TU Delft University of Technology

Effects of Fitness on Generative Graph Models with Preferential Attachment

Daan van Velzen

TUDelft Delft University of Technology

Effects of Fitness on Generative Graph Models with Preferential Attachment

by

D. P. van Velzen

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
in Applied Mathematics

with a Specialization in *Stochastics*

at the Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science

to be defended publicly on

Wednesday, November 2nd, 2022, at 10.00 a.m.

Thesis Committee

Chair & Supervisor:	Dr. J. Komjáthy	DIAM, Applied Probability
Second Reader:	Dr. N. Parolya	DIAM, Statistics

Student Information

Student Number: 4340213

Thesis Information

Keywords: *Networks, Probability, Graph Sequences, Preferential Attachment*

TU Delft

Preface

Before you is the end result of my year and a half thesis project, which marks the end of my college years. This thesis report, named *Effects of Fitness on Generative Graph Models with Preferential Attachment*, concerns itself with the qualitative behavior of *Preferential Attachment* models. After having written a bachelor's thesis about the phase transitions in Erdős–Rényi random graphs, the subject seemed a good fit, and after a year and a half of courses and an internship, I was excited to see how my new knowledge prepared me to interact with the ideas in the literature.

The preferential attachment model is a simple model abstracting an intuition which is true intuitively, i.e., *the rich get richer*. Yet, while the model itself is simple, the mathematical techniques used to prove results about them are incredibly varied and provided me with an opportunity to gain basic knowledge about urn models, reinforcement processes, branching processes and other models. To me, the joy in learning was to have my intuition proven wrong, as that is where mathematics is most interesting and where stochastics has a lot to offer. The process of learning about a phenomenon like condensation, not understanding a theorem's content, and then getting a feel for it only a short while later was very rewarding.

Having started a new job in early September, my cycles of confusion and understanding now concern public administration. The sense of wonder that comes with studying mathematics is dear to me, and something I hope to experience in working life as much as in study. My college years have shaped me to the person I am now, and I look back fondly on years of mathematics, friendship, music and life in two great university towns.

I would like to express my gratitude to all who have helped me progress towards my graduation; to Dr. Cipriani, who supervised me even after changing employment; and to Dr. Komjáthy for her patience, studying tips and clear communication. The progress towards graduation was challenging for me, but I feel I have improved a lot under your guidance. I also thank Dr. Parolya, for accepting his position as a second reader on the assessment committee. Lastly, I thank my parents, friends and Magda, for their support and confidence.

To the reader, I hope you enjoy the read!

Daan van Velzen
Leiden, October 2022

Abstract

Preferential Attachment models offer an explanation for why power laws are so common in real-world data. In these models, we start out with an initial network and add nodes one at a time. For each new node, we make m connections to existing nodes and *if* we define the attachment probability of attaching to a vertex to be proportional to its degree, we have defined the Barabási–Albert (BA) model. The tail of the degree distribution of the vertices of this model, decays like a power law, offering an explanation for why power law behavior is common in real world data.

The BA model can be adjusted by letting the probability of attaching to a vertex depend on its degree and an additional parameter for each vertex called fitness, that we randomly assign from a fitness distribution. If we adjust the attachment probability by multiplying with fitness, this model, the Bianconi–Barabási (BB) model, shows three qualitative behaviors. For flat distributions, the model behaves as the BA model, *the old get richer*; if fitness values differ, the exponent of the sublinear growth of the expected degree is larger if fitness is larger; if the tail of the fitness distribution is light enough, we find that a larger than expected proportion of half-edges connects to vertices with very high fitness values—condensation.

In this thesis we compare qualitative behaviors of various preferential attachment models. Chiefly, we concern ourselves with models with vertex fitness. For these models, we study how the fitness distribution affects qualitative behaviors, such as condensation.

First we review known results about the degree distributions of the graph. Degree distributions are an accessible and understandable property that is related to the topology of the network. We do this for the model without vertex fitness, for multiplicative fitness and for models with additive fitness and define the regimes of their qualitative behavior. For these results, we give clarification of the behavior and highlight the principles employed to derive their proofs. In doing so, we gather the commonalities of these proofs and provide intuition about their underlying principles by accenting the recurring themes.

Lay abstract—Inleiding in het Nederlands

De langste persoon op aarde is 2, 51 m lang, ongeveer de helft langer dan de gemiddelde lengte. De rijkste persoon op aarde—Elon Musk—heeft bezittingen met een totale waarde van 247, 9 mrd. dollar, ongeveer 30 miljoen keer meer dan het wereldwijde mediane vermogen.

Wie als man 1, 98 m lang is, is langer is dan 98% van de mannelijke Nederlanders (gemiddeld 1, 84 m), en zal vermoedelijk nooit iemand tegenkomen nog eens zoveel langer is. Veel minder zeldzaam is het voor een rijke om iemand te ontmoeten die twee keer zo rijk is. Dit is een eigenschap van de verdeling van vermogen volgens een ontdekking die in de 19^de^ eeuw werd gedaan door de Italiaanse econoom Pareto [1]. Hij ontdekte dat vermogen volgens een machtsverdeling verdeeld is. Iemand tegenkomen die twee keer zo rijk is, is 2^γ keer zo zeldzaam (voor een of andere waarde van γ) en de rijkste 20% van de mensen heeft 80% van het vermogen in handen. Deze machtsverdelingen blijken veel voor te komen. Je vindt ze bijvoorbeeld als je telt hoe vaak woorden in een tekst voorkomen of als je de hyperlinks die naar een webpagina gaan zou tellen.

Met de komst van het internet was er grote interesse naar de structuur van het www. Onderzoekers maakten web-crawlers, programma's die op onderzoek uitgingen om de hyperlinks op pagina's te tellen. Ook hier werder machtsverdelingen gevonden. Vlak voor het jaar 2000 bedachten twee onderzoekers in Amerika, Barabási en Albert, een verklaring voor de aanwezigheid van deze verdeling. Zij dachten dat pagina's waar al veel naar gelinkt werd, waarschijnlijk ook gemakkelijker nieuwe links zouden krijgen. Dit wordt *Preferential Attachment* genoemd, maar is ook bekend als het adagium *de rijken worden rijker*.

Het wereldwijde web, met zijn pagina's en verbindingen, vormt een netwerk. Barabási en Albert onderzochten wat de eigenschappen zijn van een netwerk dat volgens het *preferential-attachment*-principe is opgebouwd. Het bleek inderdaad dat dit principe tot machtsverdelingen leidt.

Helaas bleef bij hun netwerk een van de eerste knooppunten vaak de rijkste. Dat is niet realistisch. In de echte wereld neemt een nieuwkomer de eerste plek vaak over van een gevestigde naam. Denk bijvoorbeeld aan de zoekmachine Google, die Yahoo uit de markt dreef.

Bianconi en Barabási bedachten een aanpassing om het model realistischer te maken. De kans dat een nieuwe pagina zich met een bestaande pagina zou verbinden, werd naast het huidige aantal inkomende links, ook afhankelijk van een vermenigvuldingsfactor die de *fitness* ging heten. Deze aanpassing maakte het mogelijk dat fittere nieuwkomers een grote inhaalslag konden maken—*the fit get richer*.

Als we iedere nieuwe binnenkomer een willekeurige fitness meegeven, dan kan het afhankelijk van de spreiding van de fitness iets vreemds gebeuren met de machtsverdeling. Waar bij evenwichtige spreiding het aantal verbindingen met pagina's zich voorspelbaar gedraagt, blijkt dat bij een spreiding met maar weinig hoge fitnesswaarden een overmatig deel van de nieuwe pagina's zich met met de pagina's met hoge fitneswaardes verbindt. Dit verschijnsel, condensatie, is vernoemd naar een parallel met de kwantummechanica.

In deze scriptie geef ik een weergave van overgangen in het gedrag van deze *preferential-attachment*-modellen. We bekijken de netwerken los van hun context. Deze scriptie gaat dus niet over pagina's en hyperlinks, maar knopen en zijden. Zo kunnen we op een algemene manier resultaten presenteren over bijvoorbeeld condensatie. Daarnaast geef ik een overzicht van de manieren waarop deze resultaten bewezen kunnen worden.

Contents

Chapter 1

Introduction

In many regards, the start of modern network science begins with an urban legend. Are any two people on earth separated by only six handshakes? This is the question at the heart of the lesser known experiment by S. Milgram [2, 3]. In the experiment respondents from Nebraska and Boston were asked to send a package to an unknown target in Massachusetts by only sending the package to their acquaintances. For those packages that reached the right addressee, the number of steps to reach their destination, was on average only 5.2. This was a strong attestation to the idea that we are living in a "small world." While this phrase was not yet used in the 1960's, it would become a central concept in the field of *Complex Networks*, an interdisciplinary field that concerns itself with understanding the structure and dynamics of real-world networks.

In the sixties, however, the theory that would give an explanation of the phenomenon of small-world networks was not yet developed. Yet, several scientific fields already studied the phenomena that would be unified; the mathematical community studied random networks [4], the natural sciences had done research into the appearance of power-laws [5], the social sciences were mapping the structure of social relations [6] and the study of networks of articles had taken place [7]. These seemingly unrelated developments would later come together in the emergence of a new field of study, that of *Network Studies*. This field would study the structure of the networks in the individual disciplines, separate from the context and attempt to identify the features that unify them and the mechanisms by which these features emerge. We will begin with a historical overview of the research that predates modern network studies.

1.1 Historical overview of research preceding network science

We will begin the exposition with the early research of two concepts that would later find unification. These are that of *small worlds* and *power law* degree distributions.

1.1.1 Small worlds

Even though the research of Travers and Milgram [2] may have popularized the concept of small worlds, the idea was not originally conceived by them. In fact the concept does not originate in science but in fiction, in a short story by Hungarian author Karinthy [8]. This writer proposes that technology is making the world smaller. This sentiment strikes true to nearly anyone who at one time found out that two seemingly unrelated friends happen to know each other. The first attempt at a mathematical explanation of the phenomenon was made in the 50's by de Sola Pool and Kochen [9], but was not published until 1978. The small-world effect was however spread as an urban legend and captured the imagination of many, among whom the playwright John Guare, who wrote the 1990 play *Six Degrees of Separation*. The name of this play became another common name for the small-world effect.

In fact, the name of this play seems to have inspired a party game of the 1990's, called the six degrees of Kevin Bacon. In this game, the goal is to connect any actor or actress to Kevin Bacon by making a chain of movie collaborations between them, with five steps or fewer. Kevin Bacon seemed a good candidate for this game, because of his extensive collaborations across a variety of different styles of movie. Later it would turn out, that the existence of well connected hubs in a network is not particularly noteworthy in itself, because hubs occur in many types of networks.

Similarly, within the mathematical community, the collaboration distance to P. Erdős, a very productive author and himself an early founder of random graph theory, is often called the Erdős number and stated with considerable pride—Erdős himself, and close associates are revered for their work. However the random graphs first studied by Erdős and Rényi [4] do not make a particularly good model for real-world networks, because the degrees of the vertices in the model are Poisson-distributed.

1.1.2 Power laws

That power-law distributions, i.e. $p(k) \propto k^{-\gamma}$, occur in nature and language has been well established. It would however take computer analysis in the late 1990's to find out how common it is for the distribution of the degrees of a network to be a power law. Until that point in time the occurrence of power laws had been noted and attempts had been undertaken to explain why they appeared. We will here give a historical overview.

Two famous examples of distributions that follow a power law are the distribution of wealth in society [1], and the frequency of word usage in texts of natural languages. The linguist Zipf notes:

> In any extensive sample of connected English, it will, in all probability, be found that the most frequent word in the sample will occur on the average once in approximately every 10 words, [...], the nth. most frequent word once in every 10n words; in brief, the distribution of words in English approximates with remarkable precision an harmonic series [10]. (A)

One attempt to provide an explanation for power-law distributions comes from biology. In 1925, an article by Yule [5] explains the phenomenon that few biological genera comprise a large number of species. Moreover, the relation between the size of a genus and the frequency of genera of the size follows a power law. Yule posits that the chance of a mutation adding new species to a genus is proportional to the number of species in the genus. His assumptions about mutation (and extinction) allow the derivation of the power-law relation. The assumption that the probability of growth was proportional to the size of the genus would later turn out to be vital in network studies.

In contrast to [5], Zipf proposes that the rank-frequency distribution of words in natural languages is not the result of chance, but of optimization [11, 12]. Struck by the emergence of a harmonic series in the frequency data, he ponders whether a relation to harmonic series in natural sciences exists [12]. In [11], Zipf makes a non-mathematical attempt to argue for the optimality of the power-law. Zipf posits and argues that human behavior obeys a *principle of least effort*. In language this yields a conflict—it is optimal for a speaker to have one word that could convey any meaning, and for the listener to hear many words that bring precisely one meaning across. This antithesis is provided as a driving force that leads to the rank-frequency distribution. While Zipf does not mathematically support the occurrence of power laws in linguistics, he does provide such an argument for the sizes of human settlements. This analogy is supported by showing that an optimization problem between the size of a settlement and the proximity of resources would lead to a power law.

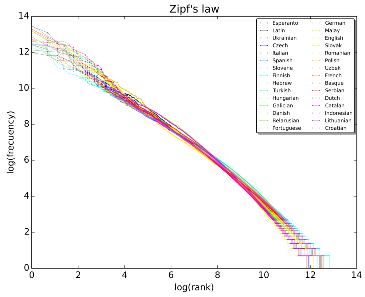


Figure 1.1: Log-log graph of word frequency vs. word rank for wikipedia pages in 30 languages. SergioJimenez, CC BY-SA 4.0, via Wikimedia Commons

The publication of the works of Zipf roughly coincides with the foundation of a new scientific discipline—that of information theory in [13]. The problem of the rank-frequency distribution of words in natural language is thus given new impetus around the turn of the fifties. Mandelbrot [14] presents a refinement on [11]; in fact reasserting that the power law is optimal in information theoretical sense. Mandelbrot's solution is optimal in that it minimizes the average cost per unit of information.

Simon leads the argument for an opposed organizational principle for rank-frequency distributions of words in natural language—chance. In [15], the author revisits the argument made by Yule [5] and refines it noting that Yule's article predates the modern science of stochastic processes. Simon simplifies the model of Yule and states

> The probability that the $k+1$-st word is a word that has already appeared exactly i times is proportional to $i \cdot f(i,k)$—that is, to the total number of words that have appeared exactly i times. [15] (B)

This model is strikingly close to the modern-day model we will study in this thesis. Interestingly, both the explanations of Zipf and Mandelbrot, and that of Simon offer an explanation for why the tails of the distribution of word frequencies follow a power law. The opposing views would lead to an eight year back and forth discussion between Mandelbrot and Simon who published their rebuttals in the journal "Information and Control". The debate concerning optimization and chance did not settle conclusively, but we will see how the side of change gained new impetus in the late 1990's, with the study of the structure of the world wide web. Before then, a model closely related to that of Simon would emerge.

The study of citation networks takes off in 1965 when de Solla Price [7] questions: "More work is urgently needed on the problem of determining whether there is a probability that the more a paper is cited the more likely it is to be cited thereafter." In a 1976 the same author formulates an answer [16]. He proposes a model of *Comparative Advantages* and asserts it to produce a degree sequence that follows a power law. Price's model would in current literature be called a directed preferential attachment graph. Price however does not originally give his model the form of a graph, instead describing it as an urn model or stochastic process.

1.2 The breakthroughs of network science

We have seen how observations in economics, sociology, linguistics and bibliometrics lead to the concepts of small worlds and the discovery of power-law distributions. The individual fields had however not been unified by finding a common language. A graph, essentially a set of points and connections between those points, is the mathematical answer to talking about network structure. Although graph theory dates back to Euler, early graph theorists did not concern themselves with statistics on networks or finding ensembles of graphs that exhibit some statistical property. Later, Erdős and Rényi [4] studied *random graphs*, but this ensemble of graphs is a poor match for networks in the real world.

1.2.1 Study of real world networks

In an overview paper of Network Science by Albert and Barabási, three concurrent developments that allow the study of structural (i.e. topological) features of real world data are named [17];

- The advancements in computing allowing automated data acquisition.
- More computing power enabling the analysis of this data.
- Increase in publicly accessible databases of research data across field boundaries.

The development of network science thus seems intimately linked to the advent of the computer age. The internet, which itself has a network structure of pages (points) and hyperlinks (connections), promised to be an exciting ingress of study. Faloutsos *et al.* [18] were among the first to discover that the both the number of links between domains and the number of connections between routers (Fig. 1.2) are distributed according to a power law with exponent between 2 and 3 (Fig. 1.3). Moreover, they discovered that this exponent appears to be constant in time. Their method, to follow connections between pages and routers with a so-called web crawler, was rapidly becoming more popular. This explains why the year 1999 sees a number of publications about networks in Nature, a very prestigious journal of the natural sciences [19, 20]. The internet is one of many examples of data sets that was studied in the 90's. We include

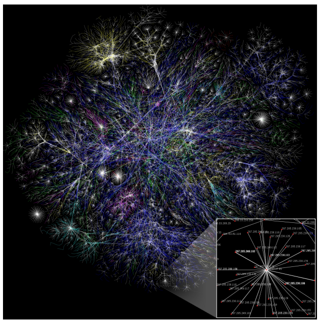
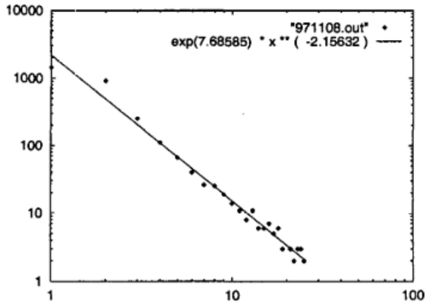
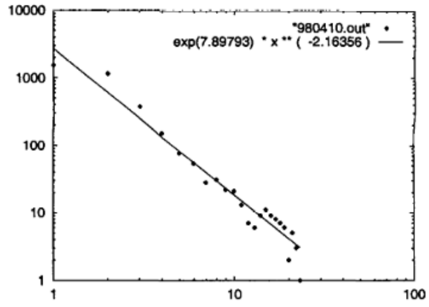


Figure 1.2: Visualization of IP-adresses and connections between IP-adresses connected over the internet. Color corresponds to geographical location, white being the internet backbone. The Opte Project / Barrett Lyon CC BY-SA 4.0

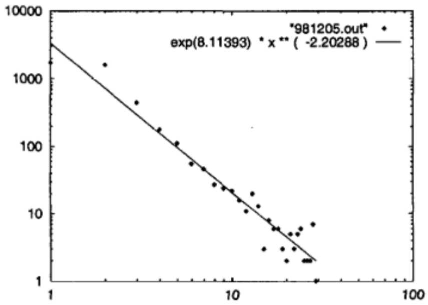
a table Table 1.1 of various data sets with their estimated power law exponents and average minimum path length. We see that the network of routers connected by internet protocol, hyperlinks between domains, metabolic relations in E. Coli, cooperation between actors and cooperation between authors in mathematics follow an approximate power law and appear to be form small world.



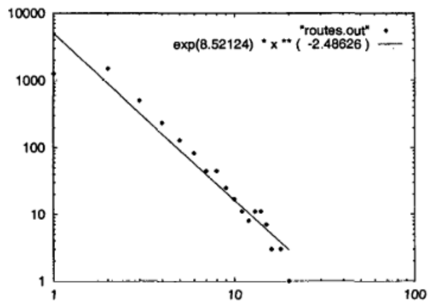
(a) The internet, November 1997



(b) The internet, April 1998



(c) The internet, December 1998



(d) Routers in 1995

Figure 1.3: Four log-log plots from Faloutsos *et al.* [18], describing the relation between the log of the total degree (x-axis, described as log of the out-degree in the article) and the log of the frequency that this degree is found (y-axis). Figures 1.3a to 1.3c, describe the number of hyperlinks between domains (not discerning the sub-domains after the /) in the internet at the respective dates. Figure 1.3d, describes the connections between routers in the internet in 1995.

Network	N	$\langle k \rangle$	L	C	γ	Source (original)
Routers	228 263	2.80	9.5	0.03	2.18	[21–23]
WWW	$\sim 2 \times 10^8$	7.5	16	0.11	2.1/2.7	[24]
Metabolic	778	3.2	7.40	0.7	2.2/2.1	[19]
Actors	225 226	61	3.65	0.79	2.3	[25, 26]
Math1999	57 516	5.00	8.46	0.15	2.47	[27, 28]

Table 1.1: Statistical features of real world networks. [29, verbatim, pp. 188] Size N, average degree $\langle k \rangle$, average shortest path length L, clustering coefficient C, power law exponent γ (in-degree/out-degree)

Although not all authors are convinced that these networks have power-law distributions—favoring log-normal distributions instead [30]—there has been wide interest in studying the principles that generate this structure. We can safely assume that the internet domain structure neither was designed, nor evolved to have a power-law tail. This suggests that the school of Simon may have been right to assert that chance gives the world this order. But if chance does so, then what generative process gives us graphs with the same structural characteristics? The answer to this is important. If we know how to generate graphs then we can answer questions about epidemiology, network reliability and the spread of rumors.

If, as an example, we want to know the rate of spread of an illness in a social network, then first mapping the exact network would be prohibitively time consuming and possibly unethical. It may instead be better to simulate a network with sufficiently realistic structure and then model the spread of disease on that. Researchers in network science have set out to find generative models that closely resemble the real world.

1.2.2 Watts–Strogatz model

Determined to describe a class of graphs that better matches real-world examples, Watts and Strogatz devised a method to interpolate between random graphs and regular lattices (Fig. 1.4). Their model turned out to be a good example of a small-world graph. In the classical random graph model by Erdős and Rényi [4] we consider a fixed number of points and consider each possible connection between pairs of points which we then link with a fixed, small probability. This type of graph is not very clustered—neighbors of a single point are only rarely connected. This is opposite to real world networks, which tend to have high clustering and a small diameter. An ensemble of random networks with these properties is given by Watts and Strogatz [25]. In this model we start out with a regular graph (Fig. 1.4) and rewire each link to a random other point with a fixed probability. The resulting network is a small world and is clustered, but there is no power law. The Watts–Strogatz model for small worlds marks a start of the current era of network modeling. Their seminal article was published in Nature and well read, leading others to follow their lead.

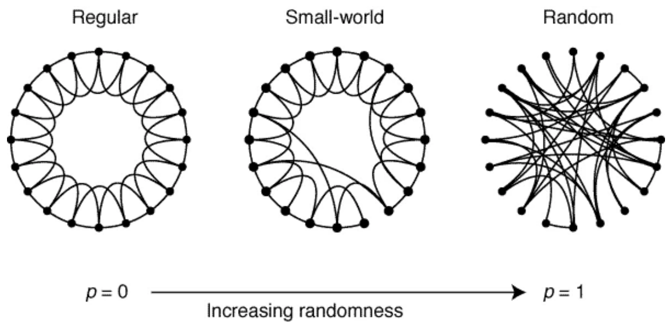


Figure 1.4: The Watts–Strogatz model [25]. Re-used with permission of publisher.

1.2.3 Scale-free networks

Shortly after the Watts and Strogatz article, we see the publication of generative models that exhibit power-law tails and the formation of hubs—two properties loosely referred to as scale-freeness—in a paper by Barabási and Albert [31]. While the model is not substantively different from the model of de Solla Price [16], it is now explicitly applied to networks and exhibits scale-freeness. This model is based on the principle of *preferential attachment* expressed in Quote B—the probability that a new node added to the graph attaches to a node i, is proportional to the degree of i.

This model has been popular to fit to real world data ever since its introduction, and many variations exist to make the fit even better. The properties of the whole of the graph are not caused by meticulous design, but are a result of the (random) way by which new nodes are added to the network. The principles that govern the evolution of the model, decide the properties of the networks it generates. This principle is that of the *Matthew effect*, *rich get richer*, or *preferential attachment*.

The preferential attachment principle does however have a shortcoming in the modeling of real world data. In a network that is solely governed by preferential attachment the older nodes will typically have more connections than the newer nodes. Many examples exist of networks in which newcomers become the best connected. A common example [32] is the popularity of the up-and-coming search engine Google at a time when Yahoo was the undisputed leader in internet searches. Here we see how newer hubs overtake older ones. In order to model this, Bianconi and Barabási [33] introduced a per node fitness, describing the intrinsic capability of a node to gain more connections. This parameter adds a new qualitative behavior described by the adage *fit get richer*. The addition of this parameter not only allows us to model the real world more accurately, but also brings interesting new qualitative behavior richer than only *fit get richer*. This behavior is called *Winner takes all* or *condensation* and has a connection to the physics of Bose–Einstein gases [34].

1.3 Goal of this thesis and outline

In this thesis we set out to give a comparison of various Preferential attachment models. Specifically we focus on the role of vertex fitness and how this vertex fitness influences degree distributions. We focus on degree distribution as an accessible object of study relating to network topology. We distinguish models without fitness, with multiplicative fitness and models with additive fitness and define the regimes of qualitative behavior that they exhibit. We provide an overview of proof techniques for known results in the literature and perform simulations to provide examples of the qualitatively different behaviors that occur. By studying proof techniques, we gather the commonalities of preferential attachment model proofs and attempt to make intuitive their underlying principles by accenting similarities. Moreover, we provide clarification to proofs and examples of model behavior.

In Chapter 2, we set out to describe the various variants of preferential attachment graphs; the seminal Barabási–Albert model, the Bianconi–Barabási model which adds fitness and the models with additive fitness and non-linear preferential attachment.

In Chapter 3 we study the various proof techniques used in proving the statistical properties of the networks defined in Chapter 2. We elaborate on the more difficult parts of some proof and attempt to highlight the usefulness and difference between techniques. Moreover, we provide examples and visualizations of graphs exhibiting behavior of the various qualitative regimes.

Chapter 4 is dedicated to reflect on the differences in behavior and proof method, highlighting similarities and the extent to which the methods can be used. Finally we look at the limits of current research and highlight avenues of further research.

Chapter 2

Model definitions

The language of network models in mathematics is that of graphs. In order to define the models that we study in this thesis, We will first introduce notation and basic definitions of graphs and related concepts. We then define the various models that we discuss in this thesis. First we visit the Barabási–Albert model and Bianconi–Barabási model—two original preferential attachment models—respectively without and with multiplicative fitness parameter. We then look at preferential attachment with additive fitness and preferential attachment with non-linear fitness.

2.1 Preliminaries

There are several notable variations of preferential attachment graph models. Each model produces a graph sequence $(\mathcal{G}_n)_{n\geqslant 1}$, by defining the probability $\mathbb{P}(v_{n+1} \to v_j \mid \mathcal{G}_n)$ that a new vertex v_{n+1} attaches to vertex v_j, with $j \in \{1, \dots, n\}$. We first define what a *graph* is.

Definition 2.1 (Graph)**.** A *graph* is an ordered pair $\mathcal{G} = (V, E)$ of a set V of *vertices* (also called *nodes*) and a set $E \subseteq V^2$ of *edges*, where an edge $e \in E$ denotes a connection between two vertices. «

Graphs can either be directed or undirected, meaning that their edges either do or do not have notion of direction;

Notation 2.1. We distinguish so called *directed* graphs $\mathcal{G} = (V, E)$, where for $u, v \in V$, with $u \neq v$, the edge $(u, v) \neq (v, u)$. An element $e \in E$ can also be called an *arrow*, or *arc*. We write $u \to v \Leftrightarrow (u, v) \in E$.

Alternatively, a graph can be *undirected*. In this case $(u, v) = (v, u)$. An edge can also be written as $e = \{u, v\}$ to accent that the order is immaterial. We write $u \to v \Leftrightarrow (u, v) \in E \vee (v, u) \in E$. «

The probability $\mathbb{P}(v_{n+1} \to v_j \mid \mathcal{G}_n)$, describing the probability that vertex v_{n+1} connects to v_j is proportional to an increasing function of the number of connections to v_j at time n is commonly called the *attachment probability*. The transition from $\mathcal{G}_n$ to $\mathcal{G}_{n+1}$ consists of two general phases.

(i) Add vertex v_{n+1} to the set of vertices.
(ii) Add m edges between v_{n+1} and other vertices to the set of edges.

We discern three general forms of the attachment probability $\mathbb{P}(v_{n+1} \to v_j \mid \mathcal{G}_n)$. These variants are dependent on a function $f(\deg^{\bullet}_{\mathcal{G}_n}(v_j))$, where $\deg^{\bullet}_{\mathcal{G}_n}(v_j)$ is the number of connections to v_j and can be the in-degree or total degree of the vertex v_j. This function, called the

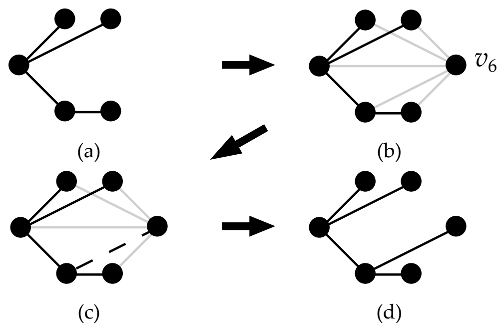


Figure 2.1: Process of growing the graph sequence $(\mathcal{G}_n)_{n\in\mathbb{Z}}$. We see the initial graph $\mathcal{G}_5$ as an example in (a). With will grow this graph as described in Items (i) and (ii) with $m = 1$. As $\mathcal{G}_5$ evolves to $\mathcal{G}_6$, we first add the new vertex v_6 (see (b), Item (i)). The grey lines represent possible edges to connect v_6. Here we assume we are in a model where self loops do not occur. In step (c), we choose one of the possible lines (see Item (ii)). In (d) this line is added to $E_{\mathcal{G}_6}$.

attachment kernel, must be increasing in the number of connections to v_j. The degrees of vertices are integer valued, and the value of $f(k)$, with $k \in \mathbb{Z}_{\geqslant 1}$ is often written as A_k. The attachment probability to vertex j is also possibly dependent on a value $\mathcal{F}^{(j)}$, called the *fitness* of vertex j. We then distinguish three forms of attachment probability for a vertex v_j of degree k:

(i) Models without fitness, in which $\mathbb{P}(v_{n+1} \to v_j \mid \mathcal{G}_n) \propto A(k)$.
(ii) Models with multiplicative fitness, in which $\mathbb{P}(v_{n+1} \to v_j \mid \mathcal{G}_n) \propto \mathcal{F}^{(j)} \cdot A(k)$.
(iii) Models with additive fitness, in which $\mathbb{P}(v_{n+1} \to v_j \mid \mathcal{G}_n) \propto A(k) + \mathcal{F}^{(j)}$.

In the theoretical study of preferential attachment models, the fitness values are often random variables drawn from a distribution μ, called the *fitness distribution*. In the modeling of real world networks, a fitness value represents characteristics intrinsic to the vertex, that influence the ability of the vertex to attract new links.

We can see that for each of the enumerated forms of attachment probability on page 9, the attachment probability of vertex v increases when v has more connections. The number of connections to v can be given as a *degree* or *in-degree*, which we will define here.

Definition 2.2 (Degree)**.** Let $\mathcal{G} = (V, E)$, be a graph. The *degree* $\deg_{\mathcal{G}}(v)$ of vertex $v \in V$ is

$$\deg_{\mathcal{G}}(v) = \sum_{e=(a,b)\in E} \Big(\mathbb{1}_{\{a\}}(v) + \mathbb{1}_{\{b\}}(v) \Big) . \tag{2.1.1}$$

«

This means, that for vertex v, we count all edges $e \in E$ for which v is their origin or endpoint. Moreover, self loops contribute doubly towards the degree. We contrast this to the *in-degree*, which is the number of edges that have v as their endpoint.

Definition 2.3 (In-degree)**.** Let $\mathcal{G} = (V, E)$, be a graph. The *in-degree* $\deg^-_{\mathcal{G}}(v)$ of vertex $v \in V$ is

$$\deg^-_{\mathcal{G}}(v) = \sum_{e=(a,b)\in E} \mathbb{1}_{\{b\}}(v). \tag{2.1.2}$$

«

The precise definition of the models under study requires that we specify the following details regarding the nature of the graph construction. In this model we only consider so called *multigraphs*.

Remark 2.2. For E the notion of a set is generally extended to mean *multiset*. This means that an edge $e \in E$ can occur multiple times, e.g., $E = \{e, e\}$ is a valid set. The number of instances of e in E is called the *multiplicity*. «

A graph $\mathcal{G} = (V, E)$, with multiset E is called a *multigraph*. Next, we must consider how the general process that is defined in the preferential attachment models is started.

Remark 2.3 (Initialization). The general "algorithm" for the construction of preferential attachment networks does not specify the initial network, typically $\mathcal{G}_1$, from which the iterative construction is started. For the derivation of results it is often needed to specify the *initialization*—the graph from which construction is started. «

Moreover, we note that graphs models, depending on their exact generation process may or may not have *self loops*—edges of the form $e = (v, v)$ for some vertex v.

Remark 2.4 (Self loops). Many previous articles have initialized the model with a graph consisting of a vertex with a single self-loop—an edge $v \to v$. The existence of self-loops depends on the precise specification of the generation process. «

Finally, models differ in the following sense.

Remark 2.5 (Intermediate updating). In the second construction step, the adding of m edges, there are two options concerning the attachment probability. We may either add m edges independently with the same probability distribution, i.e., *simultaneously*. Alternatively we can update the attachment probability based on the new degrees of the nodes after each added edge. This is called *intermediate updating*. «

In what follows, we present the most commonly studied preferential attachment models, that together give a good overview of the behaviors that these models display.

2.2 Barabási–Albert model

The Barabási–Albert (BA) model marked renewed interest into the preferential attachment paradigm [31]. Basic properties of the model had been known for longer [15], although the model was not described as a graph model then. Barabási and Albert [31] repopularized the model and stated it in the form presented here. Barabási and Albert, two physicists, also gave a proof for the degree sequence of the model. This proof is intuitively correct, but attracted attention from mathematicians, who disagreed with the proof method an attempted to improve its rigor.

The model's degree sequence was first exhaustively analyzed by Bollobás *et al.* [35]. The authors of the latter article note that the original model definition by Barabási and Albert is "rather imprecise", and they give a more precise model, which is suited to their proof method. The mathematically precise definition below has been taken from an article by Bollobás and Riordan [36] and van der Hofstad [37]. , The Barabási–Albert model is a generative graph model without fitness. The graph is initialized as a single vertex with one loop. The graph sequence $(\mathcal{G}_n)_{n \geqslant 1}$ is constructed recursively. The model is undirected, and allows both self-loops and multi edges. We first limit ourselves to the case that $m = 1$, i.e., vertices are connected by a single edge.

Definition 2.4 (Barabási–Albert model for $m = 1$)**.** Given a graph $\mathcal{G}_n$, we form the network $\mathcal{G}_{n+1}$ by appending node v_{n+1} and randomly choose a single edge $(n+1, i)$ with probability

$$\mathbb{P}(v_{n+1} \to v_i \mid \mathcal{G}_n) = \begin{cases} \frac{\deg_{\mathcal{G}_n}(v_i)}{2n+1} & \text{if } 1 \leqslant i \leqslant n, \\ \frac{1}{2n+1} & \text{if } i = n+1. \end{cases} \tag{2.2.1}$$

The chosen edge is then appended to E. «

The Barabási–Albert model with number of added edges $m > 1$ is constructed via collapsing, we here give a definition of this procedure.

Definition 2.5 (Collapsing vertices in graph $\mathcal{G}$)**.** Let $\mathcal{G}' = (V', E')$ be a graph, and let $v', w' \in V'$, be two distinct vertices. The graph $\mathcal{G} = (V, E)$, obtained by *collapsing* v' and w' satisfies the following properties:

- $V = V' \setminus \{v', w'\} \cup \{v\}$
- For all edges $e = (u'_1, u'_2) \in E'$, we add a corresponding edge to E, with the following rules.
  - If $u'_1 \in \{v', w'\} \wedge u'_2 \notin \{v', w'\}$, then (v, u'_2) is added to E.
  - If $u'_1 \notin \{v', w'\} \wedge u'_2 \in \{v', w'\}$, then (u'_1, v) is added to E.
  - If $u'_1 \in \{v', w'\} \wedge u'_2 \in \{v', w'\}$, then (v, v) is added to E.
  - If $u'_1 \notin \{v', w'\} \wedge u'_2 \notin \{v', w'\}$, then (u'_1, u'_2) is added to E. «

We note that we can collapse multiple nodes, e.g., u', v', w' by first collapsing u', v' to vertex v and then v, w' to w. This process can be extended for any number of vertices and the order of collapsing does not influence the structure of the resulting graphs. After this repeated collapsing, we can naturally associate vertices $u'_1, \ldots, u'_n$ to u_n.

We can now define the Barabási–Albert model for $m > 1$.

Definition 2.6 (Barabási–Albert model for $m > 1$)**.** In the case that $m > 1$, the graph sequence $(\mathcal{G}_n)_{1 \leqslant n \leqslant N}$ is constructed from $(\mathcal{G}'_n)_{1 \leqslant n \leqslant N \cdot m}$ where $\mathcal{G}' = (V', E')$ is constructed with $m = 1$. The vertices $v_i \in V'$, with $i = 1, \ldots, m$ are then associated with vertex v_1 of $\mathcal{G}$, vertices $v_{m+1}, \ldots, v_{2m} \in V'$ are associated with vertex v_2 of $\mathcal{G}$, and so on via the collapsing procedure given in Definition 2.5 (see Fig. 2.2). «

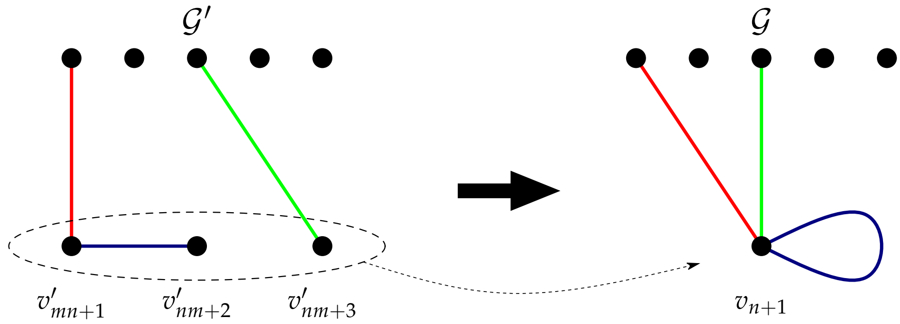


Figure 2.2: We form the graph $\mathcal{G}$ by collapsing $v'_{nm+1}, v'_{nm+2}, v'_{nm+3}$ of $\mathcal{G}'$ with the process described in Definitions 2.5 and 2.6

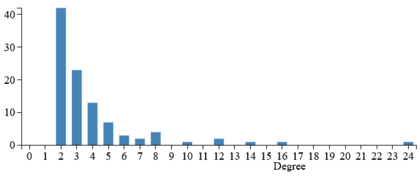
Though this model was very influential in the revival of interest in graph models to model real-world networks, it has limitations. One limitation is that the *tail exponent* of the resulting degree distribution is fixed at -3. This tail exponent, roughly corresponds to the slope of a line fitted to the log–log plot of degree vs. frequency. A precise definition of tail exponent, and a proof of its value will be presented in Chapter 3. The value of the tail exponent can however be tuned by adding a parameter δ, that models *initial attractiveness*.

Definition 2.7 (Barabási–Albert model with initial attractiveness δ)**.** Similar to Definition 2.4 and Definition 2.6, we define the *Barabási–Albert model with initial attractiveness δ*, by setting $\delta \geqslant -m$.

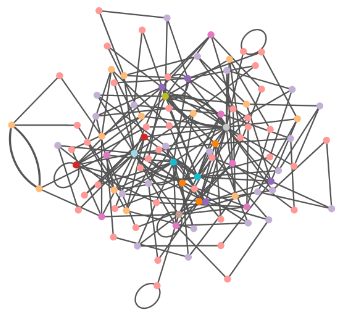
We then follow the construction as for the Barabási–Albert model, but replace Eq. (2.2.1), by the following attachment probability.

$$\mathbb{P}(v_{n+1} \to v_i \mid \mathcal{G}_n) = \begin{cases} \frac{\deg_{\mathcal{G}_n}(v_i)+\delta}{n(2+\delta)+1+\delta} & \text{if } 1 \leqslant i \leqslant n, \\ \frac{1+\delta}{n(2+\delta)+1+\delta} & \text{if } i = n+1. \end{cases} \tag{2.2.2}$$

The construction for $m = 1$, remains the same. For $m > 1$, we construct the graph by collapsing vertices from the Barabási–Albert model with initial attractiveness and $m = 1$, but we set δ', the initial attractiveness for the graph $\mathcal{G}'$ with $m = 1$, to $\frac{\delta}{m}$. «



(a) Degree vs. frequency plot



(b) Instance of the BA model, generated with $n = 100, m = 2, \delta = 1$

Figure 2.3: Graph drawing and degree vs. frequency plot of a graph generated with the BA model as defined in Definition 2.7. Fitzner, Gösgens, networkpages.nl

The parameter δ provides a tunable parameter that allows the fitting of the model to data. We note that as $\delta \to \infty$, the influence of the degrees of $v_i \in V$, become less. The attachment probability for high values of δ comes close to a uniform distribution. The value of δ thus interpolates between (and extrapolates from) the model of *uniform attachment* and the BA-model.

The BA model with initial attractiveness overcomes some difficulties in the modeling of real world networks. The desire to define models that capture the behavior of real-world networks leads to a variety of different models.

2.3 Preferential attachment with multiplicative fitness

A limitation of the BA model is that older nodes, having had the most time to acquire new links, will likely have the highest degree—a concept known as *old get richer*. This is in contrast to networks observed in real-life systems, in which late arrivals may prove competitive and overtake the older vertices in the competition for links. This competitiveness can be modeled by assigning a fitness parameter to each node [33, 34]. Bianconi and Barabási first proved a number of results on a model which we will call the *Bianconi Barabási (BB) model*. Results obtained about the model are general to a wide class of *Preferential attachment with fitness* (PAF) models. We will first give the original definition of the *Bianconi–Barabási model*. We will then present a different, more general, parametrization of PAF models taken from Dereich and Ortgiese [38].

Definition 2.8 (Bianconi–Barabási model [33, 34])**.** Let $m \in \mathbb{Z}_{\geqslant 1}$ be the number of edges added per time step. Assign an independent μ-distributed fitness $\mathcal{F}^{(i)}$ to each vertex v_i with $i \in \mathbb{Z}_{\geqslant 1}$. Initialize the graph $\mathcal{G}_2$ as the graph $V = \{v_1, v_2\}$, with $E = \{(v_1, v_2)\}$. Given graph $\mathcal{G}_n$, we construct $\mathcal{G}_{n+1}$ by adding vertex v_{n+1} and adding m (undirected) edges simultaneously, with attachment probability

$$\mathbb{P}(v_{n+1} \to v_i \mid \mathcal{G}_n) = \frac{\mathcal{F}^{(i)} \deg_{\mathcal{G}_n}(v_i)}{\sum_{i \in V_n} \mathcal{F}^{(i)} \deg_{\mathcal{G}_n}(v_i)}. \tag{2.3.1}$$

«

The distribution of μ is important, but is not specified in the original work. We will later see how this distribution influences the evolution of the degrees in the graph sequence.

Note 2.6 (Initialization). The model construction from Definition 2.8 as given by Bianconi and Barabási [34] does not specify initialization. «

2.4 Additive fitness model

The study of *preferential attachment models in which fitness is additive* (PAFA models), initially received less attention than the multiplicative model. This model exhibits different qualitative behavior than the BB model and has attracted more recent interest from mathematicians [39, 40]

We study an exposition of the subject by Lodewijks and Ortgiese [39]. We focus on the model described in the article as PAFUD, which stands for *preferential attachment with fitness and updating degree*.

The model by Lodewijks and Ortgiese [39] adds connects each added vertex v_{n+1} to the graph $\mathcal{G}_n$ by m new edges. Let $\deg^-_{n,j}(i)$ be the in-degree of vertex v_i in graph G_n of the graph process after j of m edges have been added from vertex v_{n+1}. Moreover, we define S_n as the sum of the respective fitness value for each vertex, e.g.;

$$S_n = \sum_{i=1}^{n} \mathcal{F}^{(i)}. \tag{2.4.1}$$

We can then define the PAFUD model.

Definition 2.9 (Preferential attachment with additive fitness (PAFUD))**.** Let $m \in \mathbb{Z}_{\geqslant 1}$ be the number of edges added per time step. Assign an independent μ-distributed fitness $\mathcal{F}^{(i)}$ to each vertex v_i with $i \in \mathbb{Z}_{\geqslant 1}$. Let the graph $\mathcal{G}_{n_0}$, with n_0 vertices and m_0 be given. Given graph $\mathcal{G}_n$, we construct $\mathcal{G}_{n+1}$ by adding vertex v_{n+1} and adding m directed edges, updating the fitness of each vertex after every added edge. For $j = 1, \ldots, m$, and letting the value of $\deg^-_{n,j}(i)$ be dependent on the first $j-1$ edges added, we add edge j from v_{n+1} to $v_i \in V_{\mathcal{G}_n}$ with probability

$$\mathbb{P}(v_{n+1} \to v_i \mid \mathcal{G}_n, j) = \frac{\mathcal{F}^{(i)} + \deg^-_{n,j}(i)}{m_0 + m(n-n_0) + (j-1) + S_n}. \tag{2.4.2}$$

«

This model will be analyzed in Section 3.3.

2.5 Sub- and superlinear preferential attachment

So far, we have limited the overview to models with (affine) linear attachment probability in the degree. We will later see that these models tend to have the most richness in dynamics. There has however been study into models with non-linear attachment probabilities. We here discern sub- and super-linear preferential attachment. Sub- and superlinear preferential attachment differs in that the *attachment kernel* A_k is non-linear.

Definition 2.10. In preferential attachment, the probability that vertex v_{n+1} attaches to vertex v_j with degree k is proportional to an increasing function A_k of k.

$$\mathbb{P}(v_{n+1} \to v_j \mid \deg_{\mathcal{G}_n}(v_j) = k) = \frac{A_k}{\sum_{1 \leqslant k \leqslant k_{\max}} A_k}. \tag{2.5.1}$$

Here A_k is called the *attachment function*. We speak of sub- respectively superlinear preferential attachment when A_k is sub- respectively superlinear. «

The sub- and superlinear models are not the focus of this thesis. For an overview of the different preferential attachment there are numerous sources [17, 29, 37, 41].

Articles which focus on sub- and superlinear preferential attachment are among others [42–47].

Chapter 3

Relation of fitness to degree distributions

The model definitions of preferential graphs seen in Chapter 2 were often motivated by the desire to capture some real-world behavior in the graph model. We have seen the introduction of fitness to the Bianconi–Barabási model, to allow newcomers to compete with the earlier arrivals in the graph. We also remarked with some foreshadowing that the fitness distribution of vertices in graph models was of considerable importance. In this section, we will provide results linking the distributions of vertex fitness to the distribution of degrees in the graph.

Why then are we interested in the degree distribution? The degrees of the resulting graphs are a tractable property that we can study that tells us about the *topology* of a graph. Here, we use the word *topology* loosely, we are interested in the connectedness of the graphs. Other properties that we could study are for instance graph diameter or the clustering coefficient. We will now define the degree distribution.

Definition 3.1 (Degree distribution)**.** Let n be the number of nodes in graph $\mathcal{G}_n$ of graph sequence $(\mathcal{G}_n)_{n\geqslant 1}$ and let n_k be the number of nodes with degree k. That is $n_k = \sum_{v\in V_n} \mathbb{1}_{\{\deg(v)=k\}}$. Define the *degree distribution* p_n as

$$p_n(k) = \frac{n_k}{n}. \tag{3.0.1}$$

«

Note that we may use the previous definition to study the empirical degree distribution of a network. However, the expected degree distribution can also be viewed as a topological characteristic of a generative graph model. We also note that the previous distribution is indeed a probability distribution. It describes the probability that a uniformly chosen vertex of graph $\mathcal{G}_n$ has degree k.

Definition 3.2 (Limiting degree distribution)**.** We define the measure p_∞ as the pointwise limit of the distributions p_n. That is

$$p_\infty(k) = \lim_{n\to\infty} p_n(k). \tag{3.0.2}$$

We may abbreviate $p_\infty(k)$ as $p(k)$, if the omission of the subscript does not cause confusion. «

We note that it is a-priori unclear if p_∞ exists and if it is in fact a probability distribution. In fact, we are particularly interested in the case when p_∞ does not converge to a probability distribution. We will later see that this case corresponds to what is called *condensation*.

The Bianconi–Barabási model adds fitness to the model defined by Barabási and Albert. This change is motivated by the desire, to make it possible for new vertices to overtake older vertices in preferential attachment models, something that is rare in the Barabási–Albert model. This is a departure from the adage dubbed *Matthew effect*, i.e., the "rich get richer". In the article by Bianconi and Barabási [34], the alternative phrases, "the fit get richer" and "winner takes all" are coined. We will see how to describe this mathematically, and that there is a surprising link to physics, that we now explain.

Example 3.3 (Bose–Einstein condensation)**.** A Bose–Einstein condensate is a non classical aggregation state of matter that occurs at temperatures very close to absolute zero. In the state, a non-zero fraction of bosons aggregate in the lowest quantum energy state. Bianconi and Barabási [34] liken this to the situation in which a non-zero linear fraction of added vertices to a preferential attachment model attach to a small sub-linear proportion of the vertices present in the graph. «

One breakthrough of modern network science was an observation described in [34]: A correspondence can be established between the thermodynamics of Bose gases and the degree of vertices in the BB model. This correspondence introduced the physical term *condensation* to the mathematical literature and informs the following definition.

Definition 3.4 (Condensation)**.** A graph sequence $(\mathcal{G}_n)_{n\geqslant 1}$ is said to exhibit *condensation* if a linear proportion of edges connects to a sub-linear proportion of vertices. That is, $\exists \rho > 0$, such that, for increasing n, at least $\rho|E_n|$ of the edges connect to $S_n \subset V_n$, where S_n is a minimal set of vertices such that $|S_n| = o(n)$. These vertices together form a subset called the *condensate*. «

Intuitively, the fact that a n-linear proportion of edges connects to a sub-linear proportion of vertices has an implication for the degree distribution (see Definition 3.1). In fact, the definition implies that the total degree of vertices in the condensate grows linearly. This means that the average degree within the condensate grows more rapidly than that of the vertices outside of it.

In the following section, we study the degree distribution and conditions under which condensation occurs in preferential attachment models with and without fitness and highlight the wealth of different techniques that can be employed to prove results.

3.1 Models without fitness

We will first turn to the classical BA model and results stated by Barabási and Albert [31]. Although, as stated in the introduction, they were not the first to state the model or derive its properties, their publication appeared shortly after [20]. These publications in close succession in Science and Nature—two of the worlds most highly regarded scientific journals—gave incredible momentum to the idea that topological features of networks such as the internet could emerge from simple rules for network growth. The model in Definition 2.6 is thus commonly called the BA model. The degree distribution of this model had been studied earlier by for instance de Solla Price [7], Simon [15], and de Solla Price [16] and then in 1999 by Barabási and Albert [31] and Barabási *et al.* [48], who all but Simon had a background in physics. This head start of the physicists left room for mathematicians to improve on the mathematical rigor of the results, the earlier proofs of which Bollobás *et al.* [35] calls "heuristic". The primary result for the degree sequence of the BA model is given here.

Theorem 3.1 (Tail exponent for the degree sequence of the BA model [37, pp. 263])**.** *Let $m \in \mathbb{Z}_{\geqslant 1}$ and $\delta \in (-m, \infty)$ be given and let $(\mathcal{G}_n)_{n\in\mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by the BA model with parameters m, δ defined in Definition 2.6. As n goes to infinity we find that $p_n(k) \to p_\infty(k)$, with the following tail behavior for k*

$$p_\infty(k) = c_{m,\delta} k^{-\gamma} \left(1 + \mathcal{O}(k^{-1})\right), \tag{3.1.1}$$

where

$$\gamma = 3 + \frac{\delta}{m} \qquad \text{and} \qquad c_{m,\delta} = \left(2 + \frac{\delta}{m}\right) \frac{\Gamma(m + 2 + \delta + \frac{\delta}{m})}{\Gamma(m + \delta)}. \tag{3.1.2}$$

Stronger results than this exist, as we will see in Theorem 3.4, but it is important to note that the degree distribution has a tail that decreases as a power of k.

We will now present a number of results about the models defined in Chapter 2. We are primarily interested in results about the distribution of degrees. We will first consider a heuristic derivation of tail degree exponent.

3.1.1 Heuristic derivation of tail exponent

The heuristic derivation breaks down into two parts; to establish the estimated degree of a vertex v_i and then to relate the expected degree to the distribution of degrees. We first rigorously prove the following preliminary result that explains the evolution of a degree of a single vertex.

Proposition 3.2 (Degree evolution of individual vertex ($m = 1$) [37, pp. 263])**.** *Let $\delta \in (-1, \infty)$ be given and let $(\mathcal{G}_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by the BA model with parameter δ defined in Definition 2.4. We then find the following relation for the degrees of vertex v_i, where M_i, is a random variable limit of a martingale with expectation* 1.

$$\frac{\deg_{\mathcal{G}_n}(v_i) + \delta}{n^{\frac{1}{2+\delta}}} \xrightarrow{a.s.} M_i \frac{(1+\delta)\Gamma(i - \frac{1}{2+\delta})}{\Gamma(i)}. \tag{3.1.3}$$

It is a-priori unclear, with what probability M_i is zero in this proposition. This is one of the reasons why the derivation of the degree distribution carried out by Barabási and Albert [31] was called heuristic. The proof below is a refinement of the proof in [31], refined by van der Hofstad [37] to be explicit about the mathematical leaps of faith that it uses. It is more explicit than the original method of derivation, which was to approximate the stochastic process with the ODE master equation

$$\frac{d}{dt} \deg(v_i) = \frac{\deg(v_i)}{2t}. \tag{3.1.4}$$

An analysis by this technique is called the *mean-field approach* as it ignores the variance around the expectation of the degree and is described in [48].

Proof of Proposition 3.2. Let $m = 1$, $\delta > -1$ and let $i < n + 1$. We then find that

$$\mathbb{E}[\deg_{\mathcal{G}_{n+1}}(v_i) + \delta \mid \deg_{\mathcal{G}_n}(v_i)] = \deg_{\mathcal{G}_n}(v_i) + \delta + \mathbb{E}[\deg_{\mathcal{G}_{n+1}}(v_i) - \deg_{\mathcal{G}_n}(v_i) \mid \deg_{\mathcal{G}_n}(v_i)] \tag{3.1.5a}$$

$$= \deg_{\mathcal{G}_n}(v_i) + \delta + \frac{\deg_{\mathcal{G}_n}(v_i) + \delta}{(2+\delta)n + 1 + \delta} \tag{3.1.5b}$$

$$= \left(\deg_{\mathcal{G}_n}(v_i) + \delta\right) \frac{(2+\delta)(n+1)}{(2+\delta)n + 1 + \delta}. \tag{3.1.5c}$$

For the occurrence of self-loops ($v_{n+1} \to v_{n+1}$), or $i = n + 1$, we find

$$\mathbb{E}[\deg_{\mathcal{G}_{n+1}}(v_{n+1}) + \delta] = 1 + \delta + \frac{1+\delta}{(2+\delta)n + 1 + \delta} \tag{3.1.6}$$

$$= (1+\delta) \frac{(2+\delta)(n+1)}{(2+\delta)n + 1 + \delta}. \tag{3.1.7}$$

We can now establish the expectation of $\deg_{\mathcal{G}_n}(v_i)$, by telescoping Eq. (3.1.5c), whilst keeping account of Eq. (3.1.7). Note that we determine the expectation for $\mathcal{G}_n$, changing the limits of the summation;

$$\mathbb{E}[\,\deg_{\mathcal{G}_n}(v_i)+\delta\,] = (1+\delta)\frac{(2+\delta)i}{(2+\delta)(i-1)+1+\delta}\cdot\prod_{t=i}^{n-1}\frac{(2+\delta)(t+1)}{(2+\delta)t+1+\delta}. \tag{3.1.8}$$

This simplifies to

$$\mathbb{E}[\,\deg_{\mathcal{G}_n}(v_i)+\delta\,] = (1+\delta)\prod_{t=i-1}^{n-1}\frac{(2+\delta)(t+1)}{(2+\delta)t+1+\delta}. \tag{3.1.9}$$

For the repeated product, we make use of the Gamma-function, which satisfies the identity

$$\Gamma(t+1) = t\Gamma(t).$$

This allows us to rewrite the product by telescoping it as

$$\prod_{t=i-1}^{n-1}\frac{t+1}{t+\frac{1+\delta}{2+\delta}} = \frac{\Gamma(n+1)\Gamma(i-\frac{1}{2+\delta})}{\Gamma(i)\Gamma(n+\frac{1+\delta}{2+\delta})}. \tag{3.1.10}$$

From Eqs. (3.1.9) and (3.1.10), we can see that $M_i(n)$, defined by

$$M_i(n) = \frac{\deg_{\mathcal{G}_n}(v_i)+\delta}{1+\delta}\,\frac{\Gamma(i)\Gamma(n+\frac{1+\delta}{2+\delta})}{\Gamma(n+1)\Gamma(i-\frac{1}{2+\delta})}, \tag{3.1.11}$$

is a martingale with expectation 1 (see Definition B.1). By the martingale convergence theorem (Theorem B.1), we find that the martingale converges a.s. to some random variable M_i. We can then make use of Stirling's formula

$$\frac{\Gamma(n+a)}{\Gamma(n)} = n^a\left(1+\mathcal{O}(n^{-1})\right), \tag{3.1.12}$$

to find that

$$M_i(n) = \frac{\deg_{\mathcal{G}_n}(v_i)+\delta}{1+\delta}\,\frac{\Gamma(i)}{\Gamma(i-\frac{1}{2+\delta})}\,\frac{1}{n^{\frac{1}{2+\delta}}}\,\frac{1}{1+\mathcal{O}(n^{-1})} \xrightarrow{\text{a.s.}} M_i, \tag{3.1.13}$$

where $\mathbb{E}[\,M_i\,] = 1$. We consider the distribution of M_i unknown. We do not investigate this here, but it can be shown that $M_i(n)$ is uniformly integrable (Definition B.3), which by Theorem B.2 implies that $M_i(n)$ converges to 1 in mean. Note that $\frac{1}{2+\delta} > 0$, so we can reshuffle Eq. (3.1.13) to find

$$\frac{\deg_{\mathcal{G}_n}(v_i)+\delta}{n^{\frac{1}{2+\delta}}} \xrightarrow{\text{a.s.}} M_i\frac{(1+\delta)\Gamma(i-\frac{1}{2+\delta})}{\Gamma(i)}, \tag{3.1.14}$$

proving Proposition 3.2. □

From Proposition 3.2, we can heuristically derive the same tail exponent as found in Theorem 3.1. This gives an intuition as to why the result of Theorem 3.1 should hold.

Heuristic argument for Theorem 3.1

Using Stirling's formula on Eq. (3.1.14), we find that i asymptotically decays with the same exponent as n, $\frac{1}{2+\delta}$. Write $\sim$ to denote asymptotic equivalence, i.e., $f \sim g \Leftrightarrow \lim_{n\to\infty} f(x)/g(x) = 1$. Using this notation, assuming n and i are sufficiently large, we find by taking expectation that there exists some constant $C > 0$ that is universal across i, such that

$$\mathbb{E}[\deg_{\mathcal{G}_n}(v_i)] \sim C\left(\frac{n}{i}\right)^{\frac{1}{2+\delta}}. \tag{3.1.15}$$

For $m > 1$, we note from Definition 2.7, that the graph $\mathcal{G}$ that is generated by such a model is constructed from a graph $\mathcal{G}'$ constructed with $m' = 1$ and $\delta' = \frac{\delta}{m}$. Thus we find for the degree of vertex i of graph $\mathcal{G}$, that

$$\mathbb{E}[\deg_{\mathcal{G}_n}(v_i)] = \sum_{j=1}^{m} \mathbb{E}[\deg_{\mathcal{G}'_{m\cdot n}}(v_{m(i-1)+j})]. \tag{3.1.16}$$

We can then combine Eqs. (3.1.15) and (3.1.16) to conclude that there exists some $a_{m,\delta}$, again a global constant dependent on m and δ, such that for i and t large enough

$$\mathbb{E}[\deg_{\mathcal{G}_n}(v_i)] \sim a_{m,\delta}\left(\frac{n}{i}\right)^{\frac{1}{2+\frac{\delta}{m}}}. \tag{3.1.17}$$

The above asymptotic equivalence does rely on the implicit assumption that there is a only small chance that the M_i in Eq. (3.1.15) are zero. A priori, it might however be the case that edges concentrate on a few vertices (see [49, Section 3.4] for an example of this assumption failing).

Now using this result, we can heuristically derive the tail-exponent of the power law. Let $N_n^{\geqslant}(k)$, be the number of vertices with degree greater than k. That is

$$N_n^{\geqslant}(k) = \sum_{i=1}^{n} \mathbb{1}_{\{\deg_{\mathcal{G}_n}(v_i)\geqslant k\}}. \tag{3.1.18}$$

Here the original proof by Barabási and Albert [31] makes the assumption that we can equate $\#\{v \in V \mid \deg_{\mathcal{G}_n}(v) \geqslant k\}$ and $\#\{v \in V \mid \mathbb{E}[\deg_{\mathcal{G}_n}(v) \geqslant k]\}$. Given this assumption we derive

$$N_n^{\geqslant}(k) \sim \sum_{i=1}^{n} \mathbb{1}_{\{\mathbb{E}[\deg_{\mathcal{G}_n}(v_i)]\geqslant k\}} \sim \sum_{i=1}^{n} \mathbb{1}_{\left\{a_{m,\delta}\left(\frac{n}{i}\right)^{1/(2+\delta/m)}\geqslant k\right\}} \tag{3.1.19a}$$

$$= \sum_{i=1}^{n} \mathbb{1}_{\left\{a_{m,\delta}^{2+\delta/m}\left(\frac{n}{i}\right)\geqslant k^{2+\delta/m}\right\}} = \sum_{i=1}^{n} \mathbb{1}_{\left\{i\leqslant n a_{m,\delta}^{2+\delta/m} k^{-(2+\delta/m)}\right\}} \tag{3.1.19b}$$

$$= \left\lfloor n a_{m,\delta}^{2+\delta/m} k^{-(2+\delta/m)} \right\rfloor. \tag{3.1.19c}$$

From this, we can derive that as $n \to \infty$

$$\frac{N_n^{\geqslant}}{n} \approx \frac{\left\lfloor n a_{m,\delta}^{2+\delta/m} k^{-(2+\delta/m)} \right\rfloor}{n} \to a_{m,\delta}^{2+\delta/m} k^{-(2+\delta/m)}. \tag{3.1.20}$$

Note that for the distribution function F of degrees in $\mathcal{G}_n$, we expect that

$$\frac{N_n^{\geqslant}}{n} \overset{\text{a.s.}}{\to} 1 - F(k) = a_{m,\delta}^{2+\delta/m} k^{-(2+\delta/m)}. \tag{3.1.21}$$

By differentiating we find that probability mass function $f(k) \propto k^{-(3+\delta/m)}$. This confirms the tail exponent γ of the degree distribution is equal to $3 + \frac{\delta}{m}$. Although this derivation was heuristic, it turns out that the limiting distribution is still correct. We will later see how to turn this heuristic proof into an actual proof.

3.1.2 Comparison with uniform attachment

Although various growing graph models lead to power-law distributions of the degree [37] it is important to see that this behavior is not universal and that preferential attachment is a good explanatory example of a simple model with this behavior.

We contrast preferential attachment with so-called *uniform attachment* in which the attachment probability $\mathbb{P}(v_{n+1} \to v_i \mid \mathcal{G}_n)$ is $\frac{1}{n+1}$ irrespective of the degree of vertex v_i. For this model, we find a shifted geometric distribution.

Proposition 3.3. *Let $m \in \mathbb{Z}_{\geqslant 1}$ be given and let $(\mathcal{G}_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by a graph model with uniform attachment, i.e., $\mathbb{P}(v_{n+1} \to v_i \mid \mathcal{G}_n) = \frac{1}{n+1}$ irrespective of degree. We then find that the limiting distribution as $n \to \infty$ for the distribution of in-degrees is geometric with parameter $\frac{1}{m+1}$.*

Note that we find the distribution for the total degree by shifting the geometric distribution by m, the out-degree.

We will first present an adaptation of the proof for Proposition 3.2 and use the same heuristic argument to derive the limiting distribution. We will then look at an improvement of this by Bollobás *et al.* [35].

First heuristic derivation

Let $m = 1$, and let $1 \leqslant i < n + 1$. We then find that

$$\mathbb{E}[\deg^-_{\mathcal{G}_{n+1}}(v_i)] = \mathbb{E}[\deg^-_{\mathcal{G}_n}(v_i)] + \mathbb{E}[\deg^-_{\mathcal{G}_{n+1}}(v_i) - \deg^-_{\mathcal{G}_n}(v_i)] \tag{3.1.22a}$$

$$= \mathbb{E}[\deg^-_{\mathcal{G}_n}(v_i)] + \frac{1}{n+1}. \tag{3.1.22b}$$

For the occurrence of self-loops ($v_{n+1} \to v_{n+1}$), or $i = n+1$, we find

$$\mathbb{E}[\deg^-_{\mathcal{G}_{n+1}}(v_{n+1})] = \frac{1}{n+1}. \tag{3.1.23}$$

We can now establish the expectation of $\deg^-_{\mathcal{G}_n}(v_i)$, by inductively determining the sum value of $\mathbb{E}[\deg^-_{\mathcal{G}_{n+1}}(v_i)]$. Here we use that for the harmonic series, the partial sum $H_n = \ln n + \gamma + \frac{1}{2n} + \epsilon_n$, where $0 \leqslant \epsilon_n \leqslant \frac{1}{8n^2}$ and letting γ denote the Euler–Mascheroni constant. Note that we determine the expectation for $\mathcal{G}_n$

$$\mathbb{E}[\deg^-_{\mathcal{G}_n}(v_i)] = \sum_{s=i-1}^{n-1} \frac{1}{s+1} = \ln n - \ln i + \mathcal{O}(i^{-1}). \tag{3.1.24}$$

Here we can repeat the steps taken in Eqs. (3.1.19) and (3.1.21). Again let $N_n^{\geqslant,-}(k)$, be the number of vertices with in-degree greater than k as in Eq. (3.1.18) and equate $\#\{v \in V \mid \deg_{\mathcal{G}_n}(v) \geqslant k\}$ and $\#\{v \in V \mid \mathbb{E}[\deg^-_{\mathcal{G}_n}(v) \geqslant k]\}$. We moreover assume that i is bounded away from 1 and n, such that the $\mathcal{O}(1/n)$ term is small compared to $\ln(n/i)$. Given this assumption we derive

$$N_n^{\geqslant,-}(k) = \sum_{i=1}^{n} \mathbb{1}_{\{\mathbb{E}[\deg^-_{\mathcal{G}_n}(v_i)]\geqslant k\}} \sim \sum_{i=1}^{n} \mathbb{1}_{\{\log(\frac{n}{i})+\geqslant k\}} \tag{3.1.25a}$$

$$= \sum_{i=1}^{n} \mathbb{1}_{\{i\leqslant \frac{n}{e^k}\}} = \left\lfloor \frac{n}{e^k} \right\rfloor. \tag{3.1.25b}$$

From this, we can derive that as $n \to \infty$

$$\frac{N_n^{\geqslant,-}}{n} \xrightarrow{\text{a.s.}} \frac{1}{e^k} = \frac{1}{e^k}. \tag{3.1.26}$$

We will see that the rate of decay for this functional is in fact incorrect. However, we clearly see that the $\mathbb{Z}_{\geqslant 1}{}_n^{\geqslant,-}$ is not power-law distributed. Using the law of large numbers we can say

$$\frac{N_n^{\geqslant}}{n} \xrightarrow{\text{a.s.}} 1 - F(k). \tag{3.1.27}$$

For X a geometrically distributed variable with variable p, we know that $\mathbb{P}(\text{Geo}(p) \geqslant k) = (1-p)^k$, thus the degrees are distributed geometrically with parameter $p = 1 - 1/e$. This "proves" that the degrees in the uniform attachment model follow a geometric distribution. However, as we will see below, the parameter p is incorrect.

We see that following the steps of proof that were used for the BA model requires the assumption that we bound i away from 1 and n. However we consequently need to sum over $1, \ldots, n$, which is problematic. We will show a more quantitative argument, that is due to Bollobás *et al.* [35] which uses Poisson approximation and delivers the correct parameter for the degree distribution.

Second heuristic derivation

Below we will present an improvement of the naive proof for the degree sequence of the uniform attachment model.

Proof. Let $m \in \mathbb{Z}_{\geqslant 1}$ be given, and let $i < n+1$. If we consider the directed version of the uniform attachment model, we follow Eq. (3.1.24) and find for the in-degree

$$\mathbb{E}[\deg^-_{\mathcal{G}_n}(v_i)] = m \sum_{s=i}^{n} \frac{1}{s}. \tag{3.1.28}$$

We can then again approximate this expectation by the integral

$$\mathbb{E}[\deg^-_{\mathcal{G}_n}(v_i)] \approx m \int_i^n \frac{1}{s}\, ds = m(\log(n) - \log(i)) = m \log(\frac{n}{i}) + \mathcal{O}(\frac{1}{i}). \tag{3.1.29}$$

This approximation holds well for $1 \ll i \ll n$. We observe that for large n the average in-degree is small compared to n. Moreover, the sum of the all in-degrees must add up to nm. It is therefore, by Theorem B.4 reasonable to say that for large n, the in-degrees of vertex v_i is asymptotically Poisson distributed with parameter $\lambda = m \ln(\frac{n}{i})$. However, we note that the errors in Theorem B.4 do not converge to 0 and that the probabilities p_i in Theorem B.4 are not independent. Hence, this derivation is still heuristic. Using these assumptions, we estimate

$P^-(k)$ the expected proportion of vertices with in-degree k, by integrating the probability density

$$\mathbb{E}[\, P_n^-(k)\,] = \int_0^n \frac{\lambda^k}{k!}\exp(-\lambda)\, di = \frac{m^k}{k!}\int_1^n \left(\ln(\frac{n}{i})\right)^k \left(\frac{n}{i}\right)^{-m} di. \tag{3.1.30}$$

We can solve this integral by substituting $s := \frac{i}{n}$ and letting s vary from 0 to 1. We find

$$\mathbb{E}[\, P_n^-(k)\,] = \frac{m^k}{k!}\int_0^1 (-\ln s)^k s^m\, ds. \tag{3.1.31a}$$

Substitute first $t = \ln s$, and then $u = -(m+1)t$

$$\mathbb{E}[\, P_n^-(k)\,] = \frac{m^k}{k!}\int_{-\infty}^0 (-t)^k e^{(m+1)t}\, dt = \frac{m^k}{k!(m+1)^{k+1}}\int_0^\infty u^k e^{-u}\, du. \tag{3.1.31b}$$

The integral portion of the last expression is the Γ-function $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}\, dt$ and for integer s, $\Gamma(s+1) = s!$. We thus find

$$\mathbb{E}[\, P_n^-(k)\,] = \frac{m^k}{(m+1)^{k+1}} = \left(1 - \frac{m}{m+1}\right)\left(\frac{m}{m+1}\right)^k. \tag{3.1.32}$$

Therefore $P_n^-(k)$ is asymptotically geometric with parameter $\frac{1}{m+1}$. The asymptotic probability mass function of the total degree then is the same distribution shifted by m. □

The improved proof correctly takes into account that we can not assume that the proportion of vertices with in-degree greater than k is not the same as the proportion of vertices with *expected* in-degree greater than k. The technique of Poissonization allows us to take variance into account and derives the correct result. We can however note that approximating the harmonic series by a continuous function gives great relative errors close to $i = 1$ and $i = n$. Apart from that the Poisson approximation technique is not accurate for smaller sizes of the graph $\mathcal{G}_n$, as the attachment probabilities p_i, are large.

3.1.3 Overview of proof techniques in mathematically rigorous treatment

As stated before the proofs of the degree sequence of the Barabási–Albert and uniform PA model are heuristic. They rely on the assumption that the number of vertices with degree $\geqslant k$ is asymptotically equivalent to the number of vertices with expected degree $\geqslant k$. Moreover, there actually exists a closed form expression for the limiting distribution P with $\mathbb{P}(N_k = k) = p_k$. We will state these results and an informal overview of how they are derived below.

The formal statement of the idea expressed in Section 3.1.1 is given below. The result states that for every value of degree k the expected number of vertices with degree k converges in probability to the limiting distribution p_k.

Theorem 3.4 (Van der Hofstad [37, p. 262, Theorem 8.3])**.** *Let $m \in \mathbb{Z}_{\geqslant 1}$ and $\delta \geqslant -m$ be given. There exists a constant $C = C(m, \delta)$ such that for $n \to \infty$ we find*

$$\mathbb{P}\left(\max_k |P_k(n) - p_k| > C\sqrt{\frac{\log n}{n}}\right) = o(1). \tag{3.1.33}$$

with p_k given by

$$p_k = (2 + \frac{\delta}{m})\frac{\Gamma(k+\delta)\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(k+3+\delta+\frac{\delta}{m})\Gamma(m+\delta)}. \tag{3.1.34}$$

As a matter of interest, if $\delta = 0$, the expression for p_k simplifies significantly and we find

$$p_k = \frac{2\Gamma(k)\Gamma(m+2)}{\Gamma(k+3)\Gamma(m)} = \frac{2m(m+1)}{k(k+1)(k+2)}.$$

In fact for all values of δ, we can see from Theorem 3.4 by Stirling's formula (Eq. (3.1.12)) that

$$p_k = (2 + \frac{\delta}{m}) \frac{\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)} k^{-(3+\frac{\delta}{m})}(1 + \mathcal{O}(k^{-1})), \tag{3.1.35}$$

showing that the number of vertices with degree k decays with tail exponent $3 + \frac{\delta}{m}$, and leading to Theorem 3.1.

The result in Theorem 3.4 is derived from two propositions. A result that relates N_k to $\mathbb{E}[\, N_k \,]$ and one that establishes an exact expression for $\mathbb{E}[\, N_k \,]$. The following two propositions are adapted from van der Hofstad [37] and require the following notation.

Notation 3.1. Let $(\mathcal{G}_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by the Barabási–Albert model and define

$$N_n(k) = \sum_{i=1}^{n} \mathbb{1}_{\{\deg_{\mathcal{G}_n}(v_i) = k\}} = nP_n(k) \qquad \bar{N}_n(k) = \mathbb{E}[\, N_n(k) \,] = \mathbb{E}[\, nP_n(k) \,]. \tag{3.1.36}$$

«

The first proposition shows that the number of vertices with degree k is close to $\bar{N}_n(k)$.

Proposition 3.5 (Concentration around expectation [37, p. 264, Proposition 8.4])**.** *Let $(\mathcal{G}_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by the Barabási–Albert model with $\delta > -m$ and $m \geqslant 1$. For any $C > m\sqrt{8}$, we find as $t \to \infty$*

$$\mathbb{P}\left(\max_k |N_k(n) - \mathbb{E}[\, N_k(n) \,]| > C\sqrt{n \log n} \right) = o(1). \tag{3.1.37}$$

The seconds proposition shows that the difference between the expected degree distribution $\bar{N}_n(k)$ and the limiting probability mass function is bounded.

Proposition 3.6 (Expected degree sequence [37, p. 268, Proposition 8.7])**.** *Let $(\mathcal{G}_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by the Barabási–Albert model with $\delta > -m$ and $m \geqslant 1$. Then, there exists a constant $C = C(\delta, m)$ such that, for all $n \in \mathbb{Z}_{\geqslant 1}$ and all $k \in \mathbb{Z}_{\geqslant 1}$*

$$|\bar{N}_n(k) - p_k n| \leqslant C. \tag{3.1.38}$$

The derivation of Propositions 3.5 and 3.6 depends on two specific and unrelated techniques. The proof of Proposition 3.5 relies on the definition of the *Martingale of increasing information* also known as *Doob's Martingale* (Definition B.2) of $N_n(k)$ given a filtration of σ-algebras defined by the graph's sequence $(\mathcal{G}_i)_{0 \leqslant i \leqslant t}$ from the graph sequence generated by the BA model. Thus the martingale M_t for $0 \leqslant t \leqslant n$ is defined

$$M_t = \mathbb{E}[\, N_n(k) \mid \mathcal{G}_t \,], \tag{3.1.39}$$

where we let the σ-algebra corresponding to $\mathcal{G}_0$ mean there is no information. This martingale has $M_0 = \mathbb{E}[\, N_n(k) \,]$ and $M_n = N_n(k)$. This is a martingale with bounded differences—something that holds for many generative graph models. By the Azuma–Hoeffding inequality (Theorem B.3, the absolute differences from the expectation of this martingale can be bounded, with the subsequent result in Proposition 3.5. Van der Hofstad [37] notes that this method of proof is general and can be applied to various specifications of growing graph models.

Proposition 3.6 is instead derived for $m = 1$ by defining a system of recursive relations for $\mathbb{E}[\, N_{n+1}(k)\,]$ in terms of $N_n(k)$, $N_t(k-1)$, and the possibility that vertex v_{n+1} has either degree 1 or 2. The resulting relationship looks like

$$\mathbb{E}[\, N_{n+1}(k) - N_n(k) \mid \mathcal{G}_n\,] = \frac{k-1+\delta}{(2+\delta)n+1+\delta} N_n(k-1) - \frac{k+\delta}{(2+\delta)n+1+\delta} N_n(k) + \left(1 - \frac{1+\delta}{(2+\delta)n+1+\delta}\right) \mathbb{1}_{\{k=1\}} + \frac{1+\delta}{(2+\delta)n+1+\delta} \mathbb{1}_{\{k=2\}}. \tag{3.1.40}$$

We can state an approximation to this recursive relation by attempting to eliminate n-dependence from Eq. (3.1.40). Following [37], we write $\bar{N}_{n+1}(k) - \bar{N}_n(k) \approx p_k$, $\bar{N}_n(k) \approx np_k$. The coefficients of the summands can then be approximated $n/\,(n(2+\delta)+(1+\delta)) \approx 1/(2+\delta)$ and $(1+\delta)/\,(t(2+\delta)+(1+\delta)) \approx 0$. This leads to the following relation for p_k if $k \geqslant 1$

$$p_k = \frac{k-1+\delta}{2+\delta} p_{k-1} - \frac{k+\delta}{2+\delta} p_k + \mathbb{1}_{\{k=1\}}. \tag{3.1.41}$$

The answer $\bar{N}_n(k)$ to this relation is established by deriving a result of the approximate n-independent Eq. (3.1.41) and proven to be correct by estimating the error $|N_n(k) - \bar{N}_n(k)|$. For $m > 1$ a more general method is employed which defines a linear operator T_{n+1} that operates on sequences $(Q_k)_{k\in\mathbb{Z}}$. For an expected degree sequence Q at time n the expected degree sequence at time $n+1$ is $T_{n+1}Q$. Thus the properties of this transformation can be studied to establish Proposition 3.6. This proof is rather technical and will not be included here.

We have seen in this section how the original derivation of the degree sequence in the BA model was not mathematically rigorous, but did arrive at the correct conclusion—the tail of the degree distribution decays with tail exponent $3 + \frac{\delta}{m}$. Subsequent analysis by Bollobás *et al.* [35], here presented in the notation of van der Hofstad [37] finally derives the result. We will later see how the techniques that are used in the proof for the BA model will continue to be used, but will require modification to establish degree results for other models.

3.2 The Bianconi–Barabási model

The introduction of a fitness parameter $\mathcal{F}$ contributes inhomogeneity to the preferential attachment model. As a consequence of the model change, the theories in Section 3.1 require revision. Here we note that these revisions are not merely a matter of bookkeeping. Results in the study of the Bianconi–Barabási model are usually not concerned with degree sequences for fixed fitness values, but rather focus on qualitative results that apply to models with fitness distributed according to some measure μ. We will in fact see that the BB model displays several qualitative regimes which have been named “fit get richer,” and in contrast “winner takes all,” “Bose-Einstein phase” or “innovation pays off”. Apart from the difference in question that we ask ourselves, the addition of the factor $\mathcal{F}$ that is μ-distributed comes with a technical issue. The denominator of Eq. (2.3.1) is no longer deterministic as in Eq. (2.2.1). We will give a statement of the most important results concerning the BB model, and study the implications of this change in model.

In this section, we will present developments leading to a very general result of Dereich and Ortgiese [38]. For the sake of clarity, we will relinquish some generality to aid the presentation, and focus the exposition of results on Definition 2.8 as defined in this text. The theorems that are presented in [38] instead apply to various variations of the BB model. In addition to the proof overview. We will also give an overview of the article that came before [38].

3.2.1 Original work by Bianconi and Barabási

To understand the terminology used in the study of the Bianconi–Barabási model, we first look at the two original papers co-authored by Bianconi and Barabási [33, 34]. Here the exact order of publication is a bit obfuscating; for the purpose of this thesis, we assume that the motivation of the development of the BB model is laid out in [33]. Here, the authors observe that in real-world networks, not all nodes are equally successful in gaining links. They therefore propose, that unlike the BA model, vertices may be given an individual fitness. They then prove, that the growth of the expected degree is influenced by the fitness. This is a property they describe as *multi-scaling.*

Multi-scaling in the BB model

In contrast to Eq. (3.1.17), the exponent of the degree distribution is not global, but unique to each vertex and dependent on the vertex fitness;

Theorem 3.7 (Exponent in the degree expectation [33]). *Let $(\mathcal{G}_n)_{n\in\mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by the BB model. Let $\mathcal{F}^{(i)}$ be the fitness of vertex v_i. We then find*

$$\mathbb{E}[\,\deg(v_i)\,] = m\left(\frac{t}{i}\right)^{\beta(\mathcal{F}^{(i)})},$$

for some vertex dependent exponent $\beta(\mathcal{F}^{(i)})$, which is a deterministic function taken at random value $\mathcal{F}^{(i)}$, with $0 \leqslant \beta(\mathcal{F}^{(i)}) \leqslant 1$.

In fact, subsequently, they make the observation that

Theorem 3.8. *For the BB-model such as defined in Theorem 3.7, with fitness distribution μ.*

$$\beta(x) = \frac{x}{C},$$

Where C is a global constant, that is a solution to

$$1 = \int_0^{\mathcal{F}_{\max}} \frac{1}{(C/s)-1}\,\mu(ds), \tag{3.2.1}$$

with $\mathcal{F}_{\max}$ the essential supremum of fitness distributed according to probability measure μ.

After this, for uniform and exponential fitness distributions, an argument analogous to that in [31] is made to derive a power-law distribution. For the exponential distribution, they derive that the degree-distribution does not follow a power law, but a stretched exponential. Indeed they hypothesize that not all fitness distributions may lead to power-law distributions and call the characterization of which do a "formidable challenge."

Condition for condensation—Physics connection

This challenge is picked up by the same authors. In [34] the authors compare the BB model to a Bose-gas (see Example 3.3). Through this comparison they import concepts from the realm of statistical physics into the mathematical world. The final consequence is that one of the core concepts in the study of preferential attachment models is called *condensation*. In the proposed correspondence, the fitness value $\mathcal{F}^{(i)}$ of a vertex in the BB model corresponds to an energy level $\epsilon_i = -\frac{1}{\beta}\log(\mathcal{F}^{(i)})$. Here β, is the thermodynamical temperature, a control variable that has no significance in the network setting. In the BB-model, the half edges in the graph correspond to the bosons in the gas. Thus by following the physics, we find;

Note 3.2 (Correspondence of BB model to Boson gas). The probability that a boson is at energy state ϵ_i corresponds to the probability that an edge is connected to a vertex with the corresponding fitness $\mathcal{F}^{(i)}$. «

From this correspondence, condensation gets its meaning in the preferential attachment models. As in Definition 3.4, we see that the situation in which a linear proportion of half edges connects to a sub-linear proportion of vertices corresponds to condensation in the statistical mechanics. We will now explore this connection.

In thermodynamics the Bose–Einstein distribution is the measure that under equilibrium conditions gives the probability that a certain system state occurs for a system of indistinguishable particles, and it critically depends only on the energy of the state. The probability that a system is in state i, is

$$p_i \propto \frac{1}{e^{(\epsilon_i-\mu)\beta}-1}. \tag{3.2.2}$$

This equation can also be used to calculate the relative probability, that part of the thermodynamical system is in energy state i or j, respectively. The parameter μ is the so-called chemical potential in thermodynamics.

For gases, this distribution dictates that atoms distribute themselves in the system according to the energy of their states. However as a result of quantum effects, the distribution may not always hold. In physics one example of when the principle does not hold is so called Bose-Einstein condensation, where the lowest energy state captures a fraction of atoms that is larger then predicted by Eq. (3.2.2). By the correspondence, the physical laws that govern Bose gases, predict that a larger than expected number of half edges may connect to the vertex with the highest fitness. This correspondence with Bose gases is not completely justifiable in principle. The "atoms", can not change energy state; therefore there is no easy definition of equilibrium. However, the prediction made by the Bose-Einstein statistic are accurate for the BB model, as confirmed by simulation in [34]. Moreover, the correspondence gives an intriguing condition for the occurrence of condensation; it occurs if there is no solution for the chemical potential μ to the integral equation

$$\int \frac{1}{e^{\beta(\epsilon-\mu)}-1} g(\epsilon)\, d\epsilon = 1. \tag{3.2.3}$$

In physics, the function g is the degeneracy, i.e., $g(\epsilon_i)$ corresponds to the number of states that have energy ϵ_i. In the BB model, this becomes $g(\epsilon) = \beta f(e^{-\beta\epsilon})e^{-\beta\epsilon}$, where f is the probability density corresponding to fitness measure λ. This integral equation that is lifted from physics proves to be a sharp classification for when condensation occurs in the BB model. The paper [34] does however draw heavily on analogy to physics, we will now continue with a look into more mathematically minded literature.

3.2.2 Further mathematical analysis by Borgs *et al.*

In the Borgs *et al.* article, the authors limit their research to fitnesses drawn from probability measures with finite or countably infinite support. For these models, they rigorously prove convergence to a limiting distribution and the conditions for convergence.

They do so by embedding the BB model into a *Polya's urn* model. This is a well known model that captures the idea of *reinforcement*. The model definition is that an urn is filled with arbitrary numbers of balls of two different colors. At each time step, a ball is drawn from the urn, replaced, and an additional ball of the same color is added to the urn. The convergence behavior of Polya's urn models is interesting. If the model is started with a single ball of each color—say red and blue—the ratio of red balls to the total number of balls converges almost surely to a uniform random variate on the interval $(0, 1)$.

This urn model can be generalized to allow an arbitrary number of colors, while leaving the reinforcement mechanism the same. For this model there are other convergence results. The article by Borgs specifically uses a result from Janson [51], that holds for a variant of Polya's model. The model in [51] is a *generalized* Polya's urn process (GPU);

Definition 3.5 (Generalized Polya's Urn process [51, pp. 1])**.** We define an urn model of q colors and for $n \in \mathbb{Z}_{\geqslant 1}$ write $X_n = (X_{n,1}, \ldots, X_{n,i}, \ldots, X_{n,q})$ for the vector that for each $1 \leqslant i \leqslant q$ represents the number of balls of color i at time n. In addition, let $(\xi_i) = (\xi_{i,1}, \ldots, \xi_{i,q})$ be the update vector of length q corresponding to color i, with $\xi_{i,j} > 0$. To each color we assign activity values $a_i \geqslant 0$. The evolution of the GPU model is given by two steps. In step one, choose a ball from the urn with the probability $p_{n,i}$ of taking a ball of color i at time n proportional to $a_i X_{n,i}$. Assume w.l.o.g. that we have chosen a ball of color i, we then obtain X_{n+1} as[1]

$$X_{n+1} = X_n + \vec{\xi}_i^{(n)}. \tag{3.2.4}$$

«

We see that the change to vector X_n is thus not limited to adding balls to the color of ball i. Instead the model allows changes to the count of any of the colors.

There is a convergence result that relates the limit of the $X_{n,i}$ to eigenvalue and vectors of the matrix A, with $A_{i,j} = a_i \mathbb{E}[\xi_{i,j}]$. [51, Theorem 3.21] guarantees that under some conditions, matrix A has greatest eigenvalue λ_1 and corresponding eigenvector u_1. For a vector v_1, that solves the equations $v_1 \cdot a = 1$, with a the vector of activities and $v_1 \cdot u_1 = 1$. The theorem then reads (in the words of [50])

Theorem 3.9 (Limit theorem for the GPU [51, Theorem 3.21])**.** *Assume conditions (A1–A6) of [51] are satisfied. Conditional on non-extinction (see [51]), we find* $n^{-1} X_n \xrightarrow{a.s.} \lambda_1 v_1$ *as* $n \to \infty$

We highlight the proof for the following result

Theorem 3.10 (Regimes in the BB-model with discrete fitness values)**.** *Let* $(\mathcal{G}_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ *be a graph sequence generated by the BB model with* $m = 1$*. Let* $\mathcal{F}^{(i)}$ *be the fitness of vertex* v_i*, where* $\mathcal{F}^{(i)}$ *is drawn from a bounded countably infinite probability mass function with supremum* h *such that the probability of drawing* $\mathcal{F}^{(i)}$ *is* q_i*. This graph exhibits two regimes dependent on fitness distribution*

- ***Fit get richer*** *when,*

$$\sum_{i=1}^{\infty} \frac{f_i q_i}{h - f_i} \geqslant 1; \tag{3.2.5}$$

- ***Innovation pays off*** *when,*

$$\sum_{i=1}^{\infty} \frac{f_i q_i}{h - f_i} < 1. \tag{3.2.6}$$

These regimes correspond to the regular multi-scaling as in Theorem 3.7 and the condensation outlined in Definition 3.4 respectively. We see that the condition for condensation corresponds to Eq. (3.2.3). Here q_i corresponds to $g(\epsilon)$, and f_i/h to $\exp[-\beta(\epsilon - \mu)]$. In Eq. (3.2.6) the distribution q_i of the f_i is tilted to weigh the values around h more heavily. If, despite this transformation, the weight of fitnesses is low, then the tail of the fitness distribution is light—high fitnesses are rare. Under these circumstances, the vertices with high fitness have an advantage, and their *innovation pays off*.

[1] Janson [51] allows $\xi_i^{(n)}$ to be a random variate, allowing another layer of stochasticity. A deterministic model is sufficient for Borgs *et al.* [50]

Urn embedding for BA-model

We now outline the proof structure of Theorem 3.10. The first step is to embed the model without fitness (BA) into the GPU. This is done by first fixing some value of $k \in \mathbb{Z}_{\geqslant 1}$, and then letting X_n be a count vector of $k+1$ colors. The colors $1 \leqslant i \leqslant k$ correspond to vertex degrees equal to 1 through k, color $k+1$ corresponds to vertex degree greater than k. To each of the colors i, we assign activity $a_i = 1$. The update vector $\xi_i^{(n)}$ is then given by

$$\xi_i^{(n)} = \begin{cases} 1 & \text{if } j = 1, \\ -i & \text{if } j = i, \\ i+1 & \text{if } j = i+1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.7}$$

for $1 < i < k+1$. For $i = 1$ and $i = k+1$, a slight modification can be made to respect the boundary conditions.

$$\xi_1^{(n)} = \begin{cases} 2 & \text{if } j = 2, \\ 0 & \text{otherwise,} \end{cases} \qquad \xi_{k+1}^{(n)} = \begin{cases} 1 & \text{if } j = 1, \\ 1 & \text{if } j = k+1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.8}$$

The count of each color i thus reflects the number of half-edges connected to vertices with degree i, or degree $\geqslant i$ if $i = k+1$. We see in the case that $1 < i < k+1$, that if a half edge is added to a vertex of degree i, this means that now, i fewer half edges are connected to degree i, since the vertex they connect to is now of degree $i+1$. Using this model, one can find the stationary values of a power law result as in Section 3.1.3 by use of Theorem 3.9.

Embedding for the competition of fitnesses

In order to extend this technique to the BB model, the choice of fitness distribution μ of the BB model is first restricted to distributions with a countably finite support. This means that there are $1 < j < j_{\max}$ values $\mathcal{F}^{(i)}$ that can be attained. This has to be done since the GPU model is not defined for continuous or infinite colors. Borgs *et al.* then use another urn embedding to analyze the competition between fitness values. First, they define $M_{n,j}$ as the number of half-edges at time n that are attached to a vertex with fitness value f_j. If vertex n connects to a vertex with fitness f_j, one half edge is added to color j, and one half edge is randomly assigned with probability q_ℓ to fitness f_ℓ. This corresponds to the random fitness of the added vertex. Therefore, the update vector ξ_j^n is given by

$$\xi_j^n = e_j + \text{cat}(e_1, \ldots, e_{j_{\max}})(\vec{q}), \tag{3.2.9}$$

where cat denotes the categorical distribution (multinomial distribution for $n = 1$). We can then set the activation a_j to be f_j and define λ_0 as the largest solution to

$$\sum_{j=1}^{j_{\max}} \frac{f_j q_j}{\lambda_0 - f_j} = 1, \tag{3.2.10}$$

and note that, since the LHS of Eq. (3.2.10) is decreasing in λ_0, the solution must be larger than $f_{j_{\max}}$. Let ν_j be twice the limiting probability that a uniformly chosen half-edge is connected to a vertex with fitness f_j. Then, using Theorem 3.9, we can compute v_1 and λ_1. For the BA model, the equality $\lambda_1 = \lambda_0$ holds and we find the ν_j increase as

$$\frac{M_{n,j}}{n} \xrightarrow{n} \nu_j = \lambda_0 \frac{q_j}{\lambda_0 - f_j}. \tag{3.2.11}$$

The values of ν_j, thus correspond form the limit of the empirical fitness, when samples uniformly on the half-edges. We find that, relative to base rate q_j, the high fitnesses attract a larger proportion of half-edges.

Combining the embeddings

With these embeddings that give results for competition between degrees and between fitness values, an urn representation for the entire model can be made. The discrete fitness levels then form another dimension of colors so that we have $kj_{\max}$ different colors. For this model, it is again possible to define transition rules such as Eqs. (3.2.7) to (3.2.9), that apply to the colors of a certain fitness value f_j and degree i. Instead of choosing the balls with equal activations, we let the activity a of a fitness value f_j and vertex degree k be f_j. Borgs *et al.* find the eigenvalue $\lambda_1 = \lambda_0$ as in Eq. (3.2.10). Moreover, solving for the eigenvector v_1 and left eigenvector u_1, such that $u_1 \cdot v_1 = 1$ and $a \cdot v_1 = 1$ Borgs *et al.* find

$$(u_1)_i = \frac{f_i}{\lambda_0 - f_i}, \qquad (v_1)_i = \lambda_0 \frac{q_i}{\lambda_0 + f_i}. \tag{3.2.12}$$

From this, the limiting probability $\eta_{j,k}$ of finding a vertex v with fitness f_j and degree k is established and asymptotic decay is given with big-Θ notation (see Notation B.1) as;

$$\eta_{j,k} = \nu_j \frac{1}{k} \prod_{\ell=2}^{k} \frac{\ell}{\ell + \frac{\lambda_0}{f_j}}, \quad \text{so that as } n \to \infty, \quad \frac{N_n(j,k)}{n} \to \eta_{j,k} = \Theta(\nu_j k^{-(1+\lambda_0/f_j)}). \tag{3.2.13}$$

However, this result only applies to discrete distributions with finite support. In order to extend the result to countably infinite distributions. A coupling argument is made to couple the countably infinite case to two graphs with finitely supported fitness distributions. Here, the truncated fitnesses are replaced by respectively 0 and h to make sure that the discrete fitnesses form upper and lower bounds for the values of $M_{n,j}/n$. From this, it is shown that the coupling is valid. However, whereas earlier for the finite case Eq. (3.2.13) guaranteed existence of summable $\vec{\eta}, \vec{\nu}$, in the infinite case, $\sum \nu_j$ may be less than 2. In this case, condensation occurs and the extra mass ends up with the vertices with highest fitness.

3.2.3 Limiting degree distribution—continuous case

Thus far in Sections 3.2.1 and 3.2.2, we have seen how an ingenious observation of the similarity of the BB model to the statistics of Bose gases gave a prospective condition for condensation (Eq. (3.2.3)) and how this condition was proven correct for fitness distributions with finite support. Although [50] claimed to leave out the proof for brevity, the first actual publication of this proof was given in an article by Dereich and Ortgiese [38]. This article does not rely on results for reinforcement processes such as the generalized Polya's urn in [50]. Instead, Dereich and Ortgiese [38] use a lemma from a classical paper by Robbins and Monro [52] laying the foundations of a technique called *stochastic approximation* (Appendix B.5). This technique finds its main use in applied mathematics, but can be used to derive convergence results. We first define the notation used in [38].

Definition 3.6. We define the *impact* of vertex v_i in directed graph $\mathcal{G}$ as

$$\operatorname{imp}_{\mathcal{G}}(i) := 1 + \deg^{-}_{\mathcal{G}}(i). \tag{3.2.14}$$

«

Initialize $\mathcal{G}_n$ as a single vertex v_1 with no edges. We let the attachment probability of new edges to vertex v_i be

$$\mathbb{P}(v_{n+1} \to v_i \mid \mathcal{G}_n) \propto \mathcal{F}^{(i)} \cdot \operatorname{imp}_{\mathcal{G}_n}(i). \tag{3.2.15}$$

Here m edges are added concurrently by multinomial distribution with probabilities as in Eq. (3.2.15). Without loss of generality, we let μ be defined such that $\operatorname{ess\,sup} \mu = 1$. We note that higher values of fitness only change the relative attractiveness between vertices, thus normalizing the essential supremum to 1 does not affect the generality of the proof. However, in case we allow the distribution to be unbounded, we note without proof that we always have condensation.

We note that, forgetting the order of in which edges are added to the graph, the evolution of the graph $\mathcal{G}_n$ is determined uniquely by the evolution of the impact at each time step. We can then define the following random measure.

Definition 3.7 (Random measure)**.** Let $(\Lambda_n)_{n\in\mathbb{Z}_{\geqslant 1}}$ be a sequence of random measures defined as

$$\Lambda_n = \frac{1}{n} \sum_{v_i \in V_n} \operatorname{imp}_{\mathcal{G}_n}(v_i) \cdot \delta_{\mathcal{F}^{(i)}}, \tag{3.2.16}$$

where δ_x represents the point measure that puts mass 1 on x. The measures Λ_n thus represent the (non-normalized) probability of finding a vertex of fitness $\mathcal{F}^{(i)}$ weighted by impact. «

This measure corresponds to the energies in Note 3.2 and the vector ν_j from [50]. Thus, $\Lambda_n((a,b))$ is the non-normalized probability that a half edge is connected to a vertex with fitness $\mathcal{F} \in (a,b)$. With this notation, we quote the following theorems.

Theorem 3.11 (Regimes in BB model [38, p. 390, Theorem 2.4])**.** *Let $(\mathcal{G}_n)_{n\in\mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by the BB model with i.i.d. fitnesses from fitness distribution μ. If,*

$$\int \frac{s}{1-s}\, \mu(ds) \geqslant m, \tag{3.2.17}$$

then we define $\theta^* \geqslant 1$ as the unique value such that

$$\int \frac{s}{\theta^* - s}\, \mu(ds) = m \tag{3.2.18}$$

and set $\theta^* := 1$ otherwise. We then conclude that $\lim_{n\to\infty} \bar{\mathcal{F}}_n = \theta^*$ almost surely and find two regimes:

(i) ***Fit get richer*** *If*

$$\int \frac{s}{1-s}\, \mu(ds) \geqslant m, \tag{3.2.19}$$

then (Λ_n) converges in the weak topology to Λ, where*

$$\Lambda(dx) := \frac{\theta^*}{\theta^* - x}\, \mu(dx). \tag{3.2.20}$$

(ii) ***Winner takes all*** *If*

$$\int \frac{s}{1-s}\, \mu(ds) < m, \tag{3.2.21}$$

then (Λ_n) *converges in the weak* topology to* Λ, *where*

$$\Lambda(dx) := \frac{1}{1-x}\,\mu(dx) + \left(1 + m - \int \frac{1}{1-y}\,\mu(dy)\right)\,\delta_1(dx). \tag{3.2.22}$$

Here, we have convergence of measures (μ_n) to μ in weak* topology if and only if for every continuous, bounded function $f \in C_0(\mathbb{R}^+)$, that vanishes at infinity, the sequence $\int f\,\mu_n(dx)$ converges to $\int f\,\mu(dx)$.

Behavior

From Theorem 3.11 we get a characterization of the behaviors in the BB model. We find that the condition as it held for discrete fitness has a continuous counterpart in Theorem 3.11. Moreover, finding the value of θ^* corresponds to finding λ_0, the eigenvalue of the transitions in the fitness vector from the model with discrete fitness.

Again we see that if high values of fitness are rare enough after weighting (see Fig. 3.1), condensation occurs. Here, in the limit as $n \to \infty$, a proportion equal to $1 + m - \int \frac{1}{1-y}\,\mu(dy)$ of half edges connects to vertices with fitness 1. This proportion corresponds to the proportion linear proportion of half-edges in Definition 3.4 that in the limit connect vertices with fitness 1. Moreover, the half edges, correspond to the bosons of Note 3.2 that take up the lowest energy state. All this gives a good idea of how the BB model condensates.

Proof technique

We do not go into detail about the proof structure, but will give a summary. The technique of stochastic approximation Lemma B.5 is a general technique that can be used to prove convergence for e.g., urn models [53] as in this article by Benaïm [54]. However, for continuous distributions of fitness, the urn approach can not be used. Dereich and Ortgiese instead consider the convergence of Λ_n for fitness intervals in $[a,b] \subseteq (0,1]$. We see here, that by definition Λ_n is an average over n, so the variations from $\Lambda_n([a,b])$ to $\Lambda_{n+1}([a,b])$ decrease as in Eq. (B.5.1). By analyzing using Lemma B.5, the impact measures converge if the normalization in Eq. (2.3.1) can be bounded tightly [38, pp. 397, Proposition 4.1].

In order to bound the denominator in Eq. (2.3.1), Dereich and Ortgiese consider

$$\bar{\mathcal{F}}_n := \frac{1}{mn}\sum_{i=1}^{n} \mathcal{F}^{(i)}\mathrm{imp}_{\mathcal{G}_n}(v_i). \tag{3.2.23}$$

First an operator T is defined, that given an upper bound θ for the limit $\bar{\mathcal{F}} := \lim_{n\to\infty} \bar{\mathcal{F}}_n$, gives a tighter upper bound $T(\theta)$, again using Lemma B.5 [38, pp. 399, Proposition 4.2]. Then, it is shown that a fix point θ^* of T exists, which is tight in some regard, which proves Theorem 3.11.

The proof by Dereich and Ortgiese therefore relies on first principle rather than embedding. This allows for a very general proof, that also works for, for instance, random out-degree. The authors thus call their analysis robust—technical changes to the model definition are often covered by their results.

Fitness distribution

Finally, we look at the degree distribution in the BB model with continuous fitness distributions. We first define

$$\Lambda_n^{(k)} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathrm{imp}_{\mathcal{G}_n}(v_i)=k\}}\,\delta_{\mathcal{F}^{(i)}}. \tag{3.2.24}$$

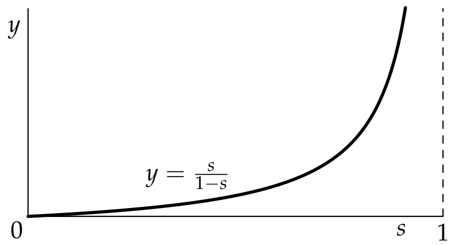


Figure 3.1: In Eq. (3.2.21), the fitness measure is transformed to an equivalent measure (Both the original and the derived measure have the same null sets). The inequality in Eq. (3.2.21) holds if the probability of finding fitnesses near 1 is very low.

This $\Lambda_n^{(k)}$, gives the non-normalized probability of finding a vertex with fitness $\mathcal{F}^{(i)}$, when impact is equal to k. Thus, $\Lambda_n^{(k)}(\{f_j\})$ corresponds to $N_n(j, k-1)/n$ in Section 3.2.2. The exact limit of measure $\Lambda_n^{(k)}$, is given as;

Theorem 3.12. *Let $(\mathcal{G}_n)_{n\in\mathbb{Z}_{\geqslant 1}}$ be a sequence of graphs generated by the Bianconi–Barabasi model with fitness distribution μ. Then let θ^* be defined as in Theorem 3.11. We then find that $\Lambda_n^{(k)}$ converges in the weak* topology to $\Lambda^{(k)}$, with*

$$\Lambda^{(k)}(ds) := \frac{1}{k+\frac{\theta^*}{s}}\frac{\theta^*}{s}\prod_{i=1}^{k-1}\frac{i}{i+\frac{\theta^*}{s}}\mu(ds). \tag{3.2.25}$$

From this, the limiting degree distribution can be derived as.

$$p_\infty(k) = \Lambda^{(k)}((0,1)) = \int_{(0,1)}\frac{1}{k+\frac{\theta^*}{s}}\frac{\theta^*}{s}\prod_{i=1}^{k-1}\frac{i}{i+\frac{\theta^*}{s}}\mu(ds). \tag{3.2.26}$$

We thus conclude this overview of the BB model and the developments that have led to a rigorous proof for the degree distribution of the BB model.

3.3 Additive Fitness

We now take a look at the model defined in Definition 2.9. In this section we first explore the qualitative behavior of the preferential attachment model with additive fitness (PAFA), by considering results from Lodewijks and Ortgiese [39]. We do not further explore this article, as there is an overlap in technique to [38]. Instead we study a proof strategy from Bhamidi [40] that is particularly powerful and rather intuitive.

3.3.1 Model Behavior

The model shows behavior that is generally similar to that of the Barabási–Albert model (Definition 2.6). The model however allows for a variety in qualitative behaviors when the fitness distribution is heavy tailed. We first define;

$$\Lambda_n := \frac{1}{n}\sum_{i=1}^{n}\deg^-_{\mathcal{G}_n}(v_i)\delta_{\mathcal{F}^{(i)}}, \qquad \Lambda_n^{(k)} := \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{\deg^-_{\mathcal{G}_n}(v_i)=k\}}\delta_{\mathcal{F}^{(i)}}, \qquad p_n(k) := \Lambda_n^{(k)}([0,\infty)). \tag{3.3.1}$$

These definitions are similar to those in Section 3.2.3. We now cite the main theorem.

Theorem 3.13 (Degree distributions in PAF models [39, pp. 5, Theorem 2.4])**.** *Consider the PAFUD model as in Definition 2.9 and suppose the fitness distributed with measure μ satisfy $\mathbb{E}[\mathcal{F}] := \int x\,\mu(dx) \leqslant \infty$. Let $\theta_m := 1 + \mathbb{E}[\mathcal{F}]/m$. Then, almost surely, for any $k \in \mathbb{Z}_{\geqslant 1}$, as $n \to \infty$,*

$$\Lambda_n \to \Lambda, \qquad \Lambda_n^{(k)} \to \Lambda^{(k)}, \qquad \text{and } p_n(k) \to p_\infty(k), \tag{3.3.2}$$

where the first two are with respect to the weak topology and the limits are*

$$\Lambda(dx) = \frac{x}{\theta_m - 1}\,\mu(dx), \qquad \Lambda^{(k)}(dx) = \frac{\theta_m}{x + \theta_m} \prod_{\ell=1}^{k} \frac{(\ell - 1) + x}{\ell + x + \theta_m}\,\mu(dx), \tag{3.3.3}$$

and

$$p_\infty(k) = \int_0^\infty \frac{\theta_m}{x + \theta_m} \prod_{\ell=1}^{k} \frac{(\ell - 1) + x}{\ell + x + \theta_m}\,\mu(dx). \tag{3.3.4}$$

This result is proven in the article by Lodewijks and Ortgiese, and makes use of the same stochastic approximation lemma as the proof for the limiting distribution of Theorem 3.11. We will omit a description of the technical proofs employed in this article, rather choosing to present an embedding technique from Bhamidi [40] in Section 3.3.2

The practical consequence of Theorem 3.13 is that the distribution of the degrees in the PAFUD model is altered depending on the fitness distribution with probability measure μ. We state the following result from Lodewijks and Ortgiese [39].

Assumption 3.14 (Assumption of distribution [39, pp. 5, Assumption 2.3])**.** *The fitness distribution is a power law with exponent $\alpha > 1$, i.e.,*

$$\mu(x, \infty) = \ell(x) x^{-(\alpha-1)}, \qquad \text{for } x > 0, \tag{3.3.5}$$

where ℓ is a slowly-varying function at infinity, i.e., for all $c > 0$, $\lim_{x\to\infty} \frac{\ell(cx)}{\ell(x)} = 1$.

Examples of slowly varying functions are functions that converge for $x \to \infty$ and $\log(x)$. We can now state the following theorem.

Theorem 3.15 (Degree distribution in PAFUD model [39, pp. 6–7, Theorem 2.6])**.** *Let $(\mathcal{G}_n)_{n\in\mathbb{Z}_{\geqslant 1}}$ be a graph sequence generated by PAFUD model (Definition 2.9) with fitness distribution μ and out-degree m. Let $p(k)$, $k \in \mathbb{Z}_{\geqslant 0}$ be as in Definition 3.1 and define $\theta_m := 1 + \frac{1}{m} \int x\,\mu(dx)$. The following behaviors occur;*

(i) ***Weak disorder*** *If $\int x^{\theta_m}\,\mu(dx) < \infty$, then as $k \to \infty$*

$$p(k) \sim Ck^{-(1+\theta_m)}, \qquad \text{where } C := \theta_m \int_0^\infty \frac{\Gamma(x + \theta_m)}{\Gamma(x)}\,\mu(dx). \tag{3.3.6}$$

(ii) ***Strong disorder*** *If probability measure μ satisfies Assumption 3.14, then,*

a) if $\alpha = 1 + \theta_m$ and $\int x^{\theta_m}\,\mu(dx) = \infty$, we find as $k \to \infty$

$$p(k) = \Theta\left(C(k) k^{-(1+\theta_m)}\right), \qquad \text{where } C(k) := \int_1^k \ell(x)/x\,dx. \tag{3.3.7}$$

Here we use big-Θ notation described in Notation B.1.

b) If $\alpha \in (2, 1+\theta_m)$, then as $k \to \infty$,

$$p(k) = \Theta(\ell(k)k^{-\alpha}). \tag{3.3.8}$$

(iii) ***Extreme disorder*** *If probability measure μ satisfies Assumption 3.14 and if $\alpha \in (1,2)$, consider a uniformly chosen vertex v in $\mathcal{G}_n$ and let $\epsilon > 0$. Then for n sufficiently large*

$$\mathbb{P}(\deg^-_{\mathcal{G}_n}(v) = \deg^-_{\mathcal{G}_1}(v)) \geqslant 1 - Cn^{-\frac{1}{\alpha}\min(2-\alpha,\alpha-1)+\epsilon}, \tag{3.3.9}$$

where $C > 0$ is some constant.

We first present several observations [39]. As a consequence of Theorem 3.15 in the weak disorder regime, the PAFUD model has the same tail exponent for the degree distribution as the BA model with initial attractiveness $\delta = \int x\,\mu(dx)$, the average fitness. In the strong disorder regime, if the fitness distribution satisfies Assumption 3.14, i.e., it is power-law distributed then depending on the power law exponent α, two situations occur. If $\alpha = 1+\theta_m$ then the model's tail exponent is asymptotically bounded above and below by the power law $k^{-(1+\theta_m)}$ multiplied with a factor of approximate order $\log k$ when k is large. This is the border case between weak and strong disorder. If α lies between 2 and $1+\theta_m$, then the θ_m-th mean of the fitness distribution is undefined. The degree distribution then is asymptotically bounded above and below by $\ell(k)k^{-\alpha}$. The extreme disorder regime occurs when μ is power-law distributed with tail exponent α between 1 and 2, such that the mean of the distribution is infinite. Under these conditions, there is a growing probability that a random vertex chosen from a graph at step n has not increased its in-degree since initialization.

3.3.2 Outline of proof technique by Bhamidi

As the derivation of Theorems 3.13 and 3.15 is too long to include here, we cite a very general result from an unpublished draft [40]. This work embeds the preferential attachment tree, i.e., preferential attachment with $m = 1$, into continuous time branching processes. In this continuous time branching process, the intensity of the associated birth process is equal to the fitness function f, which for the additive model is $f(i) = \deg^-_{\mathcal{G}_n}(v_i) + \mathcal{F}^{(i)}$. Here we note, that the in-degree here is only correct in the preferential attachment sense. In branching processes, the older, parent vertex points to the child vertex, inverting the relation in the preferential attachment model.

Once we embed the PAFA model into a multi-type CTBP, results from the general theory of branching processes become applicable and allow us to derive a result similar to Theorem 3.13.

We cite this result below.

Theorem 3.16 ([40, pp. 34, Corrolary 50]). *As before fix any sequence of finite stopping times $T_n \uparrow \infty$ almost surely. Let $N_n^{\geqslant,-}(k)$ be the number of vertices with in-degree greater than k at time n (in-degree in the original sense). Then:*

$$\lim_{n\to\infty} \frac{N_n^{\geqslant,-}(k)}{n} \xrightarrow{\mathbb{P}} \prod_0^{k-1}\left(\frac{f(i)}{\theta + f(i)}\right), \qquad \text{for } k = 1,2,3,\ldots, \tag{3.3.10}$$

where $\theta > 0$ is the solution to the equation

$$\theta \int_0^\infty e^{-\theta t}\mathbb{E}[\,\mathcal{P}(t)\,]\,dt = 1, \tag{3.3.11}$$

where $\mathcal{P}$ is the underlying birth process of the continuous time branching process.

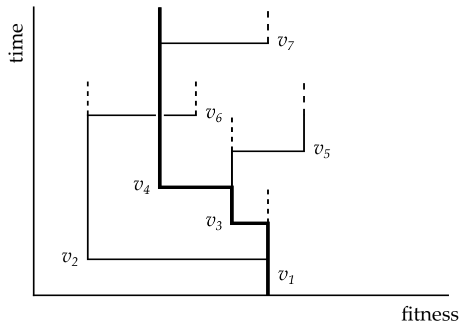


Figure 3.2: Illustration of branching process with random vertex parameter assigned to each new vertex. The genealogy of vertex v_4 is given in bold. Each vertex has an associated birth process whose arrival times mark the spawn of a new vertex. The new vertices are assigned a random fitness value, shows as a lateral jump in this figure. See also [55]

The embedding

We now turn to a description of the embedding used to prove Theorem 3.16 in [40]. In the article, the additive fitness model Definition 2.9 with $m = 1$ is embedded into a *multi-type continuous time branching process* (multi-type CTBP) (see Fig. 3.2. This model has been studied by Jagers and Nerman [55], with a result predicting that, under technical assumptions, a multi-type CTBP will have an empirical distribution that converges in probability to a stable distribution. This empirical distribution describes "a typical individual, her back- ground and future when sampling from among all those born, dead or alive." [55].

A useful result pertaining to multi-type CTBP is given by Aldous [56]. Aldous proves that the distribution of sub-trees rooted in a randomly chosen vertex v, for many tree models, converges to the distribution of sub-trees in a sin-tree—a tree with a single infinite path—as $n \to \infty$. A consequence of this is that functionals on the random trees also converge to the distribution of the functional on a sin-tree. Bhamidi uses these results to prove convergence of the empirical degree distribution as in Theorem 3.16.

At time $t = 0$, we initialize the random tree $\mathcal{T}_0$ with a single vertex v_1 with fitness $\mathcal{F}_1$ drawn from distribution μ. Each vertex v in $\mathcal{T}$ has an associated birth process[2], with rate function $r(t) = \lambda(\mathcal{F}_1 + \deg^+_{\mathcal{T}_t}(v) + 1)$, that is started when the vertex is added to the graph. We can liken these birth process to faulty alarm clocks, that go of randomly, where for $(a, b) \subseteq \mathbb{R}^+$, $\int_{(a,b)} r(s)\, ds$, is the expected number of times the alarm clock will go off in an interval (a, b).

Given vertex v_i, whose associated birth process has arrival times $T_{i,1}, T_{i,2}, \ldots$ we add a vertex as a child to v_i at each of these arrival times. Each vertex v that is added to the tree at time t is assigned a fitness from distribution μ, and an independent birth process with rate function $r(t) = \lambda(\mathcal{F}_1 + \deg^+_{\mathcal{T}_t}(v) + 1)$, started at time t. Thus, each time the alarm of v_i rings, we add a child node to v_i.

By stopping this process at time $\tau(n) = \inf\{t \geqslant 0 || \mathcal{T}| = n\}$, we obtain a tree with the same law as a tree generated by the PAFA model with fitness distribution μ and $m = 1$. We can associate the vertices in the PAFA model with those in the multi-type CTBP by labeling the vertices $1, \ldots, n$ by order of arrival in the graph. In fact, we can see that this model is related to the urn model in Section 3.2.2. If we consider the degree counts of the $\mathcal{T}_{\tau(n)}$ trees, that are stopped when the n-th vertex is added, we find the same transition properties as for the urn model.

[2]Definitions in Random Services Kyle Siegrist CC BY-SA 2.0

By using this simple embedding a wealth of probabilistic results for graph models can be proven. The unpublished article of Bhamidi [40] is therefore a worthwhile read. We omit further details about this proof.

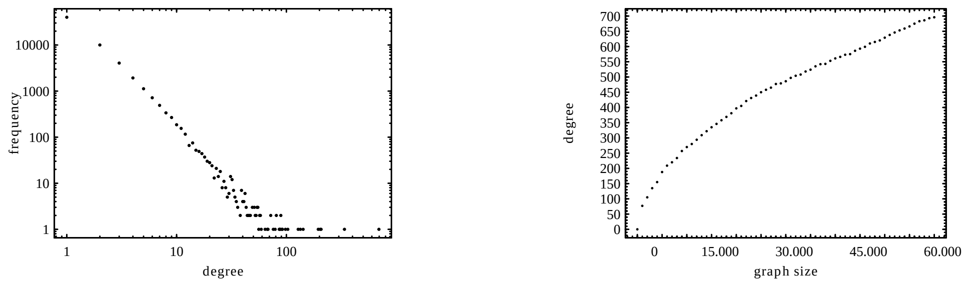
3.4 Simulation

In this section we visualize the degree distributions for the various qualitative regimes we discussed in Chapter 3. We will discuss examples fitness distribution that exhibit behavior from the respective regimes. For the purpose of this section, we consider the BA model with initial attractiveness Definition 2.7 to be a model with additive fitness, as it compares naturally to the PAFUD model.

3.4.1 Multiplicative fitness

For models with multiplicative fitness we first illustrate the ordinary BA model (Fig. 3.3), which is a baseline. We can see the inversely proportional relation (with some power γ) between degree and the frequency of occurrence from the approximately linear relation in the log-log plot Fig. 3.3a. Moreover, we see the polynomial (sub-linear) growth as predicted in Proposition 3.2 of an individual vertex v_1, which has maximum degree at $n = 60\,000$. We can see that this is a good example of *rich get richer*. However, as v_1 is also the oldest vertex in the graph, this behavior is also predictable. The BA model leads to behavior called *old get richer*.

Figure 3.3

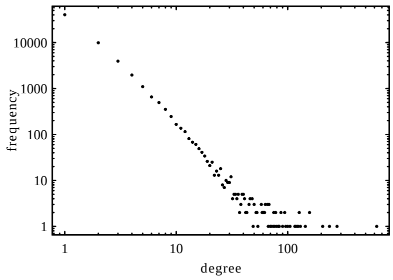


(a) Frequency vs. degree log-log plot. Generated from BA-model with $m = 1$, to size $n = 60\,000$. Graph initialized with $V = \{v_1\}$, with 1 self-loops.

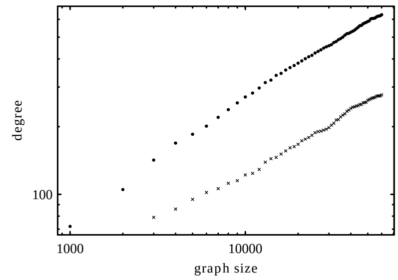
(b) Degree development of vertex v_1, which has highest degree at $n = 30\,000$.

In Fig. 3.4 we see an example of the BB model with fitnesses drawn form a uniform distribution with support $[0.6, 1]$. In this model, we still have an inversely proportional relation between frequency and degree raised to some exponent as shown in Fig. 3.4a. Moreover in Fig. 3.4b, we see the sub-linear growth for vertices of two different fitnesses. Here we note that by Theorem 3.8, the exponent of growth is linear with fitness, which is apparent from the different linear growth rates in the log-log plots. This is the phenomenon of multi-scaling. While the exponents for the growth of degrees of individual vertices varies, they all are sub linear (Fig. 3.4c).

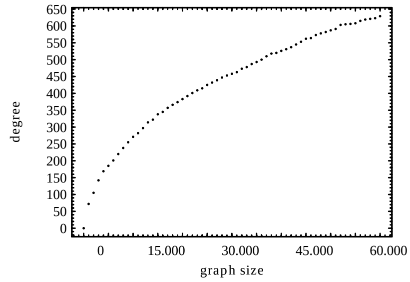
Figure 3.4



(a) Frequency vs. degree log-log plot. Generated from BB-model with $\mu = \text{Unif}(0.6, 1)$, $m = 1$, to size $n = 60\,000$. Graph initialized with $V = \{v_1\}$, with 1 self-loop.



(b) Degree development of vertex v_1, which has highest degree at $n = 60\,000$ (•), $\mathcal{F}_1 = 0.889$ and vertex v_2 (×) with fitness $\mathcal{F}_2 = 0.710$. Graph with log-log scales.



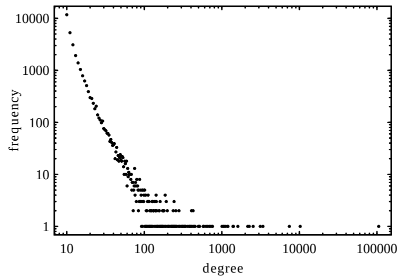
(c) Degree development of vertex v_1, which has highest degree at $n = 60\,000$, $\mathcal{F}_1 = 0.889$.

In Fig. 3.5, we see an example of condensation. This graph was generated with fitness distributed with the β distribution with density f, which is given by

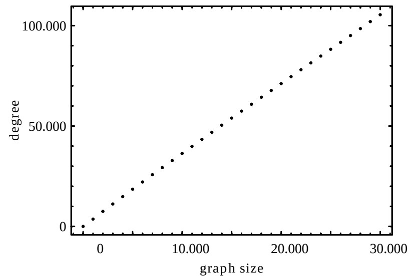
$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \tag{3.4.1}$$

with Γ denoting the Gamma-function. Calculating $\int \frac{x}{1-x}\,\mu(dx) = 4 < m = 10$, we find that we will have condensation by Theorem 3.11. We note that in Fig. 3.5a, we have large outliers in degree. In fact, the largest outlier has a degree that grows approximately linearly in n (Fig. 3.5b. We see that this vertex has a fitness that is quite far from the maximum fitness 1, but still ranks high in percentile score. Theorem 3.11 predicts that the proportion of half-edges connecting to the condensate will end up in with fitness 1 as $n \to \infty$, we find that as we simulate for finite n, we can only see a pre-asymptotic regime, in which fitness values that are high relative to the competition attract edges. Would we simulate on, we would see that Fig. 3.5b is in fact sub-linear and we are looking at a level of locality in which it appears linear. If we consider a vertex with average fitness, we see that such a vertex increases its degree sub-linearly.

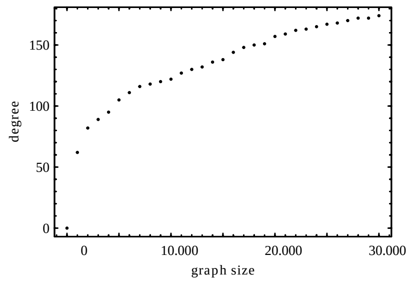
Figure 3.5



(a) Frequency vs. degree log-log plot. Generated from BB-model with $\mu = \beta(1, 5)$, $m = 10$, to size $n = 30\,000$. Graph initialized with $V = \{v_1\}$, with 10 self-loops.



(b) Degree development of vertex v_{10}, which has highest degree at $n = 30\,000$. Vertex fitness $\mathcal{F}_{10} = 0.65$ (99th percentile).



(c) Degree development of vertex with v_7, with fitness 0.19 (66th percentile).

We here choose to discuss these examples intuitively, and do not verify the results proven in Chapter 3 experimentally as this is not the focus of this thesis. We now continue with a discussion of graph models with additive fitness.

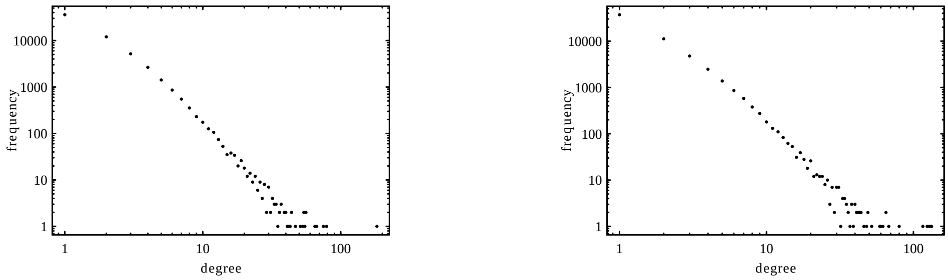
3.4.2 Additive fitness

Going back to the BA model, Theorem 3.1 predicts that the exponent of the inverse relation between frequency and degree grows as $3 + \frac{\delta}{m}$ if we add a δ as in Definition 2.7. Additionally Theorem 3.15, predicts that for additive fitness distributed with a light tail, exponent of the degree distribution is the same as for the BA model with δ equal to the average of the fitness distribution. We show this visually in Fig. 3.6, for fitness distributed with the exponential distribution, and remark that, compared to Fig. 3.3, the frequency–degree plot indeed decays at a faster rate. Moreover Figs. 3.6a and 3.6b are not visually distinguishable in this plot. The exponential distribution has density f, with,

$$f(x; \lambda) = \lambda e^{-\lambda x}. \tag{3.4.2}$$

This means that $\theta_m = 1.1$ and $\int x^{\theta_m} \mu(dx) < \infty$, which by Theorem 3.15, means that this is the weak disorder regime.

Figure 3.6



(a) Frequency vs. degree log-log plot. Generated from BA-model with δ, where $\delta = 1$, $m = 1$, to size $n = 60\,000$. Graph initialized with $V = \{v_1\}$, with 1 self-loop.

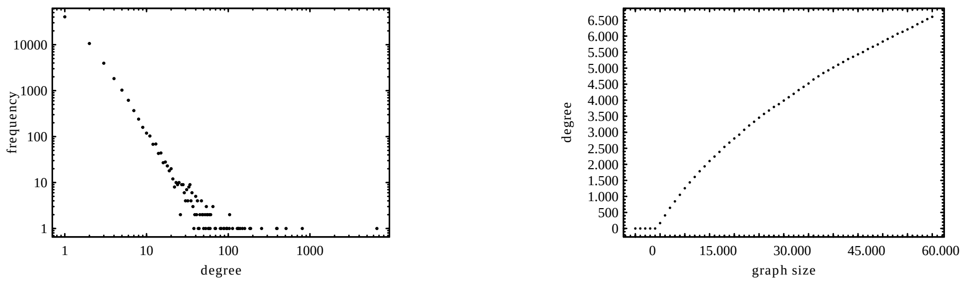
(b) Frequency vs. degree log-log plot. Generated from PAFUD model, where $m = 1$, and $\mu = \exp(1)$, such that $\mathbb{E}[\mathcal{F}] = 1$. Grown to size $n = 60\,000$. Graph initialized with $V = \{v_1\}$, with 1 self-loop.

We now turn to the strong disorder regime from Theorem 3.15. Considering the strong disorder regime (Fig. 3.7), we see that the decay of the degree distribution is not visually distinguishable from that in Fig. 3.6. This is in contrast to the prediction in Theorem 3.15, but the very heavy tail of the Pareto distribution makes simulation difficult. We finally see that the outliers in the degree distribution come from the tail of the fitness distribution. We use the Pareto distribution, with support (η, ∞), defined by density f;

$$f(x; \eta, \theta) = \frac{\theta \eta^\theta}{x^{\theta+1}}. \tag{3.4.3}$$

This distribution differs from the distribution in Assumption 3.14, where a different parametrization is used, i.e., $\alpha = \theta + 1$. Using $\theta = 1.1$, $\mathcal{F} \sim \text{Par}(1, 1.1)$, we we have $\theta_m = 2$, such that x^{θ_m} has unbounded expectation. Since $\alpha = 2.1 \in (2, 1 + \theta_m = 3)$, we are in the strong disorder regime and expect to see the degree distribution decay slower than the models in Fig. 3.6. This is not apparent from the figure. We do however see that in this regime, there are more outliers. These outliers might superficially look like vertices exhibiting condensation, but Fig. 3.7b tells otherwise. The degree of the maximum-degree vertex grows sub-linearly.

Figure 3.7

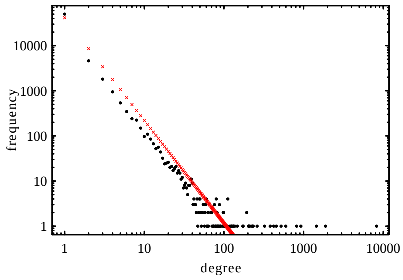


(a) Frequency vs. degree log-log plot. Generated from PAFUD model, where $m = 1$, and $11\mathcal{F} \sim \text{par}(1, 1.1)$, such that $\mathbb{E}[\mathcal{F}] = 1$. Grown to size $n = 60\,000$. Graph initialized with $V = \{v_1\}$, with 1 self-loop.

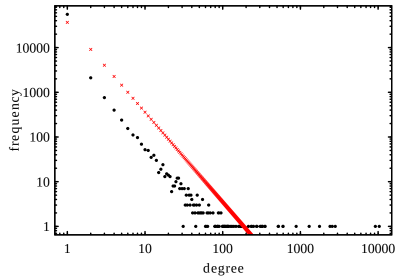
(b) Degree development of vertex v_{4375}, which has highest degree at $n = 60\,000$. Vertex v_{4375} has fitness 4603.8, the maximum fitness in this simulation.

If we use estimation of the tail exponent of degree distribution, we can generate a plot such as in Fig. 3.8, where we compare the degree distribution to a fitted discrete Pareto distribution with index estimated as in Appendix C. Here we find that the exponent of the degree distribution does in fact follow the exponent of the degree distribution, as predicted by Theorem 3.15. The differences between these measurements are not statistically significant in current form, but should be for a larger sample size. Due to computational constraints, we choose not to analyze further.

Figure 3.8



(a) Frequency vs. degree log-log plot. Graph (•) generated from the additive preferential attachment model with fitness values distributed according to $\mathcal{F}^{(i)} \sim \exp(\frac{10}{3})$, with $m = 1$ and $n = 60\,000$, such that $\mathbb{E}[\mathcal{F}] = .3$, and $\gamma = 2.3$. Compared with discrete Pareto distribution (×) with estimated parameter $\hat{\gamma} = 2.276 \pm 0.108$.



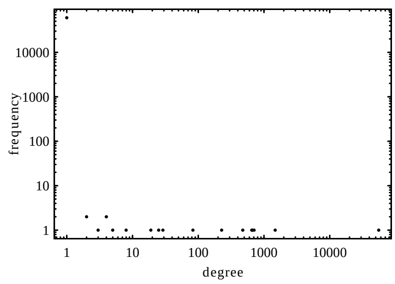
(b) Frequency vs. degree log-log plot. Graph (•) generated from the additive preferential attachment model with fitness values distributed according to $\mathcal{F}^{(i)} \sim \mathrm{Par}(\frac{3}{110}, 1.1)$, with $m = 1$ and $n = 60\,000$, such that $\mathbb{E}[\mathcal{F}] = .3$, and $\gamma = 2.1$. Compared with discrete Pareto distribution (×) with $\hat{\gamma} = 2.0095 \pm 0.108$.

Finally we consider the extreme disorder case (Fig. 3.9), here we use the Pareto distribution with infinite mean. In this case we see that only 17 vertices have attracted new edges since their initialization as expected from Theorem 3.15. We used $\alpha = 1.1$ as parameter, which would predict

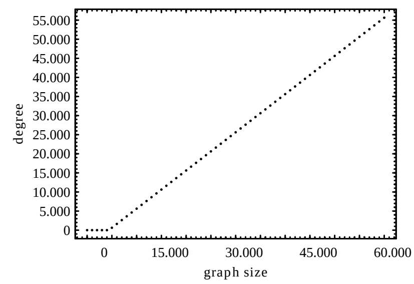
$$\mathbb{P}(\deg^-_{\mathcal{G}_n}(v) = \deg^-_{\mathcal{G}_1}(v)) \geqslant 1 - Cn^{-\frac{1}{11}+\epsilon} \approx 0.64, \tag{3.4.4}$$

when $n = 60\,000$. We see for our simulation that this percentage ends up a lot higher. In addition, it appears that the vertex with maximal fitness takes up a linear proportion of edges. However, it is not at all impossible that a vertex with higher fitness comes along. The regime name *extreme disorder* therefore seems fitting.

Figure 3.9



(a) Frequency vs. degree log-log plot. Generated from PAFUD model, where $m = 1$, and $\mathcal{F} \sim \mathrm{Par}(1, .1)$, such that $\mathbb{E}[\mathcal{F}] = \infty$. Grown to size $n = 60\,000$. Graph initialized with $V = \{v_1\}$, with 1 self-loop.



(b) Degree at time n of vertex v_{4375}, with highest degree at $n = 60\,000$. Vertex $v_{4375}{}^{\dagger}$ has fitness $5.617838 \cdot 10^{51}$, the maximum fitness in this simulation. (†) Simulation carried out with same seed as used for Fig. 3.7

Chapter 4

Discussion and conclusion

In this chapter, we present a discussion and a summary of the models discussed in this thesis. First, we outline the different qualitative behaviors that occur in preferential attachment models. We then discuss proof techniques used to prove results about the degrees of preferential attachment models and the applicability of these proof techniques. Finally we consider the limitations of preferential attachment graph models and we consider further lines of research.

Qualitative behavior of models with fitness

In Chapters 2 and 3 we first presented and then analyzed popular graph models with preferential attachment (PA) by studying existing literature. The PA model without fitness or BA-model was originally an answer to the question, "Why do data sets exhibit power-law dynamics?" The BA-model has a power-law distribution with power-law exponent $\gamma = 3$ of its degrees and functions as a source of inspiration for a *zoo* of further models.

An important variant of the BA-model is the preferential attachment with initial fitness model in Definition 2.7, this model allows for a slight tuning of the power-law distribution by adding a tunable parameter δ, acts homogeneously on attractiveness. This modification allows us to attain $\gamma = 3 + \frac{\delta}{m} \in (2, \infty)$. The tunable parameter thus allows to fit the model to real-world network data. It is being used by applied scientists as a base model for networks in the real world.

Bianconi–Barabási model

In this thesis we studied two other variants of the BA-model that both add a fitness parameter to each vertex in the model. We first considered the Bianconi–Barabási (BB) model, which changes the attachment probability by modifying it with a multiplicative fitness term. Bianconi and Barabási [33, 34] discovered that the addition of this term allows vertices to gain connections at different rates for the same degree, a phenomenon known as multi-scaling. We find this phenomenon for instance in citation networks, where articles by Balázs and Komjáthy [57] and Komjáthy and Lodewijks [58], respectively published in 2008 and 2020 were cited respectively 15 and 18 times, according to Google Scholar.

Moreover, if we consider the model where vertex fitness values are drawn from a distribution with measure μ three qualitative regimes occur depending on μ. These are respectively known as "old get richer", "fit get richer" and "fitness pays off" [50]. We summarize the behavior in the three phases below

- **Old get richer** The old get richer regime occurs when the fitness distribution is flat. Each nodes attractiveness depends only on degree and degree evolution occurs like in the BA-model.
- **Fit get richer** In the fit get richer phase, fitter vertices gain connections faster than vertices with the same degree but lower fitness. This is the multi-scaling effect described in [33].
- **Fitness pays off** When the fitness distribution is sufficiently rarefied for high values of fitness—i.e., only a few vertices have very high fitness—condensation occurs. In this regime, a sub-linear proportion of vertices gains a linear proportion of new edges.

We have seen that there exist a correspondence of the last phase to Bose-Einstein condensation in physics. We studied fitness distributed by the $\beta(1,5)$ distribution as an example of a distribution that causes condensation. Moreover, we note that distributions with countably finite support never cause condensation and distributions with unbounded support always cause condensation.

Additive fitness model

The model with additive fitness is closely related to the PA model with initial attractiveness. The parameter δ here becomes a fitness assigned per vertex and chosen randomly from a distribution μ. Much like the BB model, the properties of this distribution determine the qualitative behavior of the degree sequence. Lodewijks and Ortgiese [39] discern three regimes.

- **Weak disorder** Weak disorder is comparable to the BA-model with initial attractiveness δ equal to the mean fitness $\mathbb{E}[\mathcal{F}] = \int x\,\mu(dx)$. It occurs when the fitness distribution with probability measure μ has a tail that is lighter than the degree distribution of BA-model with $\delta = \mathbb{E}[\mathcal{F}]$ would have.
- **Strong disorder** When the fitnesses in the model are distributed according to a power-law distribution with a heavier tail than in the weak disorder regime, but with a finite mean, then the power-law exponent of the degree distribution in the PAFUD model is the same as for the fitness distribution.
- **Extreme disorder** When the fitness distribution is sufficiently heavy-tailed—it has an infinite mean—then for sufficiently high n a randomly chosen vertex has a high probability of not having attracted any connections since initialization.

Universal Proof techniques

In order to find the limiting degree distributions of graphs with preferential attachment, we have seen various strategies. First, for the BA model, we considered heuristic methods, that relied on calculation of the expected growth of the degree of a vertex as it ages. When calculating the proportion of vertices with degree greater than k, these techniques relied on substituting degree for expected degree.

Starting with Bollobás *et al.* [35], direct techniques have been used to derive the proportion of vertices degrees with degree k. This can be done by establishing bounds on the number of vertices with degree k [37]. Thus results such as Theorem 3.1 can be proven rigorously.

To derive a degree distribution for models with fitness (multiplicative and additive) we need bounds on both the number of half-edges linked to vertices with degree k and for the distributions between fitness values. Doing this directly requires bounds on the denominator of the attachment probabilities [38]. We can rather use embeddings to make use of known statistical results.

Firstly, we can make use of urn-models with reinforcement, by defining the transitions between colors representing fitness and degree. Secondly, we can embed the PA model into multi-type continuous time branching processes. Here we noted, that urn models themselves occur in these branching processes—the models are naturally related. Branching processes have powerful convergence results for functionals on sub-trees, that allow the derivation of a variety of results for PA models.

Further research

Mathematically, there is a good understanding of the models described in this thesis. However, there is active research to find models that better describe the real world. Here we note that though PA models, with $m > 1$, are generally small-worlds, they do not have realistic clustering—neighbors of neighbors are not connected more than unrelated nodes. Additionally, many real world networks have spatial characteristics, making it more likely for nodes to connect if they are close together geographically. Lastly, the model is static, i.e., there are no new connections between nodes after entry and connections do not disappear.

We note several developments in modeling;

- **Spatial models** In order to improve clustering properties, nodes are given a position in space. Connections to vertices that are close occur with higher probability, giving better clustering properties [59].
- **Rewiring** The hierarchical nature of preferential attachment models (nodes only connect to older nodes) is not realistic. By rewiring connections between existing nodes, this problem is tackled while preserving power-law distributions [60].
- **Changing Fitness** Instead of assigning a fixed fitness at birth, a node may have changing fitness. In the citation network example, we may consider that old articles are losing fitness [61].

In addition, the estimation of fitness distributions in PA models has received significant attention [62, 63]. Lastly, stochastic processes, e.g., spread of disease, are analyzed on PA models [64]. Here we note that better understanding of these processes might make epidemic management better in the future, which is just one example showing the relevance of preferential attachment models.

A. References

[1] V. Pareto, "LA LEGGE DELLA DOMANDA," *G. degli Econ.*, vol. 10 (Anno 6, pp. 59–68, 1895, ISSN: 11252855. [Online]. Available: http://www.jstor.org/stable/23219874.

[2] J. Travers and S. Milgram, "An Experimental Study of the Small World Problem," *Sociometry*, vol. 32, no. 4, p. 425, Dec. 1969. DOI: 10.2307/2786545.

[3] S. Milgram, "The small world problem," *Psychol. Today*, vol. 2, no. 1, pp. 60–67, 1967, ISSN: 00134252.

[4] P. Erdős and A. Rényi, "On random graphs I," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.

[5] G. U. Yule, "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S," *Philos. Trans. R. Soc. London. Ser. B, Contain. Pap. a Biol. Character*, vol. 213, no. 402-410, pp. 21–87, 1925. DOI: 10.1098/rstb.1925.0002.

[6] J. L. Moreno, *Who shall survive?: A new approach to the problem of human interrelations.* Washington: Nervous and Mental Disease Publishing Co, 1934. DOI: 10.1037/10648-000.

[7] D. J. de Solla Price, "Networks of Scientific Papers," *Science*, vol. 149, no. 3683, pp. 510–515, 1965. DOI: 10.1126/science.149.3683.510.

[8] F. Karinthy, "Chain-links," in *Struct. Dyn. Networks*, vol. 9781400841, Princeton University Press, Dec. 2011, pp. 21–26, ISBN: 9781400841356. DOI: 10.1515/9781400841356.21.

[9] I. de Sola Pool and M. Kochen, "Contacts and influence," *Soc. Networks*, vol. 1, no. 1, pp. 5–51, 1978. DOI: 10.1016/0378-8733(78)90011-4.

[10] G. K. Zipf, *The psycho-biology of language.* Oxford, England: Houghton, Mifflin, 1935, pp. ix, 336–ix, 336.

[11] G. K. Zipf, *Human behaviour and the principle of least effort an introduction to human ecology*. Cambridge, Mass: Addison-Wesley Press, 1949.

[12] G. K. Zipf, "The unity of nature, least-action, and natural social science.," *Sociometry*, vol. 5, no. 1, pp. 48–62, 1942. DOI: 10.2307/2784953.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. July 1928, pp. 379–423, 1948. DOI: 10.1145/584091.584093.

[14] B. Mandelbrot, "An Informational Theory of the Statistical Structure of Language," *Commun. Theory*, no. 2, 1953.

[15] H. A. Simon, "On a Class of Skew Distribution Functions," *Biometrika*, vol. 42, no. 3/4, pp. 425–440, 1955. DOI: 10.2307/2333389.

[16] D. J. de Solla Price, "A general theory of bibliometric and other cumulative advantage processes," *J. Am. Soc. Inf. Sci.*, vol. 27, no. 5, pp. 292–306, 1976. DOI: 10.1002/asi.4630270505.

[17] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, Jun. 2001. DOI: 10.1103/RevModPhys.74.47.

[18] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, Oct. 1999. DOI: 10.1145/316194.316229.

[19] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks.," eng, *Nature*, vol. 407, no. 6804, pp. 651–654, Oct. 2000. DOI: 10.1038/35036627.

[20] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the World-Wide Web," *Nature*, vol. 401, no. 6749, pp. 130–131, 1999. DOI: 10.1038/43601.

[21] R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, 2004, ISBN: 0521826985.

[22] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, "Dynamical and correlation properties of the internet," *Phys. Rev. Lett.*, vol. 87, no. 25, p. 258 701, 2001.

[23] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, "Topology, hierarchy, and correlations in Internet graphs," in *Complex networks*, Springer, 2004, pp. 425–440.

[24] A. Broder *et al.*, "Graph structure in the web," *Comput. networks*, vol. 33, no. 1-6, pp. 309–320, 2000, ISSN: 1389-1286.

[25] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998. DOI: 10.1038/30918.

[26] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," *Proc. Natl. Acad. Sci.*, vol. 97, no. 21, pp. 11 149–11 152, 2000, ISSN: 0027-8424.

[27] R. De Castro and J. W. Grossman, "Famous trails to Paul Erdős," *Math. Intell*, vol. 21, no. 5163, p. 17, 1999.

[28] J. W. Grossman and P. D. F. Ion, "On a portion of the well-known collaboration graph," *Congr. Numer.*, pp. 129–132, 1995, ISSN: 0384-9864.

[29] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, no. 4-5, pp. 175–308, 2006. DOI: 10.1016/j.physrep.2005.10.009.

[30] M. Mitzenmacher, "A Brief History of Generative Models for Power Law and Lognormal Distributions," *Internet Math.*, vol. 1, no. 2, 2004. DOI: 10.1080/15427951.2004.10129088.

[31] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999. DOI: 10.1126/science.286.5439.509.

[32] L. Zhang, M. Small, and K. Judd, "Exactly scale-free scale-free networks," *Phys. A Stat. Mech. its Appl.*, vol. 433, pp. 182–197, 2015. DOI: https://doi.org/10.1016/j.physa.2015.03.074.

[33] G. Bianconi and A.-L. Barabási, "Competition and multiscaling in evolving networks," *Europhys. Lett.*, vol. 54, no. 4, pp. 436–442, 2001. DOI: 10.1209/epl/i2001-00260-6.

[34] G. Bianconi and A.-L. Barabási, "Bose-Einstein condensation in complex networks," *Phys. Rev. Lett.*, vol. 86, no. 24, pp. 5632–5635, 2001. DOI: 10.1103/PhysRevLett.86.5632.

[35] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, "The degree sequence of a scale-free random graph process," *Random Struct. Algorithms*, vol. 18, no. 3, pp. 279–290, May 2001. DOI: 10.1002/rsa.1009.

[36] B. Bollobás and O. Riordan, "The diameter of a scale-free random graph," *Combinatorica*, vol. 24, no. 1, pp. 5–34, 2004. DOI: 10.1007/s00493-004-0002-2.
[37] R. van der Hofstad, *Random Graphs and Complex Networks* (Cambridge Series in Statistical and Probabilistic Mathematics), 1st. Cambridge: Cambridge University Press, 2017, vol. 1, ISBN: 9781316779422. DOI: 10.1017/9781316779422.
[38] S. Dereich and M. Ortgiese, "Robust analysis of preferential attachment models with fitness," *Comb. Probab. Comput.*, vol. 23, no. 3, pp. 386–411, 2014. DOI: 10.1017/S0963548314000157.
[39] B. Lodewijks and M. Ortgiese, "A phase transition for preferential attachment models with additive fitness," *Electron. J. Probab.*, vol. 25, no. none, pp. 1–54, Jan. 2020. DOI: 10.1214/20-EJP550.
[40] S. Bhamidi, "Universal techniques to analyze preferential attachment trees : Global and Local analysis," 2007. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.120.5134.
[41] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks," *Adv. Phys.*, vol. 51, no. 4, pp. 1079–1187, Jun. 2002. DOI: 10.1080/00018730110112519.
[42] P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of Growing Random Networks," *Phys. Rev. Lett.*, vol. 85, no. 21, pp. 4629–4632, Nov. 2000. DOI: 10.1103/PhysRevLett.85.4629.
[43] P. L. Krapivsky and S. Redner, "Organization of growing random networks," *Phys. Rev. E*, vol. 63, no. 6, p. 066 123, May 2001. DOI: 10.1103/PhysRevE.63.066123.
[44] P. L. Krapivsky and D. Krioukov, "Scale-free networks as preasymptotic regimes of superlinear preferential attachment," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 78, no. 2, pp. 1–11, 2008. DOI: 10.1103/PhysRevE.78.026114.
[45] S. Dereich and P. Mörters, "Random networks with sublinear preferential attachment: Degree evolutions," *Electron. J. Probab.*, vol. 14, no. none, pp. 1222–1267, Jan. 2009. DOI: 10.1214/EJP.v14-647.
[46] P. Mörters and S. Dereich, "Random networks with sublinear preferentia attachment: The giant component," *Ann. Probab.*, vol. 41, no. 1, pp. 329–384, 2013. DOI: 10.1214/11-AOP697.
[47] D. Sénizergues, "Geometry of weighted recursive and affine preferential attachment trees," *Electron. J. Probab.*, vol. 26, pp. 1–56, 2021. DOI: 10.1214/21-EJP640.
[48] A.-L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Phys. A Stat. Mech. its Appl.*, vol. 272, no. 1-2, pp. 173–187, Oct. 1999. DOI: 10.1016/S0378-4371(99)00291-5.
[49] L. Lu and F. R. K. Chung, *Complex graphs and networks* ({CBMS} regional conference series in mathematics no. 107). Providence, RI: American Mathematical Society, 2006, pp. 55–74, ISBN: 978-0-8218-3657-6.
[50] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch, "First to market is not everything," in *Proc. thirty-ninth Annu. ACM Symp. Theory Comput. - STOC '07*, New York, New York, USA: ACM Press, 2007, p. 135, ISBN: 9781595936318. DOI: 10.1145/1250790.1250812.
[51] S. Janson, "Functional limit theorems for multitype branching processes and generalized Pólya urns," *Stoch. Process. their Appl.*, vol. 110, no. 2, pp. 177–245, 2004. DOI: 10.1016/j.spa.2003.12.002.
[52] H. Robbins and S. Monro, "A Stochastic Approximation Method," *Ann. Math. Stat.*, vol. 22, no. 3, pp. 400–407, Sep. 1951. DOI: 10.1214/aoms/1177729586.
[53] R. Pemantle, "A survey of random processes with reinforcement," *Probab. Surv.*, vol. 4, no. 1, pp. 1–79, 2007. DOI: 10.1214/07-PS094.
[54] M. Benaïm, "Dynamics of stochastic approximation algorithms," vol. 33, pp. 1–68, 1999. DOI: 10.1007/bfb0096509.
[55] P. Jagers and O. Nerman, "The asymptotic composition of supercritical, multi-type branching populations," pp. 40–54, 1996. DOI: 10.1007/bfb0094640.
[56] D. Aldous, "Asymptotic Fringe Distributions for General Families of Random Trees," *Ann. Appl. Probab.*, vol. 1, no. 2, pp. 228–266, Sep. 1991, ISSN: 10505164. [Online]. Available: http://www.jstor.org.tudelft.idm.oclc.org/stable/2959767.
[57] M. Balázs and J. Komjáthy, "Order of Current Variance and Diffusivity in the Rate One Totally Asymmetric Zero Range Process," *J. Stat. Phys.*, vol. 133, no. 1, pp. 59–78, Oct. 2008. DOI: 10.1007/s10955-008-9604-1.
[58] J. Komjáthy and B. Lodewijks, "Explosion in weighted hyperbolic random graphs and geometric inhomogeneous random graphs," *Stoch. Process. their Appl.*, vol. 130, no. 3, pp. 1309–1367, Mar. 2020. DOI: 10.1016/j.spa.2019.04.014.
[59] E. Jacob and P. Mörters, "Spatial preferential attachment networks: Power laws and clustering coefficients," *Ann. Appl. Probab.*, vol. 25, no. 2, pp. 632–662, 2015. DOI: 10.1214/14-AAP1006.
[60] C. Mitra, J. Kurths, and R. V. Donner, "Rewiring hierarchical scale-free networks: Influence on synchronizability and topology," *Epl*, vol. 119, no. 3, pp. 1–7, 2017. DOI: 10.1209/0295-5075/119/30002.
[61] A. Cipriani and A. Fontanari, "Dynamical fitness models: evidence of universality classes for preferential attachment graphs," *J. Appl. Probab.*, pp. 1–22, Jun. 2022. DOI: 10.1017/jpr.2021.81.
[62] T. Pham, P. Sheridan, and H. Shimodaira, "PAFit: A statistical method for measuring preferential attachment in temporal complex networks," *PLoS One*, vol. 10, no. 9, pp. 1–18, 2015. DOI: 10.1371/journal.pone.0137796.
[63] T. Pham, P. Sheridan, and H. Shimodaira, "Joint estimation of preferential attachment and node fitness in growing complex networks," *Sci. Rep.*, vol. 6, no. August, pp. 1–13, 2016. DOI: 10.1038/srep32558.
[64] P. H. T. Schimit and F. H. Pereira, "Disease spreading in complex networks: A numerical study with Principal Component Analysis.," eng, *Expert Syst. Appl.*, vol. 97, no. January, pp. 41–50, May 2018. DOI: 10.1016/j.eswa.2017.12.021.
[65] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-Law Distributions in Empirical Data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Nov. 2009. DOI: 10.1137/070710111.
[66] J. A. Rice, *Mathematical Statistics and Data Analysis* (Advanced series), 3rd ed. Brooks/Cole, Cengage Learning, 2007.

B. Probabilistic results

In this appendix we will state some general probabilistic result that will be used without proof.

B.1 Martingales

In this thesis we limit the use of martingales to the discrete case in which $T = \mathbb{Z}_{\geqslant 1}$. We state the following general results for a fixed measurable space $(\Omega, \mathcal{F})$. In this thesis we do not explicitly state the filtration $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$, as we refer to the *natural filtration*, where $\mathcal{F}_t = \sigma(X_0, \dots, X_t)$. These results and definitions have been taken verbatim from Kyle Siegrist, CC BY-SA 2.0, Random Services.

Definition B.1 (Martingale)**.** The process $\mathbf{X} = \{X_t : t \in T\}$ is a martingale with respect to $\mathfrak{F}$ if $\mathbb{E}[\, X_t \mid \mathcal{F}_s \,] = X_s$ for all $s, t \in T$ with $s \leqslant T$. «

Definition B.2 (Doob's martingale)**.** Suppose that $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ is a filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that X is a real-valued random variable with $\mathbb{E}[\, |X| \,] < \infty$. Define $X_t = \mathbb{E}[\, X \mid \mathcal{F}_t \,]$ for $t \in T$. Then $\mathbf{X} = \{X_t : t \in T\}$ is a martingale with respect to $\mathfrak{F}$. «

Theorem B.1 (Doob's martingale convergence theorem)**.** *Suppose that $\mathbf{X} = \{X_t : t \in T\}$ is a martingale with respect to $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ and $\mathbb{E}[\, |X_t| \,]$ is bounded in $t \in T$. Then there exists a random variable X_∞ that is measurable with respect to $\mathcal{F}_\infty = \bigcup_{t \in T} \mathcal{F}_t$, such that $\mathbb{E}[\, |X_\infty| \,] < \infty$ and $X_t \overset{a.s.}{\to} X_\infty$ as $t \to \infty$.*

Definition B.3 (Uniform Integrability)**.** The collection $\mathbf{X} = \{X_i : i \in I\}$ is *uniformly integrable* if for each $\epsilon \geqslant 0$ there exists $x \geqslant 0$ such that

$$\sup_{i \in I} \mathbb{E}[\, |X_i| \mathbb{1}_{\{|X_i| \geqslant x\}} \,] \leqslant \epsilon. \quad «$$

The concept of uniform integrability is useful, because by the following theorem, it is the condition under which convergence in probability implies convergence in mean.

Theorem B.2 (Uniform Integrability)**.** *If $\{X_n : n \in \mathbb{Z}_{\geqslant 1}\}$ is uniformly integrable and $X_n \to X$ as $n \to \infty$ in probability, then $X_n \to X$ as $n \to \infty$ in mean.*

B.2 Azuma–Hoeffding inequality

Theorem B.3 (Azuma–Hoeffding inequality [37, pp. 75, Theorem 2.27])**.** *Let $(M_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ be a martingale process with the property that there exist constants $K_n \geqslant 0$ such that almost surely*

$$|M_n - M_{n-1}| \leqslant K_n, \qquad \textit{for all } n \geqslant 1, \tag{B.2.1}$$

where, by convention we define $M_0 = \mu = \mathbb{E}[\, M_n \,]$. Then, for every $a \geqslant 0$ and $n \geqslant 1$,

$$\mathbb{P}(|M_n - \mu| \geqslant a) \leqslant 2 \exp\left\{ -\frac{a^2}{2 \sum_{i=1}^{\infty} K_i^2} \right\}. \tag{B.2.2}$$

B.3 Poissonization

Theorem B.4 (Poisson limit theorem ([37, pp. 62, Theorem 2.10])). *Let* $(I_i)_{i=1}^n$ *be independent with* $I_i \sim Be(p_i)$*, and let* $\lambda = \sum_{i=1}^n p_i$*. Let* $X = \sum_{i=1}^n I_i$*, and let* Y *be a Poisson random variable with parameter* λ*. Then, there exists a coupling* $(\hat{X}, \hat{Y})$ *of* X *and* Y *such that*

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leqslant \sum_{i=1}^{n} p_i^2. \tag{B.3.1}$$

B.4 Big-Θ notation

Notation B.1. In Theorem 3.15, we let $f(x) = \Theta(g(x))$ mean that there exist $L, U \geqslant 0$ and $N \geqslant 0$ such that for all $x > N$, $Lg(x) \leqslant f(x) \leqslant Ug(x)$. «

B.5 Stochastic Approximation

Let $(X_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ be vector valued stochastic process with values in $\mathbb{R}^d$, which satisfies

$$X_{n+1} - X_n = \frac{1}{n+1} F(X_n) + R_{n+1} - R_n, \tag{B.5.1}$$

where F is a suitable vector valued function and R is a typically random error term. For a variant such processes the Dereich and Ortgiese prove Lemma B.5.

Lemma B.5 (Stochastic approximation [38, p. 394, Lemma 3.1]). *Let* $(X_n)_{n \in \mathbb{Z}_{\geqslant 1}}$ *be a non-negative stochastic process. We suppose that the following estimate holds:*

$$X_{n+1} - X_n \leqslant \frac{1}{n+1}(A_n - B_n X_n) + R_{n+1} - R_n, \tag{B.5.2}$$

where

(i) (A_n) *and* (B_n) *are almost surely convergent stochastic processes with deterministic limits* $A, B > 0$,

(ii) (R_n) *is an almost surely convergent stochastic process.*

Then we have that, almost surely,

$$\limsup_{n \to \infty} X_n \leqslant \frac{A}{B}. \tag{B.5.3}$$

Conversely, if under the same conditions (i) *and* (ii)

$$X_{n+1} - X_n \geqslant \frac{1}{n+1}(A_n - B_n X_n) + R_{n+1} - R_n, \tag{B.5.4}$$

Then we have that, almost surely,

$$\liminf_{n \to \infty} X_n \geqslant \frac{A}{B}. \tag{B.5.5}$$

C. Finding the tail exponent

In order to confirm the difference in tail exponent for the graphs simulated in Section 3.4.2, we make use of recommended techniques for the estimation of tail exponent for power-law distributions from Clauset *et al.* [65]. Clauset *et al.* advise not to use least-squares regression on the log-log plot, as this is not a good estimator. Instead they advise to use the maximum-likelihood estimator (MLE) and to verify the hypothesis that the underlying random variable is power-law distributed by using a likelihood-ratio test against alternative distributions. As we have results proving that the tail of the distribution follows a power law, we will not verify this. We will however give a short account of how the estimations in Section 3.4.2 were carried out.

Let X_i be a series of i.i.d. random variables distributed according to a discrete Pareto or Zipf distribution, which for $x, x_{\text{min}} \in \mathbb{Z}_{\geqslant 1}$, $\alpha > 1$, has density

$$f(x; \alpha, x_{\text{min}}) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\text{min}})}. \tag{C.0.1}$$

Here ζ denotes the Hurwitz ζ-function, which for $\alpha > 1$, x_{min} is defined as

$$\zeta(\alpha, x_{\text{min}}) = \sum_{i=0}^{\infty} (n + x_{\text{min}})^{-\alpha}. \tag{C.0.2}$$

From this expression, we can derive the log-likelihood $\mathcal{L}$ for a sample $X_1, \ldots, X_N$

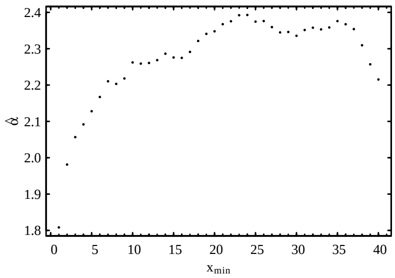
$$\mathcal{L}(\alpha) = n \ln(\zeta(\alpha, x_{\text{min}})) - \alpha \sum_{i=1}^{N} \ln(X_i). \tag{C.0.3}$$

Although it is possible to derive an exact condition for the MLE $\hat{\alpha}$, we do not gain a closed form. Moreover, since this condition contains a derivatives of the Hurwitz ζ-function, and the implementation of the `VGAM R` library does not support derivatives for non integer input, we instead choose to maximize numerically. Doing this for several values of x_{min}, we obtain a so-called Hill plot. Considering that the estimated tail exponent varies wildly, we let the tail start at 15, the value around which the estimator seems to stabilize. For this value we find $\hat{\alpha} = 2.2759$. The MLE is asymptotically efficient, thus $\text{Var}(\hat{\alpha}) \to \frac{1}{nI(\hat{\alpha})}$.[1] Clauset *et al.* [65] derive the variance by assuming that $\mathcal{L}$ is regular enough to say

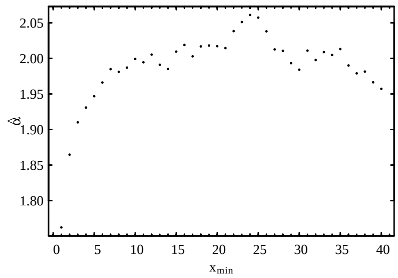
$$I(\alpha) = -\mathbb{E}[\frac{d^2\mathcal{L}(\alpha)}{d\alpha^2}]. \tag{C.0.4}$$

This derivation gives a closed-form expression for the variance, but since we cannot evaluate derivatives of the Hurwitz ζ-function, we use a numerical estimation of the second derivative to estimate the variance and find $\text{Var}(\alpha) = 0.003$. We the then use the quantile function of the normal distribution, and establish a confidence interval (95%) of $\alpha \in (2.168, 2.384)$. Similarly choosing $x_{\text{min}} = 15$, we find $\hat{\alpha} = 2.0095$ and confidence interval $(1.9015, 2.1176)$. We note that the length of the two confidence intervals is equal, but that this is coincidence. From this, we can make the plots in Section 3.4.2 and we conclude that the power-law exponents are not equal.

[1] J. A. Rice, *Mathematical Statistics and Data Analysis* (Advanced series), 3rd ed. Brooks/Cole, Cengage Learning, 2007.



(a) Hill graph generated from the additive preferential attachment model with fitness values distributed according to $\mathcal{F}^{(i)} \sim \exp(\frac{10}{3})$, with $m = 1$ and $n = 60\,000$, such that $\mathbb{E}[\mathcal{F}] = .3$, and $\gamma = 2.3$



(b) Hill graph generated from the additive preferential attachment model with fitness values distributed according to $\mathcal{F}^{(i)} \sim \text{Par}(\frac{3}{110}, 1.1)$, with $m = 1$ and $n = 60\,000$, such that $\mathbb{E}[\mathcal{F}] = .3$, and $\gamma = 2.1$

Figure C.1: Estimated tail index $\hat{\alpha}$ for several values of x_{min}. Obtained by numerically maximizing Eq. (C.0.3).