.69}$$

Finally, we need to bound the fourth term of (4.A.68). On the event $\mathcal{C}_3$ the following inequality holds for each $k \leq l$ (compare with [54, (A.1.4) and (A.1.15)]),

$$\mathbb{E}_N\big[\big|\overline{Z}_k - \mathcal{Z}_k\big|\big] \leq \max\left\{\nu - \alpha_N, \nu_N - \alpha_N\right\} \sum_{k=1}^{l} \mathbb{E}_N\big[\overline{Z}_k\big]\left(\max\left\{\nu, \nu_N\right\}\right)^{l-k}, \quad (4.A.70)$$

where

$$\alpha_N = \sum_{n=0}^{\infty} n \min\{g_n, g_n^{(N)}\}$$

and ν_N is given by (4.A.37). We bound the sum in (4.A.70), using (4.A.40), which implies

$$\begin{aligned}
\sum_{k=1}^{l} \mathbb{E}_N\big[\overline{Z}_k\big]\left(\max\left\{\nu, \nu_N\right\}\right)^{l-k} &= \sum_{k=1}^{l} \mu_N \nu_N^{m-1}\left(\max\left\{\nu, \nu_N\right\}\right)^{l-k} \\
&\leq l \max\left\{\mu, \mu_N\right\}\left(\max\left\{\nu, \nu_N\right\}\right)^{l-1} \\
&\leq l\left(1 + \mathcal{O}(N^{-\alpha_1})\right)^l \mathcal{O}\left(N^{1/2+\eta}\right) \leq N^{1/2+2\eta}
\end{aligned}$$
$$(4.A.71)$$

for N sufficiently large. On the event $\mathcal{C}_3$ we have

$$\max\left\{\nu - \alpha_N, \nu_N - \alpha_N\right\} \leq \sum_{n=1}^{\infty} n\left|g_n - g_n^{(N)}\right| \leq N^{-\alpha_1}. \quad (4.A.72)$$

Combining (4.A.70), (4.A.71) and (4.A.72), we obtain that, on $\mathcal{C}_3$,

$$\mathbb{E}_N\big[\big|\overline{Z}_k - \mathcal{Z}_k\big|\big] = \mathcal{O}\left(N^{1/2+2\eta-\alpha_1}\right).$$

Thus, all together, the left side of (4.A.68) satisfies

$$\mathbb{E}[|\underline{C}_k - \mathcal{Z}_k|\, \mathbf{1}_{\mathcal{H}}] = \mathcal{O}\left(N^{1/2-\min\left\{\delta, \alpha_2, \delta - 2\eta, \alpha_1 - 2\eta\right\}}\right). \quad (4.A.73)$$

This gives a bound for the right hand of (4.A.62) with

$$u = \min\left\{\delta, \alpha_2, \delta - 2\eta, \alpha_1 - 2\eta\right\}.$$

The second claim of (4.A.62) is evident, because $k \leq l \leq (1/2 + \eta)\log_\nu N$ and therefore

$$\mathbb{E}[\mathcal{Z}_k \mathbf{1}_{\mathcal{H}}] \leq \mathbb{E}[\mathcal{Z}_k] = \mu\nu^{k-1} \leq \mu\nu^{l-1} = \mathcal{O}\left(N^{1/2+\eta}\right). \quad (4.A.74)$$

Finally, the third claim of (4.A.62) follows from using in turn $\underline{C}_k \leq \overline{C}_k$,

$$\mathbb{E}[\underline{C}_k \mathbf{1}_{\mathcal{H}}] \leq \mathbb{E}\big[\overline{C}_k \mathbf{1}_{\mathcal{H}}\big] = \mathbb{E}\left[\sum_{v=1}^{\overline{Z}_{k-1}} \Lambda\big(\overline{M}_{k-1,v}\big)\mathbf{1}_{\mathcal{H}}\right].$$

Now by taking conditional expectation with respect to $\overline{Z}_{i-1}$ and the capacities, we obtain

$$
\mathbb{E}_N\left[\sum_{v=1}^{\overline{Z}_{k-1}} \Lambda\big(\overline{M}_{k-1,v}\big)\,\mathbf{1}_{\mathcal{H}}\,\bigg|\,\overline{Z}_{k-1}\right] = \sum_{v=1}^{\overline{Z}_{k-1}} \mathbb{E}_N\big[\Lambda\big(\overline{M}_{k-1,v}\big)\,\mathbf{1}_{\mathcal{H}}\,\big|\,\overline{Z}_{k-1}\big]
$$
$$
\leq \sum_{v=1}^{\overline{Z}_{k-1}} \mathbb{E}_N\big[\Lambda\big(\overline{M}_{k-1,v}\big)\,\mathbf{1}_{\mathcal{C}}\,\big|\,\overline{Z}_{k-1}\big]
$$
$$
= \overline{Z}_{k-1}\mathbb{E}_N[\Lambda_M\mathbf{1}_{\mathcal{C}}] \leq \overline{S}\,\underline{S}\,\overline{Z}_{k-1}\mathbf{1}_{\mathcal{C}}
$$

so that (4.A.40) implies that

$$
\mathbb{E}[\underline{C}_k\mathbf{1}_{\mathcal{H}}] \leq \mathbb{E}\left[\mathbb{E}_N\left[\sum_{v=1}^{\overline{Z}_{k-1}} \Lambda\big(\overline{M}_{k-1,v}\big)\,\mathbf{1}_{\mathcal{H}}\,\bigg|\,\overline{Z}_{k-1}\right]\right] \leq \overline{S}\,\underline{S}\,\mathbb{E}\big[\overline{Z}_{k-1}\mathbf{1}_{\mathcal{C}}\big] = \mathcal{O}\big(N^{1/2+2\eta}\big).
$$

Thus, we have shown the claims given by (4.A.62) if we restrict η to $0 < \eta < \min\{\alpha_1,\alpha_2\}/2$ due to (4.A.73). To satisfy condition (4.A.61), we restrict η to

$$
0 < \eta < \min\{\alpha_1/4, \alpha_2/2, \delta/2\}, \tag{4.A.75}
$$

and pick $\alpha_3 = (u-2\eta)/2$. Then $\alpha_3 > 0$ and condition (4.A.61) is satisfied, because

$$
u - 2\eta < \alpha_3 = (u-2\eta)/2 = \min\{\alpha_1 - 4\eta, \alpha_2 - 2\eta, \delta - 2\eta\}/2 > 0,
$$

for each $0 < \delta < 1/2$ and for each η given by (4.A.75).

We finish the proof by showing that $\mathbb{P}(\mathcal{H}^c) = \mathcal{O}\big(N^{-\xi}\big)$ for some $\xi > 0$. We bound $\mathbb{P}(\mathcal{H}^c)$ by

$$
\mathbb{P}(\mathcal{H}^c) \leq \mathbb{P}(\mathcal{H}_1^c) + \mathbb{P}(\mathcal{H}_2^c \cap \mathcal{C}) + \mathbb{P}(\mathcal{H}_3^c \cap \mathcal{C}) + \mathbb{P}(\mathcal{H}_4^c \cap \mathcal{C}) + \mathbb{P}(\mathcal{C}^c). \tag{4.A.76}
$$

The last term we bound by (4.A.34), which states $\mathbb{P}(\mathcal{C}^c) = \mathcal{O}(N^{-\varepsilon})$ for some $\varepsilon > 0$. We bound $\mathbb{P}(\mathcal{H}_1^c)$ using Proposition 4.16,

$$
\mathbb{P}(\mathcal{H}_1^c) = \mathcal{O}\big(N^{-\beta_2}\big). \tag{4.A.77}
$$

Next, we will bound $\mathbb{P}(\mathcal{H}_2^c \cap \mathcal{C})$ and $\mathbb{P}(\mathcal{H}_3^c \cap \mathcal{C})$ from above. We will show that by using the Chebychev inequality that

$$
\mathbb{P}_N\big(\mathcal{H}_{2,k}^c \cap \mathcal{C}\big) = \mathcal{O}\big(N^{-1/2+2\delta+\gamma+2\eta}\big), \text{ for } k \leq l \leq (1/2+\eta)\log_\nu N. \tag{4.A.78}
$$

Then by Boole's inequality we have,

$$
\mathbb{P}(\mathcal{H}_2^c) \leq \mathbb{P}\left(\bigcup_{k=0}^{l} \mathcal{H}_{2,k}^c \cap \mathcal{C}\right) + \mathbb{P}(\mathcal{C}^c) \leq \sum_{k=0}^{l} \mathbb{E}\big[\mathbb{P}_N\big(\mathcal{H}_{2,k}^c \cap \mathcal{C}\big)\big] + \mathbb{P}(\mathcal{C}^c)
$$
$$
\leq lN^{-1/2+2\delta+\gamma+2\eta} + \mathbb{P}(\mathcal{C}^c) = \mathcal{O}\Big((\log_\nu N)N^{-\min\{1/2-2\delta-\gamma-2\eta,\eta/2,\varepsilon\}}\Big). \tag{4.A.79}
$$

Fix k, $1 \leq k \leq l \leq (1/2 + \eta) \log_\nu N$, and denote by $V_v = \Lambda\big(\overline{M}_{k-1,v}\big) - \mathbb{E}_N[\Lambda_M]$ for $v = 1, 2, \ldots, \overline{Z}_{k-1}$. We have that $\mathbb{E}_N[V_v] = 0$, and conditionally on the marks $\Lambda_1, \ldots, \Lambda_N$ the sequence $\{V_v\}_{v=1}^{\overline{Z}_{k-1}}$ is an i.i.d. sequence. Hence, on $\mathcal{C}$ and using Lemma 4.21,

$$
\mathbb{E}_N\left[\left.\left(\sum_{v=1}^{\overline{Z}_{k-1}} V_v\right)^2 \right| \overline{Z}_{k-1}\right] = \sum_{v=1}^{\overline{Z}_{k-1}} \sum_{w=1}^{\overline{Z}_{k-1}} \mathbb{E}_N[V_v V_w]
$$

$$
= \sum_{v=1}^{\overline{Z}_{k-1}} \sum_{w=1,w\neq v}^{\overline{Z}_{k-1}} \mathbb{E}_N[V_v]\,\mathbb{E}_N[V_w] + \sum_{v=1}^{\overline{Z}_{k-1}} \mathbb{E}_N\big[V_v^2\big]
$$

$$
= \sum_{v=1}^{\overline{Z}_{k-1}} \mathbb{E}_N\big[V_v^2\big] = \overline{Z}_{k-1}\mathrm{Var}_N(\Lambda_M) \leq \overline{Z}_{k-1}N^\gamma. \quad (4.\mathrm{A}.80)
$$

Therefore, using (4.A.40), on $\mathcal{C}$,

$$
\mathrm{Var}_N\left(\sum_{v=1}^{\overline{Z}_{k-1}} V_v\right) = \mathbb{E}_N\left[\mathbb{E}_N\left[\left.\left(\sum_{v=1}^{\overline{Z}_{k-1}} V_v\right)^2 \right| \overline{Z}_{k-1}\right]\right]
$$

$$
\leq \mathbb{E}_N\big[\overline{Z}_{k-1}\big]\, N^\gamma = \mathcal{O}\Big(N^{1/2+\gamma+2\eta}\Big).
$$

Thus, by the Chebyshev inequality, on $\mathcal{C}$,

$$
\mathbb{P}_N\big(\mathcal{H}_{2,k}^c\big) = \mathbb{P}_N\left(\left|\sum_{v=1}^{\overline{Z}_{k-1}} V_v\right| \geq N^{1/2-\delta}\right)
$$

$$
\leq N^{2\delta-1}\mathrm{Var}_N\left(\sum_{v=1}^{\overline{Z}_{k-1}} V_v\right) \leq N^{-1/2+2\delta+\gamma+2\eta}.
$$

Similarly, we can show that

$$
\mathbb{P}\big(\mathcal{H}_{3,k}^c \cap \mathcal{C}\big) = \mathcal{O}\Big((\log_\nu N)N^{-\min\{1/2-2\delta-\gamma-2\eta,\eta/2\}}\Big),
$$

when we replace $\overline{Z}_{k-1}$ by $\mathcal{Z}_{k-1}$, set $V_n = X_{k-1,n} - \mathbb{E}_N[d_N]$, where $X_{k-1,n}$ is an independent copy of d_N. This, yields on $\mathcal{C}$,

$$
\mathbb{P}_N\big(\mathcal{H}_{3,k}^c\big) = \mathbb{P}_N\left(\left|\sum_{v=1}^{\mathcal{Z}_{k-1}} V_v\right| \geq N^{1/2-\delta}\right) \leq N^{2\delta-1}\mathbb{E}_N\left[\mathbb{E}_N\left[\left.\left(\sum_{v=1}^{\mathcal{Z}_{k-1}} V_v\right)^2 \right| \mathcal{Z}_{k-1}\right]\right]
$$

$$
= N^{2\delta-1}\mathbb{E}_N[\mathcal{Z}_{k-1}\mathrm{Var}_N(d_N)].
$$

Using that $\mathrm{Var}_N(d_N)$ is constant under $\mathbb{P}_N$, Lemma 4.18 and (4.A.74) we have that on $\mathcal{C}$:

$$
\mathbb{E}_N[\mathcal{Z}_{k-1}\mathrm{Var}_N(d_N)] = \mathrm{Var}_N(d_N)\,\mathbb{E}_N[\mathcal{Z}_{k-1}] = \mathcal{O}\Big(N^{1/2+\eta+\gamma}\Big).
$$

The above yields, compare with (4.A.79),

$$\mathbb{P}(\mathcal{H}_3^c) = \mathcal{O}\Big((\log_\nu N)N^{-\min\{1/2-2\delta-\gamma-2\eta,\eta/2,\varepsilon\}}\Big). \qquad (4.A.81)$$

To bound $\mathbb{P}(\mathcal{H}_4^c)$ we will use that $\mathbb{E}_N[\Lambda_M] = S_{N,2}/S_{N,1}$, where $S_{N,q} = \frac{1}{N}\sum_{i=1}^N \Lambda_i^q$, $q = 1,2$. Notice that on $\mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0}$, recall (4.2.4), that

$$\big(1 - \mathcal{O}(N^{-\alpha_0})\big) \le \frac{S_{N,1}}{\mathbb{E}[\Lambda]} \le \big(1 + \mathcal{O}(N^{-\alpha_0})\big)$$

and

$$\big(1 - \mathcal{O}(N^{-\alpha_0})\big) \le \frac{S_{N,2}}{\mathbb{E}[\Lambda^2]} \le \big(1 + \mathcal{O}(N^{-\alpha_0})\big),$$

which yields, together with $\nu = \mathbb{E}[\Lambda^2]/\mathbb{E}[\Lambda]$

$$\big(1 - \mathcal{O}(N^{-\alpha_0})\big)\nu \le \frac{S_{N,2}}{S_{N,1}} \le \big(1 + \mathcal{O}(N^{-\alpha_0})\big)\nu. \qquad (4.A.82)$$

Thus, using Lemma 4.4, the Markov inequality and $(4.A.82)$ we have

$$\begin{aligned}
\mathbb{P}(\mathcal{H}_4^c) &\le \mathbb{P}(\mathcal{H}_4^c \cap \mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0} \cap \mathcal{C}) + \mathbb{P}\big(\mathcal{S}_{1,\alpha_0}^c \cap \mathcal{C}\big) + \mathbb{P}\big(\mathcal{S}_{2,\alpha_0}^c \cap \mathcal{C}\big) + \mathbb{P}(\mathcal{C}) \\
&\le N^\delta \mathbb{E}\big[|\mathbb{E}_N[\Lambda_M] - \nu|\,\mathbf{1}_{\mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0} \cap \mathcal{C}}\big] + \mathcal{O}\Big(N^{-\min\{\beta_0,\varepsilon\}}\Big) \\
&= N^\delta \mathbb{E}\big[|S_{N,2}/S_{N,1} - \nu|\,\mathbf{1}_{\mathcal{S}_{1,\alpha_0} \cap \mathcal{S}_{2,\alpha_0} \cap \mathcal{C}}\big] + \mathcal{O}\Big(N^{-\min\{\beta_0,\varepsilon\}}\Big) \\
&= \mathcal{O}(N^{\delta-\alpha_0}) + \mathcal{O}\Big(N^{-\min\{\beta_0,\varepsilon\}}\Big) = \mathcal{O}\Big(N^{-\min\{\alpha_0-\delta,\beta_0,\varepsilon\}}\Big). \qquad (4.A.83)
\end{aligned}$$

By Lemma 4.4 we can pick $\alpha_0 > \delta/2$ and $\beta_0 > 0$ for all $\delta < 1/2$.

Combining (4.A.77), (4.A.79), (4.A.81) and (4.A.83) with (4.A.76) gives

$$\mathbb{P}(\mathcal{H}^c) = \mathcal{O}\big(N^{-\xi}\big),$$

where $\xi = \min\{(1/2 - 2\delta - \gamma - 2\eta)/2, \beta_2, \eta/2, \beta_0, \varepsilon, \alpha - \delta\}$. Remember that γ is restricted to $\frac{1}{\tau-1} < \gamma < 1/2$, that δ is restricted to $0 < \delta < 1/2$ and η is restricted by $(4.A.75)$. So, pick $\eta, \delta > 0$ such that $2\delta + \eta < 1/2 - \gamma$, because we can pick $\delta > 0$ and $\eta > 0$ arbitrary small, then $\xi > 0$. $\qquad \square$

Proof of Lemma 4.21 Consider the first claim of $(4.A.67)$. The variance of a random variable is bounded from above by its second moment. Therefore, using (4.3.2), Lemma 4.17 and (4.2.7), on $\mathcal{C}$

$$\mathrm{Var}_N(\Lambda_M) \le \mathbb{E}_N[\Lambda_M^2] = \frac{1}{L_N}\sum_{m=1}^N \Lambda_m^3 \le \frac{\Lambda_N^{(N)}}{L_N}\sum_{m=1}^N \Lambda_m^2$$

$$\le N^\gamma \frac{1}{L_N}\sum_{m=1}^N \Lambda_m^2 = N^\gamma S_{N,2} S_{N,1}^{-1}\mathbf{1}_{\mathcal{C}} \le \overline{S}\,\underline{S}^{-1}N^\gamma = \mathcal{O}(N^\gamma),$$

for N sufficiently large. We turn to the second claim of (4.A.67). Using (4.3.8) we bound (4.A.67) on $\mathcal{C}$ from above by

$$\mathrm{Var}_N(d_N)\,\mathbf{1}_{\mathcal{C}} \leq \sum_{n=1}^{\infty} n^2 g_n^{(N)} \mathbf{1}_{\mathcal{C}} = \sum_{n=1}^{\infty} \frac{n^2}{S_{N,1}N} \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\Lambda_m^{n+1}}{n!}\mathbf{1}_{\mathcal{C}}.$$

We split off the term with $n = 1$ and use (4.2.7) and the fact that for $n > 1$, we have $\frac{n}{n-1} \leq 2$, to obtain, on $\mathcal{C}$,

$$\mathrm{Var}_N(d_N) \leq \frac{1}{S_{N,1}N} \sum_{m=1}^{N} e^{-\Lambda_m}\Lambda_m^2 + \frac{1}{S_{N,1}N} \sum_{m=1}^{N} \Lambda_m^3 \sum_{n=2}^{\infty} e^{-\Lambda_m} \frac{\Lambda_m^{n-2}}{(n-2)!} \frac{n}{n-1}$$

$$\leq \frac{S_{N,2}}{\underline{S}} + \frac{2}{\underline{S}N} \sum_{m=1}^{N} \Lambda_m^3 \leq \overline{S}\,\underline{S}^{-1} + 2S_{N,2}\underline{S}^{-1}\Lambda_N^{(N)} \leq \overline{S}\,\underline{S}^{-1} + 2\overline{S}\,\underline{S}^{-1}N^{\gamma}.$$

Hence, $\mathrm{Var}_N(d_N) = \mathcal{O}(N^{\gamma})$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Chapter 5

A geometric preferential attachment model with fitness

Abstract

We study a random graph G_n, which combines aspects of geometric random graphs and preferential attachment. The resulting random graphs have power-law degree sequences with finite mean and possibly infinite variance. In particular, the power-law exponent can be any value larger than 2.

The vertices of G_n are n sequentially generated vertices chosen at random in the unit sphere in $\mathbb{R}^3$. A newly added vertex has m edges attached to it and the endpoints of these edges are connected to old vertices or to the added vertex itself. The vertices are chosen with probability proportional to their current degree plus some initial attractiveness and multiplied by a function, depending on the geometry.

5.1   Introduction

Preferential attachment models are proposed by Barabási and Albert [4] as models for large-scale networks like the Internet, electrical networks, telephone networks, and even complex biological networks. These networks grow in time, because, for example, new routers, transform houses, switchboards or proteins are added to the network. The behavior can be modeled by means of a random graph process. A random graph process is a stochastic process that describes a random graph evolving with time. At each time step, the random graph is updated using a given *rule of growth*, which will be specified later.

In literature a number of different rules of growth are explored. For example, each time step we add or remove edges/vertices [27], or, more advanced, copy parts of the graph [61]. Furthermore, there is freedom in the choice how to connect endpoints of newly added edges. Mostly, one randomly chooses the endpoint over the vertices, or proportional to the degree. Another possibility is to assign to each vertex a *fitness*. In [29, 30] *additive fitness* is explored where one chooses

proportional to the degree plus some (random) value. In [35, 71] *multiplicative fitness* is explored where each vertex has a random fitness and one chooses a vertex proportional to the degree times the fitness. In this chapter we use a constant additive fitness and a variant of multiplicative fitness, depending on the distance between vertices.

Many large networks of interest have power-law degree sequences, by which we mean that the number of vertices with degree k falls off as $k^{-\tau}$ for some exponent $\tau > 1$. The parameter τ is called the power-law exponent. Depending on the value of τ we classify the following three categories: the *infinite mean case*, *the finite mean and infinite variance case*, and the *finite variance case*, which corresponds to $\tau \in (1, 2)$, $\tau \in (2, 3)$ and $\tau > 3$, respectively.

These categories are of interest, because the behavior of the typical distance is determined by the power-law exponent τ. Results in the literature show that if $\tau \in (1, 2)$ the typical distance is bounded by some constant, if $\tau \in (2, 3)$ the typical distance is concentrated around $\log \log n$ and if $\tau > 3$ it is concentrated around $\log n$, where n is the number of vertices of the graph, see [38, 53, 54, 56, 71].

A large number of graph models have been introduced to describe complex networks, but often the underlying geometry is ignored. In general it is difficult to get rigorous results for properties like the degree distribution, typical distances or diameter, even if one disregards the geometry. However, in wireless ad-hoc networks the geometry is of great importance, since in these networks nodes are spread over some surface and nodes can only communicate with neighbors within a certain range, depending on the geometry.

In this chapter we will rely on the geometric preferential attachment (GPA) model introduced in [43] and extended in [44] by the same authors. The GPA model is a variant of the well known Barabási-Albert (BA) model. In the BA model new vertices are added to the graph one at a time. Each new vertex is connected to m of the existing vertices, where we choose to connect to an old vertex proportional to its degree. In the GPA model each vertex has a position on a surface, and we choose to connect to an old vertex proportional to its degree times a non-constant multiplicative value. This multiplicative value depends on the distance of the old vertex and the newly added vertex. For instance, let the multiplicative constant be 1 if the vertices are at distance at most r_n, and otherwise zero. The latter attachment rule essentially describes the construction of a simplified wireless ad-hoc network.

5.1.1   Definition of the model

In this section we will introduce the Geometric Preferential Attachment model with fitness (GPAF). The GPAF model is described by a random graph process $\{G_\sigma\}_{\sigma=0}^n$, which we will study for large values of n. For $0 \leq \sigma \leq n$, each vertex of the graph $G_\sigma = (V_\sigma, E_\sigma)$ is positioned on the sphere $S \subset \mathbb{R}^3$. The radius of the sphere S is taken equal to $1/(2\sqrt{\pi})$, so that, conveniently, $\text{Area}(S) = 1$. The vertices of the graph G_σ are given by $V_\sigma = \{1, 2, \ldots, \sigma\}$ and E_σ is the set of edges. The position of vertex $v \in V_\sigma$ in the graph G_σ is given by $x_v \in S$ and the degree at time σ is given by $d_\sigma(v)$.

To describe the GPAF random graph consisting of n vertices we need 3 pa-

rameters and a function $F_n : [0, \pi] \to \mathbb{R}_+$. The first parameter of the model is $m = m(n) > 0$, which is the number of edges added in every time step. The second parameter is $\alpha \geq 0$, which is a measure of the bias toward self-loops. The third one is $\delta > -m$, which is the initial attractiveness of a vertex. The value $F_n(u)$, $u \in [0, \pi]$, is an indicator of the attraction between two vertices at distance u.

Before we give the model definition, we first will explain the use of the parameter α. Assume that the graph G_σ is given, consisting of the vertices V_σ. We construct the graph $G_{\sigma+1}$ by choosing vertex $x_{\sigma+1}$ uniformly at random in S and add it to G_σ with m directed edges emanating from the vertex $x_{\sigma+1}$. Let, for $\sigma = 1, 2, \ldots, n-1$,

$$T_{\sigma,n}(u) = \sum_{v=1}^{\sigma} (d_\sigma(v) + \delta) F_n(|x_v - u|), \tag{5.1.1}$$

where $|x_v - u| \in [0, \pi]$ is the distance from u to u_0 on the geodesic through the points u and u_0. Furthermore, let the endpoints of the m emanating edges be given by the vertices $v_{\sigma+1}^{(1)}, \ldots, v_{\sigma+1}^{(m)}$. Intuitively, we would like to choose the endpoints at random (with replacement) from V_σ, such that $v \in V_\sigma$ is chosen with probability

$$\mathbb{P}_{\sigma_+} \left(v_{\sigma+1}^{(i)} = v \right) = \frac{(d_\sigma(v) + \delta) F_n(|x_v - x_{\sigma+1}|)}{T_{\sigma,n}(x_{\sigma+1})},$$

where $\mathbb{P}_{\sigma_+}(\cdot) = \mathbb{P}(\cdot \,|\, G_\sigma, x_{\sigma+1})$. However, the above given rule of growth is not well-defined. To see this, consider the simplified model for wireless ad-hoc networks, i.e., $F_n(x) = \mathbf{1}_{\{x \leq r_n\}}$. Then, for any σ, there is a positive probability that there are no vertices within reach of the newly added vertex $x_{\sigma+1}$ and therefore $T_{\sigma,n}(x_{\sigma+1}) = 0$.

To overcome this problem we could add self-loops if $T_{\sigma,n}(x_{\sigma+1}) = 0$ or if $T_{\sigma,n}(x_{\sigma+1})$ is small. We will introduce a variant of the above model, which is based on the GPA model introduced in [44]:

Rules of growth for $\alpha > 0$:

- **Initial Rule** ($\sigma = 0$): To initialize the process, we start with G_0 being the empty graph.

- **Growth Rule** (at time $\sigma+1$): We choose vertex $x_{\sigma+1}$ uniformly at random in S and add it to G_σ with m directed edges emanating from the vertex $x_{\sigma+1}$. Let the endpoints of the m emanating edges given by the vertices $v_{\sigma+1}^{(1)}, \ldots, v_{\sigma+1}^{(m)}$. We choose the endpoints at random (with replacement) from V_σ, such that $v \in V_\sigma$ is chosen with probability

$$\mathbb{P}_{\sigma_+} \left(v_{\sigma+1}^{(i)} = v \right) = \frac{(d_\sigma(v) + \delta) F_n(|x_v - x_{\sigma+1}|)}{\max\{T_{\sigma,n}(x_{\sigma+1}), \alpha \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})]/2\}}, \tag{5.1.2}$$

  and

$$\mathbb{P}_{\sigma_+} \left(v_{\sigma+1}^{(i)} = \sigma + 1 \right) = 1 - \frac{T_{\sigma,n}(x_{\sigma+1})}{\max\{T_{\sigma,n}(x_{\sigma+1}), \alpha \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})]/2\}}, \tag{5.1.3}$$

for $i \in \{1, 2, \ldots, m\}$.

The above given random graph model is well defined, since the denominator is always strictly positive. Indeed, the following lemma calculates the value of $\mathbb{E}[T_{\sigma,n}(x_{\sigma+1})]$ which is strictly positive.

Lemma 5.1 *For any fixed point $v \in S$,*

$$\int_S F_n(|v - u|) \, \mathrm{d}u = I_n, \tag{5.1.4}$$

where

$$I_n = I(F_n) = \frac{1}{2} \int_{x=0}^{\pi} F_n(x) \sin x \, \mathrm{d}x.$$

As a consequence, if U is a randomly chosen point from S, then

$$\mathbb{E}[T_{\sigma,n}(U)] = I_n(2m + \delta)\sigma. \tag{5.1.5}$$

Proof. First note that I_n does not depend on v due to rotation invariance. Thus, without loss of generality, we can assume that v is at the north pole of the sphere. Using spherical coordinates, we find $du = r_0^2 \sin \theta \, \mathrm{d}\theta \, \mathrm{d}\varphi$, where $r_0 = 1/(2\sqrt{\pi})$, and $|v - u| = \theta$, so that:

$$\int_S F_n(|v - u|) \, \mathrm{d}u = \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} r_0^2 F_n(\theta) \sin \theta \, \mathrm{d}\theta \, \mathrm{d}\varphi = 2\pi r_0^2 \int_0^{\pi} F_n(\theta) \sin \theta \, \mathrm{d}\theta = I_n.$$

For the second claim we calculate the expected value of $T_{\sigma,n}(U)$, (5.1.1), conditional on the graph G_σ:

$$\mathbb{E}[T_{\sigma,n}(U) \,|\, G_\sigma] = \sum_{v=1}^{\sigma} (d_\sigma(v) + \delta)\mathbb{E}[F_n(|x_v - U|) \,|\, G_\sigma]$$

$$= I_n \sum_{v=1}^{\sigma} (d_\sigma(v) + \delta) = I_n(2m + \delta)\sigma,$$

where we apply (5.1.4) on $\mathbb{E}[F_n(|x_v - U|) \,|\, G_\sigma]$, since $v \leq \sigma$ and, therefore,

$$\mathbb{E}[F_n(|x_v - U|) \,|\, G_\sigma] = \int_S F_n(|x_v - u|) du = I_n. \tag{5.1.6}$$

Hence, $\mathbb{E}[T_{\sigma,n}(U)] = \mathbb{E}[\mathbb{E}[T_{\sigma,n}(U) \,|\, G_\sigma]] = I_n(2m + \delta)\sigma.$ $\qquad\square$

We use the abbreviations, for $u \in S$,

$$M_{\sigma,n}(u) = \max\{T_{\sigma,n}(u), \alpha\Theta I_n\sigma\} \quad \text{and} \quad A_{\sigma,n}(u) = F_n(|u - x_{\sigma+1}|), \tag{5.1.7}$$

where

$$\Theta = \Theta(\delta, m) = (2m + \delta)/2. \tag{5.1.8}$$

As a consequence, we can rewrite the attachment rules as

$$\mathbb{P}_{\sigma_+}\left(v_{\sigma+1}^{(i)} = v\right) = \frac{(d_\sigma(v) + \delta)A_{\sigma,n}(x_v)}{M_{\sigma,n}(x_{\sigma+1})}$$

and

$$\mathbb{P}_{\sigma+} \left(v_{\sigma+1}^{(i)} = \sigma + 1 \right) = 1 - \frac{T_{\sigma,n}(x_{\sigma+1})}{M_{\sigma,n}(x_{\sigma+1})}, \tag{5.1.9}$$

for $i \in \{1, 2, \ldots, m\}$.

Remark 5.2 *In the above description we add directed edges to the graph and therefore we construct a directed graph. For questions about the connectedness and diameter of the graph we ignore the direction of the edges, but we need the direction of the edges in the proofs of the main results.*

Remark 5.3 *In this chapter we will illustrate the theorems using the canonical functions*

$$F_n^{(0)}(u) \equiv 1, \quad F_n^{(1)}(u) = \mathbf{1}_{\{|u| \le r_n\}} \quad and \quad F_n^{(2)}(u) = \frac{1}{\max\{n^{-\psi}, u\}^\beta}, \tag{5.1.10}$$

where $r_n \ge n^{\varepsilon - 1/2}$, $\varepsilon < 1/2$, $\psi < 1/2$ and $\beta \in (0, 2) \cup (2, \infty)$. The canonical function $F_n^{(0)}$ implies that the vertices are chosen proportional to the degree, and, furthermore, the geometry is ignored, the model is then equivalent to the PARID model, see [29] or Section 5.1.2. The function $F_n^{(1)}$ implies that a new vertex can only connect to vertices at distance at most r_n. Finally, canonical function $F_n^{(2)}$ implies that vertices are chosen proportional to the degree, and, in contrast to $F_n^{(0)}$, will prefer vertices close to the new vertex, since $F_n^{(2)}$ is non-increasing as a function in u. Let $I_n^{(i)} = I_n(F_n^{(i)})$, then in [44] it is shown that one can take $I_n^{(1)} \sim r_n^2/4$, $I_n^{(2)} = \mathcal{O}(1)$ if $\beta \in (0, 2)$, and $I_n^{(2)} \sim \frac{n^{\delta(\beta-2)}}{2(\beta-2)}$ if $\beta > 2$.

5.1.2   Heuristics and main results

Using the results of [44], which is a special case of our model when $\delta = 0$, together with the results of the PARID model, introduced in [29], we will predict how the power-law exponent of the degree sequence will behave.

Consider the PARID random graph process $\{G'_\sigma\}_{\sigma \ge 0}$ as introduced in the paper [29] with constant weights equal to m. For this special case, we give a brief description of the model.

The construction of the PARID graph $G'_\sigma = (V'_\sigma, E'_\sigma)$ depends on the graph $G'_{\sigma-1}$. The rule of growth is as follow: add a vertex to the graph $G'_{\sigma-1}$ and from this vertex emanates m edges. The endpoints of these m edges are chosen independently (with replacement) from the vertices of $G'_{\sigma-1}$. The probability that vertex $v \in V'_{\sigma-1}$ is chosen is proportional to the degree of vertex v plus δ, more specifically:

$$\mathbb{P}(\text{choose vertex } v \,|\, G'_\sigma) = \frac{d_\sigma(v) + \delta}{(2m + \delta)\sigma}. \tag{5.1.11}$$

If $\alpha \le 2$, $\delta > -m$ and $F_n = F_n^{(0)}$, then the GPAF model coincides with the PARID model where the weight of each vertex is set to m. Note that, for the chosen parameters,

$$A_{\sigma,n}(x_v) = F_n^{(0)}(|x_v - x_{\sigma+1}|) = 1, \qquad I_n^{(0)} = \frac{1}{2} \int_{x=0}^\pi \sin(x)\,\mathrm{d}x = 1,$$

and

$$\alpha \Theta I_n^{(0)} \sigma \le 2\Theta\sigma = (2m+\delta)\sigma = \sum_{v \in V_\sigma} (d_\sigma(v) + \delta) = T_{\sigma,n}^{(0)}(x_{\sigma+1}),$$

thus $M_{\sigma,n}^{(0)}(x_{\sigma+1}) = T_{\sigma,n}^{(0)}(x_{\sigma+1}) = (2m+\delta)\sigma$. Therefore, the equations (5.1.9) turns into (5.1.11), since

$$\mathbb{P}_{\sigma_+}\left(v_{\sigma+1}^{(1)} = v\right) = \mathbb{P}_{\sigma_+}(\text{choose vertex } v) = \frac{(d_\sigma(v)+\delta)A_{\sigma,n}(x_v)}{M_{\sigma,n}^{(0)}(x_{\sigma+1})} = \frac{d_\sigma(v)+\delta}{(2m+\delta)\sigma}.$$

Furthermore, note that for these parameters there are no self-loops, since

$$\mathbb{P}_{\sigma_+}\left(v_{\sigma+1}^{(1)} = \sigma+1\right) = 1 - \frac{T_{\sigma,n}^{(0)}(x_{\sigma+1})}{M_{\sigma,n}^{(0)}(x_{\sigma+1})} = 1 - \frac{M_{\sigma,n}^{(0)}(x_{\sigma+1})}{M_{\sigma,n}^{(0)}(x_{\sigma+1})} = 0.$$

For the PARID model, we know that the power-law exponent is $3 + \delta/m$, thus we expect that the power-law exponent in our model is $3 + \delta/m$ if $\alpha \le 2$ and $F_n = F_n^{(0)}$. For $\alpha > 2$, $\delta = 0$ and F_n satisfying some mild condition, see (5.1.12), we know from [44] that the power-law exponent is $1 + \alpha$, which is independent of F_n.

We will show in this chapter that the power-law exponent is given by $1 + \alpha(1+\delta/2m)$, which generalizes the two mentioned papers [43, 44]. More precisely, let $N_k(\sigma)$ denote the number of vertices of degree k in G_σ and let $\bar{N}_k(\sigma)$ be its expectation. We will show that:

Theorem 5.1 (Behavior of the degree sequence) *Suppose that $\alpha > 2$, $\delta > -m = m(n)$ and in addition that for $n \to \infty$,*

$$\int_{x=0}^{\pi} F_n(x)^2 \sin x \, \mathrm{d}x = \mathcal{O}\!\left(n^\theta I_n^2\right), \tag{5.1.12}$$

where $\theta < 1$ is a constant. Then there exists a constant $\gamma_1 > 0$ such that for all $k = k(n) \ge m$,

$$\bar{N}_k(n) = n e^{\phi_k(m,\alpha,\delta)} \left(\frac{m}{k}\right)^{1+\alpha(1+\delta/2m)} + \mathcal{O}(n^{1-\gamma_1}), \tag{5.1.13}$$

where $\phi_k(m,\alpha,\delta) = \mathcal{O}(1)$ tends to a constant $\phi_\infty(m,\alpha,\delta)$ as $k \to \infty$.

Furthermore, for each $\epsilon > 0$ and n sufficiently large, the random variables $N_k(n)$ satisfy the following concentration inequality

$$\mathbb{P}\!\left(|N_k(n) - \bar{N}_k(n)| \ge I_n^2 n^{\max\{1/2, 2/\alpha\}+\epsilon}\right) \le e^{-n^\epsilon}. \tag{5.1.14}$$

Remark 5.4 *Note that the power-law exponent in (5.1.13) does not depend on the choice of the function F_n. We will see in the proof, that F_n manifests itself only in the error terms.*

All the given canonical functions in Remark 5.3 do satisfy the condition given by (5.1.12): it should be evident that for $F_n^{(0)}$ the constants I_n and θ are given by $I_n^{(0)} = 1$ and $\theta^{(0)} = 0$, respectively. Furthermore, from Remark 5.3 it follows that we can take $\theta^{(1)} = 0$ for $F_n^{(1)}$ and $\theta^{(2)} = 0$ for $F_n^{(2)}$ if $\beta \in (0,2)$ and $\theta^{(2)} = 2\psi$ if $\beta > 2$.

Before we consider the connectivity and diameter of G_n, we place some additional restrictions on the function F_n. These restrictions are necessary to end up with a graph which is with high probability connected. Keep in mind the function $F_n^{(1)}(u) = \mathbf{1}_{\{|u| \leq r_n\}}$, then it should be clear that r_n should not decrease too fast, otherwise we end up with a disconnected graph.

Let $\rho_n = \rho(\mu, n)$ be such that

$$\mu I_n = \frac{1}{2} \int_{x=0}^{\rho_n} F_n(x) \sin x \, \mathrm{d}x,$$

for some $\mu \in (0, 1]$.

We will call F_n *well-behaved* (for some value of μ) if there exist constants N, L and c_3 such that for all $n \geq N$:

  (W1) F_n is monotone non-increasing;

  (W2) $n\rho_n^2 \geq L \log n$, for some constant L;

  (W3) $\rho_n^2 F_n(2\rho_n) \geq c_3 I_n$, for some constant c_3 which is bounded from below.

Remark 5.5 *In [44] the word 'smooth' is used instead of 'well-behaved'. It is generally accepted that a 'smooth function' is a function that has derivatives of all orders. Therefore, we refrain from using the word 'smooth', and use the word 'well-behaved' instead.*

Before stating the theorem, we will give an intuitive meaning of ρ_n. To that end, consider the function $F_n^{(1)}(u) = \mathbf{1}_{\{u < r_n\}}$ and use the fact that if $r_n = \mathcal{O}(n^{-1/2-\varepsilon})$ for some $\varepsilon > 0$, then the limiting graph is not connected, see [72]. It should be intuitively clear that in the limit, each newly added vertex x_n should connect to at least one other vertex. Thus, there should be at least one vertex within distance r_n of x_n. At time n there are $n-1$ vertices and the probability that at least one of these vertices is at distance at most r_n of vertex x_n, denoted by $p_c(n, r_n)$, is at most $C(n-1)r_n^2$ for some constant C. On the other hand, we see that if $r_n = \mathcal{O}(n^{-1/2-\varepsilon})$ then $p_c(n, r_n) = \mathcal{O}(n^{-2\varepsilon})$ tends to zero for large n, and, as a consequence, in the limit the graph is not connected. If, as is our assumption, $r_n > n^{\varepsilon-1/2}$, then $p_c(n, r_n) \to \infty$.

The value of ρ_n is the 'generalized radius' indicating the relevant radius for the model. The condition that $p_c(n, r_n) \to \infty$ is replaced by $n\rho_n^2 > L \log n$. Then, intuitively, condition **W2** implies that for general F_n the value $p_c(n, \rho_n)$, does not tend to zero and implies that the limiting graph is connected. The conditions **W1** and **W3** are technicalities, combined they ensure that the 'area' due to the radius ρ_n is sufficiently large: condition **W1** states that F_n is monotone non-increasing and combined with **W3** one can show that the 'area' within radius $2\rho_n$ is $(2\rho_n)^2 F_n(2\rho_n)$, which is at least $4c_3 I_n$.

Theorem 5.2 *If $\alpha \geq 2$ and F_n is well-behaved, $m \geq K \log n$, and K is a sufficiently large constant, then with high probability*

- G_n *is connected;*

- G_n *has diameter $\mathcal{O}(\log n / \rho_n)$.*

Remark 5.6 *All the canonical functions are well-behaved. It should be evident that one can take for $F_n^{(0)}$: $\mu^{(0)} = 1$, $\rho_n^{(0)} = 1$ and $c_3^{(0)} = 1$ for $F_n^{(0)}$. For $F_n^{(1)}(u)$ one can take for example $\mu^{(1)} \sim 1/4$ and $\rho_n^{(1)} = r_n/2$ and $c_3^{(1)} \sim 1$. Finally, $F_n^{(2)}$ is also well-behaved, we refer to [44] for the precise values of $\rho_n^{(2)}, \mu^{(2)}$ and $c_3^{(2)}$.*

We end with a sharper result on the diameter, however we, also, need stronger restrictions on the function F_n. We will call F_n *tame* if there exists strictly positive constants C_1 and C_2 such that

(T1) $F_n(x) \geq C_1$ for $0 \leq x \leq \pi$;

(T2) $I_n \leq C_2$.

Theorem 5.3 *If $\alpha \geq 2$, $\delta > -m$ and F_n is tame and $m \geq K \log n$, and K sufficiently large, then with high probability*

- *G_n is connected;*

- *G_n has diameter $\mathcal{O}(\log_m n)$.*

Remark 5.7 *It should be evident that the function $F_n^{(0)}$ is tame, since one can take $C_1 = C_2 = 1$. If $\beta \in (0, 2)$ then we also have that $F_n^{(2)}$ is tame, since*

$$F_n^{(2)}(x) \geq \pi^{-\beta}, \text{ for } 0 \leq x \leq \pi, \quad \text{and} \quad I_n^{(2)} = \frac{1}{2} \int_{x=0}^{\pi} x^{-\beta} \sin x \, dx \leq \frac{\pi^{2-\beta}}{2(2-\beta)}.$$

Remark 5.8 *If we consider the configuration model (CM), see Section 5.1.3, [54, 56] or the Poissonian random graph (PRG), see [37, 71], then the typical distance depends on the power-law exponent. If the power-law exponent is larger than 3, then the typical distance is of $\mathcal{O}(\log n)$, where n is the number of vertices in the graph, and if the power-law exponent is between 2 and 3 then the typical distance is of $\mathcal{O}(\log \log n)$. It is not clear if this holds for the GPAF model. Theorem 5.2 only states an upper bound on the diameter, independent of δ.*

If $F_n = F_n^{(0)} \equiv 1$ and $\delta \in (-m, 0)$ then the authors of [53] show that the diameter in the graph G_n fluctuates around $\log \log n$. If $F_n(u) = F_n^{(1)}(u) = \mathbf{1}_{\{|u| \leq r_n\}}$, then, intuitively, the diameter depends only on r_n, since r_n determines the maximal length of an edge, and we conjecture that the diameter is at least of order $\log n$.

5.1.3   Related work

In this section we consider random graph models, which are related to the Geometric Preferential Attachment model with fitness (GPAF).

As mentioned earlier the model is related to the Albert-Barábasi (BA) model. In the BA-model the power-law exponent τ is limited to the value 3, which was proven by Bollobás and Riordan.

Cooper and Frieze introduced in [27] a very general model preferential attachment model. In this model it is both possible to introduce new vertices at each time step or to introduce new edges between old vertices. Due to the weights with

which edges of the new vertices are attached to old vertices and the adding of edges between old vertices, the power-law exponent τ can obtain any value $\tau > 2$.

In [30] the authors overcome the restriction $\tau \geq 3$ in a different way, by choosing the endpoint of an edge proportional to the in-degree of a vertex plus some initial attraction $A > 0$. This is identical by choosing the endpoint of an edge proportional to the degree of a vertex plus some amount $\delta = A - m > -m$, as done in the PARID model (cf. [29]). The power-law exponent in [30] is given by $\tau = 3 + \delta/m$. Note that for $\delta = 0$ we obtain the BA model. The authors of [29] show more rigorously some of the results in [30].

Both in [27] and in [29] it is allowed to add a random number of edges W, with the introduction of a new vertex. In case the mean of W is finite the power-law exponent is given by $\tau = 3 + \delta/\mathbb{E}[W]$. Hence, if $\mathbb{P}(W = m) = 1$ for some integer $m \geq 1$ then we see that $\tau = 3 + \delta/m \geq 2$, since we can choose for δ any value in $(-m, 0)$.

In [43, 44] the authors add geometry to the BA model, which corresponds to the GPAF model, introduced above, with $\delta = 0$. Due to a technical difficulty the model has an additional parameter, called $\alpha > 2$. As a consequence of this restriction they only obtain power-law exponents greater than 3, since the power-law exponent is given by $\tau = \alpha + 1$.

By combining the GPA and PARID model, we obtain the GPAF model, introduced in this chapter. Due to the additional parameter δ, it is in this model possible to obtain any power-law exponent τ bigger than 2.

5.1.4 Overview of the chapter

The remainder of this chapter is divided into three sections. In Section 5.2 we will derive a recurrence relation for the expected number of vertices of a given degree. In Section 5.3 we will present a coupling between the graph process and an urn scheme, which will be used in Section 5.4 to show that the number of vertices with a given degree is concentrated around its mean.

5.2 Recurrence relation for the expected degree sequence

In this section we will establish a recurrence relation for $\bar{N}_k(\sigma) = \mathbb{E}[N_k(\sigma)]$, the expected number of vertices with degree k at time σ, which is claim (5.1.13) of Theorem 5.1. From this recurrence relation, we will show that

$$\bar{N}_k(\sigma) \sim \sigma p_k,$$

where $p_k \sim k^{-(1+\alpha(1+\delta/2m))}$, as $k \to \infty$. The proof of claim (5.1.13) depends on a lemma, which is crucial for the proof. This lemma states that for sufficiently large n the value $M_{\sigma,n}(x_{\sigma+1})$ is equal to $\alpha \Theta I_n \sigma$, with high probability. This is a consequence of the fact that $T_{\sigma,n}(x_{\sigma+1})$ is concentrated around its mean $\mathbb{E}[T_{\sigma,n}(x_{\sigma+1})] = 2\Theta I_n \sigma < \alpha \Theta I_n \sigma$, see (5.1.5) and (5.1.8), which is the content of the next lemma.

Lemma 5.9 *If $\alpha > 2$, $\delta > -m$ and $\sigma = 0, 1, 2, \ldots, n$, then*

$$\mathbb{P}\Big(|T_{\sigma,n}(x_{\sigma+1}) - \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})]| > \Theta I_n \left(\sigma^{2/\alpha} + \sigma^{1/2}\log\sigma\right)\log n\Big) = \mathcal{O}(n^{-2}).$$

The proof of this lemma is deferred to Section 5.4.1.

We will allow that m depends on n, thus $m = m(n)$, as already pointed out previously. In establishing the recurrence relation for $\bar{N}_k(\sigma)$, we will rely on the derivation for $\delta = 0$ in [44, Section 3.1].

At each time, we add a new vertex from which m edges are emanating, and for each of these m emanating edges we need to choose a vertex-endpoint. The first possibility for a vertex to have degree k at time $\sigma + 1$ is that the degree at time σ was equal to k and that none of the m endpoints, emanating from $x_{\sigma+1}$, attaches to the vertex. Furthermore, ignoring for the moment the effect of selecting the same vertex twice or more, the vertex could also have degree $k - 1$ at time σ and having one endpoint attached to it at time $\sigma + 1$. Finally, it is also possible that the newly added vertex $x_{\sigma+1}$ has degree k. The total number of vertex-endpoints with degree k is distributed as $\mathrm{Bin}\,(m, p_k(\sigma))$, where

$$p_k(\sigma) = \sum_{v \in D_k(\sigma)} \frac{(k+\delta)A_{\sigma,n}(x_v)}{M_{\sigma,n}(x_{\sigma+1})}, \tag{5.2.1}$$

and $D_k(\sigma) \subset V_\sigma$ is the set of vertices with degree k in the graph G_σ. Similarly, the number of vertex-endpoints with degree $k - 1$ is distributed as

$$\mathrm{Bin}\,(m, p_{k-1}(\sigma)).$$

If the newly added vertex $x_{\sigma+1}$ ends up with degree k, then this vertex has $k - m$ self-loops. The number of self-loops, $d_{\sigma+1}(\sigma + 1) - m$, is distributed as $\mathrm{Bin}\,(m, p_{\sigma,\mathrm{self}})$, where

$$p_{\sigma,\mathrm{self}} = 1 - T_{\sigma,n}(x_{\sigma+1})/M_{\sigma,n}(x_{\sigma+1}). \tag{5.2.2}$$

For $k \geq m$, this leads to

$$\begin{aligned}
\mathbb{E}[N_k(\sigma+1)|G_\sigma, x_{\sigma+1}] = {}& N_k(\sigma) - mp_k(\sigma) + mp_{k-1}(\sigma) \\
& + \mathbb{E}[\mathbf{1}_{\{d_{\sigma+1}(\sigma+1)=k\}}\,|\,G_\sigma, x_{\sigma+1}] + \mathcal{O}(m\eta_k(G_\sigma, x_{\sigma+1})),
\end{aligned} \tag{5.2.3}$$

where $\eta_k(G_\sigma, x_{\sigma+1})$ denotes the probability, conditionally on G_σ, that the same vertex-endpoint is chosen at least twice and at most k times.

Taking expectations on both sides of (5.2.3), we obtain

$$\begin{aligned}
\bar{N}_k(\sigma+1) = {}& \bar{N}_k(\sigma) - m\mathbb{E}\left[\sum_{v \in D_k(\sigma)} \frac{(k+\delta)A_{\sigma,n}(x_v)}{M_{\sigma,n}(x_{\sigma+1})}\right] \\
& + m\mathbb{E}\left[\sum_{v \in D_{k-1}(\sigma)} \frac{(k-1+\delta)A_{\sigma,n}(x_v)}{M_{\sigma,n}(x_{\sigma+1})}\right] \\
& + \mathbb{P}(d_{\sigma+1}(\sigma+1) - m = k - m) + \mathcal{O}(m\mathbb{E}[\eta_k(G_\sigma, x_{\sigma+1})]).
\end{aligned} \tag{5.2.4}$$

Let

$$\mathcal{B}_\sigma = \left\{ |T_{\sigma,n}(x_{\sigma+1}) - 2\Theta I_n \sigma| < C_1 \Theta I_n \sigma^\gamma \log n \right\}, \qquad (5.2.5)$$

where

$$\max\{2/\alpha, \theta\} < \gamma < 1, \qquad (5.2.6)$$

and C_1 is some sufficiently large constant. If

$$\sigma \geq t_0 = t_0(n) = (\log n)^{2/(1-\gamma)},$$

then $\mathcal{B}_\sigma$ implies that for sufficiently large n,

$$T_{\sigma,n}(x_{\sigma+1}) \leq 2\Theta I_n \sigma + C_1 \Theta I_n \sigma^\gamma \log n = 2\Theta I_n \sigma \left(1 + \mathcal{O}\left(\log^{-1} n\right)\right) \leq \alpha \Theta I_n \sigma,$$

since $\alpha > 2$, and, hence, with high probability

$$M_{\sigma,n}(x_{\sigma+1}) = \max\{T_{\sigma,n}(x_{\sigma+1}), \alpha \Theta I_n \sigma\} = \alpha \Theta I_n \sigma. \qquad (5.2.7)$$

Next, we consider each term on the right hand side of (5.2.4) separately, for $\sigma = 1, 2, \ldots, n$. For the first two terms on the right hand side of (5.2.4) we will use that $p_k(\sigma)$ is a probability and that $\mathbb{P}(\mathcal{B}_n^c) = \mathcal{O}(n^{-2})$, for $\sigma > t_0$, see Lemma 5.9, which yields

$$\mathbb{E}[p_k(\sigma)] = \mathbb{E}[p_k(\sigma) \,|\, \mathcal{B}_n] \mathbb{P}(\mathcal{B}_n) + \mathbb{E}[p_k(\sigma) \,|\, \mathcal{B}_n^c] \mathbb{P}(\mathcal{B}_n^c) = \mathbb{E}[p_k(\sigma) \,|\, \mathcal{B}_n] + \mathcal{O}(n^{-2}). \tag{5.2.8}$$

Also, using that $N_k(\sigma) \leq \sigma$,

$$\mathbb{E}[N_k(\sigma) \,|\, \mathcal{B}_n] = \frac{\mathbb{E}[N_k(\sigma)] - \mathbb{E}[N_k(\sigma) \,|\, \mathcal{B}_n^c]}{\mathbb{P}(\mathcal{B}_n)} = \bar{N}_k(\sigma) + \mathcal{O}(\sigma n^{-2}). \qquad (5.2.9)$$

For σ sufficiently large, using (5.1.4), (5.1.7), (5.2.1) and (5.2.9),

$$\mathbb{E}[p_k(\sigma) \,|\, \mathcal{B}_n] = \frac{k+\delta}{\alpha \Theta I_n \sigma} \mathbb{E}\left[\sum_{v \in D_k(\sigma)} A_{\sigma,n}(x_v) \,\middle|\, \mathcal{B}_n \right]$$

$$= \frac{k+\delta}{\alpha \Theta I_n \sigma} \mathbb{E}\left[\sum_{v \in D_k(\sigma)} \int_S F_n(|x_v - u|) \, \mathrm{d}u \,\middle|\, \mathcal{B}_n \right]$$

$$= \frac{(k+\delta)\mathbb{E}[N_k(\sigma) \,|\, \mathcal{B}_n]}{\alpha \Theta \sigma} = \frac{(k+\delta)\bar{N}_k(\sigma)}{\alpha \Theta \sigma} + \mathcal{O}(kn^{-2}). \qquad (5.2.10)$$

Combining (5.2.8) and (5.2.10), we obtain

$$\mathbb{E}[p_k(\sigma)] = \mathbb{E}\left[\sum_{v \in D_k(\sigma)} \frac{(k+\delta)A_{\sigma,n}(x_v)}{M_{\sigma,n}(x_{\sigma+1})} \right] = \frac{(k+\delta)\bar{N}_k(\sigma)}{\alpha \Theta \sigma} + \mathcal{O}(kn^{-2}), \quad (5.2.11)$$

for $\sigma \geq t_0 = t_0(n) = (\log n)^{2/(1-\gamma)}$. The above statement remains true when we replace k by $k-1$.

For the third term on the right hand of (5.2.4), observe that

$$\mathbb{P}(d_{\sigma+1}(x_{\sigma+1}) = k \,|\, G_\sigma, x_{\sigma+1}) = \mathbb{P}(\mathrm{Bin}(m, p) = k - m \,|\, G_\sigma, x_{\sigma+1})$$
$$= \binom{m}{k} p_{\sigma,\mathrm{self}}^{k-m} (1 - p_{\sigma,\mathrm{self}})^{2m-k}. \qquad (5.2.12)$$

Observe that on $\mathcal{B}_n$, using Lemma 5.9, (5.2.2) and (5.2.7),

$$1 - p_{\sigma,\mathrm{self}} = \frac{T_{\sigma,n}(x_{\sigma+1})}{M_{\sigma,n}} = \frac{2\Theta I_n \sigma + \Theta I_n \mathcal{O}\big((\sigma^{2/\alpha} + \sigma^{1/2}\log\sigma)\log n\big)}{\alpha \Theta I_n \sigma}$$
$$= \frac{2}{\alpha} + \mathcal{O}\big((\sigma^{2/\alpha-1} + \sigma^{1/2-1}\log\sigma)\log n\big)$$
$$= \frac{2}{\alpha} + \mathcal{O}\big(\sigma^{\gamma-1}\log n\big), \qquad (5.2.13)$$

where γ is given by (5.2.6). Therefore, combining (5.2.12) and (5.2.13), we obtain

$$\mathbb{P}(d_{\sigma+1}(x_{\sigma+1}) - m = k - m \,|\, \mathcal{B}_n)$$
$$= \binom{m}{k-m} \mathbb{E}\Big[p_{\sigma,\mathrm{self}}^{k-m} (1 - p_{\sigma,\mathrm{self}})^{2m-k} \,\Big|\, \mathcal{B}_n \Big]$$
$$= \binom{m}{k-m} \Big(1 - \frac{2}{\alpha}\Big)^{k-m} \Big(\frac{2}{\alpha}\Big)^{2k-m} \big(1 + \mathcal{O}(m\sigma^{\gamma-1}\log n)\big)$$
$$= \binom{m}{k-m} \Big(1 - \frac{2}{\alpha}\Big)^{k-m} \Big(\frac{2}{\alpha}\Big)^{2k-m} + \mathcal{O}\big(m\sigma^{\gamma-1}\log n\big).$$

It follows from Lemma 5.9 that

$$\mathbb{P}(d_{\sigma+1}(x_{\sigma+1}) = k) = \mathbb{P}(d_{\sigma+1}(x_{\sigma+1}) - m = k - m \,|\, \mathcal{B}_n)\,\mathbb{P}(\mathcal{B}_n) + \mathcal{O}(\mathbb{P}(\mathcal{B}_n^c))$$
$$= \binom{m}{k-m} \Big(1 - \frac{2}{\alpha}\Big)^{k-m} \Big(\frac{2}{\alpha}\Big)^{2k-m} + \mathcal{O}\big(m\sigma^{\gamma-1}\log n\big). \quad (5.2.14)$$

For the fourth and final term on the right hand side of (5.2.4), we use

$$\eta_k(G_\sigma, x_{\sigma+1}) = \mathcal{O}\left(\min\left\{ \sum_{i=m}^{k} \sum_{v \in D_i(\sigma)} \frac{(i + |\delta|)^2 A_{\sigma,n}(x_v)^2}{M_{\sigma,n}(x_{\sigma+1})^2}, 1 \right\} \right),$$

which generalizes Equation (5) in [44]. Using similar arguments that led to (5.2.11), one can show for

$$\sigma > t_1 = t_1(n) = n^{(\gamma+\theta)/2\gamma} \qquad \text{and} \qquad k \leq k_0 = k_0(n) = n^{(\gamma-\theta)/4}, \qquad (5.2.15)$$

that

$$\mathbb{E}[m\eta_k(G_\sigma, x_{\sigma+1})] = \mathcal{O}\Big(\frac{k^2 n^\theta}{m\sigma}\Big) = \mathcal{O}\big(\sigma^{\gamma-1}\big). \qquad (5.2.16)$$

Substituting (5.2.11), (5.2.14) and (5.2.16) in (5.2.4), we end up with the following recurrence relation:

$$
\begin{aligned}
\bar{N}_k(\sigma+1) = \bar{N}_k(\sigma) &- \frac{m}{\alpha\Theta}(k+\delta)N_k(\sigma)/\sigma + \frac{m}{\alpha\Theta}(k-1+\delta)\bar{N}_{k-1}(\sigma)/\sigma \\
&+ \mathbf{1}_{\{m\leq k\leq 2m\}}\binom{m}{k-m}\left(1-2\alpha^{-1}\right)^{k-m}\left(2\alpha^{-1}\right)^{2k-m} \\
&+ \mathcal{O}\!\left(m\sigma^{\gamma-1}\log(n)\right),
\end{aligned}
\tag{5.2.17}
$$

for $k \geq m$ and $\bar{N}_{m-1}(\sigma) = 0$ for all $\sigma \geq 0$. The above recurrence relation depends on σ and k. Consider the limiting case, i.e., $\sigma \to \infty$, and assume that for each k the limit

$$
\bar{N}_k(\sigma)/\sigma \to p_k
\tag{5.2.18}
$$

exists. If this is indeed the case, then in the limit the recurrence relation (5.2.17) yields:

$$
\begin{aligned}
p_k = \frac{m}{\alpha\Theta}(k-1+\delta)p_{k-1} &- \frac{m}{\alpha\Theta}(k+\delta)p_k \\
&+ \mathbf{1}_{\{m\leq k\leq 2m\}}\binom{m}{k-m}\left(1-2\alpha^{-1}\right)^{k-m}\left(2\alpha^{-1}\right)^{2k-m},
\end{aligned}
$$

where $k \geq m$ and $p_{m-1} = 0$. By induction, we then obtain, for $k > 2m$,

$$
p_k = \frac{\frac{m}{\alpha\Theta}(k-1+\delta)}{1+\frac{m}{\alpha\Theta}(k+\delta)}p_{k-1} = \frac{k-1+\delta}{k+\delta+\frac{\alpha\Theta}{m}}p_{k-1} = \frac{\Gamma(m+1+\delta+\frac{\alpha\Theta}{m})\Gamma(k+\delta)}{\Gamma(m+\delta)\Gamma(k+1+\delta+\frac{\alpha\Theta}{m})}p_{2m}.
$$

Using that $\Gamma(t+a)/\Gamma(t) \sim t^a$ for $a \in [0,1)$ and t large, we can rewrite the above equation as follow:

$$
p_k = \phi_k(m,\alpha,\delta)\left(\frac{m}{k}\right)^{1+\frac{m}{\alpha\Theta}} = \phi_k(m,\alpha,\delta)\left(\frac{m}{k}\right)^{1+\alpha(1+\delta/2m)},
$$

where $\phi_k(m,\alpha,\delta) = \Xi(1)$ tends to the constant $\phi_\infty(m,\alpha,\delta)$ depending only on m,α and δ as $k \to \infty$. Finally, following the proof in [44, from equation (15) up to the end of the proof], which shows that there exists a constant M independent from n, such that

$$
|\bar{N}_k(\sigma) - p_k\sigma| \leq M(n^{1-(\gamma-\theta)/4} + \sigma^\gamma\log n),
\tag{5.2.19}
$$

for all $0 \leq \sigma \leq n$ and $m \leq k \leq k_0(n)$. Thus, the assumption (5.2.18) is satisfied. By picking $\gamma_1 > 0$ sufficiently small, we can replace the right hand of (5.2.19) by $n^{1-\gamma_1}$, and one obtains the claim (5.1.13).

5.3   Coupling

In this section we make preparations for the proofs of Lemma 5.9 and the concentration result in Theorem 5.1, see (5.1.14). In this section we take $\tau \in \{1,\ldots,n\}$ fixed and we consider the graph process up to time $\tau - 1$ resulting in the graph

$G_{\tau-1}$. At time τ we apply the **Growth Rule**, see Section 5.1.1, twice on $G_{\tau-1}$, independently of each other, which results in the graphs G_τ and $\hat{G}_\tau$. The idea is to compare the graphs G_τ and $\hat{G}_\tau$ over time by considering G_σ and $\hat{G}_\sigma$ for $\tau \leq \sigma \leq n$. To this end, we will introduce two urn processes $\{U_s\}_{s \geq \tau}$ and $\{U_s\}_{s \geq \tau}$. The urns consist of weighted and numbered balls. Instead of choosing a vertex-endpoint $v \in V_{\sigma+1}$ at time $\sigma + 1$ by (5.1.2) and (5.1.3), we will draw (with replacement) a ball proportional to its weight and then the vertex-endpoint is given by the number on the ball.

Next we describe briefly how we use the urns together with growth of the graphs G_σ and $\hat{G}_\sigma$, for $\sigma \geq \tau$. For the graphs G_σ and $\hat{G}_\sigma$ we define the urns U_σ and $\hat{U}_\sigma$. Then we draw balls from these urns and using the numbers on the drawn balls, we construct the graphs $G_{\sigma+1}$ and $\hat{G}_{\sigma+1}$, which will be explained in next section. This process can be repeated until we end up with the random graphs G_n and $\hat{G}_n$ for some $n \geq \sigma$.

The coupling between the urns will be introduced in four steps. The first step is to introduce for any $\sigma \geq \tau$ two urns. Secondly, we will introduce a probabilistic coupling between the two urn processes. Thirdly, we will describe the coupling between the graph processes G_σ, $\hat{G}_\sigma$ and the two urn processes. Finally, we consider the vertex-endpoints $v_\sigma^{(i)}$ and $\hat{v}_\sigma^{(i)}$, for $i = 1, 2, \ldots, m$, in the graphs G_σ and $\hat{G}_\sigma$, respectively, and we will calculate the probability that $v_\sigma^{(i)} \neq \hat{v}_\sigma^{(i)}$.

5.3.1 The two urns

In this section we describe the contents of the urns corresponding to the graphs G_σ and $\hat{G}_\sigma$, for $\sigma = \tau, \tau + 1, \ldots, n$, and we give an alternative way of choosing the vertex-endpoints using the urns.

Fix two graph processes $\{G_s\}$ and $\{\hat{G}_s\}$ such that the graphs up to time $\tau - 1$ are identical, i.e., $G_s = \hat{G}_s$ for $s = 0, 1, 2 \ldots, \tau - 1$, and that $x_s = \hat{x}_s$, for $s = \tau + 1, \tau + 2, \ldots, n$. Thus, the points x_τ and $\hat{x}_\tau$ are sampled independent of each other, and therefore they will differ from each other, and, as a consequence, also, the edge sets E_s and $\hat{E}_s$, for $s = \tau, \tau + 1, \ldots, \sigma$, will be different. Finally, we assume, without loss of generality, that $T_{\sigma,n}(x_{\sigma+1}) \leq \hat{T}_{\sigma,n}(x_{\sigma+1})$.

Next, we will describe the contents of the urns U_σ and $\hat{U}_\sigma$ given the graphs G_σ and $\hat{G}_\sigma$, and the newly added vertex $x_{\sigma+1}$. We will use the following abbreviations:

$$T_{\sigma,n} = T_{\sigma,n}(x_{\sigma+1}) \qquad \text{and} \qquad M_{\sigma,n} = M_{\sigma,n}(x_{\sigma+1}). \qquad (5.3.1)$$

Furthermore, if e is an edge, then we denote by $\text{TO}(e)$ the endpoint of the edge. Thus, if edge e is added at time t, emanating from the vertex t, points to a vertex $s \in V_t$ then $\text{TO}(e) = s$.

Contents of the urns:

- For each edge $e \in E_\sigma$, such that $\text{TO}(e) \neq \tau$, there is a white ball in U_σ of weight $A_{\sigma,n}(x_{\text{TO}(e)})$ and numbered $\text{TO}(e)$. Similarly, for each edge in $e \in \hat{E}_\sigma$, such that $\text{TO}(e) \neq \tau$, there is a white ball in $\hat{U}_\sigma$ of weight $A_{\sigma,n}(\hat{x}_{\text{TO}(e)}) = A_{\sigma,n}(x_{\text{TO}(e)})$ and numbered $\text{TO}(e)$. Observe that $\hat{x}_{\text{TO}(e)} = x_{\text{TO}(e)}$ since $\text{TO}(e) \neq \tau$.

- For each vertex $v \in V_\sigma \backslash \{\tau\}$ there is a red ball in each of the urns U_σ and $\hat{U}_\sigma$ of weight $(m + \delta)A_{\sigma,n}(x_v)$ and numbered v.

- For the vertex τ there is in U_σ a purple ball of weight $(d_\sigma(\tau)+\delta)A_{\sigma,n}(x_\tau)$ and number τ, and in $\hat{U}_\sigma$ there is an orange ball of weight $(\hat{d}_\sigma(\tau) + \delta)A_{\sigma,n}(\hat{x}_\tau)$ and numbered τ.

- For the vertex $\sigma+1$ each of the urns U_σ and $\hat{U}_\sigma$ contain a green ball of weight $(\alpha\Theta I_n\sigma - \hat{T}_{\sigma,n})^+$, where $(\cdot)^+ = \max\{0, \cdot\}$, and numbered $\sigma+1$. Furthermore, we add only to U_σ a blue ball of weight $((\alpha\Theta I_n\sigma - T_{\sigma,n})^+ - (\alpha\Theta I_n\sigma - \hat{T}_{\sigma,n})^+)^+$ and numbered $\sigma + 1$.

Remark 5.10 *The total weight of the white and red balls in U_σ are given by*

$$\sum_{e \in E_\sigma} A_{\sigma,n}(x_{\text{TO}(e)})\mathbf{1}_{\{\text{TO}(e)\neq\tau\}} \qquad and \qquad \sum_{v \in V_\sigma\backslash\{\tau\}} (m + \delta)A_{\sigma,n}(x_v),$$

respectively, and the weight of the purple ball in U_σ can be rewritten as

$$(d_\sigma(\tau) + \delta)A_{\sigma,n}(x_\tau) = \sum_{e \in E_\sigma} A_{\sigma,n}(x_{\text{TO}(e)})\mathbf{1}_{\{\text{TO}(e)=\tau\}} + (m + \delta)A_{\sigma,n}(x_\tau).$$

Therefore, the total weight of the white, red and purple balls in U_σ is equal to:

$$\sum_{e \in E_\sigma} A_{\sigma,n}(x_{\text{TO}(e)}) + \sum_{v \in V_\sigma} (m + \delta)A_{\sigma,n}(x_v) = \sum_{v \in V_\sigma} (d_\sigma(v) + \delta)A_{\sigma,n}(x_v) = T_{\sigma,n}.$$

Furthermore, from (5.1.7), and some easy calculation, the total weight of all the balls in U_σ is $M_{\sigma,n}$. Similarly, the total weight of the white, red and orange balls in the urn $\hat{U}_\sigma$ is $\hat{T}_{\sigma,n}$ and the total weight of all the balls in $\hat{U}_\sigma$ is, precisely $\hat{M}_{\sigma,n}$.

The weight of a ball depends on the time σ, the color of the ball and the number on the ball. Let b be a ball in U_σ or $\hat{U}_\sigma$, then we define the weight function w_σ as

$$w_\sigma(b) = \begin{cases} A_{\sigma,n}(x_{\xi(b)}) & \text{if } b \text{ is white,} \\ (m + \delta)A_{\sigma,n}(x_{\xi(b)}) & \text{if } b \text{ is red,} \\ (d_\sigma(x_\tau) + \delta)A_{\sigma,n}(x_\tau) & \text{if } b \text{ is purple,} \\ (\hat{d}_\sigma(\hat{x}_\tau) + \delta)A_{\sigma,n}(\hat{x}_\tau) & \text{if } b \text{ is orange,} \\ (\alpha\Theta I_n\sigma - \hat{T}_{\sigma,n})^+ & \text{if } b \text{ is green,} \\ ((\alpha\Theta I_n\sigma - T_{\sigma,n})^+ - (\alpha\Theta I_n\sigma - \hat{T}_{\sigma,n})^+)^+ & \text{if } b \text{ is blue,} \end{cases} \quad (5.3.2)$$

where $\xi(b)$ is the number on the ball. Observe that the number and the color together determine the weight of a ball.

We identify a set $B \subset U_\sigma$ or $B \subset \hat{U}_\sigma$ of distinct balls by the set of pairs (c, k), where c denotes the color and k the number of the ball. For any set B of distinct balls, define

$$\|B\|_\sigma = \sum_{b \in B} w_\sigma(b).$$

We will draw the balls $\{b_\sigma^{(i)}\}_{i=1}^m$ with replacement from the urn U_σ proportional to the weight. Let $\{\hat{b}_\sigma^{(i)}\}_{i=1}^m$ be the sequence of balls drawn from $\hat{U}_\sigma$, then it is easy to show that

$$\mathbb{P}\big(\xi(b_{\sigma+1}^{(1)}) = v \,|\, U_\sigma\big) = \mathbb{P}\big(v_{\sigma+1}^{(1)} = v \,|\, G_\sigma, x_{\sigma+1}\big) \tag{5.3.3}$$

and $\mathbb{P}\Big(\xi(\hat{b}_{\sigma+1}^{(1)}) = v \,|\, \hat{U}_\sigma\Big) = \mathbb{P}\Big(\hat{v}_{\sigma+1}^{(1)} = v \,|\, \hat{G}_\sigma, \hat{x}_{\sigma+1}\Big)$, for $v \in V_{\sigma+1}$. As an example we will show (5.3.3) for $v \in V_{\sigma+1} \backslash \{\tau, \sigma + 1\}$. Observe that in this case the left hand side of (5.3.3) corresponds to the probability of the event that we draw the red ball numbered v or one of the $d_\sigma(v) - m$ white balls, thus

$$\mathbb{P}\big(\xi(b_{\sigma+1}^{(1)}) = v \,|\, U_\sigma\big) = \frac{(m + \delta)A_{\sigma,n}(x_v) + (d_\sigma(v) - m)A_{\sigma,n}(x_v)}{\|U_\sigma\|_\sigma}$$
$$= \frac{(d_\sigma(v) + \delta)A_{\sigma,n}(x_v)}{\|U_\sigma\|_\sigma} = \mathbb{P}\big(v_{\sigma+1}^{(1)} = v \,|\, G_\sigma, x_{\sigma+1}\big),$$

by (5.1.9), since $\|U_\sigma\|_\sigma = M_{\sigma,n}$ (see Remark 5.10).

5.3.2   The joint distribution of drawing balls

In this section we describe how we simultaneously draw the balls from the urns U_σ and $\hat{U}_\sigma$. As before, we will assume that $T_{\sigma,n} \leq \hat{T}_{\sigma,n}$, or, equivalently, $\|U_\sigma\|_\sigma \leq \|\hat{U}_\sigma\|_\sigma$, see Remark 5.10. In the last part of this section we calculate the probability of the event $\{b_\sigma^{(i)} \neq \hat{b}_\sigma^{(i)}\}$, for $i = 1, 2, \ldots, m$, and $\tau \leq \sigma \leq n$, i.e., the event that the two balls $b_\sigma^{(i)}$ and $\hat{b}_\sigma^{(i)}$ in the ith draw do not agree on number or color, which we call a *mismatch*.

Define the following sets

$$R_\sigma = U_\sigma \backslash \hat{U}_\sigma, \qquad C_\sigma = U_\sigma \cap \hat{U}_\sigma \qquad \text{and} \qquad L_\sigma = \hat{U}_\sigma \backslash U_\sigma,$$

where, as before, we compare the balls by color and number.

Remark 5.11 *By construction, L_σ only contains white and orange balls, C_σ contains only white, red and green balls, and R_σ contains only white, purple and blue balls. Furthermore, concerning the weights, we have the following relations*

$$\|C_\sigma\|_\sigma + \|R_\sigma\|_\sigma = \|U_\sigma\|_\sigma \qquad and \qquad \|C_\sigma\|_\sigma + \|L_\sigma\|_\sigma = \|\hat{U}_\sigma\|_\sigma. \tag{5.3.4}$$

Next, we give the joint distribution of drawing balls from the urns U_σ and $\hat{U}_\sigma$.

The joint distribution: Draw, with replacement, m balls $b_{\sigma+1}^{(1)}, \ldots, b_{\sigma+1}^{(m)}$ from U_σ. For convenience we write $b^{(i)} = b_{\sigma+1}^{(i)}$ for $i = 1, \ldots, m$. For each i, we define $\hat{b}^{(i)} = \hat{b}_{\sigma+1}^{(i)}$ by

- If $b^{(i)} \in C_\sigma$ then, with probability

$$\frac{\|U_\sigma\|_\sigma}{\|\hat{U}_\sigma\|_\sigma}, \tag{5.3.5}$$

we set $\hat{b}^{(i)} = b^{(i)}$, otherwise we choose $\hat{b}^{(i)}$ from L_σ, i.e., we choose $b \in L_\sigma$ with probability $w(b)/\|L_\sigma\|_\sigma$; observe that the quotient in (5.3.5) is bounded by 1, because, as remarked earlier, $\|U_\sigma\|_\sigma \leq \|\hat{U}_\sigma\|_\sigma$.

- If $b^{(i)} \in R_\sigma$, then we choose $\hat{b}^{(i)}$ from L_σ, i.e., choose $b \in L_\sigma$ with probability $w_\sigma(b)/\|L_\sigma\|_\sigma$.

The marginal distributions: Let $\hat{\mathbb{P}}^\tau$ the measure under the above coupling where we start from the fixed graphs G_τ and $\hat{G}_\tau$, i.e., $\hat{\mathbb{P}}^\tau(\,\cdot\,) = \mathbb{P}\Big(\,\cdot\,|\,G_\tau, \hat{G}_\tau\Big)$. Furthermore, define, for $\sigma \geq \tau$,

$$\hat{\mathbb{P}}^\tau_\sigma(\,\cdot\,) = \hat{\mathbb{P}}^\tau\Big(\,\cdot\,|\,G_\sigma, \hat{G}_\sigma\Big) \qquad \text{and} \qquad \hat{\mathbb{P}}^\tau_{\sigma+}(\,\cdot\,) = \hat{\mathbb{P}}^\tau\Big(\,\cdot\,|\,G_\sigma, \hat{G}_\sigma, x_{\sigma+1}\Big). \qquad (5.3.6)$$

Observe that, for $\sigma \geq \tau$,

$$\hat{\mathbb{P}}^\tau\Big(\,\cdot\,|U_\sigma, \hat{U}_\sigma\Big) = \hat{\mathbb{P}}^\tau\Big(\,\cdot\,|\,G_\sigma, \hat{G}_\sigma, x_{\sigma+1}\Big) = \hat{\mathbb{P}}^\tau_\sigma(\,\cdot\,|\,x_{\sigma+1}) = \hat{\mathbb{P}}^\tau_{\sigma+}(\,\cdot\,).$$

We will show that under the coupling

$$\hat{\mathbb{P}}^\tau_{\sigma+}(b^{(1)} = b) = \frac{w_\sigma(b)}{\|U_\sigma\|_\sigma} = \mathbb{P}(b^{(1)} = b\,|\,U_\sigma) \qquad (5.3.7)$$

and

$$\hat{\mathbb{P}}^\tau_{\sigma+}\Big(\hat{b}^{(1)} = b\Big) = \frac{w_\sigma(b)}{\|\hat{U}_\sigma\|_\sigma} = \mathbb{P}\Big(\hat{b}^{(1)} = b\,|\,\hat{U}_\sigma\Big), \qquad (5.3.8)$$

for $b \in U_\sigma$ and $b \in \hat{U}_\sigma$, respectively. The claim (5.3.7) is true by construction. For the claim (5.3.8), if $b \in \hat{U}_\sigma$, then this implies that $b \in C_\sigma$ or $b \in L_\sigma$, but not in both. Firstly, assume $b \in C_\sigma$, then

$$\hat{\mathbb{P}}^\tau_{\sigma+}\Big(\hat{b}^{(1)} = b\Big) = \hat{\mathbb{P}}^\tau_{\sigma+}\Big(\hat{b}^{(1)} = b^{(1)}|b^{(1)} = b\Big)\,\hat{\mathbb{P}}^\tau_{\sigma+}(b^{(1)} = b) = \frac{\|U_\sigma\|_\sigma}{\|\hat{U}_\sigma\|_\sigma}\frac{w_\sigma(b)}{\|U_\sigma\|_\sigma} = \frac{w_\sigma(b)}{\|\hat{U}_\sigma\|_\sigma}.$$

Secondly, if $b \in L_\sigma$, then

$$\hat{\mathbb{P}}^\tau_{\sigma+}\Big(\hat{b}^{(1)} = b\Big) = \hat{\mathbb{P}}^\tau_{\sigma+}\Big(\hat{b}^{(1)} = b|b^{(1)} \in C_\sigma\Big)\,\hat{\mathbb{P}}^\tau_{\sigma+}(b^{(1)} \in C_\sigma)$$

$$+ \hat{\mathbb{P}}^\tau_{\sigma+}\Big(\hat{b}^{(1)} = b|b^{(1)} \in R_\sigma\Big)\,\hat{\mathbb{P}}^\tau_{\sigma+}(b^{(1)} \in R_\sigma)$$

$$= \frac{w_\sigma(b)}{\|L_\sigma\|_\sigma}\left(1 - \frac{\|U_\sigma\|_\sigma}{\|\hat{U}_\sigma\|_\sigma}\right) \cdot \frac{\|C_\sigma\|_\sigma}{\|U_\sigma\|_\sigma} + \frac{w_\sigma(b)}{\|L_\sigma\|_\sigma} \cdot \frac{\|R_\sigma\|_\sigma}{\|U_\sigma\|_\sigma}$$

$$= \frac{w_\sigma(b)\Big((\|C_\sigma\|_\sigma + \|R_\sigma\|_\sigma)\,\|\hat{U}_\sigma\|_\sigma - \|U_\sigma\|_\sigma\|C_\sigma\|_\sigma\Big)}{\|L_\sigma\|_\sigma\|U_\sigma\|_\sigma\|\hat{U}_\sigma\|_\sigma} = \frac{w_\sigma(b)}{\|\hat{U}_\sigma\|_\sigma},$$

where we used in the last step the relations given by (5.3.4). Hence, again, the claim (5.3.8) is true.

5.3.3   The joint growth rule between coupled graphs

Fix $\tau \in \{1, 2, \ldots, n\}$, as before, and consider the graph process $\{G_s\}_{s=0}^{\tau-1}$. Let $\{\hat{G}_s\}_{s=0}^{\tau-1}$ be an identical copy of $\{G_s\}_{s=0}^{\tau-1}$, and choose at time τ the position x_τ and $\hat{x}_\tau$ in G_τ and $\hat{G}_\tau$, respectively, at random in S, independently of each other. Using the urns, we will describe the growth of the graphs G_σ and $\hat{G}_\sigma$ over time.

At time τ we apply the **Growth Rule**, independently, on the graphs $G_{\tau-1}$ and $\hat{G}_{\tau-1}$. Then at time $\sigma + 1$, for $\sigma \geq \tau$, let $x_{\sigma+1}$ randomly chosen from S and set $\hat{x}_{\sigma+1} = x_{\sigma+1}$. Let U_σ and $\hat{U}_\sigma$ the urns correspond to $(G_\sigma, x_{\sigma+1})$ and $(\hat{G}_\sigma, \hat{x}_{\sigma+1})$, respectively. Note that this is precisely the setting as described in Section 5.3.1 and, as a consequence, we can use the results of Section 5.3.2. Draw with replacement m balls, $\{b_{\sigma+1}^{(i)}\}_{i=1}^m$, from U_σ, then the vertex-endpoints of $x_{\sigma+1}$ are given $\{\xi(b_{\sigma+1}^{(i)})\}_{i=1}^m$. We, also, draw with replacement m balls, $\{\hat{b}_{\sigma+1}^{(i)}\}_{i=1}^m$, from $\hat{U}_\sigma$, and construct $\hat{G}_{\sigma+1}$ in the same way.

5.3.4   The probability of a mismatch

The event of a mismatch of vertex-endpoints in the graphs G_σ and $\hat{G}_\sigma$, $\sigma \geq \tau$, can be expressed in terms of drawing balls from the urns U_σ and $\hat{U}_\sigma$, since

$$\left\{v_{\sigma+1}^{(1)} \neq \hat{v}_{\sigma+1}^{(1)}\right\} = \left\{\xi(b_{\sigma+1}^{(1)}) \neq \xi(\hat{b}_{\sigma+1}^{(1)})\right\} \subset \left\{b_{\sigma+1}^{(1)} \neq \hat{b}_{\sigma+1}^{(1)}\right\}. \tag{5.3.9}$$

Thus, we will concentrate on the probability of a mismatch between the drawn balls from the urns. Without loss of generality, we assumed that $\|U_\sigma\|_\sigma \leq \|\hat{U}_\sigma\|_\sigma$ or, equivalently, $T_{\sigma,n} \leq \hat{T}_{\sigma,n}$. Using the joint distribution of the urns, see Section 5.3.2, and (5.3.4), we obtain

$$\hat{\mathbb{P}}_{\sigma_+}^\tau\left(b_{\sigma+1}^{(1)} \neq \hat{b}_{\sigma+1}^{(1)}\right) = 1 - \sum_{b \in C_\sigma} \hat{\mathbb{P}}_{\sigma_+}^\tau\left(\hat{b}_{\sigma+1}^{(1)} = b_{\sigma+1}^{(1)} | b_{\sigma+1}^{(1)} = b\right) \hat{\mathbb{P}}_{\sigma_+}^\tau(b_{\sigma+1}^{(1)} = b)$$

$$= 1 - \sum_{b \in C_\sigma} \frac{\|U_\sigma\|_\sigma}{\|\hat{U}_\sigma\|_\sigma} \cdot \frac{w(b)}{\|U_\sigma\|_\sigma} = 1 - \frac{\|C_\sigma\|_\sigma}{\|\hat{U}_\sigma\|_\sigma} = \frac{\|L_\sigma\|_\sigma}{\|\hat{U}_\sigma\|_\sigma}. \tag{5.3.10}$$

By (5.1.7) and Remark 5.10, we can bound the denominator on the right hand side of (5.3.10) from below by

$$\|\hat{U}_\sigma\|_\sigma \geq \|U_\sigma\|_\sigma = M_{\sigma,n} \geq \alpha \Theta I_n \sigma.$$

Next, we consider the numerator on the right hand side of (5.3.10). The set L_σ only contains white balls and the orange ball, see Remark 5.11. Therefore, compare (5.3.2), the total weight of L_σ can be written as

$$\|L_\sigma\|_\sigma = \sum_{h \in \mathcal{E}_\sigma} A_{\sigma,n}(x_h) + (\hat{d}_\sigma(\tau) + \delta) A_{\sigma,n}(\hat{x}_\tau),$$

where

$$\mathcal{E}_\sigma = \cup_{e \in E_\sigma \setminus \hat{E}_\sigma} \{e : \text{TO}(e) \neq \tau\}. \tag{5.3.11}$$

Thus, the probability of a mismatch between balls is bounded from above by

$$\hat{\mathbb{P}}_{\sigma_+}^\tau\left(b_{\sigma+1}^{(1)} \neq \hat{b}_{\sigma+1}^{(1)}\right) \leq \frac{\sum_{h \in \mathcal{E}_\sigma} A_{\sigma,n}(x_h) + (\hat{d}_\sigma(\tau) + \delta) A_{\sigma,n}(\hat{x}_\tau)}{\alpha \Theta I_n \sigma} \tag{5.3.12}$$

Remark 5.12 *If $T_{\sigma,n} > \hat{T}_{\sigma,n}$, then it should be clear that one can interchange the roles of G_σ and $\hat{G}_\sigma$ in Section 5.3, which implies that for this case*

$$\hat{\mathbb{P}}^\tau_{\sigma_+}\Big(b^{(1)}_{\sigma+1} \neq \hat{b}^{(1)}_{\sigma+1}\Big) = \frac{\|R_\sigma\|_\sigma}{\|U_\sigma\|_\sigma} \leq \frac{\sum_{h\in\hat{\mathcal{E}}_\sigma} A_{\sigma,n}(x_h) + (d_\sigma(\tau) + \delta)A_{\sigma,n}(x_\tau)}{\alpha\Theta I_n\sigma}, \quad (5.3.13)$$

where $\hat{\mathcal{E}}_\sigma = \cup_{e\in\hat{E}_\sigma\setminus E_\sigma}\{e : \text{TO}(e) \neq \tau\}$.

5.4 Proof of the main results

In this section we will prove the main results, i.e Theorem 5.1, 5.2 and 5.3. The diameter results, Theorem 5.2 and 5.3, can be proved almost immediately using the proofs in [44], but this is not true for Theorem 5.1. The proof of Theorem 5.1 relies on Lemma 5.9 and this takes more effort.

This section is divided into 3 parts: in the first part we will give the proof of Lemma 5.9, then, in the second part, we will give the proof of the main results, and in the last part we show a bound on the number of expected mismatches, which is necessary for the proof of Lemma 5.9. Before doing so, we will consider the number of mismatches between G_σ and $\hat{G}_\sigma$, for $\sigma \geq \tau \geq 1$, where a perturbation is made at time τ as defined in Section 5.3.3.

At each of the times $s = \tau, \tau + 1, \ldots, \sigma - 1$, we sample (with replacement) m balls from each of the urns U_s and $\hat{U}_s$ under the coupling as introduced in Section 5.3. After m draws we end up with the balls $\{b^{(i)}_s\}^m_{i=1}$ and $\{\hat{b}^{(i)}_s\}^m_{i=1}$. If the two balls $b^{(i)}_s$ and $\hat{b}^{(i)}_s$ in the ith draw do not agree on number or color, then, as before, we call the draw a mismatch, i.e., $\{b^{(i)}_s \neq \hat{b}^{(i)}_s\}$. Let Δ^τ_σ, for $\sigma \geq \tau$, the total number of mismatches between the urns U_σ and $\hat{U}_\sigma$, then

$$\Delta^\tau_\sigma = \sum_{s=\tau}^\sigma \sum_{i=1}^m \mathbf{1}\{b^{(i)}_s \neq \hat{b}^{(i)}_s\}. \qquad (5.4.1)$$

Furthermore, for $u \in S$ and $\sigma \geq \tau$ define

$$\Delta^\tau_\sigma(u) = \sum_{s=\tau}^\sigma \sum_{i=1}^m \left| F_n(|x_{\xi(b^{(i)}_s)} - u|) - F_n(|x_{\xi(\hat{b}^{(i)}_s)} - u|) \right|. \qquad (5.4.2)$$

Next, we will relate the expected values of (5.4.1) and (5.4.2). Fix any $y \in S$ and let U be randomly chosen in S, then

$$\mathbb{E}\Big[\Big| F_n(|x_{\xi(b^{(i)}_s)} - U|) - F_n(|x_{\xi(\hat{b}^{(i)}_s)} - U|)\Big|\Big]$$
$$\leq \mathbb{E}\Big[\Big(F_n(|x_{\xi(b^{(i)}_s)} - U|) + F_n(|x_{\xi(\hat{b}^{(i)}_s)} - U|)\Big)\mathbf{1}\{b^{(i)}_s \neq \hat{b}^{(i)}_s\}\Big]$$
$$= \mathbb{E}\Big[2\int_S F_n(|y - u|)\,\mathrm{d}u\mathbf{1}\{b^{(i)}_s \neq \hat{b}^{(i)}_s\}\Big] = 2I_n\mathbb{E}\big[\mathbf{1}\{b^{(i)}_s \neq \hat{b}^{(i)}_s\}\big],$$

where we used (5.1.4). Thus,

$$\mathbb{E}[\Delta^\tau_\sigma(U)] \leq 2I_n \sum_{s=\tau}^\sigma \sum_{i=1}^m \mathbb{E}\big[\mathbf{1}\{b^{(i)}_s \neq \hat{b}^{(i)}_s\}\big] = 2I_n\mathbb{E}[\Delta^\tau_\sigma]. \qquad (5.4.3)$$

The next lemma bounds the number of expected mismatches between the graphs G_σ and $\hat{G}_\sigma$.

Lemma 5.13 *Under the conditions of Theorem 5.1, let $\sigma \geq \tau \geq 1$ and U randomly chosen in S, then for some constant $C > 0$,*

$$\mathbb{E}[\Delta_\sigma^\tau(U)] \leq CmI_n \left(\frac{\sigma}{\tau}\right)^{\frac{1}{\alpha\Theta/m}} \log \sigma, \tag{5.4.4}$$

and, as a consequence,

$$\mathbb{E}[\Delta_\sigma^\tau(U)] \leq CmI_n \left(\frac{\sigma}{\tau}\right)^{\frac{2}{\alpha}},$$

since $(\Theta/m)^{-1} < 2$.

The proof of the above lemma is deferred to Section 5.4.3.

Remark 5.14 *For the proof of the main result, we need that the number of mismatches is of $o(\sigma)$, which implies that the exponent in (5.4.4) should be smaller than 1, i.e., $m/\alpha\Theta < 1$. For $\alpha > 2$ and $\delta > -m$ this is indeed the case:*

$$\frac{m}{\Theta} \leq \frac{2m}{m + (m+\delta)} < \frac{2}{m} \leq 2,$$

thus $m/\alpha\Theta < 1$.

If $\delta = 0$, which is precisely the model introduced in [44], then the condition simplifies to $1/\alpha < 1$, which is a weaker condition than the condition used in [44]: $2/\alpha < 1$. Nevertheless, we cannot get rid of the condition $\alpha > 2$, because we need that the event $\mathcal{B}_\sigma$ occurs with high probability, see (5.2.5).

5.4.1 Proof of Lemma 5.9

In this section we will prove Lemma 5.9 using the Azuma-Hoeffding inequality, which provides exponential bounds for the tails of a special class of martingales:

Lemma 5.15 *Let $\{X_\tau\}_{\tau \geq 0}$ be a martingale process with the property that, with probability 1, there exists a sequence of positive constants $\{e_\tau\}_{\tau \geq 1}$ such that*

$$|X_{\tau-1} - X_\tau| \leq e_\tau,$$

for all $\tau \geq 1$. Then, for every $\lambda > 0$,

$$\mathbb{P}(|X_\sigma - X_0| \geq \lambda) \leq 2 \exp\left\{-\frac{\lambda^2}{4\sum_{\tau=1}^\sigma e_\tau^2}\right\}.$$

For a proof of this lemma, we refer to [52].

We will apply Lemma 5.15 by taking a Doob-type martingale

$$X_\tau = \mathbb{E}[T_{\sigma,n}(x_{\sigma+1}) \,|\, G_\tau],$$

for $0 \leq \tau \leq \sigma + 1$. By, convention, we let G_0 be the empty graph, then

$$X_0 = \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})] = 2\Theta I_n \sigma \qquad \text{and} \qquad X_{\sigma+1} = T_{\sigma,n}(x_{\sigma+1}).$$

At each time step s we add a new vertex and m edges, see the **Growth Rule** in Section 5.1.1, call this an action. We call an action A *acceptable* if the action can be applied with positive probability. Furthermore, denote by $\mathcal{A}(G)$ the set of all *acceptable actions* that can be applied on the graph G.

Clearly, for $1 \le \tau \le \sigma + 1$,

$$
\begin{aligned}
|X_{\tau-1} - X_\tau| &= |\mathbb{E}[T_{\sigma,n}(x_{\sigma+1})|G_{\tau-1}] - \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})|G_\tau]| \\
&\le \sup_{G_{\tau-1}} \sup_{A,\hat{A}\in\mathcal{A}(G_{\tau-1})} \left| \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})|G_{\tau-1}(A)] - \mathbb{E}\left[T_{\sigma,n}(x_{\sigma+1})|G_{\tau-1}(\hat{A})\right] \right|,
\end{aligned}
\tag{5.4.5}
$$

where the first supremum is taken over all possible graphs $G_{\tau-1}$, and the graphs $G_{\tau-1}(A)$ and $G_{\tau-1}(A)$ are obtained by applying the action A and $\hat{A}$ on the graph $G_{\tau-1}$, respectively.

Next, we fix the graph $G_{\tau-1}$ and let $G_\tau = G_{\tau-1}(A)$ and $\hat{G}_\tau = G_{\tau-1}(\hat{A})$, then let $\hat{\mathbb{P}}^\tau$ the probability measure as introduced in Section 5.3 where we start with the graphs G_τ and $\hat{G}_\tau$, then

$$
\sup_{A,\hat{A}\in\mathcal{A}(G_{\tau-1})} \left| \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})|G_{\tau-1}(A)] - \mathbb{E}\left[\hat{T}_{\sigma,n}(x_{\sigma+1})|\hat{G}_{\tau-1}(\hat{A})\right] \right|
$$
$$
= \sup_{A,\hat{A}\in\mathcal{A}(G_{\tau-1})} \left| \hat{\mathbb{E}}^\tau\left[T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1})\right] \right| \tag{5.4.6}
$$

Using the triangle inequality, the above implies, under the coupling,

$$
e_\tau \le \sup_{G_{\tau-1}} \sup_{A,\hat{A}\in\mathcal{A}(G_{\tau-1})} \hat{\mathbb{E}}^\tau\left[\left|T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1})\right|\right]. \tag{5.4.7}
$$

We claim that, independently of $G_{\tau-1}, A$ and $\hat{A}$, and, for $\sigma \ge \tau$,

$$
\hat{\mathbb{E}}^\tau\left[\left|T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1})\right|\right] \le \tilde{C}m I_n (\sigma/\tau)^{2/\alpha}, \tag{5.4.8}
$$

where the proof of this claim is deferred to the end of this section. Thus, using (5.4.7) and (5.4.8),

$$
\sum_{\tau=1}^{\sigma} e_\tau^2 \le \tilde{C}^2 m^2 I_n^2 \sigma^{4/\alpha} \sum_{\tau=1}^{\sigma} \tau^{-4/\alpha} = \mathcal{O}\left(m^2 I_n^2 (\sigma^{4/\alpha} + \sigma \log \sigma)\right).
$$

To show the above, let $\beta = 4/\alpha$, then $\beta \in (0,2)$. If $\beta \in (1,2)$, then $\sum_{\tau=1}^{\sigma} \tau^{-\beta} < \infty$, and, if $\beta \in (0,1]$, then

$$
\sigma^\beta \sum_{\tau=1}^{\sigma} \tau^{-\beta} \le \sigma^\beta + \sum_{\tau=2}^{\sigma} (\tau/\sigma)^{-\beta} \le \sigma^\beta + \sigma \int_{x=1}^{\sigma} x^{-1}\,\mathrm{d}x = \sigma^\beta + \sigma \log \sigma.
$$

From Lemma 5.15 we then obtain for some constant C_1,

$$
\mathbb{P}(|T_{\sigma,n}(x_{\sigma+1}) - \mathbb{E}[T_{\sigma,n}(x_{\sigma+1})]| \ge \lambda) \le 2e^{-2\log n},
$$

for $\lambda = C_1 m I_n (\sigma^{2/\alpha} + \sigma^{1/2} \log \sigma)(\log n)^{1/2}$. Remember that $\Theta = (2m + \delta)/2$, therefore by taking n sufficiently large, we can replace $C_1 m (\log n)^{1/2}$ by $\Theta \log n$, which is, precisely, the statement of Lemma 5.9, given the claim (5.4.8). □

Proof of claim (5.4.8): Denote by $d_\sigma^-(v)$ the in-degree of the vertex v at time σ, and observe that

$$d_\sigma(v) = d_\sigma^-(v) + m, \tag{5.4.9}$$

since each vertex has by construction m edges pointing outward to other vertices or itself. Furthermore, we denote by $y_s^{(i)}$ the position of the ith vertex-endpoint at time s, thus

$$y_s^{(i)} = x_{v_s^{(i)}}. \tag{5.4.10}$$

By construction of G_τ and $\hat{G}_\tau$, we can apply the coupling introduced in Section 5.3.3. Rewrite $T_{\sigma,n}(x_{\sigma+1})$, see (5.1.1) and (5.1.7), using (5.4.9), (5.4.10), and the coupling, as

$$
\begin{aligned}
T_{\sigma,n}(x_{\sigma+1}) &= \sum_{v=1}^{\sigma} d_\sigma(v) A_{\sigma,n}(x_v) + \delta \sum_{v=1}^{\sigma} A_{\sigma,n}(x_v) \\
&= \sum_{v=1}^{\sigma} d_\sigma^-(v) A_{\sigma,n}(x_v) + (m + \delta) \sum_{v=1}^{\sigma} A_{\sigma,n}(x_v) \\
&= \sum_{s=1}^{\sigma} \sum_{i=1}^{m} A_{\sigma,n}(y_s^{(i)}) + (m + \delta) \sum_{v=1}^{\sigma} A_{\sigma,n}(x_v) \\
&= \sum_{s=1}^{\sigma} \sum_{i=1}^{m} A_{\sigma,n}(x_{\xi(b_s^{(i)})}) + (m + \delta) \sum_{v=1}^{\sigma} A_{\sigma,n}(x_v).
\end{aligned}
$$

Up to and including time $\tau - 1$ both graphs are identical, thus the absolute difference

$$\left| T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1}) \right|,$$

equals

$$\left| \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \left(A_{\sigma,n}(x_{\xi(b_s^{(i)})}) - A_{\sigma,n}(x_{\xi(\hat{b}_s^{(i)})}) \right) + (m + \delta) \left(A_{\sigma,n}(x_\tau) - A_{\sigma,n}(\hat{x}_\tau) \right) \right|.$$

Using the triangle inequality and (5.4.2), we obtain

$$\left| T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1}) \right| \le \Delta_\sigma^\tau(x_{\sigma+1}) + (m + \delta) \left(A_{\sigma,n}(x_\tau) + A_{\sigma,n}(\hat{x}_\tau) \right).$$

Taking expectations, under $\hat{\mathbb{E}}_\tau$, on both sides of the above display, and using (5.1.6), (5.1.7) and (5.4.3), yields, for $1 \le \tau \le \sigma$,

$$\hat{\mathbb{E}}^\tau \left[|T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1})| \right] \le 2 I_n (\hat{\mathbb{E}}^\tau[\Delta_\sigma] + (m + \delta)), \tag{5.4.11}$$

and for $\tau = \sigma + 1$,

$$\hat{\mathbb{E}}^\tau \left[|T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1})| \right] = 0,$$

since $G_\sigma = \hat{G}_\sigma$. Applying Lemma 5.13 on (5.4.11) finally results in

$$\hat{\mathbb{E}}^\tau\Big[\big|T_{\sigma,n}(x_{\sigma+1}) - \hat{T}_{\sigma,n}(x_{\sigma+1})\big|\Big] \leq \tilde{C}mI_n(\sigma/\tau)^{2/\alpha},$$

for some constant $\tilde{C}$. This is precisely the claim (5.4.8).   □

5.4.2   Proofs of the main results

In this section we show the main results. The proof of Theorem 5.1 is almost similar to the proof of Lemma 5.9. The diameter results, i.e, Theorem 5.2 and 5.3 will be proved by adapting the proofs in [44].

Proof of Theorem 5.1: The first part of Theorem 5.1, i.e., claim (5.1.13), has been proved in Section 5.2. For the second part, i.e., claim (5.1.14), we now give a proof, which is similar to the proof of Lemma 5.9. Therefore, we follow the proof of the previous Section 5.4.1, where we now choose $X_\tau = \mathbb{E}[N_k(n) \,|\, G_\tau]$ instead of $X_\tau = \mathbb{E}[T_{\sigma,n}(U) \,|\, G_\tau]$. Similar to (5.4.5), we have that

$$
\begin{aligned}
|\mathbb{E}[N_k(n)|G_{\tau-1}] &- \mathbb{E}[N_k(n)|G_\tau]| \\
&\leq \sup_{G_{\tau-1}} \sup_{A,\hat{A}\in\mathcal{A}(G_{\tau-1})} \Big|\mathbb{E}[N_k(n)|G_{\tau-1}(A)] - \mathbb{E}\Big[N_k(n)|G_{\tau-1}(\hat{A})\Big]\Big|.
\end{aligned}
$$

Using the coupling, we can bound the right hand side in the above display by twice the number of mismatches, since each mismatch can influence at most two edges. Thus,

$$|\mathbb{E}[N_k(n)|G_{\tau-1}] - \mathbb{E}[N_k(n)|G_\tau]| \leq 2\hat{\mathbb{E}}^\tau[\Delta_n^\tau].$$

Therefore, we can take $e_\tau = 2\hat{\mathbb{E}}^\tau[\Delta_n^\tau]$ and we, again, can apply Lemma 5.15, as done in the previous section, which proves claim (5.1.14) and hence Theorem 5.1. □

Proof of Theorem 5.2: The proof is almost identical to the proof of [44, Theorem 2]. To apply this proof for general $\delta > -m$, we only need to replace the constant c_3 in [44] by c_3^*, where $c_3^* = c_3/2$ and c_3 is the constant of condition **W3**, see Section 5.1.2. This will be explained in more detail now.

Pick μ and $\rho_n = \rho_n(\mu, F_n)$ such that F_n is well-behaved for μ, see conditions **W1**, **W2** and **W3**, see Section 5.1.2. Fix $u \in S$ and denote by A_{ρ_n} the spherical cap with center u and radius ρ_n, then there exists positive constants c_1 and c_2, independent of ρ_n, such that

$$A_{\rho_n} = \int_{\{w\in S \,:\, |w-u|\leq\rho_n\}} \mathrm{d}w \in [c_1\rho_n^2, c_2\rho_n^2], \tag{5.4.12}$$

which is shown in [44]. Furthermore, in [44] the authors consider the graph at certain time steps t_s, where s is a positive integer, such that the area of the spherical cap is given by

$$s/2 \leq A_{\rho_n}t_s \leq 3s/2. \tag{5.4.13}$$

In the proof of [44, Theorem 2], the essential step is the statement that the probability that v_{t_s} chooses vertex $v \in V_{t_s}$, assuming that $|x_{t_s} - x_v| \leq 2\rho_n$, is at least $\frac{2c_1 c_3}{\alpha s}$, i.e.,

$$\mathbb{P}\big(v_{t_s}^{(1)} = v \,|\, G_{t_s-1}\big) \geq \frac{2c_1 c_3}{\alpha s}.$$

In our model this is still true, when we replace c_3 by $c_3^* = c_3/2$, since, using the assumptions **W1**, **W2** and **W3**, (5.4.12) and (5.4.13),

$$\mathbb{P}\big(v_{t_s}^{(1)} = v \,|\, G_{t_s-1}\big) \geq \frac{(d_{t_s}(v) + \delta)F_n(|x_{t_s} - x_v|)}{\alpha \Theta I_n t_s} \geq \frac{(m + \delta)F_n(2\rho_n)}{\alpha \Theta I_n t_s}$$

$$\geq \frac{2(m + \delta)A_{\rho_n}F_n(2\rho_n)}{\alpha \Theta I_n s} \geq \frac{2(m + \delta)c_1\rho_n^2 F_n(2\rho_n)}{\alpha \Theta I_n s}$$

$$\geq \frac{2(m + \delta)c_1 c_3}{\alpha \Theta s} = \frac{(m + \delta)}{\Theta}\frac{2c_1 c_3}{\alpha s} > \frac{1}{2}\frac{2c_1 c_3}{\alpha s} = \frac{2c_1 c_3^*}{\alpha s},$$

where we used that $(m + \delta)/\Theta = 1 + \delta/(2m + \delta) > 1/2$ for $-m < \delta \leq 0$ and $(m + \delta)/\Theta \geq 1 > 1/2$ for $\delta > 0$. If we replace the constant c_3 by c_3^* in the proof of Theorem 2, then the proof of [44] holds without further modifications. □

Proof of Theorem 5.3: For $\delta = 0$ the proof is given by the proof of Theorem 3 in [44]. The constant $\lambda = C_1/C_2$ in the proof of [44, Theorem 3] should be replaced by $\lambda = (C_1 + \delta)/2C_2$, then the proof holds verbatim. □

5.4.3 Bounding the expected number of mismatches

In this section we will prove Lemma 5.13. In the proof of the lemma, we rely on two claims, which will be stated now. The first claim bounds for any vertex and all time steps the expected degree:

$$\mathbb{E}[d_v(\sigma) + \delta] \leq mC\left(\frac{\sigma}{v}\right)^a, \tag{5.4.14}$$

where C is some constant and

$$a = m/\alpha\Theta. \tag{5.4.15}$$

The second claim is a technical one, which bounds the expectation of

$$Q_\sigma = \sum_{h \in \mathcal{E}_\sigma} A_{\sigma,n}(x_h)\mathbf{1}_{\{T_{\sigma,n} \leq \hat{T}_{\sigma,n}\}} + \sum_{h \in \hat{\mathcal{E}}_\sigma} A_{\sigma,n}(\hat{x}_h)\mathbf{1}_{\{T_{\sigma,n} > \hat{T}_{\sigma,n}\}} \tag{5.4.16}$$

from above. More precisely, for any $\sigma \geq \tau$,

$$\mathbb{E}[Q_\sigma] \leq I_n\left(\mathbb{E}[\Delta_\sigma^\tau] + \mathbb{E}[d_\sigma(\tau) + \delta] + \mathbb{E}\big[\hat{d}_\sigma(\tau) + \delta\big] + 2\Theta\right). \tag{5.4.17}$$

Next, we will assume that the claims (5.4.14) and (5.4.17) do hold and we will show that Lemma 5.13 follows from these two claims. After the proof of Lemma 5.13, we will prove both claims separately.

Proof of Lemma 5.13: Let $\tau < \sigma \leq t$, then the number of mismatches is recursively defined as

$$\Delta_{\sigma+1}^{\tau} = \Delta_{\sigma}^{\tau} + \sum_{i=1}^{m} \mathbf{1}\big\{b_{\sigma+1}^{(i)} \neq \hat{b}_{\sigma+1}^{(i)}\big\},$$

and, therefore,

$$\hat{\mathbb{E}}^{\tau}\big[\Delta_{\sigma+1}^{\tau}\big] = \hat{\mathbb{E}}^{\tau}[\Delta_{\sigma}^{\tau}] + m\hat{\mathbb{P}}^{\tau}\Big(b_{\sigma+1}^{(1)} \neq \hat{b}_{\sigma+1}^{(1)}\Big), \qquad (5.4.18)$$

since we draw the balls with replacement.

Combining (5.3.12) and (5.3.13), yields

$$\hat{\mathbb{P}}_{\sigma_+}^{\tau}\Big(b_{\sigma+1}^{(1)} \neq \hat{b}_{\sigma+1}^{(1)}\Big) \leq \frac{Q_{\sigma} + (d_{\sigma}(\tau) + \delta)A_{\sigma,n}(x_{\tau}) + (\hat{d}_{\sigma}(\tau) + \delta)A_{\sigma,n}(\hat{x}_{\tau})}{\alpha\Theta I_n \sigma}. \quad (5.4.19)$$

Observe from (5.1.4) that

$$\hat{\mathbb{E}}_{\sigma}^{\tau}[d_{\sigma}(\tau)A_{\sigma,n}(x_{\tau})] = d_{\sigma}(\tau)\hat{\mathbb{E}}_{\sigma}^{\tau}[F(|x_{\tau} - x_{\sigma+1}|)]$$
$$= d_{\sigma}(\tau)\int_S F(|x_{\tau} - u|)du = d_{\sigma}(\tau)I_n,$$

and, hence

$$\hat{\mathbb{E}}^{\tau}[(d_{\sigma}(\tau) + \delta)A_{\sigma,n}(x_{\tau})] = \hat{\mathbb{E}}^{\tau}[d_{\sigma}(\tau) + \delta]\,I_n.$$

Similarly, $\hat{\mathbb{E}}^{\tau}\Big[(\hat{d}_{\sigma}(\tau) + \delta)A_{\sigma,n}(\hat{x}_{\tau})\Big] = \hat{\mathbb{E}}^{\tau}\Big[\hat{d}_{\sigma}(\tau) + \delta\Big]\,I_n.$

Thus, taking expectations on both sides of (5.4.19) results in

$$\mathbb{P}\Big(b_{\sigma+1}^{(1)} \neq \hat{b}_{\sigma+1}^{(1)}\Big) \leq \frac{\mathbb{E}[Q_{\sigma}] + (\mathbb{E}[d_{\sigma}(\tau) + \delta]\,I_n) + (\mathbb{E}\Big[\hat{d}_{\sigma}(\tau) + \delta\Big]\,I_n)}{\alpha\Theta I_n \sigma}, \qquad (5.4.20)$$

Substituting (5.4.14) and (5.4.17) in (5.4.20), yields

$$\mathbb{P}\Big(b_{\sigma+1}^{(1)} \neq \hat{b}_{\sigma+1}^{(1)}\Big) \leq \frac{\mathbb{E}[\Delta_{\sigma}^{\tau}] + (2\Theta + 4mC(\sigma/\tau)^a)}{\alpha\Theta\sigma},$$

for some sufficiently large constant $C > 0$. Therefore, we can bound the right hand side of equation (5.4.18) by,

$$\mathbb{E}\big[\Delta_{\sigma+1}^{\tau}\big] \leq \mathbb{E}[\Delta_{\sigma}^{\tau}]\Big(1 + \frac{a}{\sigma}\Big) + \frac{2\Theta + 4mC(\sigma/\tau)^a}{\alpha\Theta\sigma} = \mathbb{E}[\Delta_{\sigma}]\Big(1 + \frac{a}{\sigma}\Big) + \frac{1 + 8C(\sigma/\tau)^a}{\sigma},$$

where we used that $a = m/\alpha\Theta$ (5.4.15), $1/\alpha \leq 1/2$ and $m/\Theta \leq 2$. Finally, by taking the constant C larger, we can replace the above inequality by

$$\mathbb{E}\big[\Delta_{\sigma+1}^{\tau}\big] \leq \mathbb{E}[\Delta_{\sigma}^{\tau}]\Big(1 + \frac{a}{\sigma}\Big) + C\tau^{-a}\sigma^{a-1}.$$

We will now prove an upper bound for $\mathbb{E}\big[\Delta_{\sigma+1}^{\tau}\big]$. To this end, we consider the solution of the recurrence relation $q(\sigma+1) = q(\sigma)(1+a/\sigma)+b(\sigma)$, for $\sigma > \tau$, with initial condition $q(\tau) = c$. The solution is given by

$$q(\sigma) = \frac{\Gamma(\sigma+a)}{\Gamma(\sigma)}\sum_{s=\tau}^{\sigma-1}\frac{b(s)\Gamma(s+1)}{\Gamma(s+1+a)} + c\frac{\Gamma(\sigma+a)\Gamma(\tau)}{\Gamma(\sigma)\Gamma(\tau+a)}.$$

The above solution implies, that if one takes $b(s) = C\tau^{-a}s^{a-1} \le C\tau^{-a}\frac{\Gamma(s+a)}{\Gamma(s+1)}$, then

$$\mathbb{E}[\Delta_\sigma^\tau] \le q(\sigma) = \frac{\Gamma(\sigma+a)}{\Gamma(\sigma)} \sum_{s=\tau}^{\sigma-1} \frac{b(s)\Gamma(s+1)}{\Gamma(s+1+a)} \le C\tau^{-a}\frac{\Gamma(\sigma+a)}{\Gamma(\sigma)} \sum_{s=\tau}^{\sigma-1} \frac{\Gamma(s+a)}{\Gamma(s+1+a)}.$$

The summation can be bounded from above by $\log\sigma$, therefore, for some constant C,

$$\mathbb{E}[\Delta_\sigma^\tau] \le C\left(\frac{\sigma}{\tau}\right)^a \log\sigma.$$

This proves (5.4.4) and hence Lemma 5.13, given the claims (5.4.14) and (5.4.17). $\square$

Proof of (5.4.14): Note that, see (5.1.7) and (5.1.9),

$$\hat{\mathbb{E}}_{\sigma_+}^\tau[d_{\sigma+1}(v) + \delta] = (d_\sigma(v) + \delta) + \hat{\mathbb{E}}_{\sigma_+}^\tau[d_{\sigma+1}(v) - d_\sigma(v)]$$
$$= (d_\sigma(v) + \delta)\left(1 + m\frac{A_{\sigma,n}(x_v)}{M_{\sigma,n}(x_{\sigma+1})}\right)$$
$$\le (d_\sigma(v) + \delta)\left(1 + m\frac{A_{\sigma,n}(x_v)}{\alpha\Theta I_n\sigma}\right).$$

Therefore, by taking expectations on both sides in the above display, and using (5.1.6), the value of $\mathbb{E}[d_{\sigma+1}(v) + \delta]$ is bound from above by

$$\mathbb{E}\left[(d_\sigma(v) + \delta)\left(1 + m\frac{\mathbb{E}[A_{\sigma,n}(x_v)\,|\,G_\sigma]}{\alpha\Theta I_n\sigma}\right)\right] = \left(1 + \frac{m}{\alpha\Theta\sigma}\right)\mathbb{E}[(d_v(\sigma) + \delta)].$$

Thus, by induction, and using $a = m/\alpha\Theta$,

$$\mathbb{E}[d_{\sigma+1}(v) + \delta] = \left(\frac{\sigma+a}{\sigma}\right)\mathbb{E}[(d_v(\sigma) + \delta)] = \frac{\Gamma(\sigma+a+1)\Gamma(v)}{\Gamma(\sigma+1)\Gamma(v+a)}\mathbb{E}[(d_v(v) + \delta)].$$

Finally, note that $d_v(v) \le 2m$ and that δ is a constant, which implies the claim (5.4.14). $\square$

Proof of (5.4.17): Observe that for $i = 1, \ldots, m$ and $s \ge \tau$, see (5.4.10),

$$\mathbf{1}\{v_s^{(i)} \ne \hat{v}_s^{(i)}\} + \mathbf{1}\{v_s^{(i)} = \hat{v}_s^{(i)} = \tau\} = \mathbf{1}\{y_s^{(i)} \ne \hat{y}_s^{(i)}\} = \mathbf{1}\left\{x_{v_s^{(i)}} \ne \hat{x}_{\hat{v}_s^{(i)}}\right\}, \qquad (5.4.21)$$

since, $x_i = \hat{x}_i$ if $i \ne \tau$ and $x_\tau \ne \hat{x}_\tau$. The above statement is stronger than (5.3.9). Using the definition of $\mathcal{E}_\sigma$, see (5.3.11), and (5.4.21), we have that

$$\sum_{h\in\mathcal{E}_\sigma} A_{\sigma,n}(x_h) \le \sum_{s=\tau}^\sigma \sum_{i=1}^m A_{\sigma,n}(y_s^{(i)})\mathbf{1}\{y_s^{(i)} \ne \hat{y}_s^{(i)}\},$$

and we can bound $\sum_{h \in \hat{\mathcal{E}}_\sigma} A_{\sigma,n}(x_h)$ in a similar way. Hence, we can bound (5.4.16) from above by

$$
Q_\sigma \leq \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{y_s^{(i)} \neq \hat{y}_s^{(i)}\} \Big(A_{\sigma,n}(y_s^{(i)}) \mathbf{1}\{T_{\sigma,n} \leq \hat{T}_{\sigma,n}\} + A_{\sigma,n}(\hat{y}_s^{(i)}) \mathbf{1}\{T_{\sigma,n} > \hat{T}_{\sigma,n}\} \Big)
$$

$$
= \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{y_s^{(i)} \neq \hat{y}_s^{(i)}\} A_{\sigma,n}(\hat{y}_s^{(i)})
$$

$$
+ \mathbf{1}\{T_{\sigma,n} \leq \hat{T}_{\sigma,n}\} \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{y_s^{(i)} \neq \hat{y}_s^{(i)}\} \left(A_{\sigma,n}(y_s^{(i)}) - A_{\sigma,n}(\hat{y}_s^{(i)}) \right). \quad (5.4.22)
$$

Next, we will show that the rightmost double sum of (5.4.22) can be bounded by

$$
(m + \delta)(A_{\sigma,n}(\hat{x}_\tau) - A_{\sigma,n}(x_\tau)).
$$

For this, we rewrite $T_{\sigma,n}$ as

$$
T_{\sigma,n} = \sum_{s=1}^{\sigma} (d_s(\sigma) + \delta) A_{\sigma,n}(x_s) = \sum_{s=1}^{\sigma} d_s^-(\sigma) A_{\sigma,n}(x_s) + (m + \delta) \sum_{s=1}^{\sigma} A_{\sigma,n}(x_s).
$$

Note that

$$
\sum_{s=1}^{\sigma} d_s^-(\sigma) A_{\sigma,n}(x_s) = \sum_{s=1}^{\sigma} \left\{ \sum_{t=s}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{v_t^{(i)}=s\} \right\} A_{\sigma,n}(x_s)
$$

$$
= \sum_{t=1}^{\sigma} \sum_{i=1}^{m} \left\{ \sum_{s=1}^{t} \mathbf{1}\{v_t^{(i)}=s\} A_{\sigma,n}(x_s) \right\} = \sum_{t=1}^{\sigma} \sum_{i=1}^{m} A_{\sigma,n}(y_t^{(i)}).
$$

Therefore,

$$
T_{\sigma,n} = \sum_{s=1}^{\sigma} \sum_{i=1}^{m} A_{\sigma,n}(y_s^{(i)}) + (m + \delta) \sum_{s=1}^{\sigma} A_{\sigma,n}(x_s)
$$

and a similar result hold for $\hat{T}_{\sigma,n}$. The difference of these two expressions equals:

$$
T_{\sigma,n} - \hat{T}_{\sigma,n} = \sum_{s=1}^{\sigma} \sum_{i=1}^{m} (A_{\sigma,n}(y_s^{(i)}) - A_{\sigma,n}(\hat{y}_s^{(i)})) + (m + \delta)(A_{\sigma,n}(x_\tau) - A_{\sigma,n}(\hat{x}_\tau)),
$$

or

$$
\sum_{s=1}^{\sigma} \sum_{i=1}^{m} (A_{\sigma,n}(y_s^{(i)}) - A_{\sigma,n}(\hat{y}_s^{(i)})) = T_{\sigma,n} - \hat{T}_{\sigma,n} - (m + \delta)(A_{\sigma,n}(x_\tau) - A_{\sigma,n}(\hat{x}_\tau)).
$$

Hence,

$$
\mathbf{1}\{T_{\sigma,n} \leq \hat{T}_{\sigma,n}\} \sum_{s=1}^{\sigma} \sum_{i=1}^{m} (A_{\sigma,n}(y_s^{(i)}) + A_{\sigma,n}(\hat{y}_s^{(i)})) \leq (m + \delta)(A_{\sigma,n}(x_\tau) - A_{\sigma,n}(\hat{x}_\tau)).
$$

$$
(5.4.23)
$$

Substituting (5.4.23) in (5.4.22) yields,

$$Q_\sigma \leq \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{y_s^{(i)} \neq \hat{y}_s^{(i)}\} A_{\sigma,n}(\hat{y}_s^{(i)}) + (m+\delta)(A_{\sigma,n}(x_\tau) + A_{\sigma,n}(\hat{x}_\tau)).$$

Taking the conditional expectation with respect to the graphs G_σ and $\hat{G}_\sigma$ results in

$$\hat{\mathbb{E}}^\tau\Big[Q_\sigma \,|\, G_\sigma, \hat{G}_\sigma\Big] = \hat{\mathbb{E}}_\sigma^\tau[Q_\sigma] = \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{y_s^{(i)} \neq \hat{y}_s^{(i)}\} \hat{\mathbb{E}}_\sigma^\tau[A_{\sigma,n}(\hat{y}_s^{(i)})]$$

$$+ (m+\delta)\left(\hat{\mathbb{E}}_\sigma^\tau[A_{\sigma,n}(x_\tau)] + \hat{\mathbb{E}}_\sigma^\tau[A_{\sigma,n}(\hat{x}_\tau)]\right). \quad (5.4.24)$$

For any fixed value $x \in S$ and using (5.1.4), we have that

$$\hat{\mathbb{E}}_\sigma^\tau[A_{\sigma,n}(x)] = \hat{\mathbb{E}}_\sigma^\tau[F(|x - x_{\sigma+1}|)] = \hat{\mathbb{E}}_\sigma^\tau\left[\int_{y \in S} F(|x-y|)\,\mathrm{d}y\right] = I_n,$$

which, in turns, yields that (5.4.24) can be rewritten as

$$\hat{\mathbb{E}}_\sigma^\tau[Q_\sigma] = I_n \left(\sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{y_s^{(i)} \neq \hat{y}_s^{(i)}\} + 2(m+\delta)\right).$$

Note that (5.4.21) implies

$$\sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{y_s^{(i)} \neq \hat{y}_s^{(i)}\} = \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{v_s^{(i)} \neq \hat{v}_s^{(i)}\} + \sum_{s=\tau}^{\sigma} \sum_{i=1}^{m} \mathbf{1}\{v_s^{(i)} = \hat{v}_s^{(i)} = \tau\} \leq \Delta_\sigma + d_\sigma(\tau).$$

Thus,

$$\hat{\mathbb{E}}_\sigma^\tau[Q_\sigma] \leq I_n\Big(\Delta_\sigma + d_\sigma(\tau) + 2(m+\delta)\Big) = I_n\Big(\Delta_\sigma + (d_\sigma(\tau) + \delta) + (2m+\delta)\Big).$$

For convenience, we will use the following weaker statement:

$$\hat{\mathbb{E}}_\sigma^\tau[Q_\sigma] \leq I_n(\Delta_\sigma + (d_\sigma(\tau) + \delta) + (\hat{d}_\sigma(\tau) + \delta) + 2\Theta),$$

where we replaced $(2m+\delta)$ by 2Θ, see (5.1.8). Finally, by taking the expectation on both sides in the above display, we obtain the claim (5.4.17). $\qquad \square$

Bibliography

[1] L. ADAMIC AND B. HUBERMAN (2000). Power-law distribution of the world wide web. *Science* **287,** 2115a.

[2] W. AIELLO, F. CHUNG AND L. LU (2001). A random graph model for power law graphs. *Experimental Mathematics* **10,** 53–66.

[3] W. AIELLO, F. CHUNG AND L. LU (2002). Random evolution in massive graphs. *Handbook of Massive Data Sets* 97–122.

[4] R. ALBERT AND A.-L. BARABÁSI (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.

[5] R. ALBERT AND A.-L. BARABÁSI (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* **74,** 47–97.

[6] S. ASMUSSEN (1978). Some martingale methods in the limit theory of super-critical branching processes. In *Branching processes*. Advances in Probability and related Topics. pp. 1–26.

[7] T. AUSTIN, R. FAGEN, W. PENNEY AND J. RIORDAN (1959). The number of components in random linear graphs. *Annals of Mathematical Statistics* **30,** 747–754.

[8] A. D. BARBOUR AND G. REINERT (2006). Discrete small world networks. *Electronic Journal of Probability* **II,** 1234–1283.

[9] E. BENDER AND E. CANFIELD (1978). The asymptotic number of labelled graphs with given degree sequences. *Journal of Combinatorial Theory* **A24,** 296–307.

[10] N. BERGER, C. BORGS, J. CHAYES, R. D'SOUZA AND R. KLEINBERG (2005). Degree distribution of competition-induced preferential attachment graphs. *Combinatorics, Probability and Computing* **14,** 697–721.

[11] G. BIANCONI AND A.-L. BARABÁSI (2001). Bose-Einstein condensation in complex networks. *Physical Review Letters* **86,** 5632–5635.

[12] G. BIANCONI AND A.-L. BARÁBASI (2001). Competition and multiscaling in evolving networks. *Europhysics Letters* **54,** 436–442.

[13] N. Biggs, E. Lloyd and R. Wilson (1976). *Graph Theory/1736-1963*. Clarendon Press, Oxford.

[14] P. Billingsley (1968). *Convergence of Probability Measures*. Wiley, New York.

[15] B. Bollobás (1980). The asymptotic number of labelled graphs with given degree sequences. *European Journal of Combinatorics* **1**, 311–316.

[16] B. Bollobás (2001). *Random Graphs*. Cambridge University Press.

[17] B. Bollobás, C. Borgs, J. Chayes and O. Riordan (2003). Directed scale-free graphs. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*. Symposium on Discrete Algorithms. pp. 132–139.

[18] B. Bollobás, S. Janson and O. Riordan (2007). The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms* **31**, 3–122.

[19] B. Bollobás and O. Riordan (2004). The diameter of a scale-free random graph. *Combinatorica* **24**, 5–34.

[20] B. Bollobás and O. Riordan (2004). Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks*. pp. 1–34.

[21] B. Bollobás, O. Riordan, J. Spencer and G. Tusnády (2001). The degree sequence of a scale-free random graph process. *Random Structures and Algorithms* **18**, 279–290.

[22] C. Borgs, J. Chayes, D. Daskalakis and C. Roch (2007). First to market is not everything: an analysis of preferential attachment with fitness. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. pp. 135–144.

[23] T. Britton, M. Deijfen and A. Martin-Löf (2006). Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics* **124**, 1377–1397.

[24] F. Chung, S. Handjani and D. Jungreis (to appear). Generalizations of polya's urn problem. *Annals of Combinatorics*.

[25] F. Chung and L. Lu (2002). The average distances in random graphs with expected degrees. *PNAS* **99**, 15879–15882.

[26] F. Chung and L. Lu (2002). Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics* **6**, 125–145.

[27] C. Cooper and A. Frieze (2001). A general model of undirected Web graphs. *Lecture Notes in Computer Science* **2161**, 500–511.

[28] C. Cooper and A. Frieze (2003). A general model of web graphs. *Random Structures and Algorithms* **22**, 311–335.

[29] M. DEIJFEN, H. V. D. ESKER, R. V. D. HOFSTAD AND G. HOOGHIEMSTRA (to appear). A preferential attachment model with random initial degrees. *Arkiv för Matematik*.

[30] S. DOROGOVTSEV, J. MENDES AND A. SAMUKHIN (2000). Structure of growing networks with preferential linking. *Phys. Rev.* **85,** 4633–4636.

[31] R. DURRETT (2007). *Random Graph Dynamics*. Cambridge University Press, Cambridge.

[32] P. EMBRECHTS, C. KLÜPPELBERG AND T. MIKOSCH (1997). *Modelling Extremal Events: for insurance and finance*. Springer, Berlin.

[33] P. EMBRECHTS, C. KLÜPPELBERG AND T. MIKOSCH (1999). *Small Worlds, The Dynamics of Networks between Order and Randomness*. Princeton University Press, Princeton, New Jersey.

[34] P. ERDŐS AND A. RÉNYI (1959). On random graphs I. *Publicationes Mathematicae Debrecen* **6,** 290–297.

[35] G. ERGÜN AND G. RODGERS (2002). Growing random networks with fitness. *Physica A* **303,** 261–272.

[36] H. V. D. ESKER (2008). A geometric preferential attachment model with fitness. *preprint*.

[37] H. V. D. ESKER, R. V. D. HOFSTAD AND G. HOOGHIEMSTRA (2006). Universality for the distance in finite variance random graphs. *preprint*.

[38] H. V. D. ESKER, R. V. D. HOFSTAD, G. HOOGHIEMSTRA AND D. ZNAMENSKI (2006). Distances in random graphs with infinite mean degrees. *Extremes* **8,** 111–140.

[39] L. EULER (1741). Solutio problematics ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* **7,** 128–140.

[40] M. FALOUTSOS, P. FALOUTSOS AND C. FALOUTSOS (1999). On power-law relationships of the internet topology. In *SIGCOMM*. vol. 29. pp. 251–262.

[41] W. FELLER (1971). *An Introduction to Probability Theory and Its Applications Volume II, 2nd edition*. John Wiley and Sons, New York.

[42] D. FERNHOLZ AND V. RAMACHANDRAN (2007). The diameter of sparse random graphs. *Random Structures and Algorithms* **31,** 482–516.

[43] A. FLAXMAN, A. FRIEZE AND J. VERA (2004). A geometric preferential attachment model of networks. In *Algorithms and Models for the Web-Graph*. WAW. pp. 44–55.

[44] A. FLAXMAN, A. FRIEZE AND J. VERA (to appear). A geometric preferential attachment model of networks II. *Web Algorithms Workshop*.

[45] G. FORD AND G. UHLENBECK (1957). Combinatorial problems in the theory of graphs. IV. *Proceedings of the National Academy of Sciences of the United States of America* **43,** 163–167.

[46] E. GILBERT (1959). Random graphs. *Annals of Mathematical Statistics* **30,** 1141–1144.

[47] G. GRIMMETT AND D. STIRZAKER (2001). *Probability and Random Processes*. Oxford University Press, New York.

[48] A. GUT (2005). *Probability: a graduate course*. Springer, Berlin.

[49] O. HAGBERG AND C. WIUF (2006). Convergence properties of the degree distribution of some growing network models. *Bull. Math. Biol.* **68,** 1275–1291.

[50] P. HALL (1981). Order of magnitude of moments of sums of random variables. *Journal of the London Mathematical Society* **24,** 562–568.

[51] G. HARDY, J. LITTLEWOOD AND G. PÓLYA (1994). *Inequalities*. Cambridge University Press, Cambridge.

[52] W. HOEFFDING (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58,** 13–30.

[53] R. V. D. HOFSTAD AND G. HOOGHIEMSTRA (2007). Diameters in preferential attachment models. *preprint*.

[54] R. V. D. HOFSTAD, G. HOOGHIEMSTRA AND P. VAN MIEGHEM (2005). Distances in random graphs with finite variance degrees. *Random Structures and Algorithms* **27,** 76–123.

[55] R. V. D. HOFSTAD, G. HOOGHIEMSTRA AND D. ZNAMENSKI (2004). Random graphs with arbitrary i.i.d. degrees. *Mittag-Leffler Institute*.

[56] R. V. D. HOFSTAD, G. HOOGHIEMSTRA AND D. ZNAMENSKI (2007). Distances in random graphs with finite mean and infinite variance degrees. *Electronic Journal of Probability* **12,** 703–766.

[57] S. JANSON (2002). On concentration of probability. *Contemporary Combinatorics,* Bolyai Society Mathematical Studies **10,** 289–301.

[58] S. JANSON, T. LUCZAK AND A. RUCIŃSKI (2000). *Random Graphs*. John Wiley & Sons, New York.

[59] J. JORDAN (2006). The degree sequences and spectra of scale-free random graphs. *Random Structures and Algorithms* **29,** 226–242.

[60] Z. KATONA AND T. MÓRI (2006). A new class of scale free random graphs. *Statistics and Probability Letters* **76,** 1587–1593.

[61] R. KUMAR, P. RAGHAVAN, S. RAJAGOPALAN, D. SIVAKUMAR, A. TOMKINS AND E. UPFAL (2000). Stochastic models for the web graph. In *FOCS: IEEE Symposium on Foundations of Computer Science*. 41th IEEE Symposium On Foundations of Computer Science. pp. 57–65.

[62] R. KUMAR, P. RAGHAVAN, S. RAJAGOPALAN AND A. TOMKINS (1999). Trawling the web for emerging cyber communities. *Computer Networks* **31,** 1481–1493.

[63] R. LEPAGE, M. WOODROOFE AND J. ZINN (1981). Convergence to a stable distribution via order statistics. *Annals of Probability* **9,** 624–632.

[64] M. MOLLOY AND B. REED (1995). A critical point for random graphs with a given degree sequence. In *Proceedings of the sixth international seminar on Random graphs and probabilistic methods in combinatorics and computer science*. pp. 161–179.

[65] M. MOLLOY AND B. REED (1998). The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability and Computing* **7,** 295–305.

[66] M. NEWMAN (2003). The structure and function of complex networks. *SIAM Review* **45,** 167–256.

[67] M. NEWMAN (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46,** 323–351.

[68] M. NEWMAN, S. STROGATZ AND D. WATTS (2001). Random graphs with arbitrary degree distribution and their application. *Physical Review E* **64,** 026118.

[69] M. NEWMAN AND D. WATTS (1999). Scaling and percolation in the small-world network model. *Physical Review E* **60,** 7332–7342.

[70] I. NORROS AND H. REITTU (2004). On the power-law random graph model of massive data networks. *Performance Evaluation* **55,** 3–23.

[71] I. NORROS AND H. REITTU (2006). On a conditionally Poissonian graph process. *Advances in Applied Probability* **38,** 59–75.

[72] M. PENROSE (2003). *Random Geometric Graphs*. Oxford University Press.

[73] J. PICKANDS III (1968). Moment convergence of sample extremes. *The Annals of Mathematical Statistics* **39,** 881–889.

[74] B. PITTEL (1994). Note on the heights of random recursive trees and random m-ary search trees. *Random Structures and Algorithms* **5,** 337–347.

[75] A. RUDAS, B. TÓTH AND B. VALKÓ (2006). Random trees and general branching processes. *Random Structures and Algorithms* **31,** 186–202.

[76] H. Seyed-allaei, G. Bianconi and M. Marsili (2006). Scale-free networks with an exponent less than two. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* **73,** 046113.

[77] B. Söderberg (2002). A general formalism for inhomogeneous random graphs. *Physical Review E* **66,**.

[78] S. Strogatz (2001). Exploring complex networks. *Nature* **410,** 268–276.

[79] J. Szymański (2005). Concentration of vertex degrees in a scale-free random graph process. *Random Structures and Algorithms* **26,** 224–236.

[80] H. Thorisson (2000). *Coupling, Stationarity and Regeneration.* Springer, Berlin.

[81] J. Travers and S. Milgram (1969). An experimental study of the small world problem. *Sociometry* **32,** 425–443.

[82] D. Watts and S. Strogatz (1998). Collective dynamics of 'small world' networks. *Nature* **393,** 440–442.

[83] N. Wormald (1978). Doctoral thesis: Some problems in the enumeration of labelled graphs. *Newcastle University*.

Acknowledgments

Looking back, it seems a short time. Nevertheless, a lot has happened and, also, changed in the last four years. Many people deserve thanks for their help and support during my time as Ph.D. student. My mother's mother would say: "Dat heb je niet van jezelf! Je weet wat ik bedoel!" (translates to: "It is not yours! You know what I mean by that."), without naming to Whom it belongs to.

I would like to thank Michel Dekking for offering this Ph.D. position. Before I completed my master thesis in 2004, I was a teaching assistant for almost three years. I have helped students with PROSIM©, a unique application to build event driven simulations. Michel, one of the instructors of the course, suggested me to talk with Gerard Hooghiemstra about a Ph.D. position in the field of random graphs. I did and as a result I accepted this position and Gerard Hooghiemstra became my daily supervisor.

I would like to thank Gerard Hooghiemstra for his time and, especially, his patience. He often reminded me about the assumptions of the reader, but like a motorcyclist it is sometimes frustrating to obey the speed limit. Nevertheless, I accepted his 'signs' willingly, as a 'disciple' should. It was pleasant to work at the department of Statistics and Probability of the Delft University of Technology.

I wish to thank Remco van der Hofstad. His ideas and remarks are always remarkable. His insights, no matter the subject, do work enlightened. His passion for mathematics is contagious, and I hope that it will be for a long time.

I owe my gratitude to the members of the promotion committee for their useful comments on this thesis, where I keep in mind the allowed time they had to read it.

I would like to thank my mother, my brothers and sisters, my brothers-in-law and sister-in-law, my nieces and nephews, and all other family who all have shown a lot of interest into the progress of my research, although they did not grasp the subject of the research all the time.

To avoid long lists of names of friends to whom I need to show my gratitude, I just thank you all at once. Thank you!

Pets probably do confer health benefits, but I do not know precisely how. I would like to mention the deceased rats It, Dmitri, Sergei, Tarakh (aka Paplov) and Geer, and the living rats Goor and RK.

Henri van den Esker
Delft, April 3, 2008

Samenvatting

Uit onderzoek blijkt dat complexe netwerken, zoals het Internet, het World-Wide Web, sociale netwerken en andere technologische en biologische netwerken fascinerende overeenkomsten hebben. Een complex netwerk is te representeren door een **graaf**: een verzameling van **knopen** die onderling zijn verbonden met **kanten**. Het Internet wordt mogelijk gemaakt door routers die onderling verbonden zijn door kabels. De bijbehorende graaf wordt verkregen door de routers te vervangen door knopen en de kabels door kanten.

Veel complexe netwerken zijn "**kleine werelden**" (**small worlds**): de afstand tussen twee willekeurige gekozen knopen is klein. Verder blijkt dat de verdeling van de graad van een knoop een **machtswet** volgt: het aantal knopen met graad k is proportioneel met $k^{-\tau}$ voor een exponent $\tau > 1$. Zulk een netwerk wordt een **schalingsvrij** (**scale-free**) netwerk genoemd.

Aangezien de netwerken groot en complex zijn is het onmogelijk om voor zulke netwerken de corresponderende grafen exact te construeren. Dit kost te veel tijd en de benodigde opslag capaciteit is enorm. In de literatuur zijn vele modellen geïntroduceerd om complexe netwerken te simuleren door middel van **stochastische grafen**. Stochastische grafen beschrijven niet de precieze structuur van een complex netwerk, maar proberen juist de kenmerken van een complex netwerk na te bootsen.

In dit proefschrift wordt er onderscheid gemaakt in twee soorten stochastische grafen: de **statische** en de **dynamische** stochastische graaf. In een statische graaf start men met een vast aantal knopen waarbij er stochastisch kanten worden toegevoegd tussen de knopen. De resulterende grafen kun je interpreteren als het nemen van een foto van een complex netwerk op een vast tijdstip. Aan de andere kant tracht men met een dynamische random netwerk juist de groei van het complexe netwerk over de tijd te imiteren. Men start met een gegeven graaf en daarna voegt men één voor één nieuwe knopen toe aan de bestaande graaf. Deze knopen worden verbonden met de bestaande knopen. Als bestaande knopen met een hoge graad worden geprefereerd boven de rest, dan zal in de limiet de verdeling van de graad van een knoop een machtswet hebben. Dit groei mechanisme wordt "voorkeurs aanhaken" (**preferential attachment**) genoemd.

In dit proefschrift worden de volgende twee statische modellen bestudeerd: het **configuratie model** (CM) en de **inhomogene stochastische graaf**. Verder worden er twee dynamische modellen geïntroduceerd: de **preferential attachment random graaf met initiële stochastische graden**, welke afgekort wordt tot PARID model, en de **geometrische preferential attachment graaf met initiële**

aantrekkingskracht, welke afgekort wordt tot het GPAF model. In dit proefschrift wordt de verdeling van de graad van een knoop, de typische afstand, dat is de graaf afstand tussen twee willekeurige knopen, en de diameter, dat is de maximale afstand tussen elk paar van knopen, in de verschillende modellen bestudeerd. De typische afstand, en ook de diameter, wordt in veel gevallen slechts beïnvloed door de exponent τ van de machtswet. Als $\tau > 3$ dan zijn de resulterende grafen **homogeen**. Daarmee wordt bedoeld dat de structuur rond een knoop invariant is voor de positie van de knoop in de graaf. Als $\tau \in (2,3)$, dan zijn er knopen waarvan de graad hoog is ten opzichte van alle andere knopen in een graaf. Zo'n knoop fungeert als lokaal middelpunt of **hub**. Als $\tau \in (1,2)$, dan zijn er een eindig aantal knopen die als hub fungeren voor alle andere knopen in een graaf. Deze hubs beinvloeden de de typische afstanden en de diameter, omdat ze als shortcuts fungeren.

Het configuratie model is het eerste statische stochastische graaf model dat in dit proefschrift behandeld wordt. Een realisatie van het configuratie model bestaat uit een vast aantal knopen, waarbij elke knoop een stochastisch aantal half-kanten heeft. De graaf wordt geconstrueerd door steeds twee willekeurige half-kanten te verbinden tot een kant. Merk op dat het aantal half-kanten even moet zijn, anders blijft er aan het eind van het combineren een half-kant over. In dit proefschrift is de verdeling van het aantal half-kanten per knoop identiek en onafhankelijk verdeeld, waarbij de verdeling een oneindige verwachting heeft. Als het aantal knopen naar oneindig gaat, dan blijkt dat de afstand tussen twee willekeurige knopen 2 of 3. Het is mogelijk dat het aantal half-kanten van één enkele knoop het aantal knopen overtreft. Dit is voor de meeste complexe netwerken niet realistisch. Denk bijvoorbeeld aan het Internet. Daarom wordt er ook gekeken naar de volgende beperking: als er n knopen zijn, dan is het maximale aantal half-kanten per knoop maximaal n^α met α een gekozen waarde tussen 0 en 1. Het blijkt dat onder deze beperking de afstand tussen twee willekeurige knopen constant is en deze constante hangt af van de gekozen α.

De tweede statische stochastische graaf die behandeld wordt in dit proefschrift is de inhomogene stochastische graaf. In dit model ligt het aantal knopen vast. Elke knoop krijgt een gewicht, welk stochastisch is. Het aantal kanten tussen een tweetal knopen v en w is stochastisch en hangt alleen af van de gewichten van de knopen v en w. Het klassieke voorbeeld is de Erdős and Rényi stochastische graaf, welke wordt verkregen door elke knoop een identiek gewicht mee te geven. Tussen elk tweetal knopen is er maximaal één kant, en elke van de $\binom{n}{2}$ mogelijke kanten is aanwezig met kans p onafhankelijk van elkaar. In dit proefschrift worden de asymptotische fluctuaties van de typische afstand in de **Poisson stochastische graaf**, een model uit de literatuur, bestudeerd. De resultaten gelden echter ook voor de IRG, hetgeen bewezen wordt door middel van een koppeling.

Het PARID model is het eerste dynamische model wat in het proefschrift wordt geïntroduceerd. Het model wordt beschreven door een groeiproces. Initieel bestaat de graaf uit twee knopen, welke verbonden zijn door een willekeurig aantal kanten. Op elk discreet tijdstip wordt er een nieuwe knoop toegevoegd met een willekeurig aantal kanten. Elk kant wordt verbonden met een knoop uit de oude graaf, waarbij de kans dat een knoop met graad k wordt gekozen is

proportioneel aan $k + \delta$, waarbij δ een constante is. De δ is de **initiële aantrek-kingskracht** van een knoop. In dit proefschrift wordt de verdeling van de graad van een knoop afgeleid: een machtswet, waarbij de exponent elke waarde groter dan 2 kan aannemen.

Het tweede dynamische model is het GPAF model. Een stochastische graaf die ook de onderliggende geografische structuur van een complex netwerk in acht neemt. In vele modellen wordt deze structuur vergeten, maar er zijn complexe netwerken waar dit wel degelijk van belang is. In bijvoorbeeld een wireless ad-hoc netwerk, een netwerk van mobiele routers die wireless communiceren, kan de onderliggende structuur niet genegeerd worden. In zo'n ad-hoc netwerk zijn de routers verspreid over een gebied, en routers kunnen alleen met andere routers communiceren die in de buurt staan. In dit proefschrift wordt een variant op een geometrische stochastische graaf uit de literatuur gepresenteerd. Door de introductie van initiële aantrekkingskracht, zoals in het PARID model, wordt er een stochastische graaf verkregen waarbij de exponent van de machtswet elke waarde groter dan 2 kan aannemen. In het oorspronkelijke model is de onder-grens voor de exponent 3.

Summary

Empirical studies on real-life networks, such as the Internet, the World-Wide Web, social networks, and various types of technological and biological networks, show fascinating similarities. Many of these networks are **small worlds**, meaning that typical distances in the network are small, and many of them have **power-law degree sequences**, meaning that the number of nodes with degree k falls off as $k^{-\tau}$ for some exponent $\tau > 1$. Such a network is called a **scale-free** network.

In the literature many models are introduced to model real-life networks. The structure or topology of the network is modeled by a random graph consisting of **nodes** which are connected by **edges**. For example, the Internet consists of routers (nodes) which are connected to each other by wires (edges). One way to construct a random graph is to start with a fixed number of nodes and randomly add edges between pairs of nodes. The resulting graph is a snap-shot of the network at a given time instant. Using a growth model is a second way to construct a random graph. In such a model one starts with a given graph, and at each discrete time step a new node is added to the graph, from this node emanates a number of edges. The end-point of each edge is connected to one of the old nodes, where nodes with a high number of edges are preferred. This is called **preferential attachment**, and it is believed that preferential attachment is one of the underlying mechanisms, which generates scale-free random graphs.

In this thesis two types of random graphs are considered: **static random graphs** and **dynamic random graphs**. A static random graph aims to describe a network and its topology at a given time instant, and, on the other hand, a dynamical random graph aims to explain how the network came to be as it is. Dynamic random graphs often focus on the growth of the network under consideration as a way to explain the power-law degree sequences.

In this thesis **two static random graphs** are studied which produce power-law degree sequences: the **configuration model** (CM) and the **inhomogeneous random graph** (IRG). Furthermore, **two dynamic random graphs** are introduced: the **preferential attachment model with random initial degrees** (PARID) and the **geometric preferential attachment model with fitness** (GPAF). In this thesis the degree sequence, the typical distance (the asymptotic behavior of the graph distance between two uniformly chosen nodes), and the diameter for each of the models is considered. For example, the typical distance is solely influenced by the power-law exponent τ. If $\tau > 3$, then the graph is homogeneous or flat in the sense that all nodes have roughly the same kind of neighborhood. If $\tau \in (2, 3)$,

then nodes with a high degree will appear. These nodes act as hubs, and their impact on the connectivity cannot be neglected. If $\tau \in (1, 2)$, then a finite number of nodes act as super hubs for all the other nodes, the graph has a star-shaped structure. The hubs directly influences the typical distance, since they are used as short-cuts.

The first static random graph considered in this thesis is the **configuration model**. In this random graph the number of nodes is fixed, where from each node emanates a random number of stubs or half-edges. The random graph is constructed by connecting randomly pairs of stubs, where two different stubs combined give precisely one edge. Observe that the total number of stubs should be even, otherwise there will be one free stub in the end. In the thesis it is assumed that the expected number of stubs of each node is infinite. The distance between two randomly chosen nodes is either 2 or 3 as the number of nodes goes to infinity. It is possible that the number of stubs of a single node exceeds the number of available nodes, since its expectation is infinite. In real-life networks this not very likely. Consider for example the Internet, where connections between routers are identified with edges. Therefore, the same model is considered with the following restriction: if the total nodes is n, then the number of stubs of a single node is at most n^α, where α is some fixed value between 0 and 1. Under this restriction, it is shown, in the limiting case, that the distance between two randomly chosen nodes is with high probability a constant, which depends only on the value of α.

The inhomogeneous random graph (IRG) is the second static random graph that is considered in this thesis. The IRG consists of a fixed number of nodes, where each node has a random weight. The number of edges between each pair of nodes v and w is random and it only depends on the weights of the nodes v and w. The classical example is the Erdős and Rényi random graph, where each node has the same deterministic weight. In this model the number of edges between pairs of nodes is at most one, and each of the $\binom{n}{2}$ possible edges occurs independently with probability p. The asymptotically fluctuations of the typical distance is derived for the Poissonian random graph, a known random graph in the literature, and using a probabilistic coupling these results are extended to the IRG.

The first dynamic random graph introduced in the thesis is the PARID model, which is essentially a preferential attachment model. Initially, the graph is given by two nodes which are connected by a random number of edges. Then at each discrete time step a new node with a random number of stubs is added. Each stub is randomly connected to one of the old nodes by forming an edge, where nodes with a high number of outgoing edges are preferred. More precisely, an old node is chosen proportional to its degree plus some given **additive fitness** δ. By repeating this process, a graph of any size can be constructed. In this thesis the degree distribution of the PARID random graph is described, i.e., the distribution of the number of edges of a node, as the number of nodes goes to infinity. The power-law exponent can take any value bigger than 2.

Most random graphs ignore the underlying geometry of the network, which can be unrealistic. In for example wireless ad-hoc networks the geometry is of great importance, since in these networks nodes are spread over some surface

and nodes can only communicate with neighbors within a certain range, depending on the geometry. In this thesis a geometric preferential attachment model from the literature is used, which produces power-law degree sequences with exponents which can take any value bigger than 3. By extending this random graph one obtains the GPAF model, which is the second dynamic model introduced in this thesis. By introducing additive fitness, as done in the PARID model, it is shown that in the GPAF model the power-law exponent can take any value bigger than 2.

Curriculum vitae

Hendrik van den Esker was born in Zwolle on December 20th, 1980. From 1994 to 1999 he attended high school at De Heertganck, Heerde. In 1999 he went on to Delft to study Technical Mathematics at Delft University of Technology. For three years he worked as a teaching student at the department of Statistics and Probability. In 2004 he obtained his M.Sc. degree "cum laude" in Technical Mathematics. The master thesis was on overconvergence of some next-to-interpolation polynomials, mainly supervised by Dr. M.G. de Bruin. Hendrik van den Esker became acquainted with graph theory in the master course "Fractals & Random Graphs". From September of 2004 he started his research in the field of graph theory as a Ph.D. student at Delft University of Technology at the department of Statistics and Probability. Under the supervision of Prof.dr. F.M. Dekking as the promotor, and dr. G. Hooghiemstra as the daily supervisor. This research has led to the present thesis. The research was financed by by the Netherlands Organisation for Scientific Research (NWO). Hendrik van den Esker has a accepted a position at HiQ Invest.