Preferential attachment models for dynamic networks

Preferential attachment models for dynamic networks

Alessandro Garavaglia

This work is supported in part by the Netherlands Organization for Scientific Research (NWO) through the Gravitation Networks grant 024.002.003.

Preferential attachment models for dynamic networks / by Alessandro Garavaglia - Eindhoven University of Technology, 2019

A catalog record is available from Eindhoven University of Technology Library
ISBN: 978-90-386-4689-3

Printed by Gildeprint

Preferential attachment models for dynamic networks

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op dinsdag 29 januari 2019 om 16.00 uur

door

Alessandro Garavaglia

geboren te Rho, Italië

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. M.T. de Berg
1e promotor:	prof.dr. R.W. van der Hofstad
2e promotor:	prof.dr. N.V. Litvak
leden:	prof.dr. F. Caravenna (Universitá degli Studi di Milano-Bicocca)
	prof.dr. L.A. Leydesdorff (Universiteit van Amsterdam)
	prof.dr. P. Mörters (Universität zu Köln)
	prof.dr. F.C.R. Spieksma
	prof.dr.ing. G.J. Woeginger (RWTS Aachen)
adviseur(s):	dr. J. Komjáthy

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Acknowledgments

This thesis is the product of more than four years of hard work, and it would not have been possible without the help and the support of many people that I wish to thank. I will try to mention them all in the following few paragraphs.

First of all, I would like to thank my supervisors Remco and Nelly, for giving me the opportunity to pursue a PhD. Without you, I would have not been able to do any of this work, since sometimes along these years you believed in me more than myself. The fact that you both think I am a good researcher means a lot to me.

Remco, the first time you offered me a PhD position I turned it down. Luckily for me, you were patient enough to wait for me to change my mind (or maybe you already knew I would have done that!). I have to say, I felt a little ashamed, but you always told me that in this way you were sure I would be fully committed to this research project. I am grateful you wanted me as a PhD, and I am very happy you are my supervisor, not only because working with you has been a pleasure, but because you were always there to talk with me about myself and my future. And let's be honest, going to watch together a PSV match or meeting on the Alps without planning it is kind of cool.

Nelly, I don't think I ever told you, but when I had just started my PhD, you were giving a talk about PageRank in random graphs at Eurandom. Listening to you speaking, I thought that it was very cool math, but too hard for me. I never though that after two years I would write you asking to work on the same topic. You were happy to have a new mathematical challenge, so I came to visit you in Twente. That is how you became my second supervisor. I am glad you accepted to work with me, because of your enthusiasm and joy in what you do. I want to thank you also for all the time we spent talking about almost everything came up to our minds.

I want to express my gratitude to all the members of the doctoral committee: Mark de Berg, Francesco Caravenna, Julia Komjáthy, Loet Leydesdorff, Peter Mörters, Friet Spieksma and Gerhard Woeginger. I thank all of you for the time spent reading my

thesis and for the useful comments I received from you.

Gerhard, thank you for being my supervisor for the first two years of my PhD, before you moved to Aachen for a new adventure. Even though I mainly worked in Probability, you always gave me a new point of view and many suggestions that helped me writing my first paper.

Francesco, you are a great lecturer, and the course you gave when I was a student in Milan made me decide that Probability was the field I wanted to work in. When I asked you to be my master supervisor, you were happy to give me a project that would have allowed me to travel abroad for some time. In the end (after some additional years of work) we wrote together with Remco a beautiful paper. I want to thank you the support you gave me.

When I arrived at TU/e as a master student, I did not know what to expect, but the beautiful atmosphere and the nice people from the STO group made me feel at home. I would like to thank Alessandro, Britt, Enrico, Fabio, Gianmarco, Jaron, Jori, Maria Luisa, Mirko, Serban and Thomas, who were already PhDs when I first got in Eindhoven and made me discover the university (i.e., the borrel at Gewis) and the city (i.e., the Bierprofessor). Thanks Alessandro for sharing the office with me in my first months at TU/e, and for the "Junior" nickname, that sticked even though you were not there anymore.

I also want to thank all the PhDs I met afterwards for having being part of my experience as a STO member. In particular, thanks to Fiona for trusting me as a dogsitter, thinking of me every time she makes cake (the carrot one in particular is super nice), and the help she gave me when I moved to different house. I would also like to thank Clara for all the conferences we went to, specially when we went to Austin and spent some days visiting Texas, and in the end of our PhD for writing a paper together, sending to each other job positions that might be interesting and supporting each other through bureaucratic deadlines.

I shared my office with Bart, Enrico, Lorenzo and Murtaza. If I would look at working performance, the five of us together did not make the best group of people. We got carried away in totally random discussions much too easily and much too often (if I am not mistaken, once we spent two hours making a ranking of the worst dictators in history, comparing data from wikipedia pages). Our office has been recognized as the most loud and messy of the whole floor. I am glad to have shared my office and a lot of time with you. I also want to thank Liron, who is only part-time in the office. I would like to thank Dennis and Ivo, who got in the office recently and they only met me when I was stressed out for the defense. At least, we had time to get some beers together in these few months.

I want to thank the people who worked with me in the PhD/PdEng Council: Britt, Christine, Daan, Davide, Davide, Gianmarco, Jeroen, Lorenz, Murtaza, Youri, Pranav, Rien, Sergio, Stefan. It has been a wonderful experience, not only because of all the fun at meetings, the different activities and the "free" lunches, but also because I had the chance of being both The Godfather and Santa.

Being abroad also means that you make new friends. I was lucky to find people at TU/e who like me needed a beer (or an alcohol-free drink, like colacao) and

foosball (or beer pong against students, even better) with friends on Friday at 17:00, and, occasionally, on other days at other times. Summer schedule was in fact hard, but someone had do it. Thanks Alessandro, Andy, Bouke, Britt, Christine, Daan, Davide, Gustavo, Laura, Lorenz, Lorenzo, Mahdi, Meilof, Murtaza, Omer, Rien, Stefan, Thomas.

As it is possible to grasp from the names above, I met a non-negligible number of Italians during my years at TU/e. I would like to thank all of them, because I missed Italy less: Alberto, Alessandro, Alessandro, Andrea, Carlo, Davide, Enrico, Fabio, Gianmarco, Giovanni, Laura, Lorenzo, Maria Luisa, Marta, Matteo, Melania and Mirko. In this list, I want to give a special thanks to my first two housemates ever. It was scary moving abroad, but I could not have been more lucky since I had the chance to live with Maria Luisa and Melania. Mixing Roma, Campobasso and Milano was quite the adventure, but you two made me feel like having a second family in Eindhoven.

I made a lot of friends in the Netherlands, but I also have to thank all the friends I left in Italy, because they were not scared by distance and kept being part of my life. I want to thank the members of the Fagiani group, because we will make the rupe tarpea great again: Alfo, Beppe, Chiara, Cino, Ciona, Conti, Dario, Dem, Francesca, Frech, Irene, Lando, Marta, Michael, Niggia, Noemi, Penny, Vane. I want to thank Lety and Chiara for all the "hearts nights" we had together. A big thanks goes to Dacco, my best and oldest friend, for still being around after almost 28 years of friendship.

If I was even able to start a PhD, I owe it to my father Massimo, my mother Carla and my brother Francesco for everything they did to get me here. I will never be able to repay them for all the sacrifices they made to let me free to focus on my studies and to pursue my goals. They always wanted the best for me, and pushed me further, even when I wanted to stop.

At last, I want to thank Chiara. There is no need for too many words. Thank you for every day, every hour, every minute, and for being the love of my life.

May the Force be with you,
for having read the acknowledgments this far.

Amsterdam
December 15th, 2018

Contents

1 Introduction

1.1. Complex networks

The theory of *complex networks* has gained a lot of interest in the last decades due to the accessibility of data and the growth of computational power. This led to an increase in the scientific research on random graphs and their properties.

There are many systems in the real world that can be seen as large networks, since we can see them as a collection of objects linked to each other with certain connections. For instance, the World Wide Web, Internet, railways, electric systems, social networks, interactions of genes and proteins, are all real-world systems that can be interpreted as large networks.

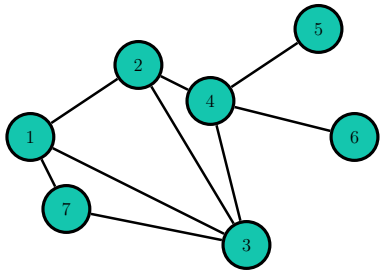


Figure 1.1: Example of a small undirected graph with 7 vertices and 9 edges.

Mathematically, a network is modeled as a *graph* G, that is a pair consisting of a set of *vertices* $V(G)$ and a set of *edges* $E(G)$. Vertices represent the objects that compose the network (for instance web pages) and edges represent the connections between vertices (for example in the world-wide web case, a hyperlink from a page to another). The graph G is *directed* when the connection is oriented (a hyperlink leads *from* a webpage *to* another), otherwise is *undirected* (a cable in the Internet networks that connects two servers has no direction).

Seeing vertices as parts of a large network allows to consider individual charac-

teristics, such as the number of connections of every vertex, as well as more global properties, for example the maximum distance in the network. Global properties are relevant since they can affect the performance of the networks itself, for instance, the flow or spread of information in the WWW, the presence of traffic in a highway system of a country or the robustness of an electric network in the case of the failure of one of its parts.

Interestingly, networks that represent very different systems share some properties, independently of the nature of the networks (social, biological, electrical, etc.). For more details of real-world networks and their common properties, we refer to [17, Chapter 1], [130, Chapters 1-5], [85, Chapter 1] and the references therein.

1.1.1. Power-law degree distribution

The degree of a vertex $v \in [n]$ is defined as the number of edges incident to v. In the undirected case this is well defined. In the directed setting, we can distinguish between the *in-degree*, given by the incident edges pointing *towards* v, and the *out-degree*, given by the incident edges pointing *away* from v.

It has been observed ([1, 17, 41, 42, 65, 155, 161], [130] and the references therein)

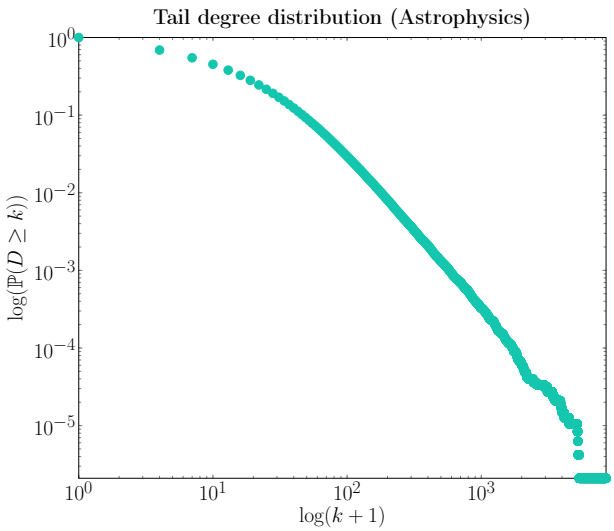


Figure 1.2: In-degree tail distribution of citations in Astrophysics citation network. The network is made by 477113 Astrophysics papers published between 1980 and 2015, and 9154818 edges. A directed edge from paper v to paper w represents a reference to w contained in paper v. The plot represents the tail of the distribution of the number of citations received by a uniformly chosen paper in the dataset in $\log\log$ form. Data from *Web Of Science*.

Network	Type	Vertices	Edges	Power-law τ
Film actors	Undirected	449913	25516482	2.3
Email messages	Directed	59812	86300	1.5/2
Telephone call graph	Undirected	47000000	80000000	2.1
WWW Alta Vista	Directed	203549046	1 466 000000	2.1/2.7
Word co-occurrence	Undirected	460902	16100000	2.7
Protein interactions	Undirected	2115	2240	2.4

TABLE 1.1: Examples of networks with type (directed/undirected), number of vertices, number of edges and estimated power-law exponent of the degree distribution (in-degree in case of directed networks). Data from [130, Table 8.1].

that many real-world networks show a *power-law degree distribution*. Such networks are called *scale free*. In a scale-free networks, the fraction of vertices with degree k decreases as $k^{-\tau}$, where τ is called the *power-law exponent*. In a loglog plot as in Figure 1.2, the scale-freeness is revealed by straight lines. In fact, denoting by n_k the number of vertices with degree k and by n the total number of vertices, then

$$\frac{n_k}{n} \approx Ck^{-\tau} \qquad \Longleftrightarrow \qquad \log(n_k/n) \approx -\tau \log k + \log C,$$

which is a line in the argument $\log k$. As a consequence, the straight line in Figure 1.2 suggests that the considered network shows a power-law degree distribution. In the undirected setting we consider the total degree, while in the directed setting in general we observe in-degree power laws.

In Table 1.1 we report information of examples of networks taken from [130]. Table 1.1 contains the estimated power-law exponent for the degree distribution. It has been observed that many real-world networks show a power-law exponent $\tau \in (2, 3)$. A power-law distribution with $\tau \in (2, 3)$ has finite mean but infinite variance. This implies that degrees in these networks show a large variability, thus allowing the existence of *hubs*, i.e., vertices with extremely high degree. These vertices play a relevant role in the network as, for example, they short distances.

1.1.2. DISTANCES

In a graph $G = (V, E)$, the distance between two vertices $v, u \in V$ is defined as the length of the shortest path made of edges connecting u and v. It has been observed [65, 85, 130, 161, 170] that in many real-world networks distances are quite short, even for very large graphs. This has been called the *small-world phenomenon*.

A way to mathematically describe this is to investigate the *typical distance* and the *diameter* of a graph. The typical distance is the distance between two vertices chosen uniformly at random in the set V. The diameter is the maximum of the distances between all pairs of vertices in the graph. If the network is not connected, then these

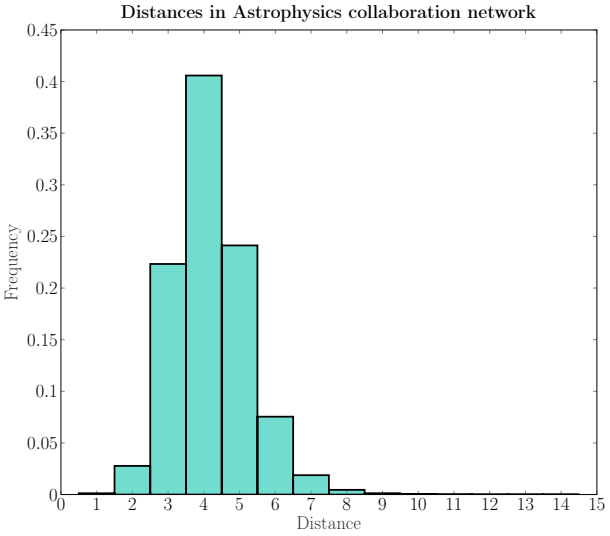


FIGURE 1.3: Distances in arXiv Astrophysics collaboration network. In this network, vertices represent authors. An edge is present between two authors if and only if they are co-authors of an Astrophysics paper that has appeared on arXiv. The network is composed of 18772 authors and 198110 co-authorships. The plot represents the frequency of values of distances. For instance, around 40% of pairs of vertices are at distance 4. The diameter of the graph is 14. Data from SNAP (Stanford University) [111].

quantities are infinite, in the sense that there are pairs of vertices with no paths connecting them. In this setting, we can look at distances within connected components.

Small-world networks have the advantage that, for example, one can send a message from one vertex to another using only a few steps. A famous example of this is the *six degrees of separation experiment* by Milgram [122, 159]. The aim of this experiment was to prove that two random people in the U.S.A. know each other through at most 6 intermediate friends. In the network framework, the vertices represent people in the U.S.A., and edges represent friendships. The six degrees of separations are interpreted as the fact that the typical distance between two vertices is 7, over a network of size the entire U.S. population.

1.1.3. Subgraphs and clustering

In many real-world networks, vertices tend to be connected to each other forming triangles and other highly-connected subgraphs. Several subgraphs were found to appear more frequently than other subgraphs [123]. Which type of subgraph appears most frequently varies for different networks, and the most frequently occurring subgraphs are believed to be correlated with the function of the network [123, 124, 174].

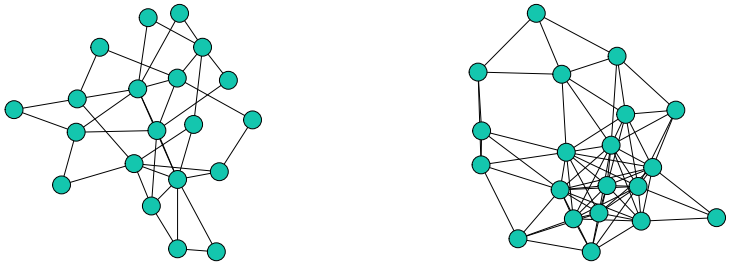


FIGURE 1.4: Two examples of graphs with 25 vertices. The graph on the left is less "clustered" than the graph on the right, since the graph on the right has more small subgraphs, such as triangles or cliques. Here by a clique we mean a subgraph of where each vertex is connected to all other vertices in the subgraph. The graph on the left is a sample of the Barabási-Albert model, while on the right is a sample of a random geometric graph in the unit square.

This property is called *clustering*. From a social networks perspective, we can say that *two people are more likely to be friends if they have a common friend*. In other words, two vertices are more likely to be connected if they are both connected to a third common vertex.

This property is measured by the *global clustering coefficient* $\mathcal{C}$, that can be defined as

$$\mathcal{C} = \frac{3 \times (\text{number of triangles})}{\text{number of connected triplets}}, \tag{1.1.1}$$

where by a connected triplet we mean three vertices connected by two edges. First introduced in networks science by [18], it has been observed in several works ([130],[131]

Network	**Vertices**	**Edges**	$\mathcal{C}$	$\mathcal{C}'$
Company directors	7673	55392	0.59	0.88
Internet	10697	31992	0.035	0.39
Power grid	4941	6594	0.1	0.08
Metabolic network	765	3686	0.09	0.67

TABLE 1.2: Examples of networks with number of vertices, number of edges, global clustering coefficient $\mathcal{C}$ and average local clustering coefficient $\mathcal{C}'$. Data from [130, Table 8.1].

and the references therein) that the clustering coefficient of real-world networks is higher than what would happen by rewiring edges uniformly at random, keeping the degrees of vertices fixed (we refer to [130, Section 7.9] for a more complete discussion).

Watts and Strogatz [171] propose a different measurement of clustering, based on the so-called *local clustering coefficient*. For a vertex i in the network, the local clustering coefficient $\mathcal{C}_i$ is defined as

$$\mathcal{C}_i = \frac{\text{number pairs of neighbors of } i \text{ that are connected}}{\text{number of pairs of neighbors of } i}. \tag{1.1.2}$$

Then, the clustering coefficient $\mathcal{C}'$ of the network is given by the average over all vertices of the terms in (1.1.2), that is

$$\mathcal{C}' = \frac{1}{n}\sum_{i=1}^{n} \mathcal{C}_i = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{d_i(d_i-1)}, \tag{1.1.3}$$

where d_i is the degree of vertex i and Δ_i is the number of triangles that include i. Notice that (1.1.3) differs from (1.1.1). In [145, 146, 160] it is empirically observed that $\mathcal{C}_i$ scales as an inverse power of the degree d_i of vertex i.

In contrast to clustered networks, there is a class of graphs that is called *tree-like networks*, where triangles are rare. More precisely, the neighborhoods of vertices are structured as trees, so short cycles are rare in the graph.

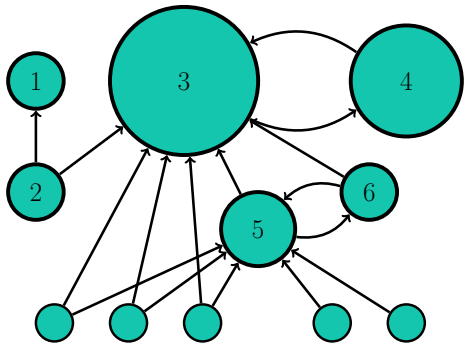


FIGURE 1.5: Explanation of PageRank scores. In this figure, the sizes of the vertices represent the PageRank score, i.e., the bigger a vertex is, the higher its PageRank score. The picture explains why PageRank gives more insight about the graph than just the in-degree. The score of vertex 3 is high since it receives a lot of edges. The score of vertex 4 is high even though it receives only one edge. This happens because the edge received by vertex 4 comes from vertex 3, that has a high score itself. Picture from *Wikipedia*.

1.1.4. Power-law PageRank distribution

PageRank, first introduced in [134], is an algorithm that generates a ranking on the vertices in finite directed graphs. The algorithm is also defined on undirected graphs, but in this case is less considered in applications. Originally introduced to rank World-Wide Web pages, PageRank has a wide range of applications including spam detection [79], citation analysis [48, 117, 165], community detection [8] or social networks analysis [15, 169]. PageRank is one of simplest, yet highly informative, networks algorithms.

PageRank is an algorithm that produces a ranking of the vertices of a graph, i.e., to every vertex PageRank associates a value that reflects the "relevance" of the vertex in the graph. PageRank is not the only ranking algorithm on graphs, in fact is part of a broader class of ranking procedures called *centrality measures*. As the name suggests, a centrality measure is a measure of the relevance or importance of vertices in a graph. For instance, (in-)degree can be seen as centrality measure, so vertices with high degree are more relevant in the network. What is interesting about PageRank is that the *score of a vertex i depends on the scores of vertices pointing towards i*. In particular, PageRank takes into account not only the *number* of edges that a vertex receives, but also the *quality*. A simple example with a heuristic explanation of PageRank is given in Figure 1.5.

More precisely, given a graph G of size n, PageRank is a deterministic vector $\boldsymbol{\pi}_n$

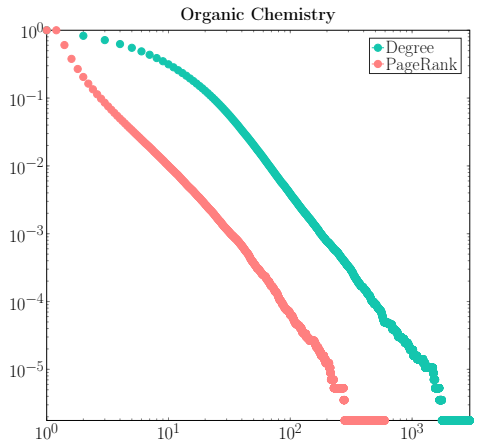


Figure 1.6: In-degree and graph-normalized PageRank in the Organic Chemistry citation network. The network is made by 567146 Organic Chemistry papers published between 1980 and 2015, and 5072139 references. We can observe that both in-degree and the graph-normalized PageRank follow a power-law distribution, with remarkably similar power-law exponent. Data from *Web Of Science*.

of length n, where π_v is the rank of vertex v. The PageRank vector $\boldsymbol{\pi}_n$ is defined as the invariant measure of a particular random walk on the graph, therefore $\boldsymbol{\pi}_n$ is a probability distribution on the vertices of the graph. Sometimes it is useful to consider the *graph-normalized version* of PageRank, that is defined as $\boldsymbol{R}(n) = n\boldsymbol{\pi}_n$. Notice that we call PageRank both the vector $\boldsymbol{R}(n)$ and the *algorithm* that produces it.

It has been observed [44, 112, 135, 163] that in real-world networks with power-law (in-) degree distributions, PageRank follows a power-law with the same exponent. In other words, the fraction of vertices r_k whose graph-normalized PageRank score exceeds k decreases as $k^{-\tau}$, where τ is the same power-law exponent of the (in-)degree distribution. Figure 1.6 illustrates this phenomenon in a citation network. The two lines, that are the PageRank tail distribution and the in-degree tail distribution in a loglog format, are parallel, indicating that both distributions obey a power law with the same exponent τ.

This led to the formulation of *the PageRank power-law hypothesis* that states that *in any graph with an (in-)degree distribution that obeys a power law with exponent τ, the PageRank distribution also obeys a power law with the same exponent τ.*

1.2. Random graphs

A mathematical way to describe real-world networks, taking into account the inhomogeneity present in the data, is through *random graphs*. Random graphs are graph models where the edges and/or the number of vertices are random. Over the years, different models of random graphs have been proposed. We now give a brief introduction to the most widely studied models.

1.2.1. Static models

Erdős-Rényi model. The first random graph model that has been investigated in the litererature is the Erdős-Rényi random graph, that wan introduced by Erdős and Rényi [63], and independently by Gilbert [76]. In the Erdős-Rényi model, we fix the size of the graph $n \in \mathbb{N}$, and a number $p \in (0, 1)$. Then, each edge (u, v) is added to the graph independently of every other edge with probability p. In general, the graph is considered undirected, but it is possible to define a directed version of this model, where (u, v) and (v, u) are two distinct directed edges.

The Erdős-Rényi model has been studied intensively. We refer to [6, 29, 100] and the references therein for early works on this model. A classical problem in the Erdős-Rényi model is the evaluation of the size of the largest connected component, i.e., the number of vertices in the largest set of connected vertices. A lot work has been done in this direction, identifying a phase transition in the behavior of the size of the largest connected component [36, 115].

In the Erdős-Rényi model all vertices are "identical", in the sense that the probability of two vertices being connected is the same for all pairs of vertices, and as a result of this there is no high inhomogeneity among vertices. This is in contrast to real-word networks. In fact, in scale-free networks, we observe many vertices of low degree connected to vertices of extremely high degree, sometimes called *hubs*.

Inhomogeneous random graphs. To incorporate inhomogeneity, the probability of an edge being present can differ from edge to edge. For example, in the Chung-Lu model [49, 50, 51] (and its variations), a *weight* w_i is assigned to every vertex i. Every edge is added independently of any other and of the weights, where the probabilty of an edge (i, j) being present is close to a linear function of the product of the weights w_i, w_j. These models are called Chung-Lu models or *inhomogeneous random graphs* (see also [26]). It is possible to obtain, for instance, power-law degree distributions by sampling the weights from a power-law distribution.

Stochastic block model. Another variation of the the Erdős-Rényi model is the *stochastic block model*, that can be defined as an inhomogeneous random graph [30]. A stochastic block model of size $n \in \mathbb{N}$ is a graph where we can fix a number $k \leq n$ of *groups*. Every vertex is assigned to one in the k groups, and the probability that an edge is present between two vertices depends *on the groups* of the two vertices. Typically, the probability that two vertices are connected is high between two vertices of the same group and low between vertices of different groups. In this way *communities* can be modeled. Here by a community we can in a simple way think of a group of vertices that is highly connected with few edges towards the outside of the group.

Configuration model. The configuration model (CM) was originally introduced by Bollobás [27] to describe random regular graphs, and it was extended by Molloy and Reed [125, 126] to general degrees. By CM now we refer to a random graph model with a prescribed degree sequence. In other words, CM is a graph where we fix the degree sequence. In particular, we can obtain a scale-free graph by *imposing* a power-law degree distribution. The graph is constructed as follows: to every vertex $i \in [n]$ are attached a number of *half-edges* equal to its degree d_i, therefore we have in total $\ell_n = d_1 + \cdots + d_n$ half-edges. We can take a half-edge, and we pair it *uniformly at random* with a different unpaired half-edge, thus creating an edge. Next, we take another unpaired half-edge and we pair it uniformly at random with an unpaired one. We continue until all half-edges are paired. The ordering we use to pick an unpaired half-edge is irrelevant since the pairing is uniform. Notice that in this case, self-loops (an edge whose endpoints are the same vertex) and multi-edges (more than one edge between the same pair of vertices) can arise. This model can be extended to the directed setting, by imposing both the in- and out-degree sequences, and then pairing an in-half-edge to an out-half-edge, thus creating a directed edge.

1.2.2. Preferential attachment models

All the previous random graphs model are *static*, in the sense that the graph is constructed given a size $n \in \mathbb{N}$. Other properties, such as the power-law degree distribution, can be obtained, for instance in inhomogeneous random graphs and configuration models, by choosing the inputs or the parameters of the model appropriately.

An attempt to explain power laws through simple dynamics was given by the

class of *preferential attachment models* (PAMs):

"Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected. A model based on these two ingredients reproduces the observed stationary scale-free distributions, which indicates that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems."

Barabási and Albert, [2, Abstract].

In their original works [2, 3], Barabási and Albert propose a mechanisms to generate a *growing* random graph model, in the sense that the graph grows over time as new vertices appear. This graph model shows properties of real-world networks *without imposing them*, but instead these properties arise naturally from the simple dynamics that defines the model.

More precisely, a PAM is a sequence of graphs $(\mathrm{PA}_t)_{t\geq 1}$, defined recursively. In this setting, the size of the graph is typically denoted by t and not n, to emphasize that the graph grows in *time*. At time $t = 1$, PA_t is a certain initial graph, often a graph consisting of a single vertex with no edge or a self-loop. Then, recursively for $t \geq 2$, a new vertex t appears. This new vertex brings one (or more) edges, that has to be connected to one of the existing vertices v in the graph. The vertex v is chosen according to the probability

$$\mathbb{P}\left(t \to v \mid \mathrm{PA}_{t-1}\right) = \frac{f(v, t-1)}{Z(t-1)}. \tag{1.2.1}$$

In (1.2.1), f is called the *PA function*, which depends on v and $t-1$, while $Z(t-1)$ is the normalization constant at time $t-1$. As one can see, the PA function plays a key role in the definition of the model. This is the reason why we can talk about a *class* of PAMs. The Barabási-Albert (BA) model is defined for $f(v, t-1) = D_v(t-1)$, where $D_v(t-1)$ is the degree of vertex v at time $t-1$. In this setting, new vertices tend to link to vertices with already a high number of connections.

We point out that a mechanism as in (1.2.1) is well defined when every vertex comes into the graph with one edge. In case of multiple edges, several definitions are possible, but each edge chooses a vertex to be attached according to a distribution that is roughly (1.2.1).

In the present work, we will focus on PAMs, and in particular, we will consider as PA function $f(v, t-1) = D_v(t-1) + \delta$, for some constant δ. As we will explain in the sequel, the parameter δ allows to tune the power-law exponent of the degree distribution. Also, we will consider a model where new vertices appear with $m \geq 1$ edges, so not necessarily $m = 1$. In this case we denote the PAM as $\mathrm{PA}_t(m, \delta)$. We point out that, once the PA function f is fixed, differences can still be made in the

precise definition of the attachment probabilities as in (1.2.1) (for instance, if self loops are allowed or not). For this reason, whenever we state a result we specify the precise PAM we consider.

The aim of the present thesis is to investigate properties of PAMs, and comparing them to what we observe in real-world networks, as discussed in Section 1.1. In particular, we prove that the PA dynamics with affine PA function can generate graphs with many of the features presented in Section 1.1, with the exception of clustering since, as it turns out, PAMs are locally tree-like. In addition, we identify other PA functions that can generate graphs with power-law degree distributions.

1.3. Degrees in PAMs

In this section, we discuss the degree distribution in PAMs. In particular, we discuss the scale-free properties in PAMs, showing that the degrees in such graphs obey a power-law distribution.

1.3.1. The degree distribution and known results

For $m \geq 1$ and in the case of PA function of the type $f(k) = k + \delta$, for a constant $\delta > -m$, PAMs are known to show power-law degree distributions. This is true for the BA model (so for $\delta = 0$), but also for other variants which differ for some detail in the attachment probabilities. In particular, these graphs show the so-called *old-get-richer effect*, since high-degree vertices are typically vertices that have appeared early in the graph.

The limiting degree distribution of PAMs has been found by several works. Bollobás, Riordan, Spencer and Tsunády [33] were the first to prove it for $\delta = 0$. Van der Hofstad [85, Theorem 8.4] states that for every $k \in \mathbb{N}$, the proportion of vertices of degree k in $\mathrm{PA}_t(m, \delta)$ converges in probability to $p_k^{(m)}$, and $p_k^{(m)}$ is given by

$$p_k^{(m)} = (2 + \delta/m) \frac{\Gamma(2 + \delta/m + m + \delta)}{\Gamma(m + \delta)} \frac{\Gamma(k + m + \delta)}{\Gamma(k + m + \delta + 3 + \delta/m)}. \tag{1.3.1}$$

where $\Gamma(\cdot)$ here denotes the Gamma function. In particular, the distribution $(p_k^{(m)})_{k \in \mathbb{N}}$ obeys a power law with exponent

$$\tau = \tau(m, \delta) = 3 + \frac{\delta}{m}. \tag{1.3.2}$$

In fact, by Stirling's formula, we have that

$$\frac{\Gamma(k + a)}{\Gamma(k + b)} = k^{a-b}(1 + O(1/k)).$$

This implies that in (1.3.1), as $k \to \infty$,

$$p_k^{(m)} \approx (k + m + \delta)^{-(3+\delta/m)} \approx k^{-(3+\delta/m)}.$$

Notice that the parameter δ allows one to tune the power-law exponent, creating different phases that impact the graph globally. For $\delta > 0$, the limiting distribution has finite variance, while for $\delta < 0$ the limiting distribution has finite mean but infinite variance. The case $\delta = 0$ is the boundary case of the Albert-Barabási model.

The formulation of (1.3.1) in terms of negative binomials is due to Ross [150]. Many properties of the degree distribution and in general of the evolution of degrees are known. Szymański [158] investigates moments of degrees of fixed vertices in the case of $\delta = 0$. Bollobás and Riordan [32] investigates the degrees for $\delta = 0$ using the relation between the PAM and the n-pairings. Jordan [103] investigates the limiting degree distribution in the case where the m edges of every vertex are added independently of each other, without updating the degrees of the chosen vertices. Hagberg and Wiuf [80], similarly to van der Hofstad [85, Chapter 8], investigate degree distribution starting from the recursive properties of $(N_k(t))_{t\in\mathbb{N}}$, i.e., the number of vertices with degree k. Deijfen et al. [52] investigate the degrees of a PAM where every vertex comes into the graph with a random number of edges. Dereich and Mörters [54], as well as Oliveira and Spencer [109], consider a PAM with a sublinear PA function of the type $f(k) \propto k^\alpha$, for some $\alpha \in (0, 1)$, proving that the corresponding degree distribution has a stretched-exponential tail, in agreement with results [9, 10, 151, 152] on PA trees.

1.3.2. Continuous-time embedding

Most proofs that the degrees in PAMs converge to the distribution $(p_k^{(m)})_{k\in\mathbb{N}}$ defined in (1.3.1) are based on a martingale argument [60, 85, 127, 128]. In particular the evolution of the degree of a vertex $v \in [t]$ can be seen as a supermartingale, that can be rescaled to a martingale with mean 1. In the case of PA trees ($m = 1$), an alternative proof is possible, using a clever relation to continuous-time branching processes (CTBPs).

CTBPs are formally introduced in Section 2.1. CTBPs are stochastic models for the evolution of a population of individuals that produce children according to i.i.d. copies of a counting process $(\xi_t)_{t\geq 0}$ called the birth process. These models have been studied for decades [9, 10, 11, 22, 152], since the first works by Jagers and Nerman [95, 96, 129]. In particular, CTBPs produce random trees, where vertices are individuals and (directed) edges are links from children to parents. These trees are dynamic, in the sense that their sizes grow as new individuals are born. This means that we can identify the sequence of random birth times $(\tau_n)_{n\in\mathbb{N}}$, i.e., the sequence at which individuals are born in the branching population.

CTBPs are known to embed PA trees (so $m = 1$), that are defined as a discrete sequence of random graphs, into continuous time [9, 10, 22, 152]. The heuristic idea is to construct a CTBP whose tree in continuous time $\mathcal{T}(t)$ coincides with a PA tree along the sequence of birth times $(\tau_n)_{n\in\mathbb{N}}$. In other words, the law of the tree $\mathcal{T}(\tau_n)$ is the law of PA_n. The advantage is that we can use the tools of CTBPs theory to investigate properties of PA trees.

The relation between a CTBP and the embedded PA tree is given by the PA function. In particular, if an individual of age t and k past children produces a new child

at rate $\lambda(t,k) = f(k)$ (so independently of the age), then the embedded PA tree is defined by the PA function f.

The limiting degree distribution in a CTBP defined by a birth process $(\xi_t)_{t\geq 0}$ is known [95, 96, 129], and it is given by

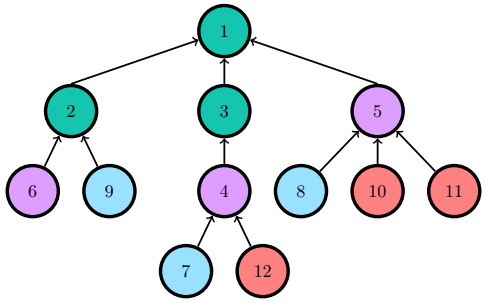
$$p_k^{(1)}(\xi) = \mathbb{P}\left(\xi_{T_{\alpha^*}} = k\right), \tag{1.3.3}$$

where T_{α^*} is an exponentially distributed random variable with mean $1/\alpha^*$, and $\alpha^* > 0$ is called the *Malthusian parameter* of the CTBP. Heuristically, α^* is the rate at which the CTBP size grows over time, in fact the population alive at time t is roughly $\mathrm{e}^{\alpha^* t}$. We refer to Section 2.1 for the definition of the Malthusian parameter and the result about the growth of a CTBP. In particular $p_k^{(1)}(\xi)$ can be interpreted as the probability that *an individual in the branching population produces k children in the random unit of time T_{α^*}*.

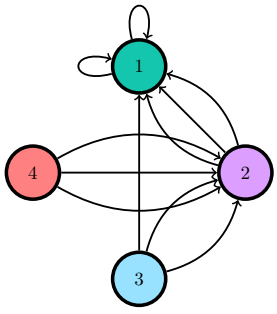
In (1.3.3) we have used the notation $(p_k^{(1)}(\xi))_{k\in\mathbb{N}}$ because (1.3.1) (for $m = 1$) and (1.3.3) coincide whenever the CTBP embeds the PA tree in continuous time.

1.3.3. Collapsed branching processes

As mentioned above, the embedding construction is limited to trees. In particular, this means that the CTBP tools can not be applied in the case $m \geq 2$. In Chapter 2, based on [70], we are able to extend the continuous-time embedding to the case $m \geq 2$, given an alternative proof of (1.3.1).



(a) Branching process tree



(b) Collapsed branching process

Figure 1.7: An example of a collapsed branching process (CBP) where vertices have fixed out-degree $m = 3$. On the left, we have a tree given by a CTBP at time τ_{12}, so the size is 12. On the right, we have the corresponding CBP where the out-degree of all vertices (except vertex 1) is 3. Notice that *individuals* 1,2 and 3 in CTBP are collapsed into *vertex* 1 in the CBP, individuals 4,5 and 6 into vertex 2, and so on.

Given a CTBP that embeds a PA tree, we define a random graph in continuous time called a *collapsed branching process* (CBP) that embeds a PA graph for $m \geq 2$. CBPs are formally defined in Definition 2.2.1, here we give the intuition behind the construction. Given a tree $\mathcal{T}(t)$ of a CTBP, we fix $m \geq 2$. Then, we divide born individuals in groups of m individuals each, according to the birth order. For instance, for $m = 1$, the first three groups are $\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}$. We can now imagine gluing together individuals belonging to the same group, thus generating *vertices*. We point out that nodes in CTBP are *individuals*, while nodes in CBP are *vertices*. We keep track of edges between individuals to draw edges between vertices. In other words, the number of edges between two vertices v_1 and v_2 is the number of edges between the two groups of individuals that have been collapsed together to generate v_1 and v_2. An example on a tree of size 12 with $m = 3$ leading to a graph of size 4 is given in Figure 1.7.

In Theorem 2.2.2, under general assumptions on $(\xi_t)_{t\geq 0}$, we show that a CBP has a limiting degree distribution $(p_k^{(m)}(\xi))_{k\in\mathbb{N}}$, that is given by

$$p_k^{(m)}(\xi) = \mathbb{P}\left(\xi^1_{T_{\alpha^*}} + \cdots + \xi^m_{T_{\alpha^*}} = k\right), \tag{1.3.4}$$

where $(\xi^1_t)_{t\geq 0}, \ldots, (\xi^m_t)_{t\geq 0}$ are m independent birth processes distributed as $(\xi_t)_{t\geq 0}$, T_{α^*} is an exponentially distributed random variable with mean $1/\alpha^*$, and $\alpha^* > 0$ is the Malthusian parameter of the underlying CTBP.

The distribution in (1.3.4) is the distribution of the sum of m independent birth processes that define the CTBP, evaluated at time T_{α^*}, similarly as the distribution in (1.3.3) which is the distribution of a single birth process at time T_{α^*}. Similarly to the tree case, we can see $(p_k^{(m)}(\xi)$ as the probability that m *individuals born at the same time produce in total* k *children in the random unit of time* T_{α^*}. This explains the notation in (1.3.3) and (1.3.4) of 1 and m, that underline the number of birth processes considered.

The proof of (1.3.4) makes extensive use of so-called *random characteristics* in CTBPs theory. A random characteristic is a bounded function that can be interpreted as a *counting function* of a certain characteristic of the individuals in the branching population. A CTBP evaluated with a characteristic is nothing more than the number of alive individuals that satisfies a certain property. For $k \in \mathbb{N}$, the function $\mathbb{1}_{\{\xi_t=k\}}$ is the characteristic that counts the number of individuals with k children in the CTBP.

In CBPs, the degree of a vertex $n \in \mathbb{N}$ is given by the sum of m different birth processes $\xi_t^{(n,1)}, \ldots, \xi_t^{(n,m)}$, where $(n, j) = m(n-1)+j$. One fundamental property is that a random characteristics Φ, when evaluated on an individual, can depend only on the properties of the individual itself and its progeny. The individuals $(n, 1), \ldots, (n, m)$ might not be related at all in the genealogy of the CTBP. In our proof we *add artificial randomness into the CTBPs*, to approximate the number of vertices in CBP with degree k with a random characteristic of the CTBP. For an explanation in a picture, we refer to Figure 1.8.

Formula (1.3.4) proved in Theorem 2.2.2 is stated for general CTBPs. We extend this to PAMs with affine PA functions. In Corollary 2.2.3 we prove that if a CTBP embeds a PAM with $m = 1$, then it satisfies the conditions of Theorem 2.2.2. In

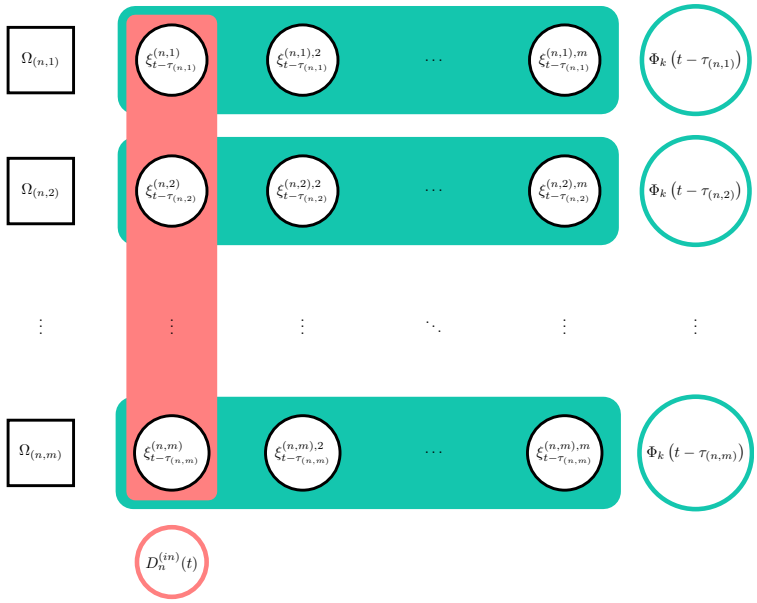


FIGURE 1.8: Heuristic of random characteristic $\Phi_k^{(m)}$. For a CTBP, to m individuals $(n,1),\dots,(n,m)$ we assign different probability spaces $\Omega_{(n,1)},\dots,\Omega_{(n,m)}$ respectively, with the birth processes $\xi^{(n,1)}_{t-\tau_{(n,1)}},\dots,\xi^{(n,m)}_{t-\tau_{(n,m)}}$. The degree of vertex n in the CBP is given by the sum of the processes as indicated by the circles within the red area. We artificially add the remaining processes in the green area to define the random characteristic $\Phi_k^{(m)}$.

Theorem 2.2.4 we show that the corresponding CBP embeds a PAM with $m \geq 2$, and in this case (1.3.4) coincides with the known distribution in (1.3.1).

The advantage of our approach is that it works *in more general settings*, for instance, for PA functions different from $f(k) = k + \delta$. As an example, we apply the result to the *random recursive tree* (RRT), that is embedded in continuous-time by a Yule process. A random recursive tree is a sequence of trees $(\mathcal{T}_n)_{n\in\mathbb{N}}$, where $\mathcal{T}_1$ consists of a single vertex, and for every $n \geq 2$, a new vertex appear with a single edge attached uniformly at random with one of the vertices $1,\dots,n-1$. We show that there exists a CTBP that embeds the RRT in continuous-time, and the corresponding CBP is a sequence of graphs $(G_n)_{n\in\mathbb{N}}$, where at every step the new vertex n has m edges attached "almost" uniformly at random to an existing vertex. This is made rigorous in Corollary 2.2.5.

A second example is given by PAMs *with aging*, where the PA function depends on the degree *and the age* of a vertex. We will introduce these models later, so we move this discussion to Section 1.9. This result is proven in Corollary 7.1.4.

1.4. PAMs can be ultra-small worlds

Chapter 3 is about the small-world phenomenon in PAMs. We investigate the behavior of the diameter of PAM *and the configuration model (CM)*. Even though the CM and PAM are quite different in nature, they are *locally* similar, because for both models the attachment probabilities are roughly proportional to the degrees. The core of our proof is a combination of *conditioning arguments* (which are particularly subtle for the preferential attachment model), with local estimates in order to derive bounds on *global* quantities, such as the diameter.

1.4.1. Typical distance and diameter

We define the *typical distance* H_n in a graph G_n as $\mathrm{dist}_{G_n}(V_1, V_2)$ (the distance between V_1 and V_2), where V_1 and V_2 are two vertices chosen uniformly at random in $[n]$. The *diameter* $\mathrm{diam}(G_n)$ of a graph G_n is defined as

$$\mathrm{diam}(G_n) := \max_{u,v\in[n]} \mathrm{dist}_{G_n}(u, v).$$

In principle, H_n and $\mathrm{diam}(G_n)$ might be infinite when the graph G_n is disconnected, i.e., when there exists a pair of vertices u, v such that there exists no path connecting u and v. In general, when investigating typical distance and diameter, we condition on the event where typical distance and diameter are finite, i.e., the graph is connected. Formally, a graph of size $n \in \mathbb{N}$ is called *small-world* whenever the typical distance and/or the diameter are of order $\log n$, and *ultra-small world* if they are of order $\log\log n$. Of course, other situations are possible (for instance, $\log n/\log\log n$), which we also call ultra-small worlds.

Typical distance and diameter have already been investigated in CM and PAM. There are more results on CM, in fact when $\tau > 3$ the typical distance is of order $\log n$ [89], while when $\tau \in (2, 3)$ H_n is sensitive to the presence of vertices with degree 1 and 2. Indeed, van der Hofstad, Hooghiemstra and Znamenski [90, 91] prove that H_n is of order $\log\log n$ when there are no vertices of degree 1 and 2, while when these are present H_n is of order $\log n$, since vertices of degree 1 and 2 form chains that increase distances. The diameter of CM shows similar behavior: it is of order $\log n$ when $\tau \in (2, 3)$ but vertices of degree 1 and 2 are present [66].

Less is known about PAMs. When $\tau \in (2, 3)$, Dereich, Mönch and Mörters [57] prove that H_n in PAMs is of order $\log\log n$, using path counting techniques we use in our result on the diameter. Van der Hofstad [86] investigates the diameter of PA trees (so $m = 1$), showing that the diameter is of order $\log n$. Dommers, van der Hofstad and Hooghiemstra [59] prove an upper bound on the diameter of PAM when $m \geq 2$, but the explicit constant in the bound (see c in (1.4.1) below) is not sharp. When $\tau = 3$ and $m \geq 2$, Bollobás and Riordan [32] prove that the diameter of PAM is of order $\log n/\log\log n$.

For a more detailed discussion of the literature about typical distances and diameter in CM and PAM, as well as other random graphs, we refer to Section 3.1.3.

1.4.2. Diameter for $\tau \in (2,3)$

We focus our analysis on the diameter of CM and PAM when the degree power-law exponent τ satisfies $\tau \in (2,3)$, which means that the degrees have finite mean but infinite variance. We consider the setting where the minimum degree of vertices in the graph in CM_n is $d_{\min} \geq 3$, and we consider $m \geq 2$ in PA_t. As mentioned, vertices of degree 1 and 2 are known to make distances longer, therefore we want to consider the case where these are not present.

We prove that

$$\frac{\mathrm{diam}(G_n)}{\log\log n} \xrightarrow{\mathbb{P}} c,$$

where G_n can be CM_n or PA_t, identifying the precise constant c for both models. In other words, we prove that, for every $\varepsilon > 0$, with high probability, as $n \to \infty$,

$$(1-\varepsilon)c\log\log n(1+o(1)) \leq \mathrm{diam}(G_n) \leq (1+\varepsilon)c\log\log n(1+o(1)). \tag{1.4.1}$$

The results are given in Theorem 3.1.3 for CM and Theorem 3.1.5 for PAM. These results, contained in Chapter 3, are based on [40].

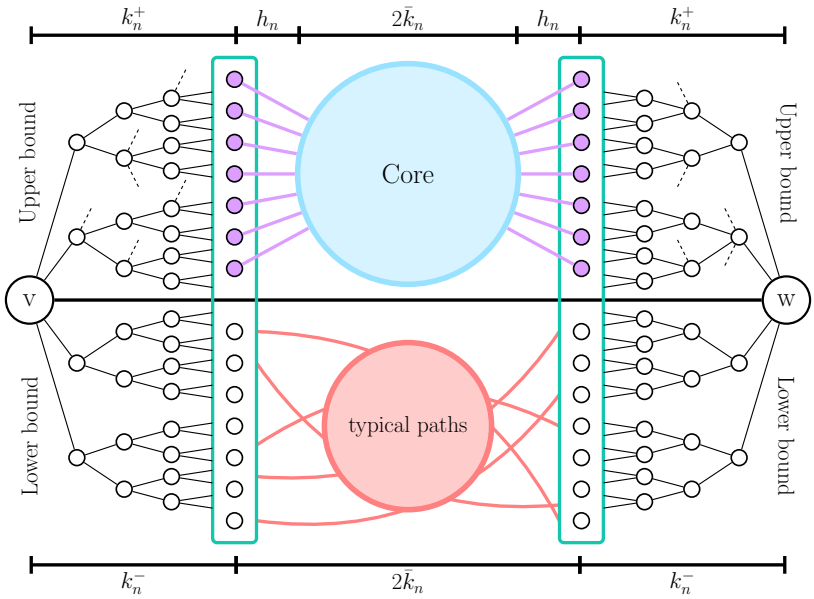


Figure 1.9: High-level explanation of the structure of the proofs of Theorem 3.1.3 and Theorem 3.1.5, that are the results about the diameter of CM and PAM respectively.

While the structures of the proofs for both models are identical, the details of the various steps are significantly different. Pairings in the configuration model are uniform, making explicit computations easy, even when already many edges have been paired. In the preferential attachment model, on the other hand, the edge statuses are highly dependent, so that we have to rely on delicate conditioning arguments. This is formalized in the notion of *factorizable events* in Definition 3.4.4.

The explicit constant c in (1.4.1) is given by

$$c = \frac{2}{\log d_{\mathrm{fwd}}} + \frac{2c_{\mathrm{dist}}}{|\log(\tau - 2)|}, \tag{1.4.2}$$

where τ is the power-law exponent, d_{fwd} is called the minimal *forward degree*, which is $d_{\min} - 1$ for CM_n and m for PA_t, and c_{dist} is a constant depending on the typical distance in the graph, that is $c_{\mathrm{dist}} = 1$ for the CM and $c_{\mathrm{dist}} = 2$ for the PAM.

The *minimal forward degree* d_{fwd} in (1.4.2) can be informally interpreted as the minimum number of vertices we can explore at each step from a vertex with minimum degree. For instance, assume that $d_{\min}$ is the minimum degree in CM. Then, starting the exploration process from a vertex $v \in [n]$, we can identify at least $d_{\min}$ vertices (assuming that we do not find multiple edges). From each one of these vertices we can find in the next step of the exploration at least $d_{\min} - 1$ new vertices, since we used a half-edge in the previous exploration step. If the neighborhood of the vertex v is a tree, the number of children we can find at each step is at least $d_{\min} - 1$. We call this the minimal forward degree since we use this in exploration processes, where we *go forward* in the exploration process by finding new vertices. The first term in (1.4.2) comes from the exploration of the *periphery* of the graph.

The second term in (1.4.2) depends on the power-law exponent, and the difference between PAM and CM lies in the value of c_{dist}. This term is related to the distances between vertices with high degrees. In fact, in PAM vertices with high-degree are typically at distance 2, while in CM they are at distance 1. This is due to the fact that the probability of connecting two high degree vertices in CM is high since we have many half-edges that we can pair to make an edge. In PAM high-degree vertices are typically old, so they might not be directly connected, but they are likely to be connected by an intermediate young vertex, thus making distances twice as big.

The proof of Theorem 3.1.3 and Theorem 3.1.5 is divided in a *lower bound* and an *upper bound* on the diameter of the graph. We now give a heuristic explanation of the proof (see Figure 1.9 for a graphical representation).

Lower bound. We prove that there are so-called *minimally-connected* vertices. These vertices are quite special, in that their neighborhoods up to distance

$$k_n^- \approx \log\log n / \log d_{\mathrm{fwd}}$$

are *trees with the minimal possible degree*, given by $d_{\mathrm{fwd}} + 1$. This explains the first term in the right hand sides of (1.4.2). Notice that k_n^- is exactly the distance up to which the explored neighborhood of a vertex v with minimum degree is a tree with high

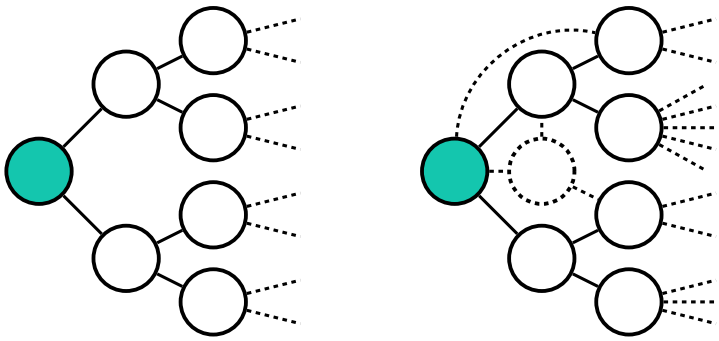


FIGURE 1.10: Examples of minimally-connected neighborhood (on the left) and exploration tree (on the right), in PAM with $m = 2$. Notice that in the minimally-connected tree at every exploration step we find exactly 2 new vertices, while the degree of all vertices is 3 (except the green vertex that has degree 2). The exploration graph is a subgraph of the neighborhood of the green vertex, where we consider only d_{fwd} edges at every exploration step. In the case of PAM, we explore the m original edges of every vertex. Notice that we can ignore both vertices and edges, that can only make distances shorter.

probability.

Equivalently, k_n^- is the exact distance such that the total number of vertices found in the exploration starting at a vertex v with minimum degree is $\log n$. In fact, since from every vertex we find d_{fwd} new vertices, the number of such vertices up to distance k_n^- from v is

$$d_{\text{fwd}}^{k_n^-} \approx d_{\text{fwd}}^{\log\log n/\log d_{\text{fwd}}} \approx \log n.$$

The tree structure of the neighborhood of a minimally-connected vertex v is a compromise between a *fast* and a *slow* structure around v. It is fast in the sense that it is a tree, i.e., a tree is the best possible structure in order to move away from vertex v. In other words, we do not waste steps in the exploration due to triangles or cycles, but we get as far away as possible from v. It is though a slow structure, since the degrees in the tree are the lowest-possible degrees in the graph.

Pairs of minimally-connected vertices are good candidates for achieving the maximal possible distance, i.e., the diameter. The tree structure with minimum degree around these vertices makes hard to reach them, or in other words, to reach a minimally-connected vertex it is necessary to go through the tree neighborhood, thus using k_n^- steps.

Since the boundaries of their tree-like neighborhoods turn out to be at distance equal to the *typical distance* $2\bar{k}_n$ in the graph, that is

$$2\bar{k}_n \approx 2c_{\text{dist}} \log\log n/|\log(\tau-2)|.$$

This leads to the second term in the right-hand side of (1.4.2). In the proof, we split the possible paths between the boundaries of two minimally-connected vertices into bad paths, which are too short, and typical paths, which have the right number of edges in them, and then show that the contribution due to bad paths vanishes. The degrees along the path determine whether a path is bad or typical. The strategy for the lower bound is depicted in the bottom part of Figure 1.9.

Upper bound. We perform a lazy-exploration from every vertex in the graph. and realize that the neighborhood up to a distance k_n^+, which is roughly the same as k_n^-, contains at least *as many vertices as the tree-like neighborhood of a minimally-connected vertex*.

In this lazy-exploration, we identify a subgraph of the k_n^+-neighborhood of vertices, that we call the *k_n^+-exploration graph*, while all possible other vertices in this neighborhood are ignored. In particular, this gives an upper bound on the distance between the starting vertex v and the boundary of such subgraph, since the ignored edges might connect v to the boundary in a shorter way.

We then show that the vertices at the boundary of these lazy neighborhoods are with high probability *quickly* connected to the core, that is by a path of $h_n = o(\log\log n)$ steps. By *core* we mean the set of all vertices with large degrees, which is known to be highly connected, with a diameter close to $2\bar{k}_n$, similar to the typical distances (see [91] for the configuration model and [59] for the preferential attachment model).

The core has diameter $2\bar{k}_n$ for the following reason. In CM, a given vertex v with large degree d is connected with high probability to at least one vertex with degree $d^{1/(\tau-2)} > d$. $\bar{k}_n$ is the exact number of steps such that

$$(d^{1/(\tau-2)})^{\bar{k}_n} \approx d^{\log n} \approx n^{\beta},$$

for some positive β. Vertices with such high degrees are at distance one, since the number of half-edges to connect two of them is so large that this happens with high probability. The factor 2 in the diameter of the core is given by the fact that we need to reach hub vertices from 2 vertices v and w, that both take $\bar{k}_n$ steps.

In PAM, this structure is similar, but the distances are doubled. In fact, to move from a vertex with degree d to a vertex with degree $d^{1/(\tau-2)}$ it is necessary to find a *young* vertex that connects the two vertices, a so-called *t-connector*.

The proof strategy for the upper bound is depicted in the top part of Figure 1.9.

1.5. Subgraphs in PAMs

In this section, we introduce our results on subgraph counts in PAMs, in particular focusing on triangles, as they are related to the clustering coefficient. The results are presented in Chapter 5, and are based on [74].

1.5.1. Subgraph counts in random graphs

To investigate which subgraphs occur more frequently than expected for a given network, the subgraph count in a given network is usually compared to the subgraph count in a random graph null model [69, 120, 124, 133]. Several random graph models could potentially serve as null models. In practice, the null model is frequently obtained by randomly switching edges while preserving the degrees. For $\tau > 3$ often very few specific subgraphs (e.g. triangles) occur, while for $\tau < 3$ this is not mathematically tractable, so that it requires simulations to estimate the subgraph count in such networks [118, 172].

Three important scale-free networks null models that are mathematically tractable are the CM, the rank-1 inhomogeneous random graph [26, 50] and the PAMs. When the power-law degree exponent satisfies $\tau < 3$, the CM results in a network with many multiple edges and self-loops [85, Chapter 7]. A possible solution is to merge all multiple edges of the CM, and thus consider the *erased* CM instead. This model is mathematically tractable, and subgraph counts for this model were derived in [93]. The rank-1 inhomogeneous random graph is closely related to the erased CM, and subgraph counts for this random graph null model show similar behavior as in the erased CM [157]. In both models, every subgraph typically occurs on vertices with degrees with specific orders of magnitude.

The advantage of considering PAMs as null model, instead of CM or rank-1 inhomogeneous random graphs is that PAMs allow us to study *directed* subgraphs in *dynamic* graphs, since most real-world network subgraphs are directed as well [124, 154].

Several studies on the behavior of specific subgraphs in the PAMs with affine PA function exist. We now briefly summarize these results. The triangle subgraph has

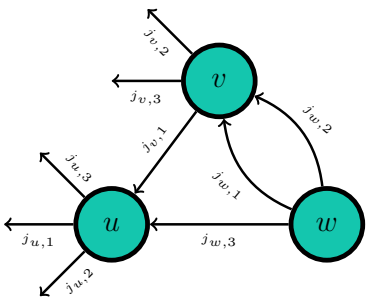
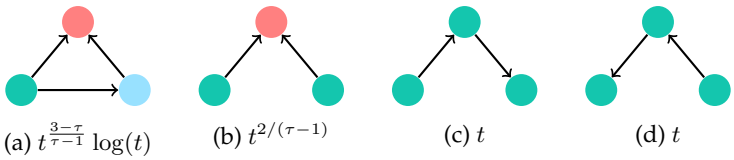


FIGURE 1.11: Example of directed ordered labeled triangles on three vertices u, v, w, with $u < v < w$. Notice that this is the unique possible ordering for a triangle. In the picture, we have two labeled triangles, one consisting of the edges $\{j_{v,1}, j_{w,1}, j_{w,3}\}$ and the other one of $\{j_{v,1}, j_{w,2}, j_{w,3}\}$.



(a) $t^{\frac{3-\tau}{\tau-1}}\log(t)$ (b) $t^{2/(\tau-1)}$ (c) t (d) t

FIGURE 1.12: Order of magnitude of $N(H)$ for all attainable directed connected graphs on 3 vertices. Vertices with degree one are green, vertices with free degree are light blue, and vertices with degree $t^{1/(\tau-1)}$ are red.

been studied in several papers, allowing to investigate clustering in the PAMs. Bollobás and Riordan [31] prove, when $\delta = 0$, that for any integer-valued function $T(t)$ there exists a PAM with $T(t)$ triangles, where t denotes the number of vertices in the PAM. They further show that the expected global clustering coefficient is of order $(\log t)^2/t$, while the expected number of triangles is of order $(\log t)^3$ and more generally, the expected number of cycles of length l scales as $(\log t)^l$.

Eggmann and Noble [61] consider $\delta > 0$, so that $\tau > 3$, and investigate the number of subgraphs for $m = 1$ (so subtrees), and for $m \geq 2$ triangles and clustering. They observe that the expected number of triangles is of order $\log t$ while the clustering coefficient is of order $\log t/t$, which is different than the results in [31].

In a series of papers [139, 140, 141] Prokhorenkova et al. proved results on the clustering coefficient and the number of triangles for a broad class of PAMs, assuming general properties on the attachment probabilities. These attachment probabilities are in a form that increases the probability of creating a triangle. They prove that the number of triangles in this setting is of order t, while the clustering coefficient behaves differently depending on the exact attachment probabilities.

1.5.2. Typical subgraph realizations in PAM

In the preferential attachment model, it is only possible for an older vertex to connect to a newer vertex but not the other way around. This puts constraints on the types of subgraphs that can be formed. For this reason, we talk about *ordered directed subgraphs*. Given a subgraph structure H on k vertices, we denote by π an ordering map, in the sense that the k vertices of a subgraph H are ordered as $\pi(u_1) < \cdots < \pi(u_k)$.

Not all subgraphs can appear in PAMs, since vertices appear in the graph one by one with $m \geq 1$ edges each. Thus, we say that a subgraph is *attainable* if it can be constructed in PAM. We formalize this idea in Definition 5.1.1 below.

An optimization problem. Given an ordered subgraph (H, π) on k vertices, our main object of interest is the probability that k uniformly chosen vertices in PAM generate (H, π) as a subgraph. To evaluate this probability, associate to (H, π) an optimization problem $B(H, \pi)$ (the precise definition is moved to Chapter 5).

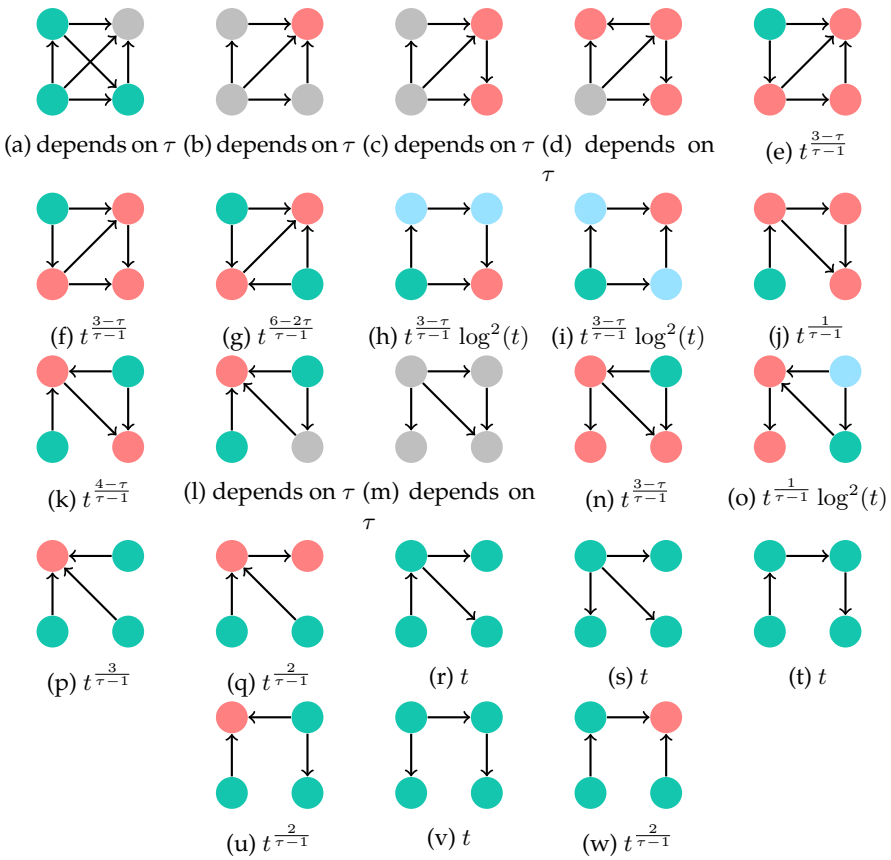


FIGURE 1.13: Order of magnitude of $N(H)$ for all attainable connected directed graphs on 4 vertices and for $2 < \tau < 3$. Vertices with degree one are green, vertices with free degrees are light blue, and vertices of degree $t^{1/(\tau-1)}$ are red. Vertices where the optimizer is not unique are gray.

The optimization problem $B(H, \pi)$ can be explained as follows. Assume that π is the identity mapping id, so that vertex 1 is the oldest vertex of H, vertex 2 the second oldest and so on. We show in Lemma 5.2.1 that the probability that an attainable subgraph is present on vertices $u_1 < u_2 < \ldots < u_k$ scales as

$$\prod_{i \in [k]} u_i^{\beta(i)}, \tag{1.5.1}$$

where

$$\beta(i) = \frac{\tau - 2}{\tau - 1}(d_H^{(\mathrm{in})}(i) - d_H^{(\mathrm{out})}(i)) - d_H^{(\mathrm{in})}(i),$$

and $d_H^{(\text{in})}(i), d_H^{(\text{out})}(i)$ denote the in- and out-degree of vertex i in the subgraph H respectively.

In other words, the probability that a set of vertices form H is proportional to the product of their indices (so their *age*) to some powers $\beta(1), \dots, \beta(k)$, that depend on the power-law exponent τ and *the internal structure of the subgraph* H. Thus, if $u_i \propto t^{\alpha_i}$ for all i, the probability that the subgraph is present scales as $t^{\sum_{i\in[k]} \alpha_i \beta(i)}$. The probability that a uniformly chosen vertex has index proportional to t^{α_i} scales as $t^{\alpha_i - 1}$. Therefore, heuristically, the number of times subgraph H occurs on vertices with indices proportional to $(t^{\alpha_i})_{i\in[k]}$ such that $\alpha_1 \le \alpha_2 \le \cdots \le \alpha_k$ scales as

$$t^{k+\sum_{i\in[k]}(\beta(i)+1)\alpha_i}. \tag{1.5.2}$$

Because the exponent is linear in α_i, the exponent is maximized for $\alpha_i \in \{0, 1\}$ for all i. Because of the extra constraint $\alpha_1 \le \alpha_2 \le \ldots \le \alpha_k$ the maximal value of the exponent is $k + B(H, \mathrm{id})$. This suggests that the number of subgraphs scales as $t^{k+B(H,\mathrm{id})}$.

Scaling of expectation of number of subgraphs Denoting the number of occurrences of the ordered subgraph (H, π) in PA_t by $N_t(H, \pi)$, we can prove that

$$C_1 \le \lim_{t\to\infty} \frac{\mathbb{E}[N_t(H,\pi)]}{t^{k+B(H,\pi)}(\log t)^{r-1}} \le C_2, \tag{1.5.3}$$

where r is the number of different optimizers of $B(H, \pi)$, or, in other words, the number of vertices in H whose value α_i is not uniquely identified in the optimization problem. This result is proved in Theorem 5.1.2, that gives the asymptotic scaling of the number of subgraphs where the order in which the vertices appeared in the preferential attachment model is known. The total number of copies of H for any ordering, $N_t(H)$, can then easily be obtained by summing over all the possible orderings π.

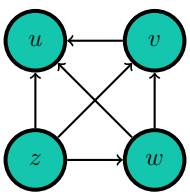


FIGURE 1.14: Complete 4-graph (4-clique) in PAM with directed edges. The unique possible ordering for this subgraph is $u < v < w < z$.

Thus, the optimization problem $B(H)$ finds the most likely configuration of a subgraph in terms of the indices of the vertices involved. If the optimum is unique, the number of subgraphs is maximized by subgraphs occurring on one set of very specific vertex indices. When the optimum is not unique, several maximizers contribute equally to the number of subgraphs, which introduces the extra logarithmic terms in (1.5.3).

Most likely degrees. As mentioned above, the optimization problem $B(H, \pi)$ finds the most likely orders of magnitude of the indices of the vertices. When the optimum is unique, it is attained by some vertices of constant index, and some vertices with index proportional to t. The vertices of constant index have degrees proportional to a constant with high probability (see [85, Theorem 8.2], and the original result in [158]), whereas the vertices with index proportional to t have degrees proportional to $t^{1/(\tau-1)}$. When the optimum is not unique, the indices of the vertices may have any range, so that the degrees of these vertices in the optimal subgraph structures have degrees ranging between 1 and $t^{1/(\tau-1)}$. Thus, the optimization problem (1.5.1) also finds the optimal subgraph structure in terms of its degrees.

The most likely degrees of all directed connected subgraphs on 3 and 4 vertices resulting from Corollary 5.1.3 and the asymptotic number of such subgraphs for $\tau \in (2, 3)$ are visualized in Figures 1.12 and 1.13. For some subgraphs, the optimum of $B(H, \pi)$ is attained by the same s and therefore the same most likely degrees for all $\tau \in (2, 3)$, while for other subgraphs the optimum may change with τ.

One such example is the complete graph of size 4 (see Figure 1.14). For the directed complete graph, there is only one attainable ordering π. For such subgraph, there is a phase transition at $\tau = 5/2$. In fact, for $\tau < 5/2$ a complete graph of size four typically contains three hub vertices of degree proportional to $t^{1/(\tau-1)}$ and one vertex of constant degree, and the number of such subgraphs scales as $t^{1-(\tau-2)/(\tau-1)}$ whereas for $\tau > 5/2$ the optimal structure contains four hub vertices instead and the number of such subgraphs scales as t.

Phase transition in the number of triangles. The result of Theorem 5.1.2 obtained through the optimization in (1.5.1) gives the order of magnitue of occurrence of subgraphs in PAM. More specific results are known for triangles [31, 61]. In Theorem 5.1.5 we are able to identify the precise constant $C > 0$ such that

$$\mathbb{E}[\Delta_t] = Ct^{3+B(\Delta)}(\log t)^{r-1}(1 + o(1)),$$

where Δ is the triangle subgraph, Δ_t is the number of triangles in PAM of size t, and r is the number of optimizers of $B(\Delta)$. In particular, $\mathbb{E}[\Delta_t]$ is of order $\log t$ for $\tau > 3$ [61], of order $(\log t)^3$ for $\tau = 3$ [31]. For $\tau \in (2, 3)$, we prove that $\mathbb{E}[\Delta_t]$ scales as $t^{(3-\tau)/(\tau-1)} \log t$. Since the expected number of connected triplets (subgraph (b)-(d) in Figure 1.12a) scales as t, this suggests that the clustering coefficient of PAMs vanishes as the size of the graph diverges. This is linked to the fact that PAMs are *locally a tree*, as we discuss next.

1.6. Treelike property and universality class of PAMs

We observe that the expected number of triangles in PAMs is sublinear with respect to the size of the graph. We now prove that this is related to the fact that the neighborhood of *most* vertices is structured as a tree (so no triangles for example) up to a finite distance. In other words, looking at a uniformly chosen vertex V_t in a PAM of size t, the subgraph composed by vertices up to a finite distance from V_t is with

high probability a tree. For this reason, PAMs are called *locally treelike graphs*. This idea can be formalized using the notion of local weak convergence.

1.6.1. Local weak convergence

Local weak convergence (LWC) is a concept introduced by Benjamini and Schramm [20] and formalized in [4, 5, 19]. Thus often in the literature LWC is called *convergence in Benjamini-Schramm sense* for undirected graphs. In this framework, a sequence of undirected random graphs, under relatively weak conditions, converges to a (possibly random) *rooted graph*, i.e., a graph where one of the vertices is labeled as the root. In simple words, the limiting graph resembles the *neighborhood of a typical vertex* in the graph sequence. This methodology has been shown to be useful to investigate *local* properties of a graph sequence – the properties that depend on the *local neighborhood* of vertices.

In the literature, limits of different types of random graphs have been investigated (Aldous and Steele give a survey in [4]). Grimmett [77] obtained the LW limit for the uniform random tree. Generalized random graphs [37, 51, 49, 64] also converge in the LW sense under some regularity conditions on the weight distribution. Convergence of undirected configuration model is proved in [86, Chapter 2]. In many random graph contexts, the LW limit is a branching process, and LW convergence provides a method to compare neighborhoods in random graphs to branching processes.

In the local weak convergence setting, a sequence of graphs $(G_n)_{n\in\mathbb{N}}$ converges to a (possibly random) rooted graph $(G, \varnothing)$ that is a *rooted graph*. Here $\varnothing \in V(G)$ denotes the root.

Heuristically, $G_n \to (G, \varnothing)$ in the LW sense when the law of the neighborhood of a typical vertex in G_n converges to the law of the neighborhood of the root in G. We now give an intuitive formulation of this concept (for a precise definition, see Section 4.1). For a vertex i in a graph G_n, denote the neighborhood of i up to distance k by $U_{\leq k}(i)$. Then, for a random rooted graph $(G, \varnothing)$ with law $\mathcal{P}$, we say that $G_n \to (G, \varnothing)$ if, for any finite rooted graph (H, y), and any $k \in \mathbb{N}$,

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{U_{\leq k}(i) \cong (H, y)\} \longrightarrow \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y)\right), \tag{1.6.1}$$

where $\mathbb{1}\{\cdot\}$ is an indicator of the event $\{\cdot\}$, and $U_{\leq k}(\varnothing)$ is the k-neighborhood of $\varnothing$ in G. The event $\{U_{\leq k}(i) \cong (H, y)\}$ means that the k neighborhood of i is structured as (H, y), ignoring the precise labeling of the vertices. Notice that the left-hand term in (1.6.1) is just the probability that the k-neighborhood of a uniformly chosen vertex in G_n is structured as (H, y).

Equation (1.6.1) is formulated for a deterministic graph sequence $(G_n)_{n\in\mathbb{N}}$. When $(G_n)_{n\in\mathbb{N}}$ is a sequence of *random* graphs, the left-hand term in (1.6.1) is a random variable. In this case there are different modes of convergence, as stated in Definition 4.1.6 below. For example, we say that $G_n \to (G, \varnothing)$ in probability if, for any finite rooted

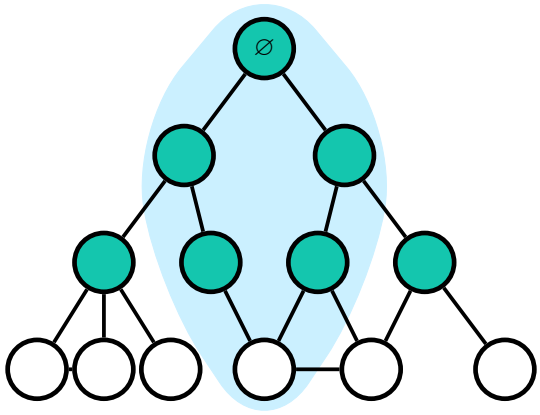


FIGURE 1.15: Example of treelike neighborhood. In this figure, the green vertices form the 2-neighborhood of the selected root $\varnothing$. Notice that this neighborhood is a tree. In treelike networks this implies that, typically, cycles are present in the graph but they are *relatively* long. For example, the vertices in the light blue area form a cycle, that is *long enough* so that we do not see it when looking at the 2-neighborhood of $\varnothing$.

graph (H, y), and any $k \in \mathbb{N}$,

$$\frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{U_{\leq k}(i) = (H, y)\} \xrightarrow{\mathbb{P}} \mathcal{P}\left(U_{\leq k}(\varnothing) = (H, y)\right).$$

The limiting rooted graph $(G, \varnothing)$ describes the typical neighborhood of G_n, i.e., the neighborhood of a typical vertex V_n. The notion of LWC then formalizes the idea of locally treelike graphs as follows: a sequence of graphs $(G_n)_{n \in \mathbb{N}}$ is locally treelike if $(G_n)_{n \in \mathbb{N}}$ converges in the LW sense (in distribution/in probability) to a random rooted *tree*. In other words, *the law of the finite neighborhood of a uniformly vertex V_n is a random rooted tree*.

1.6.2. Local weak limit of PAMs

Berger, Borgs, Chayes and Saberi [21] identify the LW limit of three versions of PAM, only in the case where the power-law exponent τ satisfies $\tau \geq 3$. The limit is called the *Pólya point graph*, which is a multi-type rooted infinite random tree. Since the Pólya point graph is actually a tree, we can thus talk about Pólya point *tree*. The Pólya point tree is a multi-type random tree where vertices have many attributes. In simple words, a Pólya point tree is a multi-type branching process with continuum

multi-dimensional type space. The definition requires heavy notation, so we refer to Section 4.3.1 for the details.

In [21], the Pólya point tree is identified as the LW limit of three versions of PAMs. These three modifications (models (d)-(f)-(g) in Section 4.3), are a mixture of PA mechanism as in (1.2.1) with $f(k) = k$, and uniform choice for the old vertex to which a new vertex is attached. In other words, fix $\alpha \in (0, 1)$. Then, every edge is attached as follows: with probability α the edge follows a PA mechanism with PA function $f(k) = k$, while with probability $1 - \alpha$ it is attached uniformly at random. These formulations are equivalent to a PA function $\hat{f}(k) = k + \hat{\delta}$, where the constant $\hat{\delta}$ satisfies $\hat{\delta} = 2(1 - \alpha)/\alpha \geq 0$.

Van der Hofstad [86, Chapter 4] claims that the LWC results holds for the sequential model presented in [21] for any value of $\delta > -m$, giving a sketch of the proof. In Chapter 4 we show how the argument of Berger et al. can be extended to different preferential attachment models. These models can differ for various reasons: different starting graph, presence or not of self-loops, edges added sequentially or independently. All the different versions of PAMs for which the convergence holds are listed in Section 4.3. Our approach using Pólya urn schemes with affine weight functions allows us to easily extend the result of Berger et al., stated for $\delta \geq 0$ ($\tau \geq 3$), to all values of $\delta > -m$ ($\tau > 2$).

In addition, we prove that for a class of PAMs the LWC to the Pólya point tree can be established only investigating the case $m = 1$, i.e., the tree setting. In other words, a preferential attachment model $(\mathrm{PA}_t(m, \delta))_{t\in\mathbb{N}}$ that, for $m \geq 2$, is defined from the tree case through a collapsing procedure as introduced in Section 1.3.3 (and Chapter 2), converges to the Pólya point tree if the same model in the tree case $(\mathrm{PA}_{mt}(1, \delta/m))_{t\in\mathbb{N}}$ converges to a Pólya point tree with different parameters. This fact highlights the relevance of the tree setting in PAMs, similarly to what we do in Chapter 2, where we investigate the degree distribution of CBPs using the properties of the underlying CTBP.

We point out that the Pólya point tree is not the unique LW limit of *all* possible preferential attachment models. In fact, Dereich and Morters [55, 54, 56] establish the LW limit in the case of preferential attachment models with *conditionally independent edges*. In this models, new vertices do not appear with a fixed $m \geq 1$ number of edges, but the new vertex $t + 1$ connects, conditionally on the current state of the graph, independently to any of the existing vertices, with probability given by $f(k)/t$. Interestingly, they identify the LW limit *for any (sub-)linear PA function* f, not only for the affine case.

1.7. Asymptotic PageRank distribution

In this section, we discuss the asymptotic distribution of PageRank scores in PAMs, and more generally in random graphs. As mentioned, PageRank is a centrality measure defined originally on directed graphs (but it has been extended to undirected

graphs). The algorithm on a graph G_n of size $n \in \mathbb{N}$ generates a vector $\boldsymbol{\nu}(n)$, where

$$\nu_i(n) = c \sum_{j \in [n]} \frac{e_{j,i}}{d_j^{(\text{out})}} \nu_j(n) + \frac{1-c}{n}. \tag{1.7.1}$$

Here $c \in (0,1)$ is called the *damping factor*, $e_{j,i}$ is the number of directed edges from j to i and $d_j^{(\text{out})}$ denotes the out-degree of vertex $j \in [n]$. The vector $\boldsymbol{\nu}(n)$ is a probability distribution on $[n]$. We will consider the rescaled version of PageRank that we call *graph-normalized PageRank* $\boldsymbol{R}(n) = n\boldsymbol{\nu}(n)$. Notice that in this setting it follows that $\mathbb{E}[R_{V_n}] = 1$, where R_{V_n} is the PageRank score of a uniformly chosen vertex $V_n \in [n]$.

Our main object of interest is the *asymptotic distribution of PageRank*. We want to investigate whether R_{V_n} converges in some sense to a distribution R, and whether R obeys a power law. In particular, whether R is a power-law distribution, we want to investigate whether the power-law exponent of the PageRank distribution coincides with the power-law exponent of the PAM.

1.7.1. Power-law PageRank and distributional equation

The PageRank power-law hypothesis presented in Section 1.1.4 has been formulated by empirical observations [38, 44, 67, 112, 135, 163]. Some works propose an explanation why PageRank shows the same power-law exponent of the in-degree distribution in a scale-free network. The equation (1.7.1) defining PageRank for the graph-normalized version has been generalized to a distributional equation of the type

$$R \stackrel{d}{=} \sum_{j=1}^{D^{(\text{in})}} \frac{C_j}{D_j^{(\text{out})}} R_j + B, \tag{1.7.2}$$

where $D^{(\text{in})}$ is the distribution of the in-degree in the network, $D^{(\text{out})}$ is the out-degree and $(C_j)_{j \in \mathbb{N}}$ and B are some random variables. Notice that (1.7.2) is more general than (1.7.1). This led to the definition of *generalized* and *personalized PageRank* algorithms. We refer to Chapter 6 for a more detailed discussion. The type of distributional equation in (1.7.2) is rather general. It appears for instance in branching process theory [101, 113, 114] and queueing theory [121].

Under some conditions on the distributions $(D^{(\text{in})}, D^{(\text{out})}, C, B)$, it is possible to show that the distribution R that solves (1.7.2) obeys a power law with exponent τ whenever $D^{(\text{in})}$ does. Using solutions of (1.7.2), Litvak and Volkovich [162] give precise result on the power-law solution of (1.7.2).

The solution of (1.7.2) is then used to prove the power-law PageRank hypothesis for some random graph models. Chen, Litvak, and Olvera-Cravioto [46] prove the result on the *directed* configuration model, assuming asymptotic independence of in- and out-degree. In a recent paper [110] Lee and Olvera-Cravioto prove the power-law hypothesis on directed inhomogeneous random graphs (we refer to Section 6.6.2 for a brief introduction of this model).

Avrachenkov and Lebedev [13] give expression for the *expected* PageRank scores

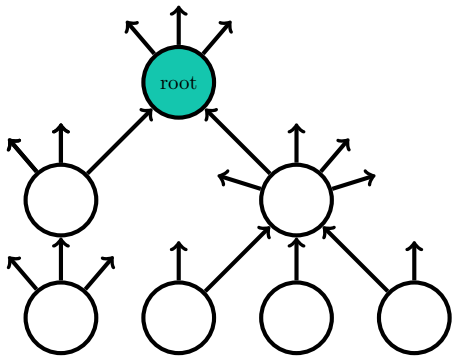


FIGURE 1.16: Example of marked rooted graph. We can visualize the difference between marks and out-degree as edges pointing to nowhere. In the LW limit, marks are not necessarily equal to the out-degrees. For instance, the LW limit of directed configuration model is a Galton-Watson tree (so out-degrees are all 1 except the root that has out-degree 0) with additional marks.

in the BA model, and they compare their result with the mean-field approach used in [16, 135]. Using heuristic arguments, they suggest that PageRank in the BA model should follow a power-law distribution with exponent $(3+c)/(1+c)$, i.e., the exponent should depend on the damping factor (see [13, Section 8]).

1.7.2. Local weak limit for PageRank

Our main contribution consists in giving a general criterion for the convergence of the PageRank distribution in a sequence of directed random graphs $(G_n)_{n\in\mathbb{N}}$. We prove that, for a sequence of directed random graphs $(G_n)_{n\in\mathbb{N}}$ that converges locally weakly, the PageRank score of a uniformly chosen vertex R_{V_n} converges to a distribution $R_\varnothing$. The result is stated in Theorem 6.2.1. This result, as well as all the results of Chapter 6, are based on [72].

Finite approximations of PageRank. The first result that we prove consists in showing that the PageRank score of a uniformly chosen vertex can be approximated with a finite number of "operations" with *arbitrarily small error*, and this error *does not depend on the graph size*.

Let us be more precise. In several works [8, 13, 23, 45] it is proven that PageRank can be written as a weighted sum of paths. In other words, the PageRank score of vertex $i \in \mathbb{N}$ is a *weighted sum of all paths ending at i of any length in G_n*. For $N \in \mathbb{N}$, a

finite approximation of PageRank is then defined as

$$R_i^{(N)}(n) = (1-c)\left(1+\sum_{k=1}^{N} c^k \sum_{\ell\in \mathrm{path}_i(k)} \prod_{h=1}^{k} \frac{e_{\ell_h,\ell_{h+1}}}{d^{(\mathrm{out})}_{\ell_h}}\right),$$

where $\mathrm{path}_i(k)$ is the set of paths of length k ending at i, and $e_{\ell_h,\ell_{h+1}}$ is the number of directed edges from e_{ℓ_h} to $e_{\ell_{h+1}}$. We prove that the PageRank score $R_{V_n}(n)$ of a uniformly chosen vertex can be approximated by $R^{(N)}_{V_n}(n)$, with an error of order c^N, for all graphs of all sizes.

Directed local weak convergence. We mentioned that LWC is defined for undirected graphs. We are able to extend the definition of LWC to directed graphs, by defining what we call *marked rooted graphs*. In our setting, to every vertex is assigned a mark $m^{(\mathrm{out})}$ larger than or equal to the out-degree. In particular, every directed graph can be seen as a marked graph just by assigning marks equal to the out-degrees. Marks are useful to keep track of outgoing edges, since in the exploration process in the graph we explore edges *only in their opposite direction*. In other words, a directed edge (u, v) is only explored from v to u.

The limiting object in this case is a marked rooted graph $(G, \varnothing, M(G))$, where G is a graph, $\varnothing$ is the root and $M(G)$ denotes the set of marks. An example is given in Figure 1.16. In particular, the limiting distribution $R_\varnothing$ of PageRank is defined as the sum of all weighted paths ending at the root $\varnothing$. Interestingly, in the limit *marks are not necessarily equal to the out-degree*, and the limiting graph *G can be of finite size*.

We prove that the finite approximations of PageRank are continuous in the directed LW topology, therefore we can pass to the limit, thus moving the analysis to

Field	Acronym	Papers	References
Astrophysics	AP	477113	9154818
Biotechnology and Applied microbiology	BT	537867	2705525
Geology	GE	56692	250668
Nuclear physics	NP	223321	1588534
Organic chemistry	OC	567146	5072139
Optics	OP	501817	2954769
Probability and Statistics	PS	185167	1053243
Sociology	SO	222416	451435

TABLE 1.3: List of citation network datasets considered. Datasets have been taken from *Web of Science*, where we considered paper classified in the above scientific fields. References are considered if an only if both the citing and cited papers are within the considered fields. References going and coming from other fields are ignored. We considered papers published between 1980 and 2015.

the directed local weak limit, where the limiting distribution $R_\varnothing$ is defined.

Application to directed PAMs. We apply Theorem 6.2.1 to directed PAMs in Section 6.4.2. We show that the directed version of PAM, where edges are directed from young to old vertices, converges in the directed LW sense to a directed version of the Pólya point graph (the undirected limit of PAM). In particular, since the out-degree in PAM is a constant $m \geq 1$, Theorem 6.2.1 implies that the PageRank score $R_{V_n}(n)$ of a uniform vertex in PAM converges to a distribution $R_\varnothing$, that is stochastically bounded by a multiple of the limiting in-degree distribution $(p_k^{(m)})_{k\in\mathbb{N}}$ given in (1.3.1). As a consequence, $\mathbb{P}(R_\varnothing > r)$ is bounded from below by a multiple of the tail of $(p_k^{(m)})_{k\in\mathbb{N}}$. In other words, $\mathbb{P}(R_\varnothing > r)$ *is at least of order* $r^{-(\tau_{\mathrm{D}}-1)}$*, where* τ_{D} *is the in-degree power-law exponent*. As a consequence, if $R_\varnothing$ obeys a power-law distribution with exponent τ_{PR}, then $\tau_{\mathrm{PR}} \leq \tau_{\mathrm{D}}$.

This partially proves the power-law PageRank hypothesis on PAMs, since we have a lower bound on $\mathbb{P}(R_\varnothing > r)$. The classical approach given by finding a solution of (1.7.2) does not apply to PAM since the directed LW limit is *finite*, therefore PageRank in PAM is not described by the solution (1.7.2) contructed in terms of weighted branching processes, as for instance in configuration model [46].

1.8. The citation networks example

So far, we have considered PAMs where the PA function f depends *only on the degree of vertices*. In particular we have focused on the case $f(k) = k + \delta$, since δ allows to tune the power-law exponent. This simple dynamics might not be feasible to represent many different types of complex networks.

We have shown that such a PA mechanism allows to generate graphs with many features that we observe in real-world networks, but the dependence of f only on the degrees shows limitations. In fact, all the different versions of PAMs with such PA function f show the *old-get-richer effect*, i.e., vertices with high degree are old vertices. This is due to the fact that all vertices are the same, and the only difference is given by the time they appear in the graph. This might not be true in real-world networks. For instance in the World-Wide Web the oldest webpage is not necessarily the page with the highest number of hyperlinks pointing towards it.

We consider now the example of *citation networks*, i.e., networks of scientific publications, where directed edges are references from a paper to another. Citation networks are a clear example of network evolving over time: at every time $t \geq 0$, a new publication can appear in the network. This new paper can contain references to the existing literature, that we see as directed edges.

We point out that in this setting, time is measured as *calendar time*, while in discrete-time PAMs time is measured by *discrete-time steps*. In other words, in discrete-time PAMs time moves from n to $n+1$ when we add a new vertex to the graph. In citation networks, time is a continuous quantity $t \in \mathbb{R}^+$.

We considered different citation network datasets taken from *Web of Science* (WoS). Details of the datasets considered are given in Table 1.3. We considered the classifi-

cation in different fields given by WoS, ignoring references coming from and going to papers that are in different fields. This division of the data might seem arbitrary, but it has two advantages. First, the data contained in WoS is too big to be analyzed simultaneously. Second, we want to consider citation networks as *finite directed graphs*, therefore we ignore references across fields.

We now present some *qualitative* analysis on the data, in order to identify characteristics of citation networks that can help us to *define a more general PA function f, that can overcome the limitation of the old-get-richer effect*. We point out that this analysis is made to identify qualitative properties of citation networks, and not quantitative features.

Exponential growth of scientific literature. Figure 1.17 shows the number of new papers published every year in the different datasets. The plot is made using a logarithmic y-axis. In this setting, a straight line in the plot implies that the number of publication grows exponentially. In fact, denote by P_t the number of papers published in year t. Then

$$\log(P_t) \approx \alpha t + b \qquad \Longleftrightarrow \qquad P_t \approx \mathrm{e}^{\alpha t + b},$$

which implies that the size of datasets grows exponentially over time. This has already been observed by Price [136, 137] and Lariviére et al. [106, 107]. In this framework *time plays a key role in the network evolution*, since to model citation networks we

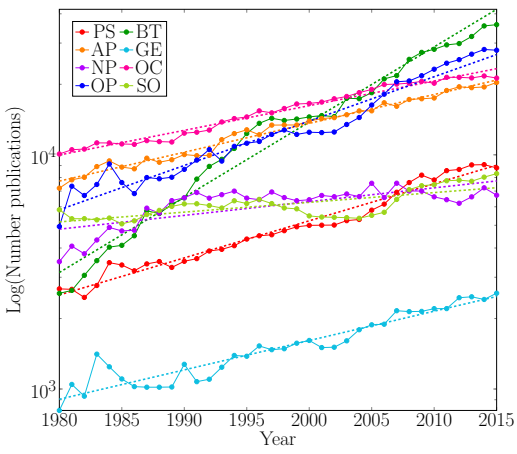


FIGURE 1.17: Number of publication per year, with logarithmic y-axis. The dashed straight lines in this plot are purely exponential functions of the type $\mathrm{e}^{\alpha t+b}$ obtained through regression, while the dots represent the actual data.

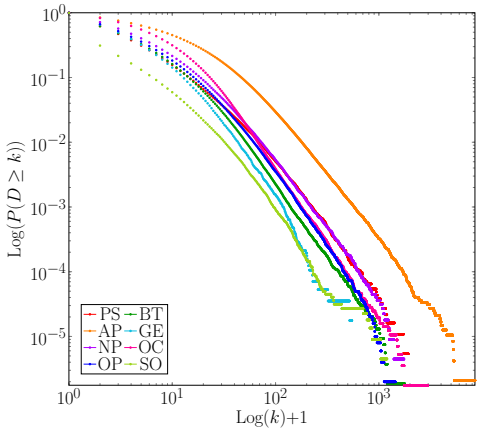


FIGURE 1.18: Loglog plot for the in-degree distribution tail in citation networks. Similarly to Figure 1.2, power-law distributions are identified by straight lines.

have to define a model where the size of the graph grows exponentially over time. As a consequence, we cannot look at discrete-time PAMs, but we have to use *continuous-time models*.

Power-law citation distribution. As in many other real-world networks, we observe power-law distributions for the number of citations (the in-degree of the networks) in the datasets. Power laws are considered to be present not only in citation networks [137], but also in other science-related networks, such as collaboration networks, productivity of authors and word occurrence [62, 116].

A discussion is still open whether a power-law distribution is the best model for citation distributions. Other than power law [137, 153], other distributions have been proposed to describe citation distributions, such as lognormal distributio [34, 144], Tsallis distribution [7, 164], Bessel distributions [142, 143].

We point out that, in Figure 1.18, the "most straight" line is the line corresponding to the Astrophysics (AP) datasets, while the one corresponding to the Geology (GE) dataset does not appear to be a power law. This might be a consequence of the "quality" of the data. In fact, the AP dataset is the *most dense dataset*, in the sense that the average degree in the AP graph in 19.18, while in the GE graph it is 4.42. In addition, the size of the AP graph is around ten times bigger than the GE graph (see Table 1.3).

Dynamical power laws. In Figure 1.19 we plot the tail degree distribution of papers published in 1985 in Organic Chemistry (OC) and Optics (OP), evaluated at different years (from 1986 to 2013).

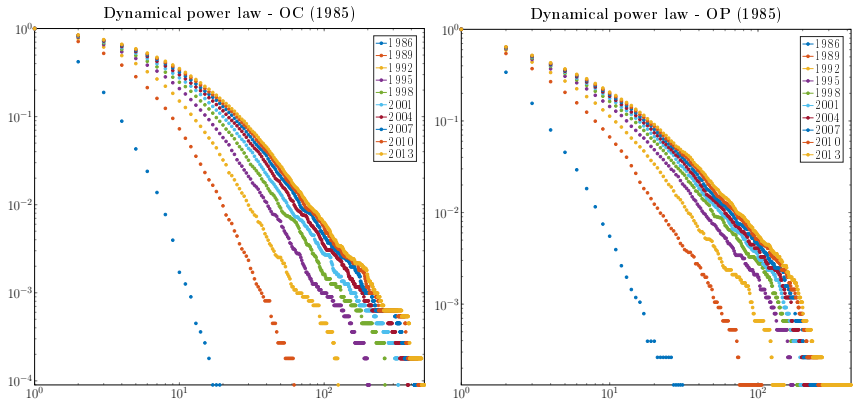


FIGURE 1.19: Evolution of the tail degree distribution of papers published in 1985 over different years. We observe that the degree distribution at time t appears to obey a power-law, with time-dependent exponent $\tau(t)$. Notice that $\tau(t)$ is decreasing, which implies that the papers have higher and higher degree.

We observe that the degree distribution seems to obey a power law, where the exponent $\tau(t)$ is *time dependent*. In particular, when we observe the degree distribution at two different times $t_1 < t_2$, we have that $\tau(t_1) > \tau(t_2)$, i.e, the power-law exponent $\tau(t)$ is decreasing. In other words, as time grows the distribution tail becomes *thicker*, i.e., papers with a larger number of citations are present. We call this phenomenon *dynamical power-law behavior*.

From a modeling point of view, we have to define a graph model where not only the graph degree distribution obeys a power law, but also the degree of fixed vertices is distributed according a power law, where the exponent is time-dependent.

Decreasing of average citations. On average, we observe that the number of citations received by papers over time decreases eventually. In Figure 1.20 we plot the average increment in citations for papers published in different years, in four different fields.

In general, we observe that in the first years after publications papers on average increase the number of citations they receive, while afterward this increment starts decreasing. This turning point is different among fields, as well as the decreasing behavior. For instance, in Nuclear Physics (NP) the maximum increment in the number of citations is around 2-4 years after publications. After that, the increment has a drastic decrease. In Probability and Statistics (PS) this maximum is reached around 4-6 years after publication, but the decreasing behavior is slower compared to the one in NP. Here, by slower we mean that in comparison to the maximum increment, the increment after 20 years in PS is higher than in the other fields.

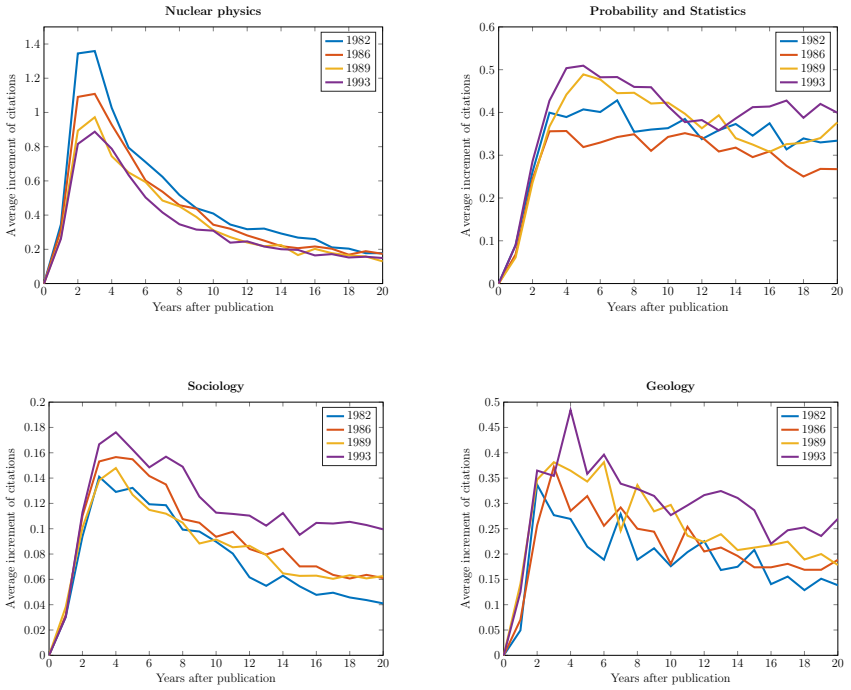


FIGURE 1.20: Average degree increment of papers published in the same year. Every line in every plot represents the average number of citations received by papers published in a particular year as a function of the years after publication. The peak of the increment in number of citations is different from field to field, as well as the *strength* of the effect of age on this increment. For example, the ratio between the maximum increment and the increment after 20 years is much larger in AP than in PS, showing that the effect of aging is stronger in AP than in PS.

Inhomogeneity of single papers. Figure 1.20 reports the average behavior of citations increment. The situation is rather different if we look at the number of citations of single papers. Figure 1.21 shows the behavior of the number of citations received by randomly selected papers published in 1985 in four different fields.

We notice that while the majority of papers stop receiving citations after some years, few papers keep being cited over a long period of time. This shows a different behavior among papers, i.e., *not all papers are the same*. This is contrast with random graph models, such as the Erdős-Rényi model, where all vertices in the networks are equivalent.

PAMs mentioned so far show different behavior from the behavior of citation networks shown in Figure 1.21. In fact, in PAMs the degree of vertices *diverges* as the size of the graph grows, and in general vertices with high degrees are the ones that appear first in the graph. This is clearly not true in citation networks, since the most cited papers are not necessarily the oldest published papers.

Affine dependence on past citations. Figure 1.22 shows the affine dependence between the past number of citations of a paper and the future ones. Each plot represents the average number of citations received by papers published in 1984 in the years 1993, 2006 and 2013 according to the initial number of citations in the same year. At least for low values of the starting number of citations k (for which there

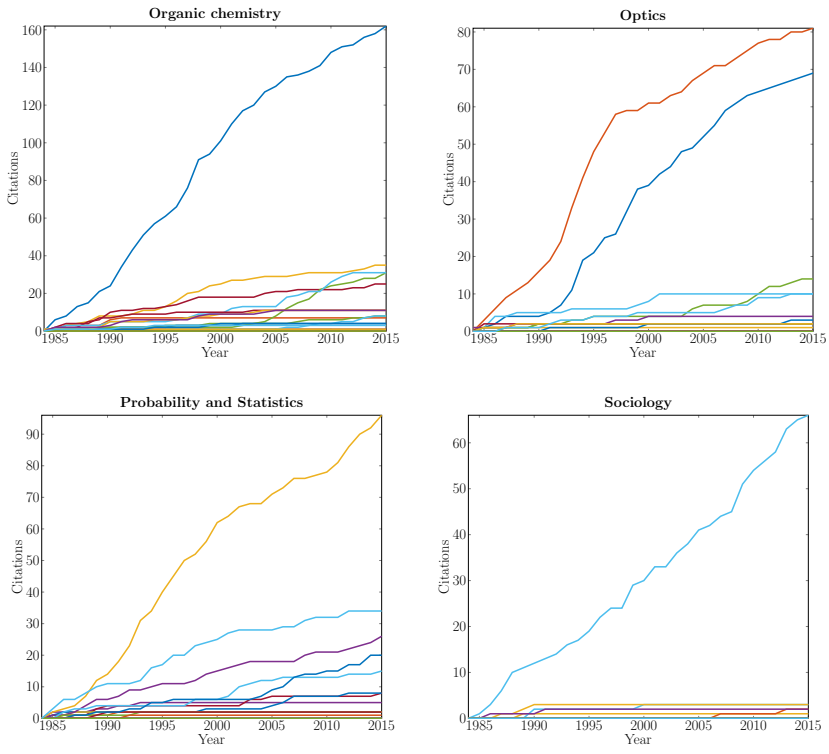


FIGURE 1.21: Evolution of the number of citations of randomly chosen papers selected in the same publication year. In particular, we observe that at individual levels papers show different behaviors. There is a substantial difference between papers that stop receiving citations and paper that still after a long period of time are cited.

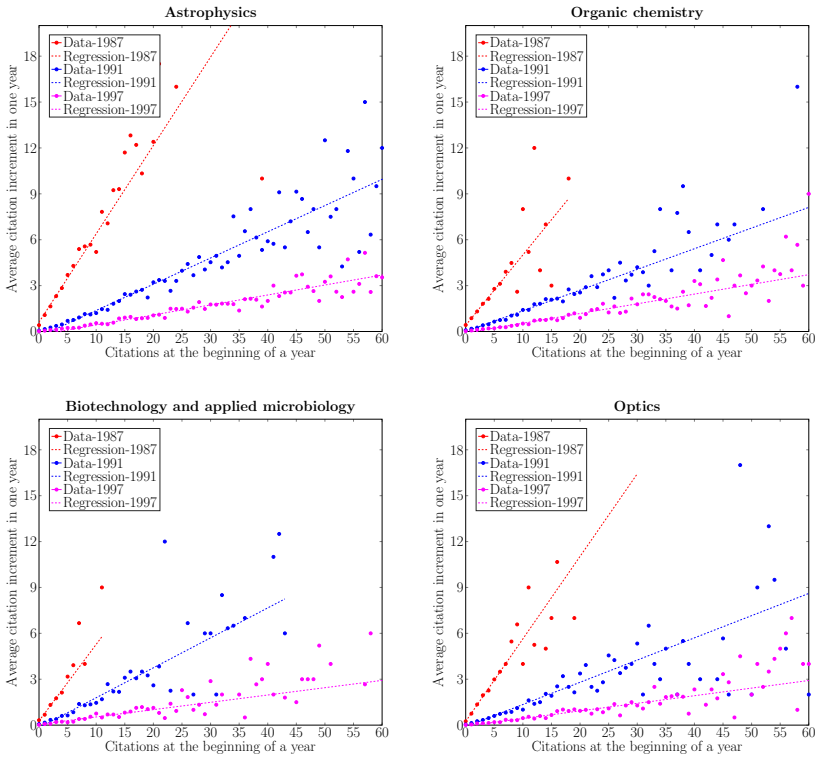


FIGURE 1.22: Linear dependence between past and future number of citations for papers from 1988. It is possible to notice that a higher number of past citations on average leads to a higher increment in the following year. This supports the assumptions that highly cited papers are more likely to be cited again.

are more papers), we see that the average number of citations received during a year grows roughly as an affine function of the type $ak + b$.

Notice that the slope a of the regression line *decreases* as time increases. This agrees with the observation in Figure 1.20, where the average increment decreases over time. In particular, papers with the same number of past citations have different average increment, according to the their age. In other words, Figure 1.22 seems to suggest that, on average, a paper with k citations published t years ago receives $a(t)k + b(t)$ new citations in the next year, for some functions $a(t), b(t)$.

Age of cited papers. Figures 1.17-1.22 show properties of papers in citation networks from the point of view of cited papers, i.e., from papers who *receive* citations. We now change this perspective, and we look at references from the point of view of

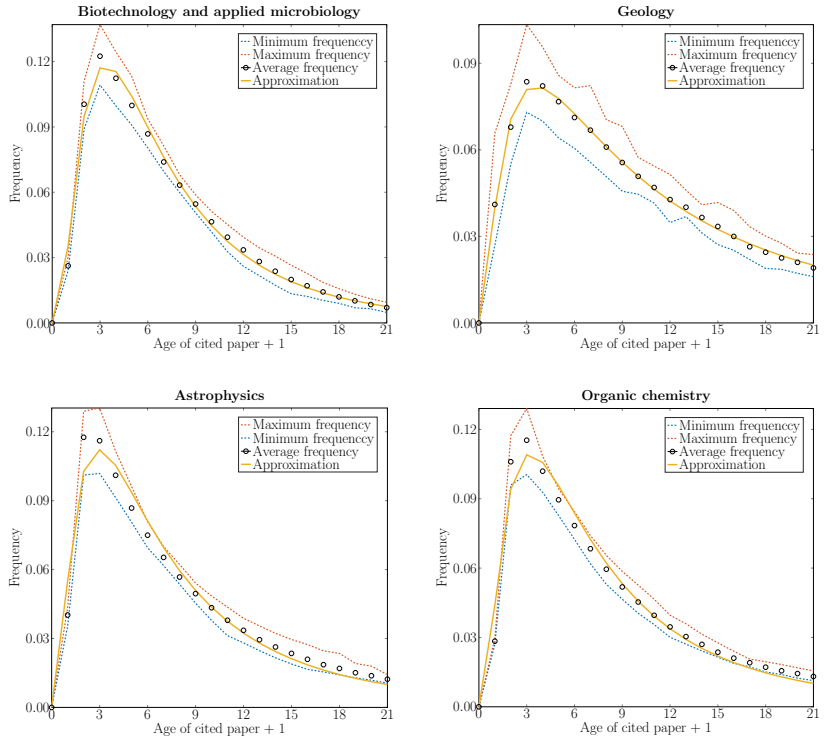


FIGURE 1.23: Distribution of the age of cited papers for different citing years. We look at a 20 years time-window to compare references contained in papers from different years.

citing papers.

In Figure 1.23 we plot the distribution of the age of cited papers for references in the datsets. We restrict our analysis on a 20-years window, in order to compare references contained in papers published in different years. In other words, the plots in Figure 1.23 represent the *distribution of the age of the paper cited by a uniformly chosen reference* in the data (conditioning on this age being smaller than 20). It appear that the distribution of the age of cited papers obeys a lognormal distribution [78, 119].

1.9. Beyond affine PA functions

As mentioned in Section 1.2.2, we can consider more general functions in (1.2.1) than the affine case. Here, motivated by the analysis of citation networks presented in Section 1.8, we describe a way to generate PA trees with more general PA functions, trying to model features that we observed in the data.

1.9.1. PA functions in the literature

Often, PAMs are defined through PA functions that depend only on the degree of vertices. Such models are called PAMs with *general PA function*. According to the asymptotics of the PA function $f(\cdot)$, the limiting degree distribution of the graph can behave rather differently.

There is an enormous body of literature showing that PAMs present power-law decay in the limiting degree distribution precisely when the PA function is affine, i.e., it is a constant plus a linear function [151]. In the sub-linear case, instead, the degree distribution is *stretched-exponential*, while in the super-linear case it collapses, in the sense that one of the first vertices will receive all the incoming new edges after a certain step [132].

As mentioned, these models show the old-get-richer effect, meaning that the vertices of highest degrees are the vertices present early in the network formation. An extension of this model is called preferential attachment models with a *random number of edges* [52], where new vertices are added to the graph with a different number of edges according to a fixed distribution, and again power-law degree sequences arise. A generalization that also gives younger vertices the chance to have high degrees is given by PAMs with *fitness* as studied in [53, 58]. Borgs et al. [35] present a complete description of the limiting degree distribution of such models, with different regimes according to the distribution of the fitness, using *generalized Pólya urns*. An interesting variant of a multitype PAM is investigated in [149], where the author considers PAMs where fitnesses are not i.i.d. across the vertices, but they are sampled according to distributions depending on the fitnesses of the ancestors.

As mentioned, we try to model *citation networks*, where vertices denote papers and the directed edges correspond to citations. For such networks, other models using preferential attachment schemes and adaptations of them have been proposed mainly in the physics literature. Aging effects, i.e., considering the *age of a vertex* in its likelihood to obtain children, have been extensively considered as the starting point to investigate their dynamics [43, 81, 82, 167, 168]. Here the idea is that old papers are less likely to be cited than new papers. Such aging has been observed in many citation network datasets and makes PAMs with weight functions depending only on the degree ill-suited for them. As mentioned above, such models could more aptly be called *old-get-richer* models, i.e., in general *old* vertices have the highest degrees. In citation networks, instead, papers with many citations appear all the time. Barabási, Wang and Song [166] investigate a model that incorporates these effects. On the basis of empirical data, they suggest a model where the aging function follows a lognormal distribution with paper-dependent parameters, and the preferential attachment function is the identity. In [166], the fitness function is estimated rather than the more classical approach, where fitnesses are assumed be i.i.d. across papers. Hazoglou, Kulkarni, Skiena and Dill in [84] propose a similar dynamics for citation evolution, but only considering the presence of aging and cumulative advantage without fitness.

1.9.2. Age and fitness in PA functions

We use trees defined by CTBPs to describe the tree case of a random graph that replicates features of citation networks. We need to define a birth process $(\xi_t)_{t\geq 0}$ that shows qualitatively the same characteristics of the number of citations of papers in the data. These results are presented in Chapter 7, based on [73].

Denote by $\lambda(t,k)$ the rate at which an individual of age t and k children generates a new child. In the *stationary* case the rate $\lambda(t,k)$ is *independent* of the age of the individual, so $\lambda(t,k) = f(k)$. PAMs with affine PA functions are part of this class.

When $\lambda(t,k)$ is *dependent* of the age of the individual, we talk about the *non-stationary* or *aging* case. In the presence of aging, we assume that age decreases the rate $\lambda(t,k)$. We model this by assuming that $\lambda(t,k) = f(k)g(t)$, for some eventually decreasing function g of time. This is the simplest way to introduce the effect of age, in the sense that f and g are two separate functions, where f depends only on the number of children already generated (so the in-degree) and g depends only on the age of the individual.

Due to the inhomogeneity we observe in the citation evolution of single papers, we assume that not all individuals are similar, but they have an *intrinsic potential* to generate more children than other individuals. We model this by assigning a *fitness* value η to every individual, independently across individuals. The fitness is considered as a multiplicative factor in the rate $\lambda(t,k)$. In this setting, an individual produce children according the rate $\lambda(t,k) = \eta f(k)g(t)$. Thus we talk about the *aging and fitness* case.

Recalling our observations of Section 1.8, we make the following assumptions of our CTBP model. First, since citation networks grows exponentially (recall Figure 1.17), we focus on CTBPs that are supercritical and Malthusian, i.e., the size of the networks grows as $\mathrm{e}^{\alpha t}$, where α denotes the Malthusian parameter of the CTBP. We refer to Section 2.1 for more details.

Second, recalling Figure 1.22, we assume that the PA function f is affine, so $f(k) = ak + b$. In terms of a PA scheme, this implies

$$\mathbb{P}\left(\text{a paper cites another with past } k \text{ citations} \mid \text{past}\right) \approx \frac{n(k)(ak+b)}{Z},$$

where $n(k)$ denotes the number of papers with k past citations, and Z is the normalization factor. Such behavior has already been observed by Price [138], Redner [147] and Barabási et al. [102].

Third, we assume that the aging function g is integrable. In fact, we start by the fact that the age of cited papers is lognormally distributed (recall Figure 1.23). By normalizing such a distribution by the average increment in the number of citations of papers in the selected time window, we identify a universal function $g(t)$. This function can be approximated by a lognormal shape with field-dependent parameters. In

particular, from the procedure used to define $g(t)$, we observe that

$$g(t) \approx \frac{\text{number of references to year } t}{\text{number of papers of age } t} \frac{\text{total number of papers considered}}{\text{total number of references considered}},$$

which means in terms of PA mechanisms that

$$\mathbb{P}\left(\text{a paper cites another of age } t \mid \text{past}\right) \approx \frac{n(t)g(t)}{Z},$$

where Z is again the normalization factor (different from before), while this time $n(t)$ is the number of papers of age t. We will consider g to be integrable, but not necessarily a lognormal function.

Fitness and aging as time rescaling. There is a nice interpretation of the presence of aging and fitness as *time-rescaled* version of the stationary process.

Let $(\xi_t)_{t\geq 0}$ be a birth process defined by the rates $\lambda(t,k) = f(k)$, for some PA function f. In this case the rate $\lambda(t,k)$ depends only on k, so the corresponding CTBP represents a PAM tree with PA function f.

It turns out that when we consider $\lambda(t,k) = \eta f(k)g(t)$, the corresponding process

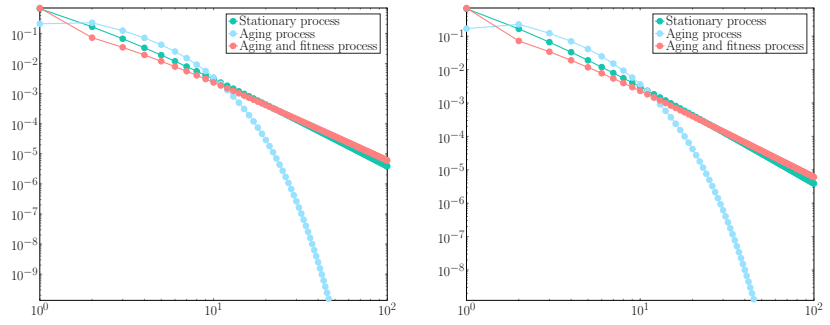


FIGURE 1.24: Examples of limiting degree distribution defined by a stationary process with affine PA sequence (red lines), two aging processes defined by the same PA function f and two different aging functions g_1 and g_2 (blue lines), and two non-stationary processes with fitness with same PA sequence, aging functions g_1 and g_2, and exponentially distributed fitness. On the left g_1 is an exponential function, while on the right g_2 is a power-law function of the type $(1+t)^{-c}$, $c > 1$. We see how the presence of an integrable aging function destroys the power-law behavior and generates distributions with a thinner tail. Here we observe that the exponential distribution for fitness restores the power-law behavior, but not with the original exponent.

$(C_t)_{t\geq 0}$ is distributed as $(\xi_{\eta G(t)})_{t\geq 0}$, where $G(t) = \int_0^t g(s)ds$. Here we use the notation C_t since this process with aging and fitness represents the number of citations of papers in our CTBP model to describe citation networks.

When the aging function is integrable, then $G(\infty) = \int_0^\infty g(s)ds < \infty$. This allows us to predict that, on average, the total number C_∞ of citation that a paper will ever receive *is finite*. In fact, assuming that we know the fitness η of a paper, we have that

$$\mathbb{E}[C_\infty \mid \eta] \approx \mathrm{e}^{\alpha\eta G(\infty)} < \infty.$$

In particular, C_∞ is always finite, but its value depends on the fitness. In other words, the higher the fitness η is, the larger C_∞ is. This agrees with Figure 1.20, where the average citation increment decreases over time, since $\mathbb{E}[C_t]$ approaches $\mathbb{E}[C_\infty] < \infty$. Also, the fact that $\mathbb{E}[C_\infty]$ depends on the fitness value agrees with Figure 1.21, where the number of citations of different papers has different behaviors. In this setting, the papers in Figure 1.21 that keep being cited are the ones with high fitness values.

The crucial point is to show that it is possible to obtain power-law degree distributions in preferential attachment trees where the birth process *is not just depending on an asymptotically linear weight sequence*, in the presence of *integrable aging* and *fitness*. Let us now briefly explain how these two effects change the behavior of the degree distribution.

Integrable aging and affine preferential attachment without fitness. In the presence of aging but without fitness, we show that the aging effect substantially slows down the birth process. In the case of affine f, aging destroys the power-law of the stationary regime, generating a limiting distribution that consists of a power law with exponential truncation. We prove this under reasonable conditions on the underlying aging function (see Lemma 7.4.1).

Integrable aging and super-linear preferential attachment without fitness. Since the aging destroys the power-law of the affine PA case, it is natural to ask whether the combination of integrable aging and *super-linear* f restores the power-law limiting degree distribution. Theorem 7.1.3 states that this is not the case, as super-linear f implies explosiveness of the branching process, which is clearly unrealistic in the setting of citation networks (here, we call a PA function f *super-linear* when $\sum_{k\geq 1} 1/f(k) < \infty$). This result is quite general, because it holds for *any* integrable aging function. Due to this, it is impossible to obtain power laws from super-linear f. This suggests that (apart from slowly-varying functions), affine PA function has the strongest possible growth, while maintaining exponential (and thus, in particular, non-explosive) growth.

Integrable aging and affine preferential attachment with unbounded fitness. In the case of aging and fitness, the asymptotic behavior of the limiting degree distribution is rather involved. We estimate the asymptotic decay of the limiting degree distribution with affine weights in Proposition 7.4.5. With the example fitness classes

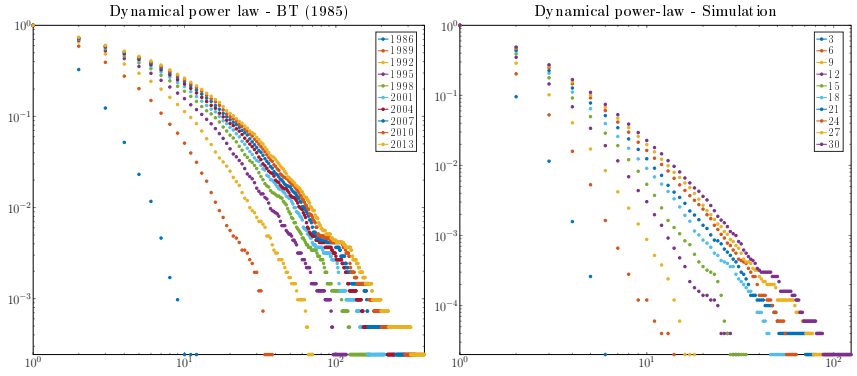


FIGURE 1.25: Dynamical power laws. Here we compare the evolution of tail distribution of the number of citations of papers in BT with simulations. The picture on the left shows the empirical tail distribution of papers published in 1985 in BT, evaluated at different years. On the right, we have simulated a number of aging and fitness birth processes, with integrable aging and exponential fitness at different time values. We observe that the two pictures show similar behaviors. On the right, our model predicts that the tail degree distribution obeys a power law with time-dependent exponent.

analyzed in Section 7.4.4, we prove that power-law tails are possible in the setting of aging and fitness, at least when the fitness has roughly exponential tail. So far, PAMs with fitness required the support of the fitness distribution to be *bounded*. The addition of aging allows the support of the fitness distribution to be unbounded, a feature that seems reasonable to us in the context of citation networks. Indeed, the relative attractiveness of one paper compared to another can be enormous, which is inconsistent with a bounded fitness distribution. We prove that we can restore the power law destroyed by the presence of aging by using fitness distributions with roughly *exponential tails*. When the fitness distribution has thinner tail, we do not restore the power law, and when it has a thicker tail than exponential the process is *explosive*.

Dynamical power laws. In the case of fitness with exponential tails, we further observe that the number of citations of a paper of age t has a power-law distribution with an exponent that depends on t. We call this a *dynamical power law*, and it is a possible explanation of the dynamical power laws observed in citation data (see Figure 1.19). In Figure 1.25 we compare simulations and data to visually inspect the similar behaviors.

Universality. An interesting and highly relevant observation in Chapter 7 is that the limiting degree distribution of preferential attachment trees with aging and fitness shows a high amount of *universality*. Indeed, for integrable aging functions, the

dependence on the precise choice of the aging function seems to be minor, except for the total integral of the aging function. Further, the dependence on fitness is quite robust as well.

1.10. Main contributions and outline of the thesis

We observed many features in complex networks. Due to their dynamic nature, we consider PAMs as a model to describe real-world networks. We recap the main content of the present thesis as follows:

Power-law degrees. Many real-world networks show power-law degree distributions. PAMs with $m \geq 1$ and PA function $f(k) = k + \delta$ are known to generate graphs with power-law degree and tunable exponent $\tau = 3 + \delta/m$. We consider the embedding in continuous-time, known for PA trees, and we extend it to $m \geq 2$. The power of our technique is that it can be applied to more general PA functions, such as in the presence of aging. This result is given in Chpater 2, based on the joint work with van der Hofstad [70].

Distances. Real-world networks are small-worlds. We investigate the diameter of PAMs with $m \geq 2$ and PA function $f(k) = k + \delta$, for $\tau \in (2, 3)$, as well as CM with scale-free degrees with $\tau \in (2, 3)$ and minimum degree larger than 3. We prove that for both models

$$\frac{\mathrm{diam}(G_n)}{\log\log n} \xrightarrow{\mathbb{P}} c,$$

where c is a constant, different for the two models, that we are identify. This shows a universality property for the proof strategy we use. These results are proven in Chapter 3, based on the joint work with Caravenna and van der Hofstad [40].

Subgraphs and clustering. Real-world networks are in general "clustered", in the sense that they show more triangles and small subgraphs than a randomized graph with the same degree distribution. Some results are already known about the number of triangles Δ_t in PAM with $m \geq 1$ and PA function $f(k) = k + \delta$. In fact, when $\tau > 3$, $\mathbb{E}[\Delta_t]$ is of order $\log t$, while when $\tau = 3$, $\mathbb{E}[\Delta_t]$ is of order $(\log t)^3$. We extend this result to $\tau \in (2, 3)$, proving that $\mathbb{E}[\Delta_t]$ is order $t^{(3-\tau)/(\tau-1)} \log t$. Our approach though is rather general, and we are able to identify the order of magnitue of every finite subgraph in PAM. These results are presented in Chpater 5, based on the joint work with Stegehuis [74].

Treelike property. PAMs are locally treelike, i.e., the finite neighborhood of most vertices is a tree. This is formalized using the local weak convergence notion, so a graph sequence $(G - n)_{n\in\mathbb{N}}$ is locally treelike if it converges to a rooted tree. In [21] the LW limit of three PAMs is identified as the Pólya point tree, when the power-law exponent τ that satisfies $\tau \geq 3$. We extend their argument, showing that the Pólya

point tree is the local weak limit of many different versions of PAMs, with power-law exponent $\tau > 2$. In particular, for PAM versions where the graph with $m \geq 2$ is defined by collapsing from the case $m = 1$, we show that the convergence holds if the tree case converges locally weakly to the Pólya point tree, with appropriately rescaled parameters. These results are presented in Chapter 4, based on a joint work with van der Hofstad [71], that is still in preparation.

Power-law PageRank. In scale-free networks with power-law exponent τ, it is empirically observed that PageRank obeys a power-law with the same exponent. We extend the notion of local weak convergence to directed graphs, showing that, for a sequence of graphs $(G_n)_{n\in\mathbb{N}}$ that converges locally weakly, the limiting PageRank distribution $R_\varnothing$ is identified by the local weak limit. In particular, in the case of PAMs with $m \geq 1$ and PA function $f(k) = k + \delta$, with our technique we prove that $R_\varnothing$ is stochastically bounded from below by a power-law distribution with exponent $\tau = 3 + \delta/m$. As a consequence, the tail of $R_\varnothing$ is at least the tail of a power-law distribution with exponent τ, solving partially the power-law PageRank hypothesis for PAMs. This results are given in Chapter 6, based on the join work with van der Hofstad and Litvak [72].

Citation networks and generalized PAMs. Motivated by an empirical analysis of citation networks, we observe that in general a PA function depending only on the degree of vertices might not be suitable for describing dynamics in complex networks. For this reason, we define a new class of CTBPs, representing directed random trees, interpreted as continuous-time PA trees, and we focus our attention on the conditions under which it is possible to generate trees with power-law degree distributions.

For PA functions f depending only on degrees, it is known that the corresponding trees show power-law behavior whenever the PA is asymptotically linear in the argument. We add the dependence on the age of the individual and on some potential attractiveness of every single individual that we call fitnesses. We prove that in the presence of aging but no fitness, we can only produce degree distributions with exponential tails. If we add fitnesses, then we can obtain a power law if and only if the fitnesses are sampled from a distribution with at most exponential tails. These results are proven in Chapter 7, based on the joint work with van der Hofstad and Woeginger [73].

In Chapter 8 we give conclusions to this thesis, and we discuss possible future research directions and open problems.

2

Degrees and scale-free property

Content and structure of the chapter

In this chapter, we address the problem of the presence of power-law degree distribution in PAM. We extend the continuous-time embedding known for PA trees to PA graphs, defining what we call *collapsed branching processes*, that are (multi)graphs defined in continuous time. Our technique has two advantages: First, it shows that the relation between continuous-time branching processes and PA graphs is not limited to trees, and second, our techniques can be applied to more general PA functions, and not only to the affine case.

The chapter is structured as follows: in Section 2.1 we introduce the theory of continuous-time branching processes, and we present known results that we need for our arguments. In Section 2.2 we formalize the definition of collapsed branching processes and we state the main results of the chapter. These results are based on Theorem 2.3.2, that is more general and it is introduced in Section 2.3, and proved in Section 2.5. Section 2.4 contains preliminary results that are necessary for the proof of Theorem 2.3.2. The novel results of this chapter are based on [70].

2.1. Preliminaries: continuous-time branching processes

Before introducing our result about the degree sequence in PAMs, it is necessary to introduce continuous-time branching processes (CTBPs) in detail, that are a fundamental ingredient of our result.

2.1.1. Definition

In CTBPs, individuals produce children according to i.i.d. copies of the same birth process. We now define birth processes in terms of point processes:

Definition 2.1.1 (Point process)**.** *A* point process ξ *is a random variable from a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *to the space of integer-valued measures on* $\mathbb{R}^+$*.*

A point process ξ is defined by a sequence of positive real-valued random variables $(T_k)_{k\in\mathbb{N}}$. With abuse of notation, we can denote the density of the point process ξ by

$$\xi(dt) = \sum_{k\in\mathbb{N}} \delta_{T_k}(dt),$$

where $\delta_x(dt)$ is the delta measure in x, and the random measure ξ evaluated on $[0, t]$ as

$$\xi(t) = \xi([0, t]) = \sum_{k\in\mathbb{N}} \mathbb{1}_{[0,t]}(T_k).$$

We suppose throughout the chapter that $T_k < T_{k+1}$ with probability 1 for every $k \in \mathbb{N}$.

Remark 2.1.2. *Equivalently, considering a sequence* $(T_k)_{k\in\mathbb{N}}$ *(where* $T_0 = 0$*) of positive real-valued random variables, such that* $T_k < T_{k+1}$ *with probability* 1*, we can define*

$$\xi(t) = \xi([0, t]) = k \qquad \textit{when} \qquad t \in [T_k, T_{k+1}).$$

We will often define a point process from the jump-times sequence of an integer-valued process $(V_t)_{t\geq 0}$*. For instance, consider* $(V_t)_{t\geq 0}$ *as a Poisson process, and denote* $T_k = \inf\{t > 0 : V_t \geq k\}$*. Then we can use the sequence* $(T_k)_{k\in\mathbb{N}}$ *to define a point process* ξ*.*

We now introduce some notation before giving the definition of CTBP.

Definition 2.1.3 (Ulam-Harris set)**.** *The Ulam-Harris set is*

$$\mathcal{U} = \bigcup_{n\in\mathbb{N}} \mathbb{N}^n.$$

For $x = x_1 \cdots x_n \in \mathbb{N}^n$ *and* $k \in \mathbb{N}$ *we denote by* $xk \in \mathbb{N}^{n+1}$ *the element* $x_1 \cdots x_n k$*. The* root *of the Ulam-Harris set is denoted by* $\varnothing \in \mathbb{N}^0$*.*

We use the elements of the Ulam-Harris set to identify individuals in the branching population. This is quite useful since the notaion in Definition 2.1.3 allows to denote the relationships between children and parents. In fact, for an individual $x \in \mathcal{U}$, we denote the k-th child of x by the element xk. This construction is well known, and has been used in other works on branching processes (see [95, 96, 129, 152] for more details).

We now are ready to define our branching process:

Definition 2.1.4 (Continuous-time branching process)**.** *Given a point process ξ, we define the* CTBP *associated to ξ as the pair of a probability space*

$$(\Omega, \mathcal{A}, \mathbb{P}) = \prod_{x \in \mathcal{U}} (\Omega_x, \mathcal{A}_x, \mathbb{P}_x)\,,$$

where $\mathcal{U}$ is the Ulam-Harris set as in Definition 2.1.3, and an infinite set $(\xi^x)_{x\in\mathcal{U}}$ of i.i.d. copies of the process ξ. We will denote the branching process by $\boldsymbol{\xi}$.

Remark 2.1.5 (Point processes and their jump times)**.** *We often define point processes in terms of jump times of processes $(V_t)_{t\geq 0}$. In order to keep the notation light, we will denote a jump process $(V_t)_{t\geq 0}$ by the corresponding point process ξ, and we write $(\xi_t)_{t\geq 0}$.*

According to Definition 2.1.4, a branching process is a pair of a probability space and a sequence of random measures. It is possible though to define an *evolution* of the branching population. At time $t = 0$, our population consists only of the root, denoted by $\varnothing$. Every time t an individual x gives birth to its k-th child, i.e., $\xi^x(t) = k + 1$, assuming that $\xi^x(t-) = k$, we start the process ξ^{xk}. Formally:

Definition 2.1.6 (Population birth times)**.** *We define the sequence of birth times for the process $\boldsymbol{\xi}$ as $\tau^{\xi}_{\varnothing} = 0$, and for $x \in \mathcal{U}$,*

$$\tau^{\xi}_{xk} = \tau^{\xi}_x + \inf\{s \geq 0 : \xi^x(s) \geq k\}\,.$$

In this way we have defined the set of individuals, their birth times and the processes according to which they reproduce. We still need a way to count how many individuals are alive at a certain time t.

Definition 2.1.7 (Random characteristic)**.** *A* random characteristic *is a real-valued process $\Phi\colon \Omega \times \mathbb{R} \to \mathbb{R}$ such that $\Phi(\omega, s) = 0$ for any $s < 0$, and $\Phi(\omega, s) = \Phi(s)$ is a deterministic bounded function for every $s \geq 0$ that only depends on ω through the birth time of the individual, as well as the birth process of its children.*

A relevant example of a random characteristic is obtained by the function $\mathbb{1}_{\mathbb{R}^+}(s)$, which measures whether the individual has been born at time s. Another example is $\mathbb{1}_{\mathbb{R}^+}(s)\mathbb{1}_{\{k\}}(\xi)$, which measures whether the individual has been born or not at time s and whether it has k children presently.

For each individual $x \in \mathcal{U}$, $\Phi_x(\omega, s)$ denotes the value of Φ evaluated on the progeny of x, regarding x as ancestor, when the age of x is s. In other words, $\Phi_x(\omega, s)$ is the evaluation of Φ on the tree rooted at x, ignoring the rest of the population. If we do not specify the individual x, then we assume that $\Phi = \Phi_{\varnothing}$. We use random characteristics to describe the properties of the branching population.

Definition 2.1.8 (Evaluated branching processes)**.** *Consider a random characteristic Φ as in Definition 2.1.7. We define the evaluated branching processes with respect to Φ at time $t \in \mathbb{R}^+$ as*

$$\boldsymbol{\xi}^{\Phi}_t = \sum_{x\in\mathcal{U}} \Phi_x(t - \tau^{\xi}_x).$$

The meaning of the evaluated branching process is clear when we consider the random characteristic $\Phi(t) = \mathbb{1}_{\mathbb{R}^+}(t)$, for which

$$\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}} = \sum_{x \in \mathcal{U}} (\mathbb{1}_{\mathbb{R}^+})_x (t - \tau_x^\xi),$$

which is the number of $x \in \mathcal{U}$ such that $t - \tau_x^\xi \geq 0$, i.e., the total number of individuals already born up to time t. Another characteristic that we consider is, for $k \in \mathbb{N}$, $\Phi_k(t) = \mathbb{1}_{\{k\}}(\xi_t)$, for which

$$\boldsymbol{\xi}_t^{\Phi_k} = \sum_{x \in \mathcal{U}} \mathbb{1}_{\{k\}} \left(\xi^x_{t - \tau_x^\xi} \right)$$

is the number of individuals with k children at time t.

As known from the literature, the properties of the branching process are determined by the behavior of the point process ξ. First of all, we need to introduce some notation. Consider a function $f : \mathbb{R}^+ \to \mathbb{R}$. We denote the Laplace transform of f by

$$\mathcal{L}(f(\cdot))(\alpha) = \int_0^\infty \mathrm{e}^{-\alpha t} f(t) dt.$$

With a slight abuse of notation, if μ is a positive measure on $\mathbb{R}^+$, then we denote

$$\mathcal{L}(\mu(d\cdot))(\alpha) = \int_0^\infty \mathrm{e}^{-\alpha t} \mu(dt).$$

We use the Laplace transform to analyze the point process ξ:

Definition 2.1.9 (Supercritical property)**.** *Consider a point process ξ on $\mathbb{R}^+$. We say ξ is* supercritical *when there exists $\alpha^* > 0$ such that*

$$\mathcal{L}(\mathbb{E}\xi(d\cdot))(\alpha^*) = \int_0^\infty \mathrm{e}^{-\alpha^* t} \mathbb{E}\xi(dt) = \sum_{k \in \mathbb{N}} \mathbb{E}\left[\int_0^\infty \mathrm{e}^{-\alpha^* t} \delta_{T_k}(dt) \right] = \sum_{k \in \mathbb{N}} \mathbb{E}\left[\mathrm{e}^{-\alpha^* T_k} \right] = 1.$$

We call α^* the Malthusian parameter *of the process ξ.*

We point out that $\mathbb{E}\xi(d\cdot)$ is an abuse of notation to denote the density of the *averaged* measure $\mathbb{E}[\xi([0,t])] = \mathbb{E}[\xi_t]$. A second fundamental property for the analysis of branching processes is the following:

Definition 2.1.10 (Malthusian property)**.** *Consider a supercritical point process ξ, with Malthusian parameter α^*. The process ξ is* Malthusian *when*

$$-\frac{d}{d\alpha} \left(\mathcal{L}(\mathbb{E}\xi(dt)) \right)(\alpha) \Big|_{\alpha^*} = \int_0^\infty t \mathrm{e}^{-\alpha^* t} \mathbb{E}\xi(d\cdot) = \sum_{k \in \mathbb{N}} \mathbb{E}\left[T_k \mathrm{e}^{-\alpha^* T_k} \right] < \infty.$$

We denote

$$\tilde{\alpha} = \inf \left\{ \alpha > 0 \ : \ \mathcal{L}\left(\mathbb{E}\xi(d\cdot) \right)(\alpha) < \infty \right\}, \tag{2.1.1}$$

and we will also assume that the process satisfies the condition

$$\lim_{\alpha \searrow \bar{\alpha}} \mathcal{L}\left(\mathbb{E}\xi(d\cdot)\right)(\alpha) > 1. \tag{2.1.2}$$

Integrating by parts, it is possible to show that, for a point process ξ,

$$\mathcal{L}\left(\mathbb{E}\xi(d\cdot)\right)(\alpha) = \mathbb{E}\left[\xi_{T_\alpha}\right],$$

where T_α is an exponentially distributed random variable independent of the process $(\xi_t)_{t\geq 0}$. Heuristically, the Laplace transform of a point process ξ is the expected number of children born at exponentially distributed time T_α. In this case the Malthusian parameter is the exponential rate α^* such that at time T_{α^*} exactly one children has been born.

These two conditions are required to prove the main result on branching processes that we rely upon:

Theorem 2.1.11 (Population exponential growth)**.** *Consider the point process ξ, and the corresponding branching process $\boldsymbol{\xi}$. Assume that ξ is supercritical and Malthusian with parameter α^*, and suppose that there exists $\bar{\alpha} < \alpha^*$ such that*

$$\int_0^\infty \mathrm{e}^{-\bar{\alpha}t}\mathbb{E}\xi(dt) < \infty.$$

Then

(1) there exists a random variable Θ such that as $t \to \infty$,

$$\mathrm{e}^{-\alpha^* t}\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}} \overset{\mathbb{P}-a.s.}{\longrightarrow} \Theta; \tag{2.1.3}$$

(2) for any two random characteristics Φ and Ψ,

$$\frac{\boldsymbol{\xi}_t^{\Phi}}{\boldsymbol{\xi}_t^{\Psi}} \overset{\mathbb{P}-a.s.}{\longrightarrow} \frac{\mathcal{L}(\mathbb{E}[\Phi(\cdot)])(\alpha^*)}{\mathcal{L}(\mathbb{E}[\Psi(\cdot)])(\alpha^*)}. \tag{2.1.4}$$

This result is given by Nerman [129, Theorem 6.3], but the actual formulation in Theorem 2.1.11 is given in [152, Theorem A].

Formula (2.1.3) implies that, $\mathbb{P}$-a.s., the population size grows exponentially with time. It is relevant though to give a description of the distribution of the random variable Θ:

Theorem 2.1.12 (Positivity of Θ)**.** *Under the hypothesis of Theorem 2.1.11, if*

$$\mathbb{E}\left[\mathcal{L}(\xi(d\cdot))(\alpha^*)\log^+\left(\mathcal{L}(\xi(d\cdot))(\alpha^*)\right)\right] < \infty, \tag{2.1.5}$$

then, on the event $\{\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}} \to \infty\}$, i.e., on the event that the branching population keeps growing in time, the random variable Θ in (2.1.3) *is positive with probability 1, and $\mathbb{E}[\Theta] = 1$. Otherwise, $\Theta = 0$ with probability 1. Condition* (2.1.5) *is called the* (xlogx) *condition.*

This result is proven in [96, Theorem 5.3], and it is the CTBPs equivalent of the Kesten-Stigum theorem for Galton-Watson processes [104, Theorem 1.1].

Formula (2.1.4) says that the ratio between the evaluation of the branching process with two different characteristics converges $\mathbb{P}$-a.s. to a constant that depends only on the two characteristics involved. In particular, if we consider, for $k \in \mathbb{N}$,

$$\Phi(t) = \mathbb{1}_{\{k\}}(\xi_t), \quad \text{and} \quad \Psi(t) = \mathbb{1}_{\mathbb{R}^+}(t),$$

then Theorem 2.1.11 gives

$$\frac{\boldsymbol{\xi}_t^{\Phi}}{\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}} \overset{\mathbb{P}-a.s}{\longrightarrow} \alpha^* \mathcal{L}(\mathbb{P}(\xi(\cdot) = k))(\alpha^*) = \mathbb{P}(\xi_{T_{\alpha^*}} = k), \tag{2.1.6}$$

since $\mathcal{L}(\mathbb{E}[\mathbb{1}_{\mathbb{R}^+}(\cdot)])(\alpha^*) = 1/\alpha^*$. The ratio in the previous formula is the fraction of individuals with k children in the whole population:

Definition 2.1.13 (limiting degree distribution for CTBP)**.** *The sequence $(p_k)_{k\in\mathbb{N}}$, where*

$$p_k = \alpha^* \mathcal{L}(\mathbb{P}(\xi(\cdot) = k))(\alpha^*) = \alpha^* \int_0^\infty \mathrm{e}^{-\alpha^* t} \mathbb{P}(\xi(t) = k)\, dt$$

is the limiting degree distribution *for the branching process $\boldsymbol{\xi}$.*

2.1.2. Stationary birth processes with no fitness

In this section we present the theory of birth processes that are called *stationary* and have deterministic rates. In particular, we give description of the affine case, which plays a central role in the present work, due to its relation to PAMs:

Definition 2.1.14 (Stationary non-fitness birth processes)**.** *Consider a non-decreasing sequence $(f_k)_{k\in\mathbb{N}}$ of positive real numbers. A* stationary non-fitness birth process *is a stochastic process $(\zeta_t)_{t\geq 0}$ such that*

(1) $\zeta_0 = 0$, and $\zeta_t \in \mathbb{N}$ for all $t \in \mathbb{R}^+$;

(2) $\zeta_t \leq \zeta_s$ for every $t \leq s$;

(3) for h small enough,

$$\mathbb{P}(\zeta_{t+h} = k + j \mid \zeta_t = k) = \begin{cases} 1 - f_k h + o(h) & \text{for } j = 0, \\ f_k h + o(h) & \text{for } j = 1, \\ o(h^2) & \text{for } j \geq 2. \end{cases} \tag{2.1.7}$$

We denote the jump times by $(T_k)_{k\in\mathbb{N}}$, i.e.,

$$T_k = \inf\{t \geq 0 : \zeta_t \geq k\}.$$

Remark 2.1.15 (Embedding birth process)**.** Often, we refer to birth processes of the type as in Definition 2.1.14 as *embedding birth processes*. The reason is that stationary non-fitness birth processes can embed in continuous-time PA trees with PA function f (see [9, 10, 22, 152]). Also, we sometimes refer to f as PA function, and to $(f_k)_{k\in\mathbb{N}}$ as PA sequence, or simply the sequence, that defines $(\zeta_t)_{t\geq 0}$. Here $f_k = f(k)$.

With abuse of notation, we denote the point process corresponding to $(\zeta_t)_{t\geq 0}$ by ζ. In this case, $(\zeta_t)_{t\geq 0}$ is an inhomogeneous Poisson process, and for every $k \in \mathbb{N}$, $T_{k+1} - T_k$ has exponential law with parameter f_k independent of $(T_{h+1} - T_h)_{h=0}^{k-1}$. It is possible to show the following proposition:

Proposition 2.1.16 (Probabilities for $(\zeta_t)_{t\geq 0}$)**.** *Consider a stationary non-fitness birth process $(\zeta_t)_{t\geq 0}$. Denote, for every $k \in \mathbb{N}$, $\mathbb{P}(\zeta_t = k) = P_k[\zeta](t)$. Then*

$$P_0[\zeta](t) = \exp\left(-f_0 t\right), \tag{2.1.8}$$

and, for $k \geq 1$,

$$P_k[\zeta](t) = f_{k-1}\exp\left(-f_k t\right) \int_0^t \exp\left(f_k x\right) P_{k-1}[\zeta](x)dx. \tag{2.1.9}$$

For a proof, see [11, Chapter 3, Section 2]. From the jump times, it is easy to compute the explicit expression for the Laplace transform of ζ as

$$\mathcal{L}(\mathbb{E}\zeta(d\cdot))(\alpha) = \sum_{k\in\mathbb{N}} \mathbb{E}\left[\int_0^\infty \mathrm{e}^{-\alpha t}\delta_{T_k}(dt)\right] = \sum_{k\in\mathbb{N}} \mathbb{E}\left[\mathrm{e}^{-\alpha T_k}\right] = \sum_{k\in\mathbb{N}}\prod_{i=0}^{k-1}\frac{f_i}{\alpha + f_i},$$

since every T_k can be seen as a sum of independent exponential random variables with parameters given by the sequence $(f_k)_{k\in\mathbb{N}}$. Assuming now that ζ is supercritical and Malthusian with parameter α^*, we have the explicit expression for the limit distribution $(p_k)_{k\in\mathbb{N}}$, that is

$$p_k^{(1)} = \frac{\alpha^*}{\alpha^* + f_k}\prod_{i=0}^{k-1}\frac{f_i}{\alpha^* + f_i}. \tag{2.1.10}$$

An analysis of the behavior of the limit distribution of branching processes is presented in [9, 151], where the authors prove that $(p_k)_{k\in\mathbb{N}}$ has a power-law tail only if the sequence of rates $(f_k)_{k\in\mathbb{N}}$ is asymptotically linear with respect to k.

Proposition 2.1.17 (Characterization of stationary and linear process ζ)**.** *Consider the sequence $f_k = ak + b$. Then:*

(1) for every $\alpha \in \mathbb{R}^+$,

$$\mathcal{L}(\mathbb{E}\zeta(d\cdot))(\alpha) = \frac{\Gamma(\alpha^*/a + b/a)}{\Gamma(b/a)}\sum_{k\in\mathbb{N}}\frac{\Gamma(k + b/a)}{\Gamma(k + b/a + \alpha/a)} = \frac{b}{\alpha - a}.$$

(2) *The Malthusian parameter is* $\alpha^* = a + b$, *and* $\tilde{\alpha} = a$, *where* $\tilde{\alpha}$ *is defined as in* (2.1.1).

(3) *The derivative of the Laplace transform is*

$$-\frac{b}{(\alpha - a)^2},$$

which is finite whenever $\alpha > a$;

(4) *The process* $(\zeta_t)_{t\geq 0}$ *satisfies the* (xlogx) *condition* (2.1.5).

Proof. The proof can be found in [152, Theorem 2], or [9, Theorem 2.6]. □

For affine PA weights $(f_k)_{k\in\mathbb{N}} = (ak + b)_{k\in\mathbb{N}}$, the Malthusian parameter α^* exists. Since $\alpha^* = a + b$, the limiting degree distribution of the branching process $\boldsymbol{\zeta}$ is given by

$$p_k = (1 + b/a)\frac{\Gamma(1 + 2b/a)}{\Gamma(b/a)}\frac{\Gamma(k + b/a)}{\Gamma(k + b/a + 2 + b/a)}. \tag{2.1.11}$$

Notice that p_k has a power-law decay with exponent $\tau = 2 + \frac{b}{a}$.

2.2. Collapsed branching processes (CBPs)

The main result of this chapter is the definition of multigraphs from continuous-time branching processes (CTBP), through a procedure that we call *collapsing*. We analyze the case where we collapse a fixed number $m \in \mathbb{N}$ of individuals. The heuristic idea is to consider the tree defined by the branching process, and collapse or merge together m different nodes in the tree to create a vertex in the multigraph. Throughout this chapter, we will consider an *individual* to be a node in the tree of the branching process, while a *vertex* is a node in the multigraph obtained by collapsing.

Fix $m \in \mathbb{N}$. We denote $(n, j) = m(n - 1) + j$, for $j = 1, \dots, m$. We now give the precise definition of the collapsed branching process:

Definition 2.2.1 (Collapsed branching process)**.** *Consider a branching process* $\boldsymbol{\xi}$. *Then, a* collapsed branching process *is a random process* $(\mathrm{CBP}_t^{(m)})_{t\geq 0}$, *for which, for every* $t \geq 0$, $\mathrm{CBP}_t^{(m)}$ *is a directed multigraph with adjacency matrix* $(g_{x,y}(t))_{x,y\in\mathbb{N}}$, *where*

$$g_{x,y}(t) = \sum_{j=1}^{m} \mathbb{1}_{\{(x,j)\to(y,1),\dots,(y,m)\}}\, \mathbb{1}_{\mathbb{R}^+}(t - \tau_{(x,j)}), \tag{2.2.1}$$

and $\{(x, j) \to (y, 1), \cdots, (y, m)\}$ *is the event that there is a directed edge between individual* (x, j) *and one of the individuals* $(y, 1), \dots, (y, m)$ *in the tree defined by the branching process at time* t. *We denote the size of* $\mathrm{CBP}_t^{(m)}$ *by* $N^{(m)}(t)$.

As the reader can see from the definition, the collapsing procedure combines m individuals together with their edges to create a vertex, and there is an edge between two vertices if and only if there is an edge between a pair of individuals collapsed to create the two vertices. $\mathrm{CBP}_t^{(m)}$ is a graph where every vertex (except vertex 1)

has out-degree m. Self-loops and multiple edges are allowed (see Figure 1.7 for an example of CBP).

We consider the birth time of the vertex n in the multigraph to be $\tau_{(n,1)} = \tau_{m(n-1)+1}$. Thus, vertex n in $\mathrm{CBP}^{(m)}$ is considered alive when $(n,1)$ is alive in $\boldsymbol{\xi}$. Notice that when n is born, it has only one out-edge, because the other individuals $(n,2),\dots,(n,m)$ are not yet alive. Clearly, the in-degree at time t of a vertex n in $\mathrm{CBP}^{(m)}$ is given by

$$D_n^{(\mathrm{in})}(t) = \sum_{j=1}^{m} \xi^{(n,j)}_{t-\tau_{(n,j)}}.$$

The main difference between CBPs and PAMs is that CBPs are defined in continuous-time, while time in PAMs is discrete. Heuristically, discrete time in PAMs is described as the time unit at which a new vertex is added to the graph (see for instance [3, 33], [85, Chapter 8]), while in CBPs time is continuous and new vertices are born at exponential rate (recall Theorem 2.1.11).

Our results are a first attempt to create a link between trees and multigraphs in continuous time. The collapsing procedure creates difficulties though. For instance, we consider different individuals to create a vertex, each one of them having its own birth time. This has to be taken into account to investigate the degree evolution of a vertex in CBP.

Here we state the result on the limiting degree distribution of CBPs, relying on properties of CTBPs as formulated in Theorem 2.3.2 below:

Theorem 2.2.2 (Limiting degree distribution of CBPs)**.** *Consider a branching process $\boldsymbol{\xi}$, and fix $m \in \mathbb{N}$. Denote the size of $\mathrm{CBP}^{(m)}_t$ by $N^{(m)}(t)$ and the number of vertices with degree k by $N^{(m)}_k(t)$. Under the hypotheses of Theorem 2.3.2, as $t \to \infty$,*

$$\frac{N^{(m)}_k(t)}{N^{(m)}(t)} \xrightarrow{\mathbb{P}} p^{(m)}_k = \mathbb{P}\left(\xi^1_{T_{\alpha^*}} + \cdots + \xi^m_{T_{\alpha^*}} = k\right), \tag{2.2.2}$$

where $(\xi^1_t)_{t\geq 0},\dots,(\xi^m_t)_{t\geq 0}$ are m independent copies of the birth process $(\xi_t)_{t\geq 0}$, α^* is the Malthusian parameter of $\boldsymbol{\xi}$, and T_{α^*} is an exponentially distributed random variable with parameter α^*.

The hypotheses of Theorem 2.3.2 are technical, and they are deferred to later. Theorem 2.2.2 is part of Theorem 2.3.2.

2.2.1. Embedding PAMs

In discrete time, PAMs are defined by the PA function f:

$$\mathbb{P}\left(n \overset{j+1}{\to} i \mid G_{(n,j)}\right) = \frac{f(D_i(n,j))}{\sum_{h=1}^{n} f(D_h(n,j))}, \tag{2.2.3}$$

where $D_i(n,j)$ denotes the degree of the vertex i in $\mathcal{G}_{(n,j)}$. When f is affine, it is possible to define the model with out-degree $m \geq 2$ from the tree case where the

out-degree is 1 (we refer to [85, Chapter 8, Section 8.2] for the precise definition). In particular, the collapsing procedure we have introduced in Definition 2.2.1 mimics the construction of PAMs with affine attachment function.

Several works in the literature [9, 10, 22, 152] use CTBPs to investigate the degree distribution of PA trees. In particular, embedding theorems are proved between discrete and continuous time (see [9, Theorem 3.3], [10, Theorem 2.1]). These results are based on the fact that all intervals between two jumps in every copy of the birth process $(\xi_t)_{t\geq 0}$ are exponentially distributed. This means that, conditionally on the present state of the tree, the probability that a new vertex is attached to the i-th vertex already present is just the ratio between the PA function of the degree of vertex i and the total weight of the tree. Also PAMs with out-degree $m \geq 2$ have been investigated, but not through embedding of CTBPs.

PAMs are embedded in continuous time by the processes in Definition 2.1.14, and this construction in used in [9, 10, 152]. It allows to embed PA trees in continuous time where the PA function is given by f. Embedding birth processes allow us to describe PAMs with out-degree $m \geq 2$ and affine f using CBPs. In fact, an immediate application of [9, Theorem 3.3] and [10, Theorem 2.1] is enough to prove that the transition probability in CBP from $\mathrm{CBP}^{(m)}_{\tau_{(n,j)}}$ to $\mathrm{CBP}^{(m)}_{\tau_{(n,j+1)}}$ are exactly given by (2.2.3), with the restriction that the first edge of every vertex cannot be a self-loop. In particular, this yields the following result:

Corollary 2.2.3 (Continuous-time PAM)**.** *Fix $m \geq 2$ and $\delta > -m$. Let $(\xi_t)_{k\in\mathbb{N}}$ be an embedding birth process defined by the sequence $(k+1+\delta/m)_{k\in\mathbb{N}}$. Then, the corresponding CBP embeds the PAM in continuous time with PA function $f(k) = k + \delta$, and satisfies Theorem 2.2.2 (and Theorem 2.3.2). As a consequence, the limiting degree distribution is given by* (1.3.1).

Corollary 2.2.3 is the application of Theorem 2.2.2 to the case of the CTBPs that embed PAMs in continuous time. Indeed, the CBP observed at times $(\tau_n)_{n\in\mathbb{N}}$ (the sequence of birth times of the CTBP) corresponds to the discrete-time PAM. However, since the ratio $N^{(m)}_k(t)/N^{(m)}(t)$ converges in probability, Theorem 2.2.2 does not imply the convergence along the sequence $(\tau_n)_{n\in\mathbb{N}}$. To prove that the convergence holds also in discrete time, a more detailed analysis is necessary, therefore we state it as a separate result:

Theorem 2.2.4 (Discrete-time PAMs)**.** *Fix $m \geq 2$ and $\delta > -m$. Let $(\xi_t)_{k\in\mathbb{N}}$ be an embedding birth process defined by the sequence $(k+1+\delta/m)_{k\in\mathbb{N}}$. Consider the corresponding discrete-time PAM defined as $\mathrm{PA}_{n,j}(m,\delta) = \mathrm{CBP}^{(m)}_{\tau_{(n,j)}}$, for $n \in \mathbb{N}$ and $j \in [m]$. Then, for every $k \in \mathbb{N}$, the fraction of vertices with degree k in $\mathrm{PA}_{n,j}(m,\delta)$ converges in probability to $p^{(m)}_k$ as in* (1.3.1).

While CTBP arguments have been used a lot in the context of PA trees (for which $m = 1$), Theorem 2.2.4 provides the first example where they are applied beyond the tree setting. Thus, our results offer the opportunity to use the powerful CTBP tools in order to study PAMs.

The PAM version that is embedded as in Theorem 2.2.4 is a modification of [85, Model (b)], where the starting graph PA_2 is a graph with a single edge between them (instead of 2 as in [85, Model (b)]). In this case, a new vertex appears with m edges that are attached sequentially with degrees update, but the *first* edge is not allowed to be a self loop, thus creating a connected graph.

To show the universality of our collapsing construction, we apply Theorem 2.2.2 to another classical random graph model. A random recursive tree (RRT) is a sequence of PA trees where the attachment function f is equal to one. At every step, a vertex is added to the tree and attached uniformly to one existing vertex (see [88, 97] for an introduction). We also consider a graph version of the RRT. In this case we obtain the following result, which could be interpreted as the $\delta = \infty$ version of Theorem 2.2.4, and grows a graph by uniform attachments:

Corollary 2.2.5 (Random recursive graph)**.** *Fix $m \geq 2$. Let $(\xi_t)_{k\in\mathbb{N}}$ be an embedding birth process defined by the sequence $\lambda_k = 1$ for every $k \in \mathbb{N}$. Then, the corresponding CBP defines a sequence of random graphs which transition probabilities are given by*

$$\mathbb{P}\left(n \overset{j+1}{\to} i \mid \mathrm{CBP}^{(m)}_{\tau_{(n,j)}}\right) = \begin{cases} \frac{1}{(n-1)+j/m} & \text{if } i \neq n, \\ \\ \frac{j/m}{(n-1)+j/m} & \text{if } i = n. \end{cases} \tag{2.2.4}$$

We call the sequence of random graphs defined by (2.2.4) random recursive graph*. As a consequence, the limiting degree distribution is given by*

$$p_k^{(m)} = \frac{1}{m+1}\left(1 + \frac{1}{m}\right)^{-k}. \tag{2.2.5}$$

Consequently, the same result also holds in discrete time.

In this case the CBP can be seen as the generalization of the RRT to the case where the out-degree is $m \geq 2$. In particular, when $m = 1$ the distribution in (2.2.5) reduces to $p_k^{(1)} = 2^{-(k+1)}$, which is the known limiting degree distribution for the RRT [97].

2.3. Growth and scaling property of CBPs

Our main result requires the following condition:

Condition 2.3.1 (Lipschitz)**.** *Assume that a birth process $(\xi_t)_{\geq 0}$ is supercritical and Malthusian. The* Lipschitz *condition is that, for every $k \in \mathbb{N}$, there exists a constant $0 < \ell(k) < \infty$ such that the function $P_k[\xi](t) = \mathbb{P}(\xi_t = k)$ is Lipschitz with constant $\ell(k)$.*

Condition 2.3.1 requires that the functions $(P_k[\xi](t))_{k\in\mathbb{N}}$ associated to the birth process $(\xi_t)_{t\geq 0}$ are smooth, in the sense that they do not have dramatic changes over time. We can now state the main result of the chapter:

Theorem 2.3.2 (Growth and convergence of CBPs)**.** *Let $(\xi_t)_{t\geq 0}$ be a supercritical and Malthusian birth process that satisfies Condition 2.3.1. Let $(\mathrm{CBP}^{(m)}_t)_{t\geq 0}$ be the corresponding collapsed branching process. Let Θ and μ be as in Theorem 2.1.11. Denote the size of*

$\mathrm{CBP}_t^{(m)}$ *by* $N^{(m)}(t)$*, and the number of vertices with degree* k *by* $N_k^{(m)}(t)$*. Then, as* $t \to \infty$*,*

$$m\mathrm{e}^{-\alpha^* t} N^{(m)}(t) \xrightarrow{\mathbb{P}-a.s.} \frac{1}{\mu\alpha^*}\Theta. \tag{2.3.1}$$

Further, for every $k \in \mathbb{N}$*, there exists* $p_k^{(m)}$ *such that,*

$$m\mathrm{e}^{-\alpha^* t} N_k^{(m)}(t) \xrightarrow{\mathbb{P}} \frac{1}{\mu\alpha^*} p_k^{(m)}\Theta. \tag{2.3.2}$$

As a consequence,

$$\frac{N_k^{(m)}(t)}{N^{(m)}(t)} \xrightarrow{\mathbb{P}} p_k^{(m)}. \tag{2.3.3}$$

The sequence $(p_k^{(m)})_{k\in\mathbb{N}}$ is called the *limiting degree distribution* of $(\mathrm{CBP}_t^{(m)})_{t\geq 0}$, and is given by

$$p_k^{(m)} = \alpha^* \mathcal{L}\left(P[\xi](\cdot)_k^{*m}\right)(\alpha^*) = \mathbb{E}\left[P[\xi](T_{\alpha^*})_k^{*m}\right], \tag{2.3.4}$$

where $P_k[\xi](t) = \mathbb{P}(\xi_t = k)$, T_{α^*} is an exponentially distributed random variable with parameter α^*, and

$$P[\xi](t)_k^{*m} = \sum_{k_1+\cdots+k_m=k} P_{k_1}[\xi](t)\cdots P_{k_m}[\xi](t) \tag{2.3.5}$$

is the k-th element of the m-fold convolution of the sequence $(P_k[\xi](t))_{k\in\mathbb{N}}$.

We now comment on Theorem 2.3.2 (for comparison with CTBPs, we refer to Theorem 2.1.11). Equation (2.3.1) assures us that the size of a CBP grows at exponential rate α^* as for the underlying CTBP. Even the size of $\mathrm{CBP}_t^{(m)}$, up to the constant m, scales exactly as the size of the CTBP, and the limiting random variable Θ is the same. This means that the collapsing procedure does not destroy the exponential growth of the graph.

(2.3.2) assures that, for every $k \in \mathbb{N}$, the number of vertices with in-degree k scales exponentially and also in this case we have a limiting random variable. (2.3.3) tells us that there exists a *deterministic* limiting degree distribution for a CBP.

The expression for $(p_k^{(m)})_{k\in\mathbb{N}}$ can be explained in terms of CTBPs. In fact, for a CTBP $\boldsymbol{\xi}$, the limiting degree distribution is given by $p_k^{(1)} = \mathbb{E}\left[P_k[\xi](T_{\alpha^*})\right]$, with α^* the Malthusian parameter of $\boldsymbol{\xi}$. We can see T_{α^*} as a *time unit* that a process $(\xi_t)_{t\geq 0}$ takes to generate, on average, 1 individual. Then, $p_k^{(1)}$ can be seen as the probability that $(\xi_t)_{t\geq 0}$ generates k individuals instead of the average 1. Using the same heuristic, the limiting degree distribution of CBP can be seen as the probability that m different individuals produce k children in total in the time unit T_{α^*}. Notice that in the expression of $(p_k^{(m)})_{k\in\mathbb{N}}$ the Malthusian parameter α^* is that of the branching process $\boldsymbol{\xi}$.

Unfortunately, the size of CBP and the number of vertices with degree $k \in \mathbb{N}$ are not the evaluation of a CTBP with a random characteristic as in Definition 2.1.7. For example the degree of a vertex in CBP is the sum of the degrees of m different

individuals. The solution for the size of CBP and the number of vertices with degree k is different. From Definition 2.2.1, it is obvious that

$$N^{(m)}(t) = \left\lceil \frac{\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}}{m} \right\rceil . \tag{2.3.6}$$

Using then (2.1.3), the proof of (2.3.1) is immediate.

The proof of (2.3.2) is harder, and it requires a conditional second moment method on $N_k^{(m)}(t)$. Before stating the result, we need a preliminary discussion. We use artificial randomness that we add to the branching process to rewrite the degree of a vertex in CBP in terms of a random characteristic. In the population space in the definition of CTBPs, we consider a single birth process $(\xi_t^x)_{t\geq 0}$ for every individual x in the population. We instead consider on every Ω_x a vector of birth processes $(\xi_t^{x,1}, \dots, \xi_t^{x,m})$, where $\xi_t^{x,1}, \cdots, \xi_t^{x,m}$ are i.i.d. copies of the birth process, defined on the space corresponding to the individual x. With this notation, the standard branching processes defined by $(\xi_t)_{t\geq 0}$ is the branching process where we consider as birth process the first component of every vector associated to every individual.

Now, for $k \in \mathbb{N}$, we consider the random characteristic

$$\Phi_k^{(m)}(t) = \mathbb{1}_{\{k\}} \left(\xi_{t-\tau_x}^{x,1} + \cdots + \xi_{t-\tau_x}^{x,m} \right), \tag{2.3.7}$$

which corresponds to the event that the sum of the components of the vector associated to the individual x when its age is $t - \tau_x$ is equal to k. This is a random characteristic that depends only on the randomness defined on the space Ω_x.

The crucial observation is that

$$\begin{aligned} \mathbb{P}\left(D_n^{(\mathrm{in})}(t) = k\right) &= \mathbb{P}\left(\xi_{t-\tau_{(n,1)}}^{(n,1)} + \cdots + \xi_{t-\tau_{(n,m)}}^{(n,m)} = k \right) \\ &\approx \frac{1}{m} \sum_{j=1}^{m} \mathbb{P}\left(\xi_{t-\tau_{(n,j)}}^{(n,j),1} + \cdots + \xi_{t-\tau_{(n,j)}}^{(n,j),m} = k \right) \\ &= \frac{1}{m} \sum_{j=1}^{m} \mathbb{E}\left[\Phi_k^{(m)}(t - \tau_{(n,j)}) \right] + (\text{error}), \end{aligned} \tag{2.3.8}$$

when we assume that the difference between the birth times $\tau_{(n,1)}, \tau_{(n,2)}, \dots, \tau_{(n,m)}$ is very small. The approximation in (2.3.8) can be explained by the fact that all the components of the vectors $(\xi_t^{n,1}, \dots, \xi_t^{n,m})$ are i.i.d. and $\tau_{(n,1)} \approx \tau_{(n,m)}$. In fact, on the left side of (2.3.8) we have the probability that the sum of m independent copies of $(\xi_t)_{t\geq 0}$, evaluated at different times, is equal to k. Assuming that the differences between the birth times $\tau_{(n,1)}, \tau_{(n,2)}, \dots, \tau_{(n,m)}$ are small, we can just evaluate the m different processes at time $\tau_{(n,1)}$, with a negligible error.

The proof of this, based on Condition 2.3.1, is given in Proposition 2.4.3. It gives the bound on the error term with the difference between the birth times of the indi-

viduals collapsed to generate the vertex, i.e., the error term is bounded by

$$m\ell|\tau_{(n,m)} - \tau_{(n,1)}|, \quad \text{where} \quad \ell = \max_{i\in[k]}\{\ell(i)\}.$$

The use of artificial additional randomness might not seem intuitive. The point is that the equality in expectation between the random characteristic $\Phi_k^{(m)}(t - \tau_{(n,1)})$ and $D_n^{(\mathrm{in})}(t)$ is enough. This relies on the fact that, conditionally on the first stages of the branching process, the contribution to the number of vertices with degree k given by the latter individuals is almost deterministic. Let us formalize this idea:

Definition 2.3.3 (x-bulk filtration)**.** *Consider a branching process $\boldsymbol{\xi}$, and its natural filtration $(\mathcal{F}_t)_{t\geq 0}$. Consider an increasing function $x(t) : \mathbb{R}^+ \to \mathbb{R}^+$. We call $(\mathcal{F}_{x(t)})_{t\geq 0}$ the* x-bulk filtration *of $\boldsymbol{\xi}$. At every time $t \geq 0$, a random variable measurable with respect to $\mathcal{F}_{x(t)}$ is called* x-bulk measurable*.*

If we consider $x(t)$ to be $o(t)$, then the x-bulk filtration heuristically contains information only on the early stage of the CTBP. Nevertheless, the information contained in $\mathcal{F}_{x(t)}$ is enough to estimate the behavior of the CTBP:

Proposition 2.3.4 (Conditional moments of $N_k^{(m)}(t)$)**.** *Assume that x is a monotonic function such that, as $t \to \infty$, $x(t) \to \infty$ and $x(t) = o(t)$. Then, under the conditions of Theorem 2.3.2, as $t \to \infty$,*

(1)

$$m\mathrm{e}^{-\alpha^* t}\mathbb{E}\left[N_k^{(m)}(t)\middle|\mathcal{F}_{x(t)}\right] \overset{\mathbb{P}-a.s.}{\longrightarrow} \frac{1}{\mu}\mathcal{L}\left(\Phi_k^{(m)}(\cdot)\right)(\alpha^*)\Theta; \tag{2.3.9}$$

(2)

$$\mathrm{e}^{-2\alpha^* t}\mathbb{E}\left[N_k^{(m)}(t)^2\middle|\mathcal{F}_{x(t)}\right] \overset{\mathbb{P}-a.s.}{\leq} \left(\mathrm{e}^{-\alpha^* t}\mathbb{E}\left[N_k^{(m)}(t)\middle|\mathcal{F}_{x(t)}\right]\right)^2 + o(1). \tag{2.3.10}$$

We point out that if $X \leq Y + o(1)$, then $o(1)$ is a term that converges almost surely to 0. The proof of Proposition 2.3.4 is moved to Section 2.5. With Proposition 2.3.4 in hand, we can prove (2.3.2). We bound $\left|m\mathrm{e}^{-\alpha^* t}N_k^{(m)}(t) - \frac{1}{\mu}\mathcal{L}(\Phi_k^{(m)}(\cdot))(\alpha^*)\Theta\right|$ by

$$\begin{aligned}&\left|m\mathrm{e}^{-\alpha^* t}N_k^{(m)}(t) - m\mathrm{e}^{-\alpha^* t}\mathbb{E}[N_k^{(m)}(t)|\mathcal{F}_{x(t)}]\right| \\ &\quad + \left|m\mathrm{e}^{-\alpha^* t}\mathbb{E}[N_k^{(m)}(t)|\mathcal{F}_{x(t)}] - \frac{1}{\mu}\mathcal{L}(\Phi_k^{(m)}(\cdot))(\alpha^*)\Theta\right|.\end{aligned} \tag{2.3.11}$$

As a consequence, (2.3.2) holds if both terms in (2.3.11) converges $\mathbb{P}$-a.s. to zero. For the second term this is true by (2.3.9). For the first term, we use (2.3.9) and (2.3.10) to conclude that $\mathrm{Var}\left(m\mathrm{e}^{-\alpha^* t}N_k^{(m)}(t)|\mathcal{F}_{x(t)}\right) = o_{a.s.}(1)$, so that

$$\left|m\mathrm{e}^{-\alpha^* t}N_k^{(m)}(t) - m\mathrm{e}^{-\alpha^* t}\mathbb{E}[N_k^{(m)}(t)|\mathcal{F}_{x(t)}]\right| \overset{\mathbb{P}}{\longrightarrow} 0. \tag{2.3.12}$$

This concludes the proof of (2.3.2). (2.3.3) follows immediately.

Remark 2.3.5 (Times and bulk sigma-field)**.** We have proved Proposition 2.3.4 (and thus Theorem 2.3.2) by looking at the CTBP at time t, considering the $x(t)$-bulk sigma-field. We can extend the argument as follows. Consider $s \geq 0$, and let $y\colon \mathbb{R}^+ \to \mathbb{R}^+$ be a monotonic function of s such that $y(s)/s \to \infty$ as $s \to \infty$. In this case, looking at the graph at time $y(s)$, and considering the s-bulk sigma-field, Proposition 2.3.4 still holds. More generally, as suggested by (2.5.1) below, conditionally on the s-bulk sigma-field, the evolution of a CTBP is almost deterministic. This implies that Proposition 2.3.4 even holds when we consider a *random* process $Y(s)$ such that $Y(s)/s \overset{a.s.}{\to} \infty$, under the assumption that $Y(s)$ is s-bulk measurable for every $s \geq 0$. These observations will be useful when extending our results to discrete time in Section 2.6.

2.4. Preliminaries on birth times

2.4.1. Bound on the difference in time

In this section, we prove the fact that the error term in (2.3.8) can be bounded by the difference of birth times of the considered individuals. We introduce the definition of convolution, as well as the bound we are interested in:

Definition 2.4.1 (Convolution)**.** *We define convolution between two sequences $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ as*

$$(a * b)_k := \sum_{l=0}^{k} a_l b_{k-l}. \tag{2.4.1}$$

Lemma 2.4.2 (Difference in times)**.** *Consider the sequence of functions $(P[\xi]_k(t))_{k\in\mathbb{N}}$. If $(\xi_t)_{t\geq 0}$ satisfies Condition 2.3.1, then, for every $x \in \mathbb{R}^+$, and for every $h_i \leq x$ for $i \in [m]$,*

$$|(P[\xi](x-h_1) * \cdots * P[\xi](x-h_m))_k - (P[\xi](x-h_1)^{*m})_k| \leq \ell \sum_{j=2}^{m} |h_1 - h_j|, \tag{2.4.2}$$

where $\ell = \max_{i\in[k]} \ell(i)$.

Proof. Without loss of generality, assume $0 \leq h_1 \leq \ldots \leq h_m$. We prove Lemma 2.4.2 by induction on m. We start the induction with $m = 2$, so

$$(P[\xi](x-h_1) * P[\xi](t-h_2))_k = \sum_{l=0}^{k} P[\xi]_l(x-h_1) P[\xi]_{k-l}(x-h_2). \tag{2.4.3}$$

We now use Condition 2.3.1 to bound $|P[\xi]_{k-l}(x-h_2) - P[\xi]_{k-l}(x-h_1)| \leq \ell(k -$

$l)(h_2 - h_1)$. Using this in (2.4.3), then we obtain, for $\ell = \max_{i\in[k]} \ell(i)$,

$$\begin{aligned}\Big|\big(P[\xi](x-h_1) * P[\xi](t-h_2)\big)_k - \big(P[\xi](x-h_1)^{*2}\big)_k\Big| \\ \leq \ell \sum_{l=0}^{k} P[\xi]_{k-l}(x-h_1)\,|(h_2-h_1)|\,.\end{aligned} \tag{2.4.4}$$

Since $\sum_{l=0}^{k} P_l[\xi](x-h_1) = P[\xi]_{\leq k}(x-h_1) \leq 1$,

$$\big|(P[\xi](x-h_1) * P[\xi](t-h_2))_k - \big(P[\xi](x-h_1)^{*2}\big)_k\big| \leq \ell|h_2-h_1|,$$

so (2.4.2) holds for $m = 2$. We now advance the induction hypothesis, so suppose that (2.4.2) holds for $m-1$. We can write

$$\begin{aligned}&\big(P[\xi](x-h_1) * \cdots * P[\xi](x-h_m)\big)_k \\ &\qquad = \sum_{l=0}^{k} \left(P[\xi](x-h_1) * \cdots * P[\xi](x-h_{m-1})\right)_l P[\xi]_{k-l}(x-h_m).\end{aligned} \tag{2.4.5}$$

Notice that we can apply (2.4.2) to the first terms in the sum in (2.4.5) thanks to the induction hypothesis, since it is now the convolution of $m-1$ functions. We just need to replace $P[\xi]_{k-l}(x-h_m)$ by $P[\xi]_{k-l}(x-h_1)$. It is easy to do this using a similar argument used to prove the bound in (2.4.4), which implies again the use of Condition 2.3.1. In the end, we have

$$\begin{aligned}\Big|\big(P[\xi](x-h_1) * \cdots * P[\xi](x-h_m)\big)_k - \left(P[\xi](x-h_1)^{*m}\right)_k\Big| \\ \leq \ell \sum_{j=2}^{m-1} |h_1-h_j| + \ell|h_m-h_1|,\end{aligned}$$

where the $m-1$ terms comes from the induction hypothesis, and the last one from the approximation of $P[\xi]_{k-l}(x-h_m)$. This completes the proof. □

Lemma 2.4.2 holds for every time x and $h_1, \ldots, h_m$ that we consider. We can now prove the bound on the error term in (2.3.8):

Proposition 2.4.3 (Approximation at fixed time)**.** *Consider* $(\mathrm{CBP}_t^{(m)})_{t\geq 0}$ *obtained from a branching process* $\boldsymbol{\xi}$*. Assume that* $(\xi_t)_{t\geq 0}$ *satisfies Condition 2.3.1. Then, for every* $k \in \mathbb{N}$*, with* ℓ *as in Lemma 2.4.2,* $\mathbb{P}$*-a.s. for every* $n \in \mathbb{N}$*,*

$$\left|\mathbb{P}\left(D_n^{(\mathrm{in})}(t) = k \mid \tau_{(n,1)}, \ldots, \tau_{(n,m)}\right) - \left(P[\xi](t-\tau_{(n,1)})^{*m}\right)_k\right| \leq \ell m|\tau_{(n,m)} - \tau_{(n,1)}|.$$

Proof. Conditionally on the birth times, the processes $(\xi_t^{(n,1)})_{t\geq 0}, \ldots, (\xi_t^{(n,m)})_{t\geq 0}$ are

independent. As a consequence,

$$\begin{aligned}\mathbb{P}\left(D_n^{(\mathrm{in})}(t)=k \mid \tau_{(n,1)},\dots,\tau_{(n,m)}\right)\\ =\left(P[\xi](t-\tau_{(n,1)})*\cdots*P[\xi](t-\tau_{(n,m)})\right)_k.\end{aligned} \tag{2.4.6}$$

Then the statement follows immediately from Lemma 2.4.2, where we consider $h_1 = \tau_{(n,1)},\dots,h_m = \tau_{(n,m)}$, and the fact that $\tau_{(n,j)} - \tau_{(n,1)} \le \tau_{(n,m)} - \tau_{(n,1)}$ for every $j = 1,\dots,m$. □

2.4.2. Replacing birth times with $\mathcal{F}_t$-measurable approximations

Recall that $\mathcal{F}_t$ denotes the natural filtration of the CTBP up to time t. It is possible to rewrite (2.1.3) as

$$n\mathrm{e}^{-\alpha^*\tau_n} \overset{\mathbb{P}-a.s.}{\longrightarrow} \frac{1}{\mu\alpha^*}\Theta.$$

As a consequence, as $n\to\infty$,

$$-\tau_n + \frac{1}{\alpha^*}\log n \overset{\mathbb{P}-a.s.}{\longrightarrow} \frac{1}{\alpha^*}\log\left(\frac{1}{\mu\alpha^*}\Theta\right). \tag{2.4.7}$$

Notice that on the event $\{\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}\to\infty\}$, Θ is positive with probability 1, so $\log\left(\frac{1}{\mu\alpha^*}\Theta\right)$ is well defined. Define, for $n \ge \boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}$,

$$\sigma_n(t) := \frac{1}{\alpha^*}\log n - \frac{1}{\alpha^*}\log\left(\frac{1}{\mu\alpha^*}\Theta_t\right), \quad \text{where} \quad \Theta_t = \mu\alpha^*\mathrm{e}^{-\alpha^* t}\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}. \tag{2.4.8}$$

Then $\sigma_n(t)$ is an approximation of τ_n given the information up to time t, where the factor Θ_t includes the stochastic fluctuation of the size of the branching process. What is interesting is that the random variable $\sigma_n(t)$ is an approximation of τ_n *measurable with respect to* $\mathcal{F}_t$. We now prove that $(\sigma_n(t))_{t\ge0}$ is an acceptable approximation of τ_n:

Lemma 2.4.4 (Error of $(\sigma_n(t))_{t\ge0}$)**.** *$\mathbb{P}$-a.s., as $t\to\infty$,*

$$\sup_{n\ge\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}} |\sigma_n(t)-\tau_n| \to 0. \tag{2.4.9}$$

Proof. For every $t\ge0$ and $n\ge\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}$ we write

$$\begin{aligned}|\sigma_n(t)-\tau_n| \le &\left|\frac{1}{\alpha^*}\log n - \tau_n - \frac{1}{\alpha^*}\log\left(\frac{1}{\mu\alpha^*}\Theta\right)\right|\\ &+\frac{1}{\alpha^*}\left|\log\left(\frac{1}{\mu\alpha^*}\Theta\right)-\log\left(\frac{1}{\mu\alpha^*}\Theta_t\right)\right|.\end{aligned} \tag{2.4.10}$$

Using (2.4.10) in (2.4.9), we can bound

$$\begin{aligned}\sup_{n\geq \boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}} |\sigma_n(t)-\tau_n| \leq & \frac{1}{\alpha^*}\left|\log\left(\frac{1}{\mu\alpha^*}\Theta\right)-\log\left(\frac{1}{\mu\alpha^*}\Theta_t\right)\right| \\ & + \sup_{n\geq \boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}}}\left|\frac{1}{\alpha^*}\log n-\tau_n-\frac{1}{\alpha^*}\log\left(\frac{1}{\mu\alpha^*}\Theta\right)\right|.\end{aligned} \tag{2.4.11}$$

First of all, from (2.1.3) we know $\Theta_t/(\mu\alpha^*) = \mathrm{e}^{-\alpha^* t}\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}} \to \Theta/(\mu\alpha^*)$. As a consequence, the first term in the right hand side of (2.4.11) converges $\mathbb{P}$-a.s. to zero. For the second term, we use (2.4.7) and the fact that the supremum decreases as $\boldsymbol{\xi}_t^{\mathbb{1}_{\mathbb{R}^+}} \to \infty$. This completes the proof. □

Lemma 2.4.4 suggests that, conditionally on $\mathcal{F}_t$, we can replace the birth sequence $(\tau_n)_{n\geq \boldsymbol{\xi}_t^{\mathbb{1}}}$ with the sequence $(\sigma_n(t))_{n\geq \boldsymbol{\xi}_t^{\mathbb{1}}}$ when evaluating random characteristics.

2.5. Second moment method: proof of Proposition 2.3.4

2.5.1. First conditional moment asymptotics

In this section, we investigate the first conditional moment of $N_k^{(m)}(t)$ with respect to the bulk filtration. In particular, consider a function x such that, as $t\to\infty$, $x(t)\to\infty$ and $x(t)=o(t)$. Heuristically, we want to show that

$$m\mathbb{E}\left[N_k^{(m)}(t)\mid \mathcal{F}_{x(t)}\right] \approx N^{(m)}(x(t))\mathbb{E}\left[\boldsymbol{\xi}_{t-x(t)}^{\Phi_k^{(m)}}\right]. \tag{2.5.1}$$

Equation (2.5.1) shows that, conditionally on the information up to time $x(t)$, at time t we have $N^{(m)}(x(t))$ processes, each one producing the expected number of vertices with degree k at time $t-x(t)$. This follows from the fact that all the individual processes in $\boldsymbol{\xi}$ are independent from each other once we condition on the birth times.

We start writing $N_k^{(m)}(t)$ as sum of indicator functions, i.e.,

$$\mathbb{E}\left[N_k^{(m)}(t)\mid \mathcal{F}_{x(t)}\right] = \mathbb{E}\left[\sum_{n=1}^{N^{(m)}(x(t))}\mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}} + \sum_{n=N^{(m)}(x(t))+1}^{\infty}\mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}}\,\middle|\,\mathcal{F}_{x(t)}\right].$$

We can ignore the first sum in the conditional expectation, since

$$0\leq \mathrm{e}^{-\alpha^* t}\mathbb{E}\left[\sum_{n=1}^{N^{(m)}(x(t))}\mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}}\,\middle|\,\mathcal{F}_{x(t)}\right] \leq \mathrm{e}^{-\alpha^* t}N^{(m)}(x(t)), \tag{2.5.2}$$

and, using Theorem 2.1.11 and the fact that $x(t)=o(t)$,

$$\mathrm{e}^{-\alpha^*(t-x(t))}\mathrm{e}^{-\alpha^* x(t)}N^{(m)}(x(t)) \xrightarrow{\mathbb{P}-a.s.} 0. \tag{2.5.3}$$

Consider the sequence $(\sigma_n(x(t)))_{n\in\mathbb{N}}^{t\geq 0}$ as defined in Section 2.4.2. This is a sequence of random variables that approximates $(\tau_n)_{n\in\mathbb{N}}$ and it is measurable with respect to the bulk filtration. This means that we can write, for any $n \geq N^{(m)}(x(t))$,

$$D_n^{(\mathrm{in})}(t) = \xi^{(n,1)}(t - \sigma_{(n,1)}(x(t))) + \cdots + \xi^{(n,m)}(t - \sigma_{(n,m)}(x(t))) + o_{a.s.}(1).$$

Now, conditionally on the birth times $\sigma_{(n,1)}(x(t)), \dots, \sigma_{(n,m)}(x(t))$, the m processes related to the n-th vertex $(\xi_t^{(n,1)})_{t\geq 0}, \dots, (\xi_t^{(n,m)})_{t\geq 0}$ are independent, so the probability that the sum is equal to k is

$$\left(P[\xi](t - \sigma_{(n,1)}(x(t))) * \cdots * P[\xi](t - \sigma_{(n,m)}(x(t))) \right)_k, \tag{2.5.4}$$

which is a x-bulk measurable random variable. As a consequence,

$$\begin{aligned} &\mathbb{E}\left[\sum_{n=N^{(m)}(x(t))+1}^{\infty} \mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}} \,\middle|\, \mathcal{F}_{x(t)} \right] \\ &= \sum_{n=N^{(m)}(x(t))+1}^{\infty} \left(P[\xi](t - \sigma_{(n,1)}(x(t))) * \cdots * P[\xi](t - \sigma_{(n,m)}(x(t))) \right)_k. \end{aligned} \tag{2.5.5}$$

For any $k \in \mathbb{N}$, the function $u \mapsto P_k[\xi](u)$ is zero for negative argument. As a consequence, the sum in (2.5.5) is taken only over indices n such that $\sigma_{(n,j)}(x(t)) < t$. From the definition of $\sigma_{(n,j)}(x(t))$ as in (2.4.8) and the fact that $(n, j) = m(n-1) + j$, it follows that $\sigma_{(n,j)}(x(t)) < t$ if and only if

$$n < 1 - j/m + \mathrm{e}^{\alpha^*(t-x(t))} \boldsymbol{\xi}_{x(t)}^{\mathbb{1}_{\mathbb{R}^+}}/m = \mathrm{e}^{\alpha^*(t-x(t))} N^{(m)}(x(t))(1 + o_{a.s.}(1)), \tag{2.5.6}$$

where $o_{a.s.}(1)$ denotes a term that converges $\mathbb{P}$-a.s. to zero. Using (2.5.6) and then applying Proposition 2.4.3, for ℓ as in Lemma 2.4.2, we obtain

$$\begin{aligned} &\sum_{n=N^{(m)}(x(t))+1}^{N^{(m)}(x(t))\mathrm{e}^{\alpha^*(t-x(t))}} P[\xi](t - \sigma_{(n,1)}(x(t)))_k^{*m} \\ &\qquad + \ell m \sum_{n=N^{(m)}(x(t))+1}^{N^{(m)}(x(t))\mathrm{e}^{\alpha^*(t-x(t))}} \sigma_{(n,m)}(x(t)) - \sigma_{(n,1)}(x(t)), \end{aligned} \tag{2.5.7}$$

where the difference between (2.5.5) and the first sum in (2.5.7) is bounded in absolute value by the second sum in (2.5.7).

Consider the difference $t - \sigma_{(n,1)}(x(t))$. Using the definition of the sequence $(\sigma_n(x(t)))_{n\in\mathbb{N}}$, and recalling that $mN^{(m)}(x(t)) = \boldsymbol{\xi}_{x(t)}^{\mathbb{1}_{\mathbb{R}^+}}(1 + o_{a.s.}(1))$ (see (2.3.6)), it fol-

lows that $t - \sigma_{(N^{(m)}(x(t)),1)}(x(t)) = (t - x(t))(1 + o_{a.s.}(1))$. As a consequence,

$$\begin{aligned} t - \sigma_{(n,1)}(x(t)) &= t - \sigma_{(N^{(m)}(x(t)),1)}(x(t)) - \left(\sigma_{(n,1)}(x(t)) - \sigma_{(N^{(m)}(x(t)),1)}(x(t))\right) \\ &= t - x(t) + \frac{1}{\alpha^*} \log\left(\frac{m(n-1)+1}{mN^{(m)}(x(t))}\right) + o_{a.s.}(1) \\ &= t - x(t) + \frac{1}{\alpha^*} \log\left(\frac{n}{N^{(m)}(x(t))}\right) + o_{a.s.}(1). \end{aligned} \tag{2.5.8}$$

The second sum in the right hand side of (2.5.7) is bounded by a telescopic sum, since $\sigma_{(n,1)}(x(t)) \geq \sigma_{(n-1,m)}(x(t))$, which implies that we can bound it with the difference between the last and the first term. Using (2.5.8) in (2.5.7), for $s = t - x(t)$, it leads to

$$\begin{aligned} &\sum_{n=N^{(m)}(x(t))+1}^{N^{(m)}(x(t))\mathrm{e}^{\alpha^* s}} P[\xi](s - \frac{1}{\alpha^*} \log\left(\frac{m(n-1)+1}{mN^{(m)}(x(t))}\right))_k^{*m} + \frac{m\ell}{\alpha^*} \log\left(\frac{mN^{(m)}(x(t))\mathrm{e}^{\alpha^* s}}{mN^{(m)}(x(t))}\right) \\ &= \sum_{p=1}^{\mathrm{e}^{\alpha^* s}} \sum_{q=1}^{N^{(m)}(x(t))} P[\xi](s - \frac{1}{\alpha^*} \log\left(p + q/N^{(m)}(x(t))\right))_k^{*m} + m\ell(t - x(t)) \\ &= N^{(m)}(x(t)) \sum_{p=1}^{\mathrm{e}^{\alpha^* s}} P[\xi]\left(s - \frac{1}{\alpha^*} \log(p)\right)_k^{*m} + m\ell(t - x(t)) \\ &= N^{(m)}(x(t)) \sum_{p=1}^{\mathrm{e}^{\alpha^* s}} \mathbb{E}\left[\Phi_k^{(m)}\left(s - \frac{1}{\alpha^*} \log(p)\right)\right] + m\ell(t - x(t)). \end{aligned} \tag{2.5.9}$$

The contribute of the term $m\ell(t-x(t))$ is negligible, since $\mathrm{e}^{-\alpha^* t} m\ell(t-x(t)) = o(1)$. To analyze the remaining sum, we introduce two measures γ_1 and γ_2 on $\mathbb{R}^+$. For $v \geq 0$,

$$\begin{aligned} \gamma_1([0,v]) &= \int_0^v \sum_{p \in \mathbb{N}} \delta_{\{1/\alpha^* \log p\}}(du) = \mathrm{e}^{\alpha^* v}, \\ \gamma_2([0,v]) &= \mathbb{E}\left[\int_0^v \sum_{n \in \mathbb{N}} \delta_{\{\tau_n\}}(du)\right] = \mathbb{E}\left[\boldsymbol{\xi}_v^{\mathbb{1}_{\mathbb{R}^+}}\right]. \end{aligned}$$

Notice that γ_2 is the average measure of the random measure given by the branching process size. From Theorem 2.1.11 we know that

$$\gamma_2([0,v]) = \mathbb{E}[\boldsymbol{\xi}_v^{\mathbb{1}_{\mathbb{R}^+}}] = (1/\mu\alpha^*)\mathrm{e}^{\alpha^* v}(1 + o(1)).$$

This means that, asymptotically in v, $\gamma_1([0,v]) = \mu\alpha^* \gamma_2([0,v])(1 + o(1))$. Using these two measures it is possible to write

$$\sum_{p=1}^{\mathrm{e}^{\alpha^* s}} \mathbb{E}\left[\Phi_k^{(m)}\left(s - \frac{1}{\alpha^*} \log(p)\right)\right] = \int_0^s \mathbb{E}[\Phi_k^{(m)}(s-u)]\gamma_1(du)$$

$$= \mu\alpha^* \int_0^s \mathbb{E}[\Phi_k^{(m)}(s-u)]\gamma_2(du) + o(1) \qquad (2.5.10)$$
$$= \mu\alpha^* \mathbb{E}\left[\boldsymbol{\xi}_s^{\Phi_k^{(m)}}\right] + o(1).$$

Using (2.5.10) in (2.5.9), we conclude that

$$\begin{aligned}
&\mathrm{e}^{-\alpha^* t}\mathbb{E}\left[N_k^{(m)}(t) \mid \mathcal{F}_{x(t)}\right] \\
&= \mathrm{e}^{-\alpha^* t}\mu\alpha^* N^{(m)}(x(t))\mathbb{E}\left[\boldsymbol{\xi}_{t-x(t)}^{\Phi_k^{(m)}}\right] + o_{a.s.}(1) \\
&= \left(\mu\alpha^* \mathrm{e}^{-\alpha^* x(t)} N^{(m)}(x(t))\right)\left(\mathrm{e}^{-\alpha^*(t-x(t))}\mathbb{E}\left[\boldsymbol{\xi}_{t-x(t)}^{\Phi_k^{(m)}}\right]\right) + o_{a.s.}(1).
\end{aligned}$$

Applying (2.1.3) it follows that, as $t \to \infty$, $\mu\alpha^* \mathrm{e}^{-\alpha^* x(t)} N(x(t))$ converges $\mathbb{P}$-a.s. to Θ, while $\mu\alpha^* \mathrm{e}^{-\alpha^*(t-x(t))}\mathbb{E}\left[\boldsymbol{\xi}_{t-x(t)}^{\Phi_k^{(m)}}\right]$ converges to $\mathcal{L}(\Phi_k^{(m)}(\cdot))(\alpha^*)/\mu$. This completes the proof of (2.3.9).

2.5.2. Conditional second moment asymptotics

In this section, we prove (2.3.10), i.e., the result on the conditional second moment of $N_k^{(m)}(t)$. We again write $N_k^{(m)}(t)$ as sum of indicator functions, which means

$$\begin{aligned}
&\mathrm{e}^{-2\alpha^* t}\mathbb{E}\left[N_k^{(m)}(t)^2 \middle| \mathcal{F}_{x(t)}\right] \\
&= \mathrm{e}^{-2\alpha^* t}\mathbb{E}\left[\sum_{n,n' \in \mathbb{N}} \mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}} \mathbb{1}_{\{D_{n'}^{(\mathrm{in})}(t)=k\}} \middle| \mathcal{F}_{x(t)}\right].
\end{aligned}$$

We now divide the sum in different sums, according to the indices n and n', as

$$\begin{aligned}
&\sum_{n,n' \le N^{(m)}(x(t))} \mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}} \mathbb{1}_{\{D_{n'}^{(\mathrm{in})}(t)=k\}} \\
&+ \sum_{n,n' > N^{(m)}(x(t))} \mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}} \mathbb{1}_{\{D_{n'}^{(\mathrm{in})}(t)=k\}} \qquad (2.5.11) \\
&+ 2 \sum_{n \le N^{(m)}(x(t)), n' > N^{(m)}(x(t))} \mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}} \mathbb{1}_{\{D_{n'}^{(\mathrm{in})}(t)=k\}}.
\end{aligned}$$

For the first sum in (2.5.11), we use (2.5.2) as bound, and by (2.5.3) it is $o_{a.s.}(1)$. For the second sum in (2.5.11), we again use the sequence $(\sigma_n(x(t)))_{n\in\mathbb{N}}$ to approximate the birth times. Using similar arguments as in Section 2.5.1, and the fact that conditionally on the birth times all the birth processes are independent, we write, for $n \neq n'$ and $n, n' > N^{(m)}(x(t))$,

$$\mathbb{P}\left(D_n^{(\mathrm{in})}(t) = k, D_{n'}^{(\mathrm{in})}(t) = k \mid \mathcal{F}_{x(t)}\right)$$

$$= \Big[P[\xi](t-\sigma_{(n,1)}(x(t))) * \cdots * P[\xi](t-\sigma_{(n,m)}(x(t)))\Big]_k \tag{2.5.12}$$
$$\times \Big[P[\xi](t-\sigma_{(n',1)}(x(t))) * \cdots * P[\xi](t-\sigma_{(n',m)}(x(t)))\Big]_k .$$

We can use (2.5.12) to bound the conditional expectation of the second sum in (2.5.11). In fact, adding the missing terms we can write

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{n,n'>N^{(m)}(x(t))} \mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}}\mathbb{1}_{\{D_{n'}^{(\mathrm{in})}(t)=k\}}\Big|\mathcal{F}_{x(t)}\Big] \\
&\le \Big(\sum_{n>N^{(m)}(x(t))} \big(P[\xi](t-\sigma_{(n,1)}(x(t))) * \cdots * P[\xi](t-\sigma_{(n,m)}(x(t)))\big)_k\Big)^2 \\
&\quad + \mathbb{E}\Big[N_k^{(m)}(t)\Big|\mathcal{F}_{x(t)}\Big] \qquad (2.5.13)\\
&= \mathbb{E}\Big[\sum_{n>N^{(m)}(x(t))} \mathbb{1}_{\{D_n^{(\mathrm{in})}(t)=k\}}\Big|\mathcal{F}_{x(t)}\Big]^2 + \mathbb{E}\left[N_k^{(m)}(t)\middle|\,\mathcal{F}_{x(t)}\right] \\
&\le \mathbb{E}\left[N_k^{(m)}(t)\middle|\,\mathcal{F}_{x(t)}\right]^2 + \mathbb{E}\left[N_k^{(m)}(t)\middle|\,\mathcal{F}_{x(t)}\right].
\end{aligned}$$

The third sum in (2.5.11) can be easily bounded by $2N^{(m)}(x(t))\mathbb{E}[N_k^{(m)}(t)|\mathcal{F}_{x(t)}]$. Using together the three bounds we obtained, we have that $\mathrm{e}^{-2\alpha^* t}\mathbb{E}\left[N_k^{(m)}(t)^2|\mathcal{F}_{x(t)}\right]$ can be bounded by

$$\mathrm{e}^{-2\alpha^* t}\mathbb{E}[N_k(t)|\mathcal{F}_{x(t)}]^2 + \mathrm{e}^{-2\alpha^* t}\left(2N(x(t))+1\right)\mathbb{E}[N_k(t)|\mathcal{F}_{x(t)}] + o_{a.s.}(1). \tag{2.5.14}$$

The result follows since the second term in (2.5.14) is again $o_{a.s.}(1)$, similarly to the first term in (2.5.11).

2.5.3. Affine and constant PA functions: Proofs of corollaries 2.2.3 and 2.2.5

In Section 2.2.1 we already showed that CBPs defined by birth processes as in Definition 2.1.14 embeds the PAM in continuous-time and what we called random recursive graph. We just need to show that Condition 2.3.1 is satisfied. In general, processes defined as in Definition 2.1.14 are differentiable and satisfy a recursive property (see [11, Section 3.2]). In particular, the derivatives of the family of functions $(P_k[\xi](t))_{k\in\mathbb{N}}$ are given by (2.1.8) and (2.1.9).

Since in general we consider a non-decreasing sequence $(f_k)_{k\in\mathbb{N}}$, it is possible to see that if we set $\ell(k) = f_k$ then Condition 2.3.1 is satisfied. Hence, the limiting degree distribution $(p_k^{(m)})_{k\in\mathbb{N}}$ is the distribution of the sum of m independent copies of $(\xi_t)_{t\ge 0}$ at exponential time T_{α^*}, for α^* Malthusian parameter of the CTBP.

In the case of the PAM embedding, the sum of m birth processes is distributed as an embedded birth process defined by the PA rule $\bar{f}_k = k + m + \delta$ (it is easy to prove

this by induction over the distribution of birth times). This implies that we can use known results on this type of birth processes [9, 152] to write

$$p_k^{(m)} = \mathbb{P}\left(\xi^1_{T_{\alpha^*}} + \cdots + \xi^m_{T_{\alpha^*}} = k\right) = \frac{\alpha^*}{\alpha^* + k + m + \delta} \prod_{i=0}^{k-1} \frac{i + m + \delta}{\alpha^* + i + m + \delta},$$

that can be rewritten as in (1.3.1) using Γ functions, since in this case $\alpha^* = 1 + \delta/m$ (see [152, Section 4.2], [73, Proposition 3.15]).

For the random recursive graph, calculations are easier. It is easy to show that in this case $\alpha^* = 1$. Since the sum of m Poisson processes (PP) with parameter 1 is a PP with parameter m, the limiting degree distribution is the distribution of a PP at an exponentially distributed time with parameter 1. Then

$$p_k^{(m)} = \mathbb{E}\left[\mathrm{e}^{-mT_1} \frac{(mT_1)^k}{k!}\right] = \frac{1}{m+1}\left(1 + \frac{1}{m}\right)^{-k}. \tag{2.5.15}$$

As mentioned, for $m = 1$ (so without collapsing) the random recursive graph reduces to the random recursive tree, and the limiting distribution is just $p_k^{(1)} = 2^{-(k+1)}$ (see [97]).

2.6. Discrete-time processes: proof of Theorem 2.2.4

The convergence result given in Theorem 2.2.2 assures that in continuous time, the proportion of vertices in CBP with degree k converges in probability to $p_k^{(m)}$. When considering a CTBP in the presence of aging, this result is enough since these types of CBPs are defined only in continuous time.

When we instead consider embedding processes as in Definition 2.1.14, we can consider a discrete-time sequence of random graphs $(\mathrm{CBP}^{(m)}_{\tau_n})_{n\in\mathbb{N}}$, where $(\tau_n)_{n\in\mathbb{N}}$ is the sequence of birth times of the corresponding CTBP. This is the way the PAM is usually defined. In particular, the sequence $(\tau_n)_{n\in\mathbb{N}}$ corresponds to the sequence of times at which a new edge appears in the CBP. In this setting, the convergence in probability given in Theorem 2.2.2 does not imply the convergence in probability of $(m\mathrm{e}^{-\alpha^*\tau_n} N_k^{(m)}(\tau_n))_{n\in\mathbb{N}}$. Here, we will prove that $\mathrm{e}^{-\alpha^*\tau_n} N_k^{(m)}(\tau_n)$ converges in probability to $p_k^{(m)}\Theta/(\mu\alpha^*)$, and that this further implies that $N_k^{(m)}(\tau_{mn})/n$ converges in probability to $p_k^{(m)}$, as required.

Recall the t-bulk sigma-field. We denote, as in (2.4.8), for $n \geq \boldsymbol{\xi}_t^{\mathbb{1}}$,

$$\sigma_n = \sigma_n(t) = \frac{1}{\alpha^*}\log n - \frac{1}{\mu\alpha^*}\Theta_t.$$

Take $t = t_n = (\log n)^{1/2}$. Then, define the sequence $(\tau'_n)_{n\in\mathbb{N}}$, where $\tau'_n := \sigma_n(t_n)$. Notice that τ'_n is t_n-bulk-measurable. Further, $\tau'_n \overset{a.s.}{\to} \infty$ and

$$\frac{t_n}{\tau'_n} = \frac{(\log n)^{1/2}}{\frac{1}{\alpha^*}\log n - \frac{1}{\mu\alpha^*}\log\Theta_{t_n}} = \frac{(\log n)^{1/2}}{\log n(1/\alpha^* - \log\Theta_{t_n}/(\mu\alpha^*\log n))} \overset{a.s.}{\to} 0.$$

By Remark 2.3.5, Proposition 2.3.4 holds for $m\mathrm{e}^{-\alpha^*\tau'_n}N_k^{(m)}(\tau'_n)$, so that

$$m\mathrm{e}^{-\alpha^*\tau'_n}N_k^{(m)}(\tau'_n) \xrightarrow{\mathbb{P}} p_k^{(m)}\Theta/(\mu\alpha^*).$$

The advantage of the sequence $(\tau'_n)_{n\in\mathbb{N}}$, other than being t_n-bulk measurable, is that it is a good approximation of the sequence $(\tau_n)_{n\in\mathbb{N}}$. Indeed,

$$|\tau_n - \tau'_n| \leq \left|\tau_n - \frac{1}{\alpha^*}\log n - \frac{1}{\mu\alpha^*}\log\Theta\right| + \left|\frac{1}{\mu\alpha^*}\log\Theta - \frac{1}{\mu\alpha^*}\log\Theta_{t_n}\right|, \tag{2.6.1}$$

so that $|\tau_n - \tau'_n| \xrightarrow{a.s.} 0$. As a consequence, also $m\mathrm{e}^{-\alpha^*\tau_n}N_k^{(m)}(\tau_n) \xrightarrow{\mathbb{P}} p_k^{(m)}\Theta/(\mu\alpha^*)$.

By Theorem 2.3.2, we further know that $m\mathrm{e}^{-\alpha^* t}N^{(m)}(t) \xrightarrow{a.s.} \Theta/(\mu\alpha^*)$, so this holds also for $m\mathrm{e}^{-\alpha^*\tau_n}N^{(m)}(\tau_n)$. As a consequence,

$$\frac{m\mathrm{e}^{-\alpha^*\tau_n}N_k^{(m)}(\tau_n)}{m\mathrm{e}^{-\alpha^*\tau_n}N^{(m)}(\tau_n)} = \frac{N_k^{(m)}(\tau_n)}{N^{(m)}(\tau_n)} = \frac{m}{n}N_k^{(m)}(\tau_n) \xrightarrow{\mathbb{P}} p_k^{(m)}. \tag{2.6.2}$$

Consequently, $N_k^{(m)}(\tau_{mn})/n \xrightarrow{\mathbb{P}} p_k^{(m)}$. This completes the proof of Theorem 2.2.4. □

3

Diameter with infinite variance degrees

Content and structure of the chapter

In this chapter, we investigate the diameter of the configuration model (CM) and PAM in the case of power-law degree distribution with finite mean and infinite variance, with minimal degree larger than 3 in CM and larger than 2 in PAM. We prove that in both models, the diameter divided by loglog the size of the graph converges in probability to an explicit constant that we are able to compute, and that depends on the parameters of the models. In particular, the same proof structure works for both models, showing similarities of the two models and universality of our argument.

The chapter is structured as follows: in Section 3.1 we define the two models that we consider in the chapter and we state the convergence result on the diameters, namely Theorem 3.1.9 for CM and Theorem 3.1.12 for PAM. In Section 3.2 we present the high-level structure of the proofs of Theorem 3.1.9 and 3.1.12. The proofs are divided in lower and upper bounds. For CM we prove the lower bound in Section 3.3 and the upper bound in Section 3.5. For PAM, we prove the lower bound in Section 3.4 and the upper bound in Section 3.6. The novel results of this chapter are based on [40].

3.1. Introduction and statements of results

3.1.1. Configuration model and main result

The configuration model CM_n is a random graph with vertex set $[n] := \{1, 2, \dots, n\}$ and with prescribed degrees. Let $\boldsymbol{d} = (d_1, d_2, \dots, d_n)$ be a given *degree sequence*, i.e., a sequence of n positive integers with total degree

$$\ell_n = \sum_{i \in [n]} d_i, \tag{3.1.1}$$

assumed to be even. The configuration model (CM) on n vertices with degree sequence $\boldsymbol{d}$ is constructed as follows: Start with n vertices and d_i half-edges adjacent to vertex $i \in [n]$. Randomly choose pairs of half-edges and match the chosen pairs together to form edges. Although self-loops may occur, these become rare as $n \to \infty$ (see e.g. [29, Theorem 2.16], [98, 99]). We denote the resulting multi-graph on $[n]$ by CM_n, with corresponding edge set $\mathcal{E}_n$. We often omit the dependence on the degree sequence $\boldsymbol{d}$, and write CM_n for $\mathrm{CM}_n(\boldsymbol{d})$.

Regularity of vertex degrees. Let us now describe our regularity assumptions. For each $n \in \mathbb{N}$ we have a degree sequence $\boldsymbol{d}^{(n)} = (d_1^{(n)}, \dots, d_n^{(n)})$. To lighten notation, we omit the superscript (n) and write $\boldsymbol{d}$ instead of $\boldsymbol{d}^{(n)}$ or $(\boldsymbol{d}^{(n)})_{n \in \mathbb{N}}$ and d_i instead of $d_i^{(n)}$. Let $(p_k)_{k \in \mathbb{N}}$ be a probability mass function on $\mathbb{N}$. We introduce the *empirical degree distribution* of the graph as

$$p_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{d_i = k\}}. \tag{3.1.2}$$

We can define now the *degree regularity conditions*:

Condition 3.1.1 (Degree regularity conditions)**.** *Let CM_n be a configuration model, then we say that $\boldsymbol{d}$ satisfies the degrees regularity conditions* (a), (b)*, with respect to $(p_k)_{k \in \mathbb{N}}$ if:*

(a) *for every $k \in \mathbb{N}$, as $n \to \infty$*

$$p_k^{(n)} \longrightarrow p_k. \tag{3.1.3}$$

(b) *$\sum_k k p_k < \infty$, and as $n \to \infty$*

$$\sum_{k \in \mathbb{N}} k p_k^{(n)} \longrightarrow \sum_{k \in \mathbb{N}} k p_k. \tag{3.1.4}$$

As notation, we write that $\boldsymbol{d}$ satisfies the d.r.c. (a), (b).

Let $F_{\boldsymbol{d},n}$ be the distribution function of $(p_k^{(n)})_{k\in\mathbb{N}}$, that is, for $k \in \mathbb{N}$,

$$F_{\boldsymbol{d},n}(k) = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{d_i \le k\}}. \tag{3.1.5}$$

We suppose that $\boldsymbol{d}$ satisfies the d.r.c. (a) and (b) with respect to some probability mass function $(p_k)_{k\in\mathbb{N}}$, corresponding to a distribution function F.

Condition 3.1.2 (Polynomial distribution condition)**.** *We say that $\boldsymbol{d}$ satisfies the polynomial distribution condition with exponent $\tau \in (2,3)$ if for all $\delta > 0$ there exist $\alpha = \alpha(\delta) > \frac{1}{2}$, $c_1(\delta) > 0$ and $c_2(\delta) > 0$ such that, for every $n \in \mathbb{N}$, the lower bound*

$$1 - F_{\boldsymbol{d},n}(x) \ge c_1 x^{-(\tau-1+\delta)} \tag{3.1.6}$$

holds for all $x \le n^\alpha$, and the upper bound

$$1 - F_{\boldsymbol{d},n}(x) \le c_2 x^{-(\tau-1-\delta)} \tag{3.1.7}$$

holds for all $x \ge 1$.

There are two examples that explain Condition 3.1.2. Consider the case of i.i.d. degrees with $\mathbb{P}(D_i > x) = cx^{-(\tau-1)}$, then the degree sequence satisfies Condition 3.1.2 a.s. A second case is when the number of vertices of degree k is $n_k = \lceil nF(k)\rceil - \lceil nF(k-1)\rceil$, and $1-F(x) = cx^{-(\tau-1)}$. Condition 3.1.2 allows for more flexible degree sequences than just these examples.

If we fix $\beta < \min\{\alpha, \frac{1}{\tau-1+\delta}\}$, the lower bound (3.1.6) ensures that the number of vertices of degree higher than $x = n^\beta$ is at least $n^{1-\beta(\tau-1+\delta)}$, which diverges as a positive power of n. If we take $\beta > \frac{1}{2}$, these vertices with high probability form a complete graph. This will be essential for proving our main results. The precise value of β is irrelevant in the sequel of this chapter.

For an asymptotic degree distribution with asymptotic probability mass function $(p_k)_{k\in\mathbb{N}}$, we say that

$$d_{\min} = \min\{k \in \mathbb{N} \colon p_k > 0\} \tag{3.1.8}$$

is the minimal degree of the probability given by $(p_k)_{k\in\mathbb{N}}$. With these technical requests, we can state the main result for the configuration model:

Theorem 3.1.3 (Diameter of CM_n for $\tau \in (2,3)$)**.** *Let $\boldsymbol{d}$ be a sequence satisfying Condition 3.1.1 with asymptotic degree distribution $(p_k)_k$ with $d_{\min} \ge 3$. Suppose that $\boldsymbol{d}$ satisfies Condition 3.1.2 with $\tau \in (2,3)$ and $d_i \ge d_{\min}$ for all $i \in [n]$. Then*

$$\frac{\mathrm{diam}(\mathrm{CM}_n)}{\log\log n} \xrightarrow[n\to\infty]{\mathbb{P}} \frac{2}{\log(d_{\min}-1)} + \frac{2}{|\log(\tau-2)|}, \tag{3.1.9}$$

where $\xrightarrow[n\to\infty]{\mathbb{P}}$ denotes convergence in probability as $n \to \infty$.

In fact, the result turns out to be false when $p_1 + p_2 > 0$, as shown by Fernholz and Ramachandran [66] (see also [91]), since then there are long strings of vertices with low degrees that are of logarithmic length.

3.1.2. Preferential attachment model and main result

In the present chapter, we consider the following version of PAM:

Definition 3.1.4 (Preferential attachment model)**.** *Fix $m \in \mathbb{N}$, $\delta \in (-m, \infty)$. Denote by $\{t \xrightarrow{j} v\}$ the event that the j-th edge of vertex $t \in \mathbb{N}$ is attached to vertex $v \in [t]$ (for $1 \le j \le m$). The preferential attachment model with parameters (m, δ) is defined by the attachment probabilities*

$$\mathbb{P}\left(t \xrightarrow{j} v \,\middle|\, \mathrm{PA}_{t,j-1}\right) = \begin{cases} \dfrac{D_v(t, j-1) + 1 + j\delta/m}{c_{t,j}} & \text{for } v = t, \\ \dfrac{D_v(t, j-1) + \delta}{c_{t,j}} & \text{for } v < t, \end{cases} \tag{3.1.10}$$

where $\mathrm{PA}_{t,j-1}$ is the graph after the first $j-1$ edges of vertex t have been attached, and correspondingly $D_v(t, j-1)$ is the degree of vertex v. The normalizing constant $c_{t,j}$ in (3.1.10) *is*

$$c_{t,j} := [m(t-1) + (j-1)]\,(2 + \delta/m) + 1 + \delta/m\,. \tag{3.1.11}$$

We refer to Section 3.4.1 for more details and explanations on the construction of the model (in particular, for the reason behind the factor $j\delta/m$ in the first line of (3.1.10)).

For the preferential attachment model, our main result is the following:

Theorem 3.1.5 (Diameter of the preferential attachment model)**.** *Let $(\mathrm{PA}_t)_{t\ge1}$ be a preferential attachment model with $m \ge 2$ and $\delta \in (-m, 0)$. Then*

$$\frac{\mathrm{diam}(\mathrm{PA}_t)}{\log\log t} \xrightarrow[t\to\infty]{\mathbb{P}} \frac{2}{\log m} + \frac{4}{|\log(\tau - 2)|}, \tag{3.1.12}$$

where $\tau = 3 + \delta/m \in (2, 3)$.

In the proof of Theorem 3.1.5 we are also able to identify the typical distances in PA_t:

Theorem 3.1.6 (Typical distance in the preferential attachment model)**.** *Let V_1^t and V_2^t be two independent uniform random vertices in $[t]$. Denote the distance between V_1^t and V_2^t in PA_t by H_t. Then*

$$\frac{H_t}{\log\log t} \xrightarrow[t\to\infty]{\mathbb{P}} \frac{4}{|\log(\tau - 2)|}. \tag{3.1.13}$$

Theorems 3.1.5–3.1.6 prove [91, Conjecture 1.8].

3.1.3. Related works

There are several other works that have already studied typical distances and diameters of CM and PAM. Van der Hofstad, Hooghiemstra and Znamenski [90] analyze typical distances in CM_n for $\tau \in (2,3)$, while Van der Hofstad, Hooghiemstra and Van Mieghem [89] study $\tau > 3$. They prove that for $\tau \in (2,3)$ typical distances are of order $\log\log n$, while for $\tau > 3$ is of order $\log n$, and it presents the explicit constants of asymptotic growth. Van der Hofstad, Hooghiemstra and Znamensky [91] shows for $\tau > 2$ and when vertices of degree 1 or 2 are present, that with high probability the diameter of CM_n is bounded from below by a constant times $\log n$, while when $\tau \in (2,3)$ and the minimal degree is 3, the diameter is bounded from above by a constant times $\log\log n$. In [92], Van der Hofstad and Komjáthy investigate typical distances for configuration models and $\tau \in (2,3)$ in great generality, extending the results in [91] beyond the setting of i.i.d. degrees.

Fernholz and Ramachandran [66] prove that the diameter of CM_n is equal to an explicit constant times $\log n$ plus $o(\log n)$ when $\tau \in (2,3)$ but there are vertices of degree 1 or 2 present in the graph, by studying the longest paths in the configuration model that are not part of the 2-core (which is the part of the graph for which all vertices have degree at least 2). Since our minimal degree is at least 3, the 2-core is whp the entire graph, and thus this logarithmic phase vanishes. Dereich, Mönch and Mörters [57] prove that typical distances in PA_t are asymptotically equal to an explicit constant times $\log\log t$, using path counting techniques. We use such path counting techniques as well, now for the lower bound on the diameters. Van der Hofstad [86] studies the diameter of PA_t when $m = 1$, and proves that the diameter still has logarithmic growth. Dommers, van der Hofstad and Hooghiemstra [59] prove an upper bound on the diameter of PA_t, but the explicit constant is not sharp.

CM_n and PA_t are two different models, in the sense that CM_n is a static model while PA_t is a dynamic model. It is interesting to notice that the main strategy to prove Theorems 3.1.3 and 3.1.5 is the same. In fact, all the statements formulated in Section 3.2 are general and hold for both models. Also the explicit constants appearing in (3.1.9) and (3.1.12) are highly similar, which reflects the same structure of the proofs. The differences consist in a factor 2 in the terms containing τ and in the presence of $d_{\min} - 1$ and m in the remaining term. The factor 2 can be understood by noting that in CM_n pairs of vertices with high degree are likely to be at distance 1, while in PA_t they are at distance 2. The difference in $d_{\min} - 1$ and m is due to the fact that $d_{\min} - 1$ and m play the same role in the two models, i.e., they are the minimal forward degree (or "number of children") of a vertex that is part of a tree contained in the graph. We refer to Section 3.2 for more details.

Typical distances and diameters have been studied for other random graphs models as well, showing $\log\log$ behavior. Bloznelis [25] investigates the typical distance in power-law intersection random graphs, where such distance, conditioning on being finite, is of order $\log\log n$, while results on diameter are missing. Chung and Lu [49, 50] present results respectively for random graphs with given expected degrees and Erdős and Rényi random graphs $G(n,p)$, see also van den Esker, van der Hof-

stad and Hooghiemstra [64] for the rank-1 setting. The setting of the configuration model with finite-variance degrees is studied in [66]. In [50] it is proved that for the power-law regime with exponent $\tau \in (2,3)$, the diameter is $\Theta(\log n)$, while typical distances are of order $\log\log n$. This can be understood from the existence of a positive proportion of vertices with degree 2, creating long, but thin, paths. In [49], the authors investigate the different behavior of the diameter according to the parameter p.

3.2. General structure of the proofs

We split the proof of Theorems 3.1.3 and 3.1.5 into a lower and an upper bound. Remarkably, the strategy is the same for both models despite the inherent difference in the models. In this section we explain the strategy in detail, formulating general statements that will be proved for each model separately in the next sections.

Throughout this section, we assume that the assumptions of Theorems 3.1.3 and 3.1.5 are satisfied and, to keep unified notation, we denote the size of the preferential attachment model by $n \in \mathbb{N}$, instead of the more usual $t \in \mathbb{N}$.

Throughout the chapter, we treat real numbers as integers when we consider graph distances. By this, we mean that we round real numbers to the closest integer. To keep the notation light and make the chapter easier to read, we omit the rounding operation.

3.2.1. Lower bound

We start with the structure of the proof of the lower bound in (3.1.9) and (3.1.12). The key notion is that of a *minimally-k-connected vertex*, defined as a vertex whose k-neighborhood (i.e., the neighborhood up to distance k) is essentially *a regular tree with the smallest possible degree*, equal to $d_{\min}$ for the configuration model and to $m+1$ for the preferential attachment model. Due to technical reasons, the precise definition of minimally-k-connected vertex is slightly different for the two models (see Definitions 3.3.2 and 3.4.2).

Henceforth we fix $\varepsilon > 0$ and define, for $n \in \mathbb{N}$,

$$k_n^- = (1-\varepsilon)\frac{\log\log n}{\log(d_{\mathrm{fwd}})}, \tag{3.2.1}$$

where d_{fwd} denotes the *forward degree*, or "number of children":

$$d_{\mathrm{fwd}} = \begin{cases} d_{\min} - 1 & \text{for } \mathrm{CM}_n; \\ m & \text{for } \mathrm{PA}_n. \end{cases} \tag{3.2.2}$$

Our first goal is to prove that the number of minimally-k_n^--connected vertices is large enough, as formulated in the following statement:

Statement 3.2.1 (Moments of $M_{k_n^-}$)**.** *Denote by $M_{k_n^-}$ the number of minimally-k_n^--connected*

vertices in the graph (either CM_n *or* PA_n*). Then, as* $n \to \infty$*,*

$$\mathbb{E}\left[M_{k_n^-}\right] \to \infty, \qquad \mathrm{Var}\left(M_{k_n^-}\right) = o\left(\mathbb{E}\left[M_{k_n^-}\right]^2\right), \tag{3.2.3}$$

where $\mathrm{Var}(X) := \mathbb{E}[X^2] - \mathbb{E}[X]^2$ *denotes the variance of the random variable* X*.*

The proof for the preferential attachment model makes use of conditioning arguments. Indeed, we describe as much information as necessary to be able to bound probabilities that vertices are minimally-k connected. Particularly in the variance estimate, these arguments are quite delicate, and crucial for our purposes.

The bounds in (3.2.3) show that $M_{k_n^-} \xrightarrow{\mathbb{P}} \infty$ as $n \to \infty$. This will imply that *there is a pair of minimally-k_n^--connected vertices with disjoint k_n^--neighborhoods*,[1] hence the diameter of the graph is at least $2k_n^-$, which explains the first term in (3.1.9) and (3.1.12). Our next aim is to prove that these minimally connected trees are typically at distance $2c_{\mathrm{dist}} \log\log n/|\log(\tau-2)|$, where $c_{\mathrm{dist}} = 1$ for the configuration model and $c_{\mathrm{dist}} = 2$ for the preferential attachment model.

For this, let us now define

$$\bar{k}_n = (1-\varepsilon)\frac{c_{\mathrm{dist}} \log\log n}{|\log(\tau-2)|}, \tag{3.2.4}$$

where

$$c_{\mathrm{dist}} = \begin{cases} 1 & \text{for } \mathrm{CM}_n; \\ 2 & \text{for } \mathrm{PA}_n. \end{cases} \tag{3.2.5}$$

The difference in the definition of c_{dist} is due to fact that in CM_n vertices with high degree are likely at distance 1, while in PA_n are at distance 2. We explain the origin of this effect in more detail in the proofs.

It turns out that the distance between the k_n^--neighborhoods of two minimally-k_n^--connected vertices is at least $2\bar{k}_n$. More precisely, we have the following statement:

Statement 3.2.2 (Distance between neighborhoods)**.** *Let W_1^n and W_2^n be two random vertices chosen independently and uniformly among the minimally-k_n^--connected ones. Denoting by $\tilde{H}_n$ the distance between the k_n^--neighborhoods of W_1^n and W_2^n, we have $\tilde{H}_n \geq 2\bar{k}_n$ with high probability.*

It follows immediately from Statement 3.2.2 that the distance between the vertices W_1^n and W_2^n is at least $2k_n^- + 2\bar{k}_n$, with high probability. This proves the lower bound in (3.1.9) and (3.1.12).

It is known from the literature that $2\bar{k}_n$, see (3.2.4), represents the *typical distance* between two vertices chosen independently and uniformly in the graph. In order to

[1] A justification for this fact is provided by the following Statement 3.2.2 (the randomly chosen vertices W_1^n and W_2^n have disjoint k_n^--neighborhoods, because $\tilde{H}_n > 0$ with high probability). For a more direct justification, see Remark 3.3.6 for the configuration model and Remark 3.4.7 for the preferential attachment model.

prove Statement 3.2.2, we collapse the k_n^--neighborhoods of W_1^n and W_2^n into single vertices and show that their distance is roughly equal to the typical distance $2\bar{k}_n$. This is a delicate point, because the collapsed vertices have a relatively large degree and thus *could* be closer than the typical distance. The crucial point why they are not closer is that the degree of the boundary grows only polylogarithmically. The required justification is provided by the next statement:

Statement 3.2.3 (Bound on distances)**.** *Let us introduce the set*

$$V_n := \begin{cases} \{v \in [n] \colon\ d_v \le \log n\} & \text{for } \mathrm{CM}_n; \\ \{v \in [n] \colon\ v \ge \frac{n}{(\log n)^2}\} & \text{for } \mathrm{PA}_n. \end{cases} \tag{3.2.6}$$

Then, denoting the distance in the graph of size n by dist_n,

$$\max_{a,b \in V_n} \mathbb{P}\left(\mathrm{dist}_n(a,b) \le 2\bar{k}_n\right) = O\left(\frac{1}{(\log n)^2}\right). \tag{3.2.7}$$

The proof of Statement 3.2.3 is based on *path counting techniques.* These are different for the two models, but the idea is the same: We split the possible paths between the vertices a and b into two sets, called *good paths* and *bad paths*. Here *good* means that the degrees of vertices along the path increase, but *not too much*. We then separately and directly estimate the contribution of each set. The details are described in the proof.

3.2.2. Upper bound

We now describe the structure of the proof for the upper bound, which is based on two key concepts: the *core of the graph* and the *k-exploration graph* of a vertex.

We start by introducing some notation. First of all, fix a constant $\sigma \in (1/(3-\tau), \infty)$. We define Core_n as the set of vertices in the graph of size n with degree larger than $(\log n)^\sigma$. More precisely, denoting by $D_v(t) = D_v(t,m)$ the degree of vertex v in the preferential attachment model after time t, i.e. in the graph PA_t (see the discussion after (3.1.10)), we let

$$\mathrm{Core}_n := \begin{cases} \{v \in [n] \colon d_v \ge (\log n)^\sigma\} & \text{for } \mathrm{CM}_n; \\ \{v \in [n] \colon D_v(n/2) \ge (\log n)^\sigma\} & \text{for } \mathrm{PA}_n. \end{cases} \tag{3.2.8}$$

The fact that we evaluate the degrees at time $n/2$ for the PAM is quite crucial in the proof of Statement 3.2.4 below. In Section 3.6, we also give bounds on $D_v(n)$ for $v \in \mathrm{Core}_n$, as well as for $v \notin \mathrm{Core}_n$, that show that the degrees cannot grow too much between time $n/2$ and n. The first statement, that we formulate for completeness, upper bounds the diameter of Core_n and is already known from the literature for both models:

Statement 3.2.4. *Define c_{dist} as in* (3.2.5)*. Then, for every $\varepsilon > 0$, with high probability*

$$\frac{\mathrm{diam}(\mathrm{Core}_n)}{\log\log n} \leq (1+\varepsilon)\frac{2c_{\mathrm{dist}}}{|\log(\tau-2)|}. \tag{3.2.9}$$

Statement 3.2.4 for CM_n is [91, Proposition 3.1], for PA_n it is [59, Theorem 3.1].

Next we bound the distance between a vertex and Core_n. For $k \in \mathbb{N}$, we define the *k-exploration graph* of a vertex v as a suitable subgraph of its k-neighborhood, built as follows: we consider the usual exploration process starting at v, but instead of exploring all the edges incident to a vertex, we only explore a *fixed* number of them, namely d_{fwd} defined in (3.2.2). (The choice of which edges to explore is a natural one, and it will be explained in more detail in the proofs.)

We stress that it is possible to explore vertices that have already been explored, leading to what we call a *collision*. If there are no collisions, then the k-exploration graph is a tree. In presence of collisions, the k-exploration graph is not a tree, and it is clear that every collision reduces the number of vertices in the k-exploration graph.

Henceforth we fix $\varepsilon > 0$ and, in analogy with (3.2.1), we define, for $n \in \mathbb{N}$,

$$k_n^+ = (1+\varepsilon)\,\frac{\log\log n}{\log(d_{\mathrm{fwd}})}. \tag{3.2.10}$$

Our second statement for the upper bound shows that the k_n^+-exploration graph of *every* vertex in the graph either intersects Core_n, or it has a bounded number of collisions:

Statement 3.2.5 (Bound on collisions)**.** *There is a constant $c < \infty$ such that, with high probability, the k_n^+-exploration graph of* every *vertex in the graph has at most c collisions before hitting Core_n. As a consequence, for some constant $s > 0$, the k_n^+-exploration graph of* every *vertex in the graph either intersects Core_n, or its boundary has cardinality at least*

$$s(d_{\mathrm{fwd}})^{k_n^+} = (\log n)^{1+\varepsilon+o(1)}. \tag{3.2.11}$$

With a bounded number of collisions, the k_n^+-exploration graph is not far from being a tree, which explains the lower bound (3.2.11) on the cardinality of its boundary. Having enough vertices on its boundary, the k_n^+-exploration is likely to be connected to Core_n *fast*, which for our purpose means in $o(\log\log n)$ steps. This is the content of our last statement:

Statement 3.2.6. *There are constants $B, C < \infty$ such that, with high probability, the k_n^+-exploration graph of every vertex in the graph is at distance at most $h_n = \lceil B\log\log\log n + C\rceil$ from Core_n.*

The proof for this is novel. For example, for the configuration model, we grow the $k_n^+ + h_n$ neighborhood of a vertex, and then show that there are so many half-edges at its boundary that it is very likely to connect immediately to the core. The proof

for the preferential attachment model is slightly different, but the conclusion is the same. This shows that the vertex is indeed at most at distance $k_n^+ + h_n$ away from the core.

In conclusion, with high probability, the diameter of the graph is at most

$$(k_n^+ + h_n) + \mathrm{diam}(\mathrm{Core}_n) + (k_n^+ + h_n),$$

which gives us the expressions in (3.1.9) and (3.1.12) and completes the proof of the upper bound.

3.3. Lower bound for CM

In this section we prove Statements 3.2.1, 3.2.2 and 3.2.3 for the configuration model. By the discussion in Section 3.2.1, this completes the proof of the lower bound in Theorem 3.1.3.

In our proof, it will be convenient to choose a particular order to pair the half-edges. This is made precise in the following remark:

Remark 3.3.1 (Exchangeability in half-edge pairing)**.** Given a degree sequence $\boldsymbol{d} = (d_1, \ldots, d_n)$ such that $\ell_n = d_1 + \ldots + d_n$ is even, the configuration model CM_n can be built iteratively as follows:

- ▷ start with d_i half-edges attached to each vertex $i \in [n] = \{1, 2, \ldots, n\}$;
- ▷ choose an *arbitrary* half-edge and pair it to a uniformly chosen half-edge;
- ▷ choose an *arbitrary* half-edge, among the $\ell_n - 2$ that are still unpaired, and pair it to a uniformly chosen half-edge; and so on.

The *order* in which the arbitrary half-edges are chosen does not matter in the above, by exchangeability (see also [85, Chapter 7]).

3.3.1. Proof of Statement 3.2.1

With a slight abuse of notation (see (3.1.8)), in this section we set

$$d_{\min} = \min\{d_1, \ldots, d_n\}\,.$$

Given a vertex $v \in [n]$ and $k \in \mathbb{N}$, we denote the set of vertices at distance at most k from v (in the graph CM_n) by $U_{\leq k}(v)$ and we call it the *k-neighborhood of v*.

Definition 3.3.2 (Minimally-k-connected vertex)**.** *For $k \in \mathbb{N}_0$, a vertex $v \in [n]$ is called* minimally-k-connected *when all the vertices in $U_{\leq k}(v)$ have minimal degree, i.e.,*

$$d_i = d_{\min} \qquad \textit{for all } i \in U_{\leq k}(v)\,,$$

and furthermore there are no self-loops, multiple edges or cycles in $U_{\leq k}(v)$. Equivalently, v is minimally-k-connected when the graph $U_{\leq k}(v)$ is a regular tree with degree $d_{\min}$.

We denote the (random) set of minimally-k-connected vertices by $\mathcal{M}_k \subseteq [n]$, and its cardinality by $M_k = |\mathcal{M}_k|$, i.e. M_k denotes the number of minimally-k-connected vertices.

Remark 3.3.3 (The volume of the k-neighborhood of k-minimally connected vertices)**.** For a minimally-k-connected vertex v, since $U_{\leq k}(v)$ is a tree with degree $d_{\min}$, the number of *edges* inside $U_{\leq k}(v)$ equals (assuming $d_{\min} \geq 2$)

$$i_k = \sum_{l=1}^{k} d_{\min}(d_{\min}-1)^{l-1} = \begin{cases} d_{\min}\, k & \text{if } d_{\min} = 2; \\ d_{\min}\dfrac{(d_{\min}-1)^k - 1}{d_{\min}-2} & \text{if } d_{\min} \geq 3. \end{cases} \tag{3.3.1}$$

Moreover, the number of *vertices* inside $U_{\leq k}(v)$ equals $i_k + 1$. By (3.3.1), it is clear why $d_{\min} > 2$, or $d_{\min} \geq 3$, is crucial. Indeed, this implies that the volume of neighborhoods of minimally-k-connected vertices grows exponentially in k.

Remark 3.3.4 (Collapsing minimally-k connected trees)**.** By Remarks 3.3.1 and 3.3.3, conditionally on the event $\{v \in \mathcal{M}_k\}$ that a given vertex v is minimally-k-connected, the random graph obtained from CM_n by collapsing $U_{\leq k}(v)$ to a single vertex, called a, *is still a configuration model* with $n - i_k$ vertices and with ℓ_n replaced by $\ell_n - 2i_k$, where the new vertex a has degree $d_{\min}(d_{\min}-1)^k$.

Analogously, conditionally on the event

$$\{v \in \mathcal{M}_k,\, w \in \mathcal{M}_m,\, U_{\leq k}(v) \cap U_{\leq m}(w) = \varnothing\}$$

that two given vertices v and w are minimally-k and minimally-m-connected with *disjoint* neighborhoods, collapsing $U_{\leq k}(v)$ and $U_{\leq m}(w)$ to single vertices a and b yields again a configuration model with $n - i_k - i_m$ vertices, where ℓ_n is replaced by $\ell_n - 2i_k - 2i_m$ and where the new vertices a and b have degrees equal to $d_{\min}(d_{\min}-1)^k$ and $d_{\min}(d_{\min}-1)^m$, respectively.

We denote the number of vertices of degree k in the graph by n_k, i.e.,

$$n_k = \sum_{i \in [n]} \mathbb{1}_{\{d_i = k\}}.$$

We now study the first two moments of M_k, where we recall that the total degree ℓ_n is defined by (3.1.1):

Proposition 3.3.5 (Moments of M_k)**.** *Let CM_n be a configuration model such that $d_{\min} \geq 2$. Then, for all $k \in \mathbb{N}$,*

$$\mathbb{E}[M_k] = n_{d_{\min}} \prod_{i=1}^{i_k} \frac{d_{\min}(n_{d_{\min}} - i)}{\ell_n - 2i + 1}, \tag{3.3.2}$$

where i_k is defined in (3.3.1)*. When, furthermore, $\ell_n > 4i_k$,*

$$\mathbb{E}[M_k^2] \le \mathbb{E}[M_k]^2 + \mathbb{E}[M_k]\left((i_k+1) + i_{2k}\, d_{\min}\, \frac{n_{d_{\min}}}{\ell_n - 4i_k}\right). \tag{3.3.3}$$

Before proving Proposition 3.3.5, let us complete the proof of Statement 3.2.1 subject to it. We are working under the assumptions of Theorem 3.1.3, hence $d_{\min} \ge 3$ and the degree sequence $\boldsymbol{d}$ satisfies the degree regularity condition Condition 3.1.1, as well as the polynomial distribution condition Condition 3.1.2 with exponent $\tau \in (2,3)$. Recalling (3.1.1)-(3.1.2), we can write $n_{d_{\min}} = n\, p^{(n)}_{d_{\min}}$ and $\ell_n = n\sum_{k\in\mathbb{N}} k p_k^{(n)}$, so that, as $n \to \infty$,

$$\begin{aligned} n_{d_{\min}} &= n\, p_{d_{\min}}(1+o(1))\,, \qquad & \ell_n &= n\,\mu(1+o(1))\,, \\ & & p_{d_{\min}} &> 0\,, \qquad & \mu &:= \sum_{k\in\mathbb{N}} k p_k < \infty. \end{aligned} \tag{3.3.4}$$

Recalling the definition (3.2.1) of k_n^- and (3.3.1), for $k = k_n^-$,

$$i_{k_n^-} = d_{\min}\frac{(d_{\min}-1)^{k_n^-}-1}{d_{\min}-2} = \frac{d_{\min}}{d_{\min}-2}(\log n)^{1-\varepsilon}(1+o(1)),$$

hence

$$i_{2k_n^-} = O((\log n)^{2(1-\varepsilon)}). \tag{3.3.5}$$

Bounding $\mathbb{E}[M_k] \le n$, it follows by (3.3.3) that

$$\mathrm{Var}[M_{k_n^-}] \le \mathbb{E}[M_{k_n^-}]\left(O(i_{k_n^-}) + O(i_{2k_n^-})\right) \le n\, O((\log n)^{2(1-\varepsilon)}) = n^{1+o(1)}\,. \tag{3.3.6}$$

On the other hand, applying (3.3.2), for some $c \in (0,1)$ one has

$$\mathbb{E}[M_{k_n^-}] \ge n\, p_{d_{\min}}\left(\frac{d_{\min}\, p_{d_{\min}}}{\mu} + o(1)\right)^{i_{k_n^-}} \ge n\, p_{d_{\min}}\, c^{(\log n)^{1-\varepsilon}} = n^{1-o(1)}\,. \tag{3.3.7}$$

Relations (3.3.6) and (3.3.7) show that (3.2.3) holds, completing the proof of Statement 3.2.1. □

Remark 3.3.6 (Disjoint neighborhoods)**.** Let us show that, with high probability, there are vertices $v, w \in \mathcal{M}_{k_n^-}$ with $U_{\le k_n^-}(v) \cap U_{\le k_n^-}(w) = \varnothing$. We proceed by contradiction: fix $v \in \mathcal{M}_{k_n^-}$ and assume that, for every vertex $w \in \mathcal{M}_{k_n^-}$, one has $U_{\le k_n^-}(v) \cap U_{\le k_n^-}(w) \ne \varnothing$. Then, for any $w \in \mathcal{M}_{k_n^-}$ there must exist *a self-avoiding path from v to w of length $\le 2k_n^-$ which only visits vertices with degree $d_{\min}$* (recall that $U_{\le k_n^-}(v)$ and $U_{\le k_n^-}(w)$ are regular trees). However, for fixed v, the number of such paths is $O((d_{\min}-1)^{2k_n^-}) = O((\log n)^{2(1-\epsilon)})$, see (3.2.1), while by Statement 3.2.1 the number of vertices $w \in \mathcal{M}_{k_n^-}$ is much larger, since $M_{k_n^-} \sim \mathbb{E}[M_{k_n^-}] = n^{1-o(1)}$, see (3.3.7).

Proof of Proposition 3.3.5. To prove (3.3.2) we write

$$M_k = \sum_{v\in[n]:\ d_v=d_{\min}} \mathbb{1}_{\{v\in\mathcal{M}_k\}}, \tag{3.3.8}$$

and since every vertex in the sum has the same probability of being minimally-k-connected,

$$\mathbb{E}\left[M_k\right] = n_{d_{\min}}\mathbb{P}(v\in\mathcal{M}_k). \tag{3.3.9}$$

A vertex v with $d_v = d_{\min}$ is in $\mathcal{M}_k$ when all the half-edges in $U_{\le k}(v)$ are paired to half-edges incident to distinct vertices having minimal degree, without generating cycles. By Remark 3.3.1, we can start pairing a half-edge incident to v to a half-edge incident to another vertex of degree $d_{\min}$. Since there are $n_{d_{\min}}-1$ such vertices, this event has probability

$$\frac{d_{\min}(n_{d_{\min}}-1)}{\ell_n-1}$$

We iterate this procedure, and suppose that we have already successfully paired $(i-1)$ couples of half-edges; then the next half-edge can be paired to a distinct vertex of degree $d_{\min}$ with probability

$$\frac{d_{\min}(n_{d_{\min}}-i)}{\ell_n-2(i-1)-1} = \frac{d_{\min}(n_{d_{\min}}-i)}{\ell_n-2i+1}. \tag{3.3.10}$$

Indeed, every time that we use a half-edge of a vertex of degree $d_{\min}$, we cannot use its remaining half-edges, and every step we make reduces the total number of possible half-edges by two. By (3.3.1), exactly i_k couples of half-edges need to be paired for v to be minimally-k-connected, so that

$$\mathbb{E}[M_k] = n_{d_{\min}}\mathbb{P}(v\in\mathcal{M}_k) = n_{d_{\min}}\prod_{i=1}^{i_k}\frac{d_{\min}(n_{d_{\min}}-i)}{\ell_n-2i+1}. \tag{3.3.11}$$

which proves (3.3.2). If $n_{d_{\min}} \le i_k$ the right hand side vanishes, in agreement with the fact that there cannot be any minimally-k-connected vertex in this case (recall (3.3.1)).

To prove (3.3.3), we notice that

$$\mathbb{E}[M_k^2] = \sum_{v,w\in[n]:\ d_v=d_w=d_{\min}} \mathbb{P}(v,w\in\mathcal{M}_k). \tag{3.3.12}$$

We distinguish different cases: the k-neighborhoods of v and w might be disjoint or they may overlap, in which case w can be included in $U_{\le k}(v)$ or not. Introducing the events

$$A_{v,w} = \{U_{\le k}(v)\cap U_{\le k}(w)\neq\varnothing\}, \qquad B_{v,w} = \{w\in U_{\le k}(v)\}, \tag{3.3.13}$$

we can write the right hand side of (3.3.12) as

$$\sum_{\substack{v,w\in[n]\\ d_v=d_w=d_{\min}}} \big[\mathbb{P}\left(v,w\in\mathcal{M}_k,A^c_{v,w}\right) + \mathbb{P}\left(v,w\in\mathcal{M}_k,A_{v,w},B_{v,w}\right)+\mathbb{P}\left(v,w\in\mathcal{M}_k,A_{v,w},B^c_{v,w}\right)\big]. \tag{3.3.14}$$

Let us look at the first term in (3.3.14). By Remarks 3.3.3 and 3.3.4, conditionally on $\{v\in\mathcal{M}_k\}$, the probability of $\{w\in\mathcal{M}_k,A^c_{v,w}\}$ equals the probability that w is minimally-k-connected in a new configuration model, with ℓ_n replaced by ℓ_n-2i_k and with the number of vertices with minimal degree reduced from $n_{d_{\min}}$ to $n_{d_{\min}}-(i_k+1)$. Then, by the previous analysis (see (3.3.11)),

$$\mathbb{P}\left(v,w\in\mathcal{M}_k,A^c_{v,w}\right)=\prod_{i=1}^{i_k}\frac{d_{\min}(n_{d_{\min}}-i-i_k-1)}{\ell_n-2i-2i_k+1}\,\mathbb{P}\left(v\in\mathcal{M}_k\right). \tag{3.3.15}$$

By direct computation, the ratio in the right hand side of (3.3.15) is always maximized for $i_k=0$ (provided $\ell_n\geq 2n_{d_{\min}}-3$, which is satisfied since $\ell_n\geq d_{\min}n_{d_{\min}}\geq 3n_{d_{\min}}$ by assumption). Therefore, setting $i_k=0$ in the ratio and recalling (3.3.11), we get the upper bound

$$\mathbb{P}\left(v,w\in\mathcal{M}_k,A^c_{v,w}\right)\leq\left[\prod_{i=1}^{i_k}\frac{d_{\min}(n_{d_{\min}}-i)}{\ell_n-2i+1}\right]\mathbb{P}(v\in\mathcal{M}_k)=\mathbb{P}\left(v\in\mathcal{M}_k\right)^2. \tag{3.3.16}$$

Since there are at most $n^2_{d_{\min}}$ pairs of vertices of degree $d_{\min}$, it follows from (3.3.16) that

$$\sum_{\substack{v,w\in[n]\\ d_v=d_w=d_{\min}}}\mathbb{P}\left(v,w\in\mathcal{M}_k,A^c_{v,w}\right)\leq n^2_{d_{\min}}\,\mathbb{P}\left(v\in\mathcal{M}_k\right)^2=\mathbb{E}[M_k]^2, \tag{3.3.17}$$

which explains the first term in (3.3.3).

For the second term in (3.3.14), v and w are minimally-k-connected with overlapping neighborhoods, and $w\in U_{\leq k}(v)$. Since

$$\{v,w\in\mathcal{M}_k\}\cap A_{v,w}\cap B_{v,w}\subseteq\{v\in\mathcal{M}_k\}\cap B_{v,w},$$

we can bound

$$\sum_{\substack{v,w\in[n]\\ d_v=d_w=d_{\min}}}\mathbb{P}\Big(v,w\in\mathcal{M}_k,A_{v,w},B_{v,w}\Big)\leq\mathbb{E}\Big[\sum_{v\in[n]\colon d_v=d_{\min}}\mathbb{1}_{\{v\in\mathcal{M}_k\}}\sum_{w\in[n]\colon d_w=d_{\min}}\mathbb{1}_{B_{v,w}}\Big],$$

and note that $\sum_{w\in[n]} \mathbb{1}_{B_{v,w}} = |U_{\le k}(v)| = i_k + 1$, by Remark 3.3.3. Therefore

$$\sum_{\substack{v,w\in[n]\\ d_v=d_w=d_{\min}}} \mathbb{P}\left(v,w\in\mathcal{M}_k, A_{v,w}, B_{v,w}\right) \le \mathbb{E}[M_k]\,(i_k+1), \tag{3.3.18}$$

which explains the second term in (3.3.3).

For the third term in (3.3.14), v and w are minimally-k-connected vertices with overlapping neighborhoods, but $w \notin U_{\le k}(v)$. This means that $\mathrm{dist}(v,w) = l+1$ for some $l \in \{k,\dots,2k-1\}$, so that $U_{\le k}(v)\cap U_{\le l-k}(w) = \varnothing$ and, moreover, a half-edge on the boundary of $U_{\le (l-k)}(w)$ is paired to a half-edge on the boundary of $U_{\le k}(v)$, an event that we call $F_{v,w;l,k}$. Therefore

$$\begin{aligned}&\{w\in\mathcal{M}_k\}\cap A_{v,w}\cap B^c_{v,w}\\ &\qquad\subseteq \bigcup_{l=k}^{2k-1} \{w\in\mathcal{M}_{l-k}\}\cap\{U_{\le k}(v)\cap U_{\le l-k}(w)=\varnothing\}\cap F_{v,w;l,k}.\end{aligned} \tag{3.3.19}$$

and we stress that in the right hand side w is only minimally-$(l-k)$-connected (in case $l=k$ this just means that $d_w = d_{\min}$). Then

$$\begin{aligned}&\mathbb{P}\Big(v,w\in\mathcal{M}_k, A_{v,w}, B^c_{v,w}\Big)\\ &\qquad\le \sum_{l=k}^{2k-1} \mathbb{E}\left[\mathbb{1}_{\{v\in\mathcal{M}_k,\, w\in\mathcal{M}_{l-k},\, U_{\le k}(v)\cap U_{\le l-k}(w)=\varnothing\}}\,\mathbb{1}_{F_{v,w;l,k}}\right].\end{aligned} \tag{3.3.20}$$

By Remark 3.3.4, conditionally on $\{v\in\mathcal{M}_k,\, w\in\mathcal{M}_{l-k},\, U_{\le k}(v)\cap U_{\le l-k}(w)=\varnothing\}$, we can collapse $U_{\le k}(v)$ and $U_{\le l-k}(w)$ to single vertices a and b with degrees respectively $d_{\min}(d_{\min}-1)^k$ and $d_{\min}(d_{\min}-1)^{l-k}$, getting a new configuration model with ℓ_n replaced by $\ell_n - 2i_k - 2i_{l-k}$. Bounding the probability that a half-edge of a is paired to a half-edge of b, we get

$$\begin{aligned}&\mathbb{P}(F_{v,w;l,k}\,|\,v\in\mathcal{M}_k,\, w\in\mathcal{M}_{l-k},\, U_{\le k}(v)\cap U_{\le l-k}(w)=\varnothing)\\ &\qquad\le \frac{d_{\min}(d_{\min}-1)^k d_{\min}(d_{\min}-1)^{l-k}}{\ell_n - 2i_k - 2i_{l-k} - 1} \le \frac{d^2_{\min}(d_{\min}-1)^l}{\ell_n - 4i_k},\end{aligned} \tag{3.3.21}$$

because $l \le 2k-1$ and, consequently, $i_{l-k} \le i_{k-1} \le i_k - 1$. Plugging (3.3.21) into (3.3.20), and then forgetting the event $\{w\in\mathcal{M}_{l-k},\, U_{\le k}(v)\cap U_{\le l-k}(w)=\varnothing\}$, leads to

$$\begin{aligned}&\sum_{\substack{v,w\in[n]\\ d_v=d_w=d_{\min}}} \mathbb{P}\Big(v,w\in\mathcal{M}_k,\, A_{v,w},\, B^c_{v,w}\Big)\\ &\qquad\le \left(\sum_{l=k}^{2k-1} \frac{d^2_{\min}(d_{\min}-1)^l}{\ell_n - 4i_k}\right) \sum_{\substack{v,w\in[n]\\ d_v=d_w=d_{\min}}} \mathbb{P}(v\in\mathcal{M}_k)\end{aligned} \tag{3.3.22}$$

$$\leq \frac{d_{\min}(d_{\min}-1)}{\ell_n - 4i_k}\, i_{2k-1}\, n_{d_{\min}}\, \mathbb{E}[M_k]\,,$$

where we have used the definition (3.3.1) of i_{2k-1}. Since $(d_{\min}-1)i_{2k-1} \leq i_{2k}$, again by (3.3.1), we have obtained the third term in (3.3.3). □

3.3.2. Proof of Statement 3.2.2

We recall that W_1^n and W_2^n are two independent random vertices chosen uniformly in $\mathcal{M}_{k_n^-}$ (the set of minimally-k_n^--connected vertices), assuming that $\mathcal{M}_{k_n^-} \neq \varnothing$ (which, as we have shown, occurs with high probability). Our goal is to show that

$$\lim_{n\to\infty} \mathbb{P}(E_n) = 0,$$

where we set

$$\begin{aligned} E_n :&= \Big\{ \mathrm{dist}\big(U_{\leq k_n^-}(W_1^n), U_{\leq k_n^-}(W_2^n)\big) \leq 2\bar{k}_n \Big\} \\ &= \Big\{ \mathrm{dist}(W_1^n, W_2^n) \leq 2k_n^- + 2\bar{k}_n \Big\}. \end{aligned}$$

We know from Statement 3.2.1 that as $n \to \infty$

$$\mathbb{P}\left(M_{k_n^-} \leq \frac{1}{2}\mathbb{E}[M_{k_n^-}] \right) \leq \mathbb{P}\left(|M_{k_n^-} - \mathbb{E}[M_{k_n^-}]| > \frac{1}{2}\mathbb{E}[M_{k_n^-}] \right) \leq \frac{\mathrm{Var}[M_{k_n^-}]}{\frac{1}{4}\mathbb{E}[M_{k_n^-}]^2} = o(1).$$

Therefore,

$$\begin{aligned} &\mathbb{P}(E_n) \\ &= \mathbb{P}\Big(E_n \cap \{ M_{k_n^-} > \tfrac{1}{2}\mathbb{E}[M_{k_n^-}] \} \Big) + o(1) \\ &= \mathbb{E}\Big[\sum_{v_1,v_2\in[n]} \mathbb{1}_{\{W_1^n = v_1, W_2^n = v_2\}} \mathbb{1}_{\{\mathrm{dist}(v_1,v_2) \leq 2k_n^- + 2\bar{k}_n\}} \mathbb{1}_{\{M_{k_n^-} > \frac{1}{2}\mathbb{E}[M_{k_n^-}]\}} \Big] + o(1) \\ &\leq \mathbb{E}\Big[\sum_{v_1,v_2\in[n]} \frac{\mathbb{1}_{\{v_1 \in \mathcal{M}_{k_n^-}, v_2 \in \mathcal{M}_{k_n^-}\}}}{M_{k_n^-}^2} \mathbb{1}_{\{\mathrm{dist}(v_1,v_2) \leq 2k_n^- + 2\bar{k}_n\}} \mathbb{1}_{\{M_{k_n^-} > \frac{1}{2}\mathbb{E}[M_{k_n^-}]\}} \Big] + o(1) \\ &\leq \sum_{v_1,v_2\in[n]} \frac{\mathbb{P}\Big(v_1, v_2 \in \mathcal{M}_{k_n^-},\, \mathrm{dist}(v_1,v_2) \leq 2k_n^- + 2\bar{k}_n \Big)}{\frac{1}{4}\mathbb{E}[M_{k_n^-}]^2} + o(1). \end{aligned} \tag{3.3.23}$$

In analogy with (3.3.13), we introduce the event

$$A_{v_1,v_2} := \big\{ U_{\leq k_n^-}(v_1) \cap U_{\leq k_n^-}(v_2) \neq \varnothing \big\}\,,$$

and show that it gives a negligible contribution. Recalling the proof of Proposition 3.3.5, in particular (3.3.18) and (3.3.22), the sum restricted to A_{v_1,v_2} leads precisely

to the second term in the right hand side of (3.3.3):

$$\begin{aligned}\sum_{v_1,v_2\in[n]} \frac{\mathbb{P}\left(v_1,v_2\in\mathcal{M}_{k_n^-},\, A_{v_1,v_2}\right)}{\frac{1}{4}\mathbb{E}[M_{k_n^-}]^2} &\le \frac{\mathbb{E}[M_{k_n^-}]\left((i_{k_n^-}+1)+i_{2k_n^-}\frac{d_{\min}\,n_{d_{\min}}}{\ell_n-4i_{k_n^-}}\right)}{\frac{1}{4}\mathbb{E}[M_{k_n^-}]^2}\\ &= \frac{O(i_{k_n^-})+O(i_{2k_n^-})}{\mathbb{E}[M_{k_n^-}]}\\ &= \frac{O((\log n)^2)}{n^{1-o(1)}} = o(1)\,,\end{aligned} \tag{3.3.24}$$

where we have used (3.3.5) and (3.3.7) (see also (3.3.4)).

We can thus focus on the event $A^c_{v_1,v_2} = \{U_{\le k_n^-}(v_1)\cap U_{\le k_n^-}(v_2)=\varnothing\}$. By Remark 3.3.4,

$$\mathbb{P}\left(\mathrm{dist}(v_1,v_2)\le 2k_n^- + 2\bar{k}_n \mid v_1,v_2\in\mathcal{M}_{k_n^-},\, A^c_{v_1,v_2}\right) = \hat{\mathbb{P}}\left(\mathrm{dist}(a,b)\le 2\bar{k}_n\right), \tag{3.3.25}$$

where $\hat{\mathbb{P}}$ is the law of the new configuration model which results from collapsing the neighborhoods $U_{\le k_n^-}(v_1)$ and $U_{\le k_n^-}(v_2)$ to single vertices a and b, with degrees $d_{\min}(d_{\min}-1)^{k_n^-} = O(\log n)$ (recall (3.2.1)-(3.2.2)). The degree sequence $\hat{\boldsymbol{d}}$ of this new configuration model is a slight modification of the original degree sequence $\boldsymbol{d}$: two new vertices of degree $O(\log n)$ have been added, while $2(i_{k_n^-}+1)=O(\log n)$ vertices with degree $d_{\min}$ have been removed (recall (3.3.5)). Consequently $\hat{\boldsymbol{d}}$ still satisfies the assumptions of Theorem 3.1.3, hence Statement 3.2.3 (to be proved in Section 3.3.3) holds for $\hat{\mathbb{P}}$ and we obtain

$$\hat{\mathbb{P}}\left(\mathrm{dist}(a,b)\le 2\bar{k}_n\right) = o(1). \tag{3.3.26}$$

We are ready to conclude the proof of Statement 3.2.2. By (3.3.23)-(3.3.24)-(3.3.25),

$$\begin{aligned}\mathbb{P}(E_n) &= \sum_{v_1,v_2\in[n]} \frac{\mathbb{P}\left(v_1,v_2\in\mathcal{M}_{k_n^-},\,\mathrm{dist}(v_1,v_2)\le 2k_n^-+2\bar{k}_n,\,A^c_{v_1,v_2}\right)}{\frac{1}{4}\mathbb{E}[M_{k_n^-}]^2} + o(1)\\ &\le \hat{\mathbb{P}}\left(\mathrm{dist}(a,b)\le 2\bar{k}_n\right)\sum_{v_1,v_2\in[n]}\frac{\mathbb{P}\left(v_1,v_2\in\mathcal{M}_{k_n^-}\right)}{\frac{1}{4}\mathbb{E}[M_{k_n^-}]^2} + o(1)\\ &= \hat{\mathbb{P}}\left(\mathrm{dist}(a,b)\le 2\bar{k}_n\right)\frac{\mathbb{E}[(M_{k_n^-})^2]}{\frac{1}{4}\mathbb{E}[M_{k_n^-}]^2} + o(1).\end{aligned}$$

Observe that $\mathbb{E}[(M_{k_n^-})^2] = \mathbb{E}[M_{k_n^-}]^2 + \mathrm{Var}(M_{k_n^-}) = O(\mathbb{E}[M_{k_n^-}]^2)$, by the second relation in (3.2.3). Applying (3.3.26), it follows that $\mathbb{P}(E_n)=o(1)$, completing the proof of Statement 3.2.2. □

3.3.3. Proof of Statement 3.2.3

In this section, we give a self-contained proof of Statement 3.2.3 for CM_n, as used in the proof of Statement 3.2.2.

Given two vertices $a, b \in [n]$, let $\mathcal{P}_k(a,b)$ be the set of all self-avoiding paths of length k from a to b, i.e. of all sequences $(\pi_0, \pi_1, \dots, \pi_k) \in [n]^{k+1}$ with $\pi_0 = a$, $\pi_k = b$ and such that (π_{i-1}, π_i) is an edge in the graph, for all $i = 1, \dots, k$. Analogously, let $\mathcal{P}_k(a) = \cup_{b\in[n]}\mathcal{P}_k(a,b)$ denote the set of all paths of length k starting at a.

Let us fix an arbitrary increasing sequence $(g_l)_{l\in\mathbb{N}_0}$ (that will be specified later). Define, for $a, b \in \mathbb{R}$, $a \wedge b := \min\{a, b\}$. We say that a path $\pi \in \mathcal{P}_k(a,b)$ is *good* when $d_{\pi_l} \le g_l \wedge g_{k-l}$ for every $l = 0, \dots, k$, and *bad* otherwise. In other words, a path is good when the degrees along the path do not increase too much from π_0 to $\pi_{k/2}$, and similarly they do not increase too much in the backward direction, from π_k to $\pi_{k/2}$.

For $k \in \mathbb{N}_0$, we introduce the event

$$\mathcal{E}_k(a,b) = \{\exists \pi \in \mathcal{P}_k(a,b) : \ \pi \text{ is a good path}\}\,. \tag{3.3.27}$$

To deal with bad paths, we define

$$\mathcal{F}_k(a) = \{\exists \pi \in \mathcal{P}_k(a) : \ d_{\pi_k} > g_k \ \text{ but } \ d_{\pi_i} \le g_i \ \forall i \le k-1\}\,. \tag{3.3.28}$$

If $\mathrm{dist}_{\mathrm{CM}_n}(a,b) \le 2\bar{k}$, then there must be a path in $\mathcal{P}_k(a,b)$ for some $k \le \bar{k}$, and this path might be good or bad. This leads to the simple bound

$$\mathbb{P}(\mathrm{dist}_{\mathrm{CM}_n}(a,b) \le 2\bar{k}) \le \sum_{k=0}^{2\bar{k}} \mathbb{P}(\mathcal{E}_k(a,b)) + \sum_{k=0}^{\bar{k}} \left[\mathbb{P}(\mathcal{F}_k(a)) + \mathbb{P}(\mathcal{F}_k(b))\right]. \tag{3.3.29}$$

We give explicit estimates for the two sums in the right hand side. We introduce the *size-biased distribution function* F_n^* associated to the degree sequence $\boldsymbol{d} = (d_1, \dots, d_n)$ by

$$F_n^*(t) = \frac{1}{\ell_n} \sum_{v\in[n]} d_v \, \mathbb{1}_{\{d_v \le t\}}. \tag{3.3.30}$$

If we choose uniformly one of the ℓ_n half-edges in the graph, and call D_n^* the degree of the vertex incident to this half-edge, then $F_n^*(t) = \mathbb{P}(D_n^* \le t)$. We also define the truncated mean

$$\nu_n(t) = \mathbb{E}\big[(D_n^* - 1)\mathbb{1}_{\{D_n^* \le t\}}\big] = \frac{1}{\ell_n} \sum_{v\in[n]} d_v(d_v - 1)\mathbb{1}_{\{d_v \le t\}}. \tag{3.3.31}$$

Now we are ready to bound (3.3.29).

Proposition 3.3.7 (Path counting for configuration model)**.** *Fix $\boldsymbol{d} = (d_1, \dots, d_n)$ (such that $\ell_n = d_1 + \ldots + d_n$ is even) and an increasing sequence $(g_l)_{l\in\mathbb{N}_0}$. For all distinct vertices*

$a, b \in [n]$ with $d_a \le g_0$, $d_b \le g_0$, and for all $\bar{k} \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}\Big(\mathrm{dist}_{\mathrm{CM}_n}(a,b) \le 2\bar{k}\Big) \le & \frac{d_a d_b}{\ell_n} \sum_{k=1}^{2\bar{k}} \left(1 - \frac{2k}{\ell_n}\right)^{-k} \prod_{l=1}^{k-1} \nu_n(g_l \wedge g_{h-l}) \\ & + (d_a + d_b) \sum_{k=1}^{\bar{k}} \left(1 - \frac{2k}{\ell_n}\right)^{-k} (1 - F_n^*(g_k)) \prod_{l=1}^{k-1} \nu_n(g_l).\end{aligned} \tag{3.3.32}$$

Proof. Fix an arbitrary sequence of vertices $\pi = (\pi_i)_{0 \le i \le k} \in [n]^{k+1}$. The probability that vertex π_0 is connected to π_1 is at most

$$\frac{d_{\pi_0} d_{\pi_1}}{\ell_n - 1},$$

because there are $d_{\pi_0} d_{\pi_1}$ ordered couples of half-edges, each of which can be paired with probability $1/(\ell_n - 1)$ (recall Remark 3.3.1), and we use the union bound. By similar arguments, conditionally on a specific half-edge incident to π_0 being paired to a specific half-edge incident to π_1, the probability that another half-edge incident to π_1 is paired to a half-edge incident to π_2 is by the union bound bounded from above by

$$\frac{(d_{\pi_1} - 1) d_{\pi_2}}{\ell_n - 3}.$$

Iterating the argument, the probability that π is a path in CM_n is at most

$$\frac{d_{\pi_0} d_{\pi_1}}{\ell_n - 1} \frac{(d_{\pi_1} - 1) d_{\pi_2}}{\ell_n - 3} \frac{(d_{\pi_2} - 1) d_{\pi_3}}{\ell_n - 5} \cdots \frac{(d_{\pi_{k-1}} - 1) d_{\pi_k}}{\ell_n - (2k-1)}. \tag{3.3.33}$$

Let us now fix $a, b \in [n]$ with $a \neq b$. Recalling (3.3.27)-(3.3.31), choosing $\pi_0 = a$, $\pi_k = b$ and summing (3.3.33) over all vertices $\pi_1, \dots, \pi_{k-1}$ satisfying $d_{\pi_i} \le g_i \wedge g_{k-i}$ yields

$$\mathbb{P}(\mathcal{E}_k(a,b)) \le d_a d_b \frac{(\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} \left(\prod_{i=1}^{k-1} \ell_n \, \nu_n(g_i \wedge g_{k-i}) \right). \tag{3.3.34}$$

Bounding $(\ell_n - 2k - 1)!!/(\ell_n - 1)!! \le (\ell_n - 2k)^{-k}$ yields the first term in the right hand side of (3.3.32). The bound for $\mathbb{P}(\mathcal{F}_k(a))$ is similar. Recalling (3.3.28)-(3.3.30), choosing $\pi_0 = a$ and summing (3.3.33) over vertices $\pi_1, \dots, \pi_{k-1}, \pi_k$ such that $d_{\pi_i} \le g_i$ for $i \le k - 1$ while $d_{\pi_k} > g_k$ gives

$$\mathbb{P}(\mathcal{F}_k(a)) \le d_a \frac{(\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} \left(\prod_{i=1}^{k-1} \ell_n \, \nu_n(g_i) \right) \big\{ \ell_n \left(1 - F_n^*(g_k)\right) \big\}, \tag{3.3.35}$$

and the same holds for $\mathbb{P}(\mathcal{F}_k(b))$. Plugging (3.3.34) and (3.3.35) into (3.3.29) proves (3.3.32). □

In order to exploit (3.3.32), we need estimates on F_n^* and ν_n, provided by the next lemma:

Lemma 3.3.8 (Tail and truncated mean bounds for D_n^*)**.** *Assume that Condition 3.1.2 holds. Fix $\eta > 0$, then there exist two constants $C_1 = C_1(\eta)$ and $C_2 = C_2(\eta)$ such that, for every $x \geq 0$,*

$$1 - F_n^*(x) \leq C_1 x^{-(\tau-2-\eta)}, \qquad \nu_n(x) \leq C_2 x^{(3-\tau+\eta)}.$$

Proof. For every $x \geq 0$ and $t \geq 0$ we can see that

$$1 - F_n^*(x) = \frac{1}{\ell_n} \sum_{v\in[n]} d_v \mathbb{1}_{\{d_v > x\}} = \frac{n}{\ell_n}\Big[\frac{1}{n} \sum_{v\in[n]} d_v \mathbb{1}_{\{d_v > x\}}\Big] = \frac{n}{\ell_n}\mathbb{E}\left[D_n \mathbb{1}_{\{D_n > x\}}\right],$$

where we recall that D_n is the degree of a uniformly chosen vertex. This means that

$$\begin{aligned}
\frac{n}{\ell_n}\mathbb{E}[D_n \mathbb{1}_{\{D_n>x\}}] =& \frac{n}{\ell_n} \sum_{j=0}^{\infty} \mathbb{P}\left(D_n \mathbb{1}_{\{D_n>x\}} > j\right) = \frac{n}{\ell_n} \sum_{j=0}^{\infty} \mathbb{P}\left(D_n > j, D_n > x\right) \\
&= \frac{n}{\ell_n} \sum_{j=0}^{\infty} \mathbb{P}\left(D_n > j \vee x\right) = \frac{n}{\ell_n} \sum_{j=0}^{\infty} \left(1 - F_{\boldsymbol{d},n}(j \vee x)\right) \\
&= \frac{n}{\ell_n}\Big[x(1 - F_{\boldsymbol{d},n}(x)) + \sum_{j=x}^{\infty} \left(1 - F_{\boldsymbol{d},n}(j)\right)\Big] \\
&\leq \frac{n}{\ell_n} C\Big[x^{-(\tau-2-\eta)} + \sum_{j=x}^{\infty} j^{-(\tau-1-\eta)}\Big] \leq C_1 x^{-(\tau-2-\eta)},
\end{aligned}$$

where we have used Condition 3.1.2 in the second last step (recall that $2 < \tau < 3$).

For ν_n, we can instead write

$$\begin{aligned}
\nu_n(x) &= \frac{1}{\ell_n} \sum_{v\in[n]} d_v(d_v - 1) \mathbb{1}_{\{d_v \leq x\}} = \frac{n}{\ell_n}\Big[\frac{1}{n} \sum_{v\in[n]} d_v(d_v - 1) \mathbb{1}_{\{d_v \leq x\}}\Big] \\
&= \frac{n}{\ell_n}\mathbb{E}\left[D_n(D_n - 1) \mathbb{1}_{\{D_n \leq x\}}\right] \leq \frac{n}{\ell_n}\mathbb{E}\left[D_n^2 \mathbb{1}_{\{D_n \leq x\}}\right],
\end{aligned}$$

where D_n is again the degree of a uniformly chosen vertex. The claim now follows from

$$\begin{aligned}
\frac{n}{\ell_n}\mathbb{E}\left[D_n^2 \mathbb{1}_{\{D_n \leq x\}}\right] &= \frac{n}{\ell_n} \sum_{j=0}^{\infty} (2j+1) \mathbb{P}\left(D_n \mathbb{1}_{\{D_n \leq x\}} > j\right) \\
&= \frac{n}{\ell_n} \sum_{j=0}^{\infty} (2j+1) \mathbb{P}\left(D_n > j, D_n \leq x\right) \qquad (3.3.36) \\
&\leq \frac{n}{\ell_n} \sum_{j=0}^{x-1} (2j+1) \mathbb{P}\left(D_n > j\right)
\end{aligned}$$

$$= \frac{n}{\ell_n} \sum_{j=0}^{x-1} (2j+1)[1 - F_{\boldsymbol{d},n}(j)]$$
$$\leq \frac{n}{\ell_n} \sum_{j=0}^{x-1} C j^{-(\tau-2-\eta)} \leq \frac{n}{\ell_n} C_2 x^{3-\tau+\eta}.$$

□

We are finally ready to complete the proof of Statement 3.2.3:

Proof of Statement 3.2.3. As in (3.2.4), we take

$$\bar{k}_n = (1-\varepsilon) \frac{\log\log n}{|\log(\tau-2)|}\,, \tag{3.3.37}$$

and our goal is to show that, as $n \to \infty$,

$$\max_{a,b\in[n]:\, d_a, d_b \leq \log n} \mathbb{P}\left(\mathrm{dist}_{\mathrm{CM}_n}(a,b) \leq 2\bar{k}_n\right) \longrightarrow 0.$$

We stress that $\tau \in (2,3)$ and $\varepsilon > 0$ are fixed. Then we choose $\eta > 0$ so small that

$$2\eta < \tau - 2 \qquad \text{and} \qquad \frac{|\log(\tau-2-2\eta)|}{|\log|\log(\tau-2)|} \leq \frac{1-\varepsilon/2}{1-\varepsilon}. \tag{3.3.38}$$

We use the inequality (3.3.32) given by Proposition 3.3.7, with the following choice of $(g_k)_{k\in\mathbb{N}_0}$:

$$g_k := (g_0)^{p^k}\,, \qquad \text{where} \qquad \begin{cases} g_0 := (\log n)^{\log\log n}; \\ p := \frac{1}{\tau-2-2\eta} > 1. \end{cases} \tag{3.3.39}$$

Let us focus on the first term in the right hand side of (3.3.32), that is

$$\frac{d_a d_b}{\ell_n} \sum_{k=1}^{2\bar{k}} \left(1 - \frac{2k}{\ell_n}\right)^{-k} \prod_{l=1}^{k-1} \nu_n(g_l \wedge g_{h-l})\,. \tag{3.3.40}$$

Since $\ell_n = \mu n(1+o(1))$ by (3.3.4), for $k \leq 2\bar{k}_n$ we have

$$\begin{aligned} \left(1 - \frac{2k}{\ell_n}\right)^{-k} \leq \left(1 - \frac{4\bar{k}_n}{\ell_n}\right)^{-2\bar{k}_n} &= 1 + O\left(\frac{\bar{k}_n^2}{\ell_n}\right) \\ &= 1 + O\left(\frac{(\log\log n)^2}{n}\right) = 1 + o(1)\,. \end{aligned} \tag{3.3.41}$$

Then observe that, by Lemma 3.3.8 and (3.3.39), for $k \le 2\bar{k}_n$

$$\begin{aligned}\prod_{l=1}^{k-1} \nu_n(g_l \wedge g_{k-l}) &= \prod_{l=1}^{k/2} \nu_n(g_l)^2 \\ &\le C_2^{k/2} \prod_{l=1}^{k/2} (g_l)^{2(3-\tau+\eta)} \\ &= C_2^{k/2} (g_0)^{2(3-\tau+\eta)\sum_{l=1}^{k/2} p^l} \le C_2^{\bar{k}_n} (g_0)^{2(3-\tau+\eta)C\, p^{\bar{k}_n}},\end{aligned} \tag{3.3.42}$$

with $C = \frac{p}{p-1}$. Note that $C_2^{\bar{k}_n} = O((\log n)^c)$ for some $c \in (0,\infty)$, see (3.3.37), while by (3.3.38)

$$\begin{aligned}p^{\bar{k}_n} &= \exp\Big(|\log(\tau-2-2\eta)|(1-\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}\Big) \\ &= (\log n)^{(1-\varepsilon)\frac{|\log(\tau-2-2\eta)|}{|\log(\tau-2)|}} \\ &\le (\log n)^{(1-\varepsilon/2)},\end{aligned}$$

hence the right hand side of (3.3.42) is $n^{o(1)}$ (since $g_0 = (\log n)^{\log\log n}$). Then, for $d_a, d_b \le \log n$,

$$(3.3.40) \le \frac{(\log n)^2}{\ell_n}(2\bar{k}_n)\big(1+o(1)\big)\, n^{o(1)} = O\bigg(\frac{(\log n)^2}{n}(\log\log n)\, n^{o(1)}\bigg) = o(1)\,.$$

It remains to look at the second sum in (3.3.32):

$$(d_a+d_b)\sum_{k=1}^{\bar{k}_n}\bigg(1-\frac{2k}{\ell_n}\bigg)^{-k}(1-F_n^*(g_k))\prod_{l=1}^{k-1}\nu_n(g_l). \tag{3.3.43}$$

By Lemma 3.3.8 ,we can bound $1-F_n^*(g_k) \le C_1(g_k)^{-(\tau-2-\eta)}$. By (3.3.41) and $C_1^{\bar{k}_n} = O((\log n)^c)$ for some $c \in (0,\infty)$, see (3.3.37), bounding the product in (3.3.43) like we did in (3.3.42) yields

$$O\big((\log n)^c\big)\,(d_a+d_b)\sum_{k=1}^{\bar{k}_n}(g_k)^{-(\tau-2-\eta)}(g_0)^{(3-\tau+\eta)Cp^{k-1}}, \tag{3.3.44}$$

where $p = 1/(\tau-2-2\eta)$ and $C = \frac{p}{p-1}$. By (3.3.39)

$$(g_k)^{-(\tau-2-\eta)}(g_0)^{-\frac{p}{p-1}(3-\tau+\eta)p^{k-1}} = (g_{k-1})^{-p(\tau-2-\eta)}(g_{k-1})^{\frac{p}{p-1}(3-\tau+\eta)},$$

where

$$p(\tau-2-\eta) = \frac{\tau-2-\eta}{\tau-2-2\eta} > 1,$$

and

$$\frac{p}{p-1}(3-\tau+\eta) = \frac{3-\tau+\eta}{3-\tau+2\eta} < 1.$$

This means that, setting $D := p(\tau-2-\eta) - \frac{p}{p-1}(3-\tau+\eta) > 0$, by (3.3.39),

$$(3.3.44) = O\big((\log n)^c\big)\,(d_a+d_b)\sum_{k=1}^{\bar{k}_n}(g_0)^{-Dp^{k-1}} \le O\big((\log n)^c\big)\,\frac{d_a+d_b}{(g_0)^D}\,. \qquad (3.3.45)$$

Since $g_0 = (\log n)^{\log\log n}$ while $d_a, d_b \le \log n$, the right hand side of (3.3.45) is $o(1)$. □

3.4. Lower bound for PAM

In this section we prove Statements 3.2.1, 3.2.2 and 3.2.3 for the preferential attachment model. By the discussion in Section 3.2.1, this completes the proof of the lower bound in Theorem 3.1.5.

We recall that, given $m \in \mathbb{N}$ and $\delta \in (-m, \infty)$, the preferential attachment model PA_t is a random graph with vertex set $[t] = \{1, 2, \dots, t\}$, where each vertex w has m outgoing edges, which are attached to vertices $v \in [w]$ with probabilities given in (3.1.10). In the next subsection we give a more detailed construction using random variables. This equivalent reformulation will be used in a few places, when we need to describe carefully some complicated events. However, for most of the exposition we will stick to the intuitive description given in Section 3.1.2.

3.4.1. Alternative construction of PAM

We introduce random variables $\xi_{w,j}$ to represent the vertex to which the j-th edge of vertex w is attached, i.e.

$$\xi_{w,j} = v \qquad \Longleftrightarrow \qquad w \overset{j}{\to} v\,. \qquad (3.4.1)$$

The graph PA_t is a *deterministic* function of these random variables: two vertices $v, w \in [t]$ with $v \le w$ are connected in PA_t if and only if $\xi_{w,j} = v$ for some $j \in [m]$. In particular, the degree of a vertex v after the k-th edge of vertex t has been attached, denoted by $D_{t,k}(v)$, is

$$D_v(t,k) := \sum_{(s,i)\le(t,k)} \big(\mathbb{1}_{\{\xi_{s,i}=v\}} + \mathbb{1}_{\{s=v\}}\big)\,, \qquad (3.4.2)$$

where we use the natural order relation

$$(s,i) \le (t,j) \qquad \Longleftrightarrow \qquad s < t \quad \text{or} \quad s = t,\ i \le j\,.$$

We point out this is the same notation as we used in Chapter 2. In fact, the model in Definition 3.1.4 is defined by collapsing together vertices of the discrete-time PA tree (see [85, Chapter 8]).

Defining the preferential attachment model amounts to giving a joint law for the sequence $\xi = (\xi_{w,j})_{(w,j)\in\mathbb{N}\times[m]}$. In agreement with (3.1.10), we set $\xi_{1,j} = 1$ for all $j \in [m]$, and for $t \geq 2$

$$\mathbb{P}\left(\xi_{t,j} = v \,|\, \xi_{\leq(t,j-1)}\right) = \begin{cases} \dfrac{D_v(t,j-1) + 1 + j\delta/m}{c_{t,j}} & \text{if } v = t; \\ \dfrac{D_v(t,j-1) + \delta}{c_{t,j}} & \text{if } v < t, \end{cases} \tag{3.4.3}$$

where $\xi_{\leq(t,i-1)}$ is a shorthand for the vector $(\xi_{s,i})_{(s,i)\leq(t,i-1)}$ (and we agree that $(t,0) := (t-1,m)$). The normalizing constant $c_{t,j}$ in (3.4.3) is indeed given by (3.1.11), because by (3.4.2),

$$\sum_{v\in[t]} D_v(t,j-1) = \sum_{(s,i)\leq(t,j-1)} (1+1) = 2((t-1)m + (j-1)).$$

The factor $j\delta/m$ in the first line of (3.4.3) is commonly used in the literature (instead of the possibly more natural δ). The reason is that, with such a definition, the graph $\mathrm{PA}_t(m,\delta)$ can be obtained from the special case $m = 1$, where every vertex has only one outgoing edge: one first generates the random graph $\mathrm{PA}_{mt}(1,\delta/m)$, whose vertex set is $[mt]$, and then collapses the block of vertices $[m(i-1)+1, mi)$ into a single vertex $i \in [t]$ (see also [85, Chapter 8]).

Remark 3.4.1. It is clear from the construction that PA_t is a *labeled directed graph*, because any edge connecting sites v, w, say with $v \leq w$, carries a label $j \in [m]$ and a direction, from the newer vertex w to the older one v (see (3.4.1)). Even though our final result, the asymptotic behavior of the diameter, only depends on the underlying undirected graph, it will be convenient to exploit the labeled directed structure of the graph in the proofs.

3.4.2. Proof of Statement 3.2.1

We denote by $U_{\leq k}(v)$ the k-neighborhood in PA_t of a vertex $v \in [t]$, i.e. the set of vertices at distance at most k from v, viewed as a labeled directed subgraph (see Remark 3.4.1). We denote by $D_v(t) = D_v(t,m)$ the degree of vertex v after time t, i.e. in the graph PA_t (recall (3.4.2)).

We define the notion of *minimally-k-connected vertex* in analogy with the configuration model (see Definition 3.3.2), up to minor technical restrictions made for later convenience.

Definition 3.4.2 (Minimally-k-connected vertex)**.** *For $k \in \mathbb{N}_0$, a vertex $v \in [t] \setminus [t/2]$ is called* minimally-k-connected *when $D_v(t) = m$, all the other vertices $i \in U_{\leq k}(v)$ are in $[t/2] \setminus [t/4]$ and have degree $D_i(t) = m+1$, and there are no self-loops, multiple edges or cycles in $U_{\leq k}(v)$. The graph $U_{\leq k}(v)$ is thus a tree with degree $m+1$, except for the root v which has degree m.*

We denote the (random) set of minimally-k-connected vertices by $\mathcal{M}_k \subseteq [t] \setminus [t/2]$, and its cardinality by $M_k = |\mathcal{M}_k|$.

For the construction of a minimally-k-connected neighborhood in the preferential attachment model we remind that the vertices are added to the graph at different times, so that the vertex degrees change while the graph grows. The relevant degree for Definition 3.4.2 is the one at the final time t. To build a minimally-k-connected neighborhood, we need

$$i_k = 1 + \sum_{i=1}^{k} m^i = \frac{m^{k+1} - 1}{m - 1} \tag{3.4.4}$$

many vertices. The center v of the neighborhood is the youngest vertex in $U_{\leq k}(v)$, and it has degree m, while all the other vertices have degree $m + 1$.

Our first goal is to evaluate the probability $\mathbb{P}(v \in \mathcal{M}_k)$ that a given vertex $v \in [t] \setminus [t/2]$ is minimally-k-connected. The analogous question for the configuration model could be answered quite easily in Proposition 3.3.5, because the configuration model can be built exploring its vertices in an arbitrary order, in particular starting from v, see Remark 3.3.1. This is no longer true for the preferential attachment model, whose vertices have an order, the chronological one, along which the conditional probabilities take the explicit form (3.1.10) or (3.4.3). This is why the proofs for the preferential attachment model are harder than for the configuration model.

As it will be clear in a moment, to get explicit formulas it is convenient to evaluate the probability $\mathbb{P}(v \in \mathcal{M}_k,\, U_{\leq k}(v) = H)$, where H is a fixed *labeled directed* subgraph, i.e. it comes with the specification of which edges are attached to which vertices. To avoid trivialities, we restrict to those H for which the probability does not vanish, i.e. which satisfy the constraints in Definition 3.4.2, and we call them *admissible*.

Let us denote by $H^o := H \setminus \partial H$ the set of vertices in H that are not on the boundary (i.e. they are at distance at most $k - 1$ from v). With this notation, we have the following result:

Lemma 3.4.3. *Let $\{\mathrm{PA}_t\}_{t\in\mathbb{N}}$ be a preferential attachment model. For any vertex $v \in [t]\setminus[t/2]$ and any directed labeled graph H which is admissible,*

$$\mathbb{P}\left(v \in \mathcal{M}_k, U_{\leq k}(v) = H\right) = L_1(H)\, L_2(H)\,, \tag{3.4.5}$$

where

$$L_1(H) \;:=\; \prod_{u\in H^o} \prod_{j=1}^{m} \frac{m + \delta}{c_{u,j}}, \tag{3.4.6}$$

$$L_2(H) \;:=\; \prod_{u\notin H^o} \prod_{j=1}^{m} \left[1 - \frac{D_H(u-1) + |H \cap [u-1]|\delta}{c_{u,j}}\right], \tag{3.4.7}$$

and $D_H(u-1) = \sum_{w\in H} D_w(u-1)$ is the total degree of H before vertex u is added to the

graph, and the normalization constant $c_{u,j}$ is defined in (3.1.11).

Proof. We recall that $\{a \xrightarrow{i} b\}$ denotes the event that the i-th edge of a is attached to b (see (3.4.1)). Since H is an admissible labeled directed subgraph, for all $u \in H^o$ and $j \in [m]$, the j-th edge of u is connected to a vertex in H, that we denote by $\theta_j^H(u)$. We can then write

$$\{v \in \mathcal{M}_k, U_{\le k}(v) = H\} = \Big(\bigcap_{u \in H^o} \bigcap_{j=1}^{m} \{u \xrightarrow{j} \theta_j^H(u)\} \Big) \cap \Big(\bigcap_{u \notin H^o} \bigcap_{j=1}^{m} \{u \not\xrightarrow{j} H\} \Big), \quad (3.4.8)$$

where of course $\{u \not\xrightarrow{j} H\} := \bigcup_{w \notin H} \{u \xrightarrow{j} w\}$. The first term in (3.4.8) is exactly the event that the edges present in H are connected in PA_t as they should be. The second term is the event that the vertices $u \notin H^o$ are not attached to H, so that $U_{\le k}(v) = H$. Notice that in (3.4.8) every vertex and every edge of the graph appears. For a vertex $u \in H^o$, by (3.1.10)

$$\mathbb{P}\Big(u \xrightarrow{j} \theta_j^H(u) \mid \mathrm{PA}_{u,j-1}\Big) = \frac{m+\delta}{c_{u,j}}, \quad (3.4.9)$$

because the vertex $\theta_j^H(u)$ has degree precisely m (when u is not already present in the graph). For $u \notin H^o$, we have to evaluate the probability that its edges do no attach to H, which is

$$\mathbb{P}\Big(u \not\xrightarrow{j} H \mid \mathrm{PA}_{u-1,j-1}\Big) = 1 - \frac{D_H(u-1) + |H \cap [u-1]|\delta}{c_{u,j}}. \quad (3.4.10)$$

Using conditional expectation iteratively, we obtain (3.4.9) or (3.4.10) for every edge in the graph, depending on whether the edge is part of H or not. This proves (3.4.6) and (3.4.7). □

The event $\{v \in \mathcal{M}_k, U_{\le k}(v) = H\}$ is an example of a class of events, called *factorizable*, that will be used throughout this section and Section 3.6. For this reason we define it precisely.

It is convenient to use the random variable $\xi_{w,j}$, introduced in Section 3.4.1, to denote the vertex to which the j-th edge of vertex w is attached (see (3.4.1)). Any event A for PA_t can be characterized iteratively, specifying a set $A_{s,i} \subseteq [s]$ of values for $\xi_{s,i}$, for all $(s,i) \le (t,m)$:

$$A = \bigcap_{(s,i) \le (t,m)} \big\{\xi_{s,i} \in A_{s,i}\big\}.$$

Of course, the set $A_{s,i}$ is allowed to depend on the "past", i.e. $A_{s,i} = A_{s,i}\big(\xi_{\le (s,i-1)}\big)$, or equivalently $A_{s,i} = A_{s,i}\big(\mathrm{PA}_{s,i-1}\big)$. Let us set $A_{\le (s,i)} := \bigcap_{(u,j) \le (s,i)} A_{u,j}$.

Definition 3.4.4 (Factorizable events)**.** *An event A for PA_t is called* factorizable *when the conditional probabilities of the events $\{\xi_{s,i} \in A_{s,i}\}$, given the past, are deterministic. More*

precisely, for any (s,i) there is a (non-random) $p_{s,i} \in [0,1]$ such that

$$\mathbb{P}\left(\xi_{s,i} \in A_{s,i} \,|\, \xi_{\leq (s,i-1)}\right) = p_{s,i} \tag{3.4.11}$$

on the event $\xi_{\leq(s,i-1)} \in A_{\leq(s,i-1)}$. As a consequence, the chain rule for probabilities yields

$$\mathbb{P}(A) = \prod_{(s,i)\leq(t,m)} p_{s,i}\,.$$

Remark 3.4.5. Relations (3.4.9) and (3.4.10) show that $A = \{v \in \mathcal{M}_k, U_{\leq k}(v) = H\}$ is a factorizable event. In fact, $A_{s,i}$ is either the single vertex $\theta_i^H(s)$ (if $s \in H^o$) or the set $[s-1] \setminus H$ (if $s \notin H^o$). In both cases, the set $A_{s,i} \subseteq [s-1]$ has *a fixed total degree and a fixed cardinality*, hence the conditional probabilities (3.4.11) are specified in a deterministic way (recall (3.4.3)).

Note that the event $\{v \in \mathcal{M}_k\}$ is not factorizable. This is the reason for specifying the realization of the k-neighborhood $U_{\leq k}(v) = H$.

Henceforth we fix $\varepsilon > 0$. We recall that k_n^- was defined in (3.2.1). Using the more customary t instead of n, we have

$$k_t^- = (1-\varepsilon)\frac{\log\log t}{\log m}. \tag{3.4.12}$$

We recall that $M_{k_t^-} = |\mathcal{M}_{k_t^-}|$ denotes the number of minimally-k_t^--connected vertices in PA_t (see Definition 3.4.2). We can now prove half of Statement 3.2.1 for the preferential attachment model, more precisely the first relation in equation (3.2.3).

Proposition 3.4.6 (First moment of $M_{k_t^-}$)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. Then, for k_t^- as in* (3.4.12)*, as $t \to \infty$,*

$$\mathbb{E}[M_{k_t^-}] \longrightarrow \infty.$$

Proof. Similarly to the proof of (3.3.2), we write

$$\mathbb{E}[M_k] = \sum_{v\in[t]\setminus[t(2]} \mathbb{P}\left(v \in \mathcal{M}_k\right) = \sum_{v\in[t]\setminus[t/2]} \sum_{H\subseteq[t]\setminus[t/4]} \mathbb{P}\big(v \in \mathcal{M}_k, U_{\leq k}(v) = H\big),$$

where the sum is implicitly restricted to admissible H (i.e., to H that are possible realizations of $U_{\leq k}(v)$).

Since we will use (3.4.5), we need a lower bound on (3.4.6) and (3.4.7). Recalling (3.1.11), it is easy to show, since the number of vertices in H^o equals $i_k - m^k = i_{k-1}$, and $u \leq v$ for $u \in H^o$,

$$L_1(H) \geq \left[\frac{m+\delta}{v(2m+\delta)+1+\delta/m}\right]^{mi_{k-1}}.$$

Note that for $u \leq t/4$ all the factors in the product in (3.4.7) equal 1, because $H \subseteq$

$[t]\setminus[t/4]$. Restricting to $u > t/4$ and bounding $D_H(u-1)+|H\cap[u-1]|\delta \le (m+1+\delta)i_k$, we get

$$L_2(H) \ge \left[1 - \frac{(m+1+\delta)i_k}{\frac{t}{4}(2m+\delta)}\right]^{3mt/4}.$$

Let us write $H = \{v\} \cup H'$ where H' is a subset of $[t/2] \setminus [t/4]$ with $|H'| = i_k - 1$. Clearly, for any such subset there is at least one way to order the vertices to generate an admissible H. The number of possible subsets in $[t/2]\setminus[t/4]$ is at least $\binom{t/4}{i_k-1}$. Then, we obtain

$$\mathbb{E}[M_k] \ge \sum_{v\in[t]\setminus[t/2]} \binom{t/4}{i_k-1} \left[\frac{m+\delta}{v(2m+\delta)+1+\delta/m}\right]^{mi_{k-1}} \left[1 - \frac{(m+1+\delta)i_k}{\frac{t}{4}(2m+\delta)}\right]^{3mt/4}.$$

Recalling that

$$\binom{t/4}{i_k-1} = \frac{t^{i_k}}{4^{i_k}(i_k-1)!}(1+o(1)),$$

since $mi_{k-1} \le i_k$, we obtain

$$\mathbb{E}[M_k] \ge \frac{t}{2}\frac{t^{i_k}}{4^{i_k}(i_k-1)!} \left[\frac{m+\delta}{t(2m+\delta)+1+\delta/m}\right]^{i_k} \left[1 - \frac{(m+1+\delta)i_k}{\frac{t}{4}(2m+\delta)}\right]^{3mt/4}.$$

Choosing $k = k_t^-$ as in (3.4.12) and bounding $1 - x \ge \mathrm{e}^{-2x}$ for x small, as well as $m+1 \le 2m$, we obtain

$$\mathbb{E}[M_{k_t^-}] \ge \frac{t}{2}\,\frac{t^{i_{k_t^-}}}{4^{i_{k_t^-}}\, i_{k_t^-}!}\left(\frac{m}{C\,t}\right)^{i_{k_t^-}} \exp\left(-3\,c\,m\,i_{k_t^-}\right) \ge \frac{1}{(C')^{i_{k_t^-}}}\,\frac{t}{2\,i_{k_t^-}!}\exp\left(-3\,c\,m\,i_{k_t^-}\right),$$

where C is a constant and $C' = 4C/m$. Recalling that i_k is given by (3.4.4), and k_t^- by (3.4.12), hence $i_{k_t^-} = \frac{m}{m-1}m^{k_t^-}(1+o(1)) \le 2(\log t)^{1-\varepsilon}$, hence

$$i_{k_t^-}! \le \lfloor 2(\log t)^{1-\varepsilon}\rfloor! \le \left[2(\log t)^{1-\varepsilon}\right]^{2(\log t)^{1-\varepsilon}} = t^{o(1)},$$

and also $(C'\mathrm{e}^{3\,c\,m})^{i_{k_t^-}} = t^{o(1)}$. This implies that $\mathbb{E}[M_k] \to \infty$, as required. □

Remark 3.4.7 (Disjoint neighborhoods for minimally k-connected pairs)**.** We observe that, on the event $\{v, w \in \mathcal{M}_k\}$ with $v \ne w$, necessarily

$$U_{\le k}(v) \cap U_{\le k}(w) = \varnothing,$$

because if a vertex x is in $U_{\le k}(v)\cap U_{\le k}(w)$ and $x \ne v, w$, this means that $D_x(t) = m+2$, because in addition to its original m outgoing edges, vertex x has one incident edge from a younger vertex in $U_{\le k}(v)$ and one incident edge from a younger vertex in $U_{\le k}(u)$, which gives a contradiction. Similar arguments apply when $x = v$ or $x = w$.

We use the previous remark to prove the second relation in Statement 3.2.1 for the preferential attachment model.

Proposition 3.4.8 (Second moment of $M_{k_t^-}$)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. Then, for $k \in \mathbb{N}$,*

$$\mathbb{E}[M_k^2] \leq \exp\left(32mi_k^2/t\right) \mathbb{E}[M_k]^2 + \mathbb{E}[M_k]. \tag{3.4.13}$$

Consequently, for $k = k_t^-$ as in (3.4.12)*, as $t \to \infty$,*

$$\mathbb{E}[M_{k_t^-}^2] \leq (1 + o(1))\, \mathbb{E}[M_{k_t^-}]^2 \,. \tag{3.4.14}$$

Proof. We write

$$\mathbb{E}\left[M_k^2\right] = \sum_{v,w\in[t]\setminus[t/2]} \mathbb{P}\left(v, w \in \mathcal{M}_k\right) = \sum_{v\neq w} \mathbb{P}\left(v, w \in \mathcal{M}_k\right) + \mathbb{E}[M_k]. \tag{3.4.15}$$

By Remark 3.4.7, for $v \neq w$ we can write

$$\mathbb{P}\left(v, w \in \mathcal{M}_k\right) = \sum_{H_v \cap H_w = \varnothing} \mathbb{P}\left(v, w \in \mathcal{M}_k, U_{\leq k}(v) = H_v, U_{\leq k}(w) = H_w\right).$$

The crucial observation is that the event

$$\{v, w \in \mathcal{M}_k, U_{\leq k}(v) = H_v, U_{\leq k}(w) = H_w\}$$

is factorizable (recall Definition 3.4.4 and Remark 3.4.5). More precisely, in analogy with (3.4.6) and (3.4.7):

$$\mathbb{P}\left(v, w \in \mathcal{M}_k, U_{\leq k}(v) = H_v, U_{\leq k}(w) = H_w\right) = L_1(H_v, H_w) L_2(H_v, H_w), \tag{3.4.16}$$

where now

$$L_1(H_v, H_w) = \prod_{x \in H_v^o \cup H_w^o} \prod_{j=1}^{m} \frac{m+\delta}{c_{x,j}}, \tag{3.4.17}$$

$$L_2(H_v, H_w) = \prod_{x \notin H_v^o \cup H_w^o} \prod_{j=1}^{m} \left[1 - \frac{D_{H_v \cup H_w}(x-1) + |(H_v \cup H_w) \cap [x-1]|\delta}{c_{x,j}}\right].$$

To prove (3.4.16), notice that in (3.4.17) and every edge and every vertex of the graph appear. Further, the first line in (3.4.17) is the probability of the event $\{U_{\leq k}(v) = H_v, U_{\leq k}(w) = H_w\}$, while the second line is the probability that all vertices not in the two neighborhoods do not attach to the two trees.

A look at (3.4.6) shows that $L_1(H_v, H_w) = L_1(H_v) L_1(H_w)$. We now show that analogous factorization holds approximately also for L_2. Since, for every $a, b \in [0, 1]$,

with $a+b<1$, it is true that $1-(a+b)\le(1-a)(1-b)$, we can bound

$$\begin{aligned}&\left[1-\frac{D_{H_v\cup H_w}(x-1)+|(H_v\cup H_w)\cap[x-1]|\delta}{c_{x,j}}\right]\\&\quad\le\left[1-\frac{D_{H_v}(x-1)+|H_v\cap[x-1]|\delta}{c_{x,j}}\right]\left[1-\frac{D_{H_w}(x-1)+|H_w\cap[x-1]|\delta}{c_{x,j}}\right].\end{aligned}\tag{3.4.18}$$

When we plug (3.4.18) into the second line of (3.4.17), we obtain $L_2(H_v)L_2(H_w)$ (recall (3.4.7)) times the following terms:

$$\begin{aligned}&\left(\prod_{x\in H_w^o}\left[1-\frac{D_{H_v}(x-1)+|H_v\cap[x-1]|\delta}{c_{x,j}}\right]\right)^{-1}\\&\qquad\times\left(\prod_{x\in H_v^o}\left[1-\frac{D_{H_w}(x-1)+|H_w\cap[x-1]|\delta}{c_{x,j}}\right]\right)^{-1}.\end{aligned}\tag{3.4.19}$$

We can bound $D_{H_v}(x-1)+|H_v\cap[x-1]|\delta\le D_{H_v}(x-1)\le(m+1)i_k$ (recall that $\delta<0$) and analogously for H_w. The square brackets in (3.4.19) equal 1 for $x\le t/4$ (since $H_v,H_w\subseteq[t]\setminus[t/4]$ by construction), and for $x>t/4$ we have $c_{x,j}\ge\frac{t}{4}(2m+\delta)\ge\frac{m}{4}t$ by (3.1.11) and $\delta>-m$. We can thus write

$$\begin{aligned}L_2(H_v,H_w)&\le L_2(H_v)\,L_2(H_w)\prod_{x\in H_v^o\cup H_w^o}\prod_{j=1}^{m}\left[1-\frac{(m+1)i_k}{\frac{m}{4}t}\right]^{-1}\\&\le L_2(H_v)\,L_2(H_w)\exp\left(2(2i_k)m\frac{(m+1)i_k}{\frac{m}{4}t}\right),\end{aligned}$$

where we have used the bound $1-z\ge\mathrm{e}^{-2z}$ for small $z>0$. Since $m+1\le 2m$, we obtain

$$\begin{aligned}&\sum_{v\ne w}\left[\sum_{H_v\cap H_w=\varnothing}\mathbb{P}\left(v,w\in\mathcal{M}_k,U_{\le k}(v)=H_v,U_{\le k}(w)=H_w\right)\right]\\&\le\exp\left(32mi_k^2/t\right)\sum_{v\in[t]\setminus[t/2]}\sum_{H_v}L_1(H_v)L_2(H_v)\sum_{w\in[t]\setminus[t/2]}\sum_{H_w}L_1(H_w)L_2(H_w)\\&=\exp\left(32mi_k^2/t\right)\mathbb{E}[M_k]^2.\end{aligned}\tag{3.4.20}$$

Substituting (3.4.20) in (3.4.15) completes the proof of (3.4.13).

Finally, for $k=k_t^-$ as in (3.4.12) we have $i_{k_t^-}\le 2(\log t)^{1-\varepsilon}$ (recall that i_k is given by (3.4.4)). We have already shown in Proposition 3.4.6 that $\mathbb{E}[M_{k_t^-}]\to\infty$, hence (3.4.14) follows. □

Together, Propositions 3.4.6 and 3.4.8 prove Statement 3.2.1. This means, as for the configuration model, since $\mathrm{Var}(M_{k_t^-}^2)=o(\mathbb{E}[M_{k_t^-}]^2)$, that $M_{k_t^-}/\mathbb{E}[M_{k_t^-}]\xrightarrow[t\to\infty]{\mathbb{P}}1$,

so in particular $M_{k_t^-} \xrightarrow[t\to\infty]{\mathbb{P}} \infty$. □

3.4.3. Proof of Statement 3.2.3

Fix $\varepsilon > 0$ and define, as in (3.2.4),

$$\bar{k}_t = (1-\varepsilon)\frac{2\log\log t}{|\log(\tau-2)|}. \tag{3.4.21}$$

Statement 3.2.3 follows from the following result on distances between not too early vertices:

Proposition 3.4.9 (Lower bound on distances)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. Then, there exists a constant $p > 0$ such that*

$$\max_{x,y\geq \frac{t}{(\log t)^2}} \mathbb{P}\left(\mathrm{dist}_{\mathrm{PA}_t}(x,y) \leq 2\bar{k}_t\right) \leq \frac{p}{(\log t)^2}. \tag{3.4.22}$$

Inequality (3.4.22) is an adaptation of a result proved in [57, Section 4.1]. Consequently we just give a sketch of the proof (the complete proof can be found in Section 3.7).

Let us denote by $u \leftrightarrow v$ the event that vertices u, v are neighbors in PA_t, that is

$$\{u \leftrightarrow v\} = \bigcup_{j=1}^{m} \left(\{u \xrightarrow{j} v\} \cup \{v \xrightarrow{j} u\}\right).$$

(As a matter of fact, $\{v \xrightarrow{j} u\}$ is only possible if $v > u$, while $\{u \xrightarrow{j} v\}$ is only possibly if $v < u$.) Given a sequence $\pi = (\pi_0, \pi_1, \ldots, \pi_k) \in [t]^{k+1}$ of distinct vertices, we denote by $\{\pi \subseteq \mathrm{PA}_t\}$ the event that π *is a path in* PA_t, that is

$$\left\{\pi \subseteq \mathrm{PA}_t\right\} = \{\pi_0 \leftrightarrow \pi_1 \leftrightarrow \pi_2 \cdots \leftrightarrow \pi_k\} = \bigcap_{i=1}^{k} \{\pi_{i-1} \leftrightarrow \pi_i\}.$$

The proof of Proposition 3.4.9 requires the following bound on the probability of connection between two vertices from [59, Lemma 2.2]: for $\gamma = m/(2m+\delta) \in (\frac{1}{2}, 1)$, there exists $c \in (0, \infty)$ such that, for all vertices $u, v \in [t]$.

$$\mathbb{P}(u \leftrightarrow v) \leq c(u \vee v)^{\gamma-1}(u \wedge v)^{-\gamma}.$$

From [59, Corollary 2.3] we know, for any sequence $\pi = (\pi_0, \pi_1, \ldots, \pi_k) \in [t]^{k+1}$ of distinct vertices,

$$\mathbb{P}\big(\pi \subseteq \mathrm{PA}_t\big) \leq p(\pi_0, \pi_1, \ldots, \pi_k) := \prod_{i=0}^{k-1} \frac{Cm}{(\pi_i \wedge \pi_{i+1})^{\gamma}(\pi_i \vee \pi_{i+1})^{1-\gamma}}, \tag{3.4.23}$$

where C is an absolute constant. The history of (3.4.23) is that it was first proved by

Bollobás and Riordan [32] for $\delta = 0$ (so that $\gamma = 1 - \gamma = 1/2$), and the argument was extended to all δ in [59, Corollary 2.3].

Remark 3.4.10. *Proposition 3.4.9 holds for every random graphs that satisfies* (3.4.23).

We proceed in a similar way as in Section 3.3.3. Given two vertices $x, y \in [t]$, we consider paths $\pi = (\pi_0, \pi_1, \dots, \pi_k)$ between $x = \pi_0$ and $y = \pi_k$. We fix a decreasing sequence of numbers $(g_l)_{l \in \mathbb{N}_0}$ that serve as truncation values for the *age* of vertices along the path (rather than the degrees as for the configuration model). We say that a path π is *good* when $\pi_l \geq g_l \wedge g_{k-l}$ for every $l = 0, \dots, k$, and *bad* otherwise. In other words, a path is good when the age of vertices does not decrease too much from π_0 to $\pi_{k/2}$ and, backwards, from π_k to $\pi_{k/2}$. Intuitively, this also means that their degrees do not grow too fast. This means that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t}(x, y) \leq 2\bar{k}_t) \leq \sum_{k=1}^{2\bar{k}_t} \mathbb{P}(\mathcal{E}_k(x, y)) + \sum_{k=1}^{\bar{k}_t} \left[\mathbb{P}(\mathcal{F}_k(x)) + \mathbb{P}(\mathcal{F}_k(y))\right], \tag{3.4.24}$$

where $\mathcal{E}_k(x, y)$ is the event of there being a good path of length k, as in (3.3.27), while $\mathcal{F}_k(x)$ is the event of there being a path π with $\pi_i \geq g_i$ for $i \leq k - 1$ but $\pi_k < g_k$, in analogy with (3.3.28).

Recalling the definition of $p(\pi_0, \pi_1, \dots, \pi_k)$ in (3.4.23), we define for $l \in \mathbb{N}$,

$$f_{l,t}(x, w) = \mathbb{1}_{\{x \geq g_0\}} \sum_{\pi_1 = g_1}^{t} \sum_{\pi_2 = g_2}^{t} \cdots \sum_{\pi_{l-1} = g_{l-1}}^{t} p(x, \pi_1, \dots, \pi_{l-1}, w), \tag{3.4.25}$$

setting $f_{0,t}(x, w) = \mathbb{1}_{\{x \geq g_0\}}$ and $f_{1,t}(x, w) = \mathbb{1}_{\{x \geq g_0\}} p(x, w)$. From (3.4.24) we then obtain

$$\begin{aligned} \mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t}(x, y) \leq 2\bar{k}_t) \leq \; & \sum_{k=1}^{2\bar{k}_t} \sum_{l = g_{\lfloor k/2 \rfloor}}^{t} f_{\lfloor k/2 \rfloor, t}(x, l) f_{\lceil k/2 \rceil, t}(y, l) \\ & + \sum_{k=1}^{\bar{k}_t} \sum_{l=1}^{g_k - 1} f_{k,t}(x, l) + \sum_{k=1}^{\bar{k}_t} \sum_{l=1}^{g_k - 1} f_{k,t}(y, l). \end{aligned} \tag{3.4.26}$$

This is the starting point of the proof of Proposition 3.4.9. We will show in Section 3.7 that the following recursive bound holds

$$f_{k,t}(x, l) \leq \alpha_k l^{-\gamma} + \mathbb{1}_{\{l > g_{k-1}\}} \beta_k l^{\gamma - 1}, \tag{3.4.27}$$

for suitable sequences $(\alpha_k)_{k \in \mathbb{N}}$, $(\beta_k)_{k \in \mathbb{N}}$ and $(g_k)_{k \in \mathbb{N}}$ (see Definition 3.7.2) We prove recursive bounds on these sequences that guarantee that the sums in (3.4.26) satisfy the required bounds. This proof, mostly based on detailed calculations, is moved to Section 3.7.

3.4.4. Proof of Statement 3.2.2

Consider now two independent random vertices W_1^t and W_2^t that are uniformly distributed in the set of minimally-k_t^--connected vertices $\mathcal{M}_{k_t^-}$. We set

$$E_t := \left\{ \mathrm{dist}\big(U_{\leq k_t^-}(W_1^t), U_{\leq k_t^-}(W_2^t)\big) \leq 2\bar{k}_t \right\} = \left\{ \mathrm{dist}(W_1^t, W_2^t) \leq 2k_t^- + 2\bar{k}_t \right\}$$

and, in analogy with Section 3.3.2, our goal is to show that

$$\lim_{t\to\infty} \mathbb{P}(E_t) = 0.$$

We know from Statement 3.2.1 that, as $t \to \infty$,

$$\mathbb{P}\left(M_{k_t^-} \leq \frac{1}{2}\mathbb{E}[M_{k_t^-}] \right) \leq \mathbb{P}\left(|M_{k_t^-} - \mathbb{E}[M_{k_t^-}]| > \frac{1}{2}\mathbb{E}[M_{k_t^-}] \right) \leq \frac{\mathrm{Var}(M_{k_t^-})}{\frac{1}{4}\mathbb{E}[M_{k_t^-}]^2} = o(1).$$

We also define the event

$$B_t := \left\{ \max_{v\in[t]} D_v(t) \leq \sqrt{t} \right\} \tag{3.4.28}$$

and note that it is known (see [85, Theorem 8.13]) that $\lim_{t\to\infty} \mathbb{P}(B_t) = 1$. Therefore,

$$\begin{aligned}
&\mathbb{P}(E_t) \\
&= \mathbb{P}\left(E_t \cap \{M_{k_t^-} > \tfrac{1}{2}\mathbb{E}[M_{k_t^-}]\} \cap B_t \right) + o(1) \\
&= \mathbb{E}\Big[\sum_{v_1,v_2\in[t]} \mathbb{1}_{\{W_1^t = v_1, W_2^t = v_2\}} \mathbb{1}_{\{\mathrm{dist}(v_1,v_2)\leq 2k_t^- + 2\bar{k}_t\}} \mathbb{1}_{\{M_{k_t^-} > \frac{1}{2}\mathbb{E}[M_{k_t^-}]\}} \mathbb{1}_{B_t} \Big] + o(1) \\
&\leq \mathbb{E}\Big[\sum_{v_1,v_2\in[t]\setminus[t/2]} \frac{1}{M_{k_t^-}^2} \mathbb{1}_{\{v_1\in\mathcal{M}_{k_t^-}, v_2\in\mathcal{M}_{k_t^-}\}} \\
&\qquad\qquad \times \mathbb{1}_{\{\mathrm{dist}(v_1,v_2)\leq 2k_t^- + 2\bar{k}_t\}} \mathbb{1}_{\{M_{k_t^-} > \frac{1}{2}\mathbb{E}[M_{k_t^-}]\}} \mathbb{1}_{B_t} \Big] + o(1) \\
&\leq \sum_{v_1,v_2\in[t]\setminus[t/2]} \frac{\mathbb{P}\left(v_1, v_2 \in \mathcal{M}_{k_t^-},\, \mathrm{dist}(v_1,v_2) \leq 2k_t^- + 2\bar{k}_t,\, B_t \right)}{\frac{1}{4}\mathbb{E}[M_{k_t^-}]^2} + o(1).
\end{aligned} \tag{3.4.29}$$

The contribution of the terms with $v_1 = v_2$ is negligible, since it gives

$$\frac{\sum_{v_1\in[t]\setminus[t/2]} \mathbb{P}\left(v_1 \in \mathcal{M}_{k_t^-} \right)}{\frac{1}{4}\mathbb{E}[M_{k_t^-}]^2} = \frac{4}{\mathbb{E}[M_{k_t^-}]} = o(1),$$

because $\mathbb{E}[M_{k_t^-}] \to \infty$ by Proposition 3.4.6. Henceforth we restrict the sum in (3.4.29) to $v_1 \neq v_2$. Summing over the realizations H_1 and H_2 of the random neighborhoods

$U_{\le k_t^-}(v_1)$ and $U_{\le k_t^-}(v_2)$, and over paths π from an arbitrary vertex $x \in \partial H_1$ to an arbitrary vertex $y \in \partial H_2$, we obtain

$$\mathbb{P}(E_t) \le \frac{4}{\mathbb{E}[M_{k_t^-}]^2} \sum_{\substack{v_1,v_2\in[t]\setminus[t/2]\\ v_1\neq v_2}} \sum_{H_1,\,H_2\subseteq[t]\setminus[t/4]} \sum_{x\in\partial H_1,\,y\in\partial H_2} \sum_{\substack{\pi:x\to y\\ |\pi|\le 2k_t}} \mathbb{P}\left(U_{\le k_t^-}(v_1)=H_1,\, U_{\le k_t^-}(v_2)=H_2,\, \pi\subseteq \mathrm{PA}_t,\, B_t\right) \,+\, o(1). \tag{3.4.30}$$

The next proposition, proved below, decouples the probability appearing in the last expression:

Proposition 3.4.11. *There is a constant $q \in (1,\infty)$ such that, for all v_1, v_2, H_1, H_2 and π,*

$$\mathbb{P}\Big(U_{\le k_t^-}(v_1)=H_1, U_{\le k_t^-}(v_2)=H_2,\, \pi\subseteq \mathrm{PA}_t,\, B_t\Big) \le q\,\mathbb{P}\left(U_{\le k_t^-}(v_1)=H_1,\, U_{\le k_t^-}(v_2)=H_2\right)\mathbb{P}\left(\pi\subseteq \mathrm{PA}_t\right). \tag{3.4.31}$$

The proof of Proposition 3.4.11 reveals that we can take $q = 2$ for t sufficiently large. Using (3.4.31) in (3.4.30), we obtain

$$\mathbb{P}(E_t) \le \frac{4q}{\mathbb{E}[M_{k_t^-}]^2} \sum_{v_1,v_2\in[t]\setminus[t/2]} \sum_{H_1,\,H_2\subseteq[t]\setminus[t/4]} \mathbb{P}\left(U_{\le k}(v_1)=H_1, U_{\le k}(v_2)=H_2\right) \times \left\{ \sum_{x\in\partial H_1,\,y\in\partial H_2} \sum_{\substack{\pi:x\to y\\ |\pi|\le 2k_t}} \mathbb{P}\left(\pi\subseteq \mathrm{PA}_t\right) \right\}. \tag{3.4.32}$$

If we bound $\mathbb{P}(\pi \subseteq \mathrm{PA}_t) \le p(\pi)$ in (3.4.32), as in (3.4.23), *the sum over π can be rewritten as the right hand side of* (3.4.26) (recall (3.4.24)-(3.4.25)). We can thus apply Proposition 3.4.9 —because the proof of Proposition 3.4.9 really gives a bound on (3.4.26)— concluding that the sum over π is at most $p/(\log t)^2$, where the constant p is defined in Proposition 3.4.9. Since $|\partial H_1| = |\partial H_2| = m^{k_t^-} = (\log t)^{1-\varepsilon}$ (recall (3.4.12)), we finally obtain

$$\mathbb{P}(E_t) \le \frac{4q}{\mathbb{E}[M_{k_t^-}]^2}\,\frac{p(\log t)^{2(1-\varepsilon)}}{(\log t)^2}\,\mathbb{E}[M_{k_t^-}^2] = \big(1+o(1)\big)\frac{4pq}{(\log t)^{2\varepsilon}},$$

where the last step uses Proposition 3.4.8. This completes the proof that $\mathbb{P}(E_t) = o(1)$. □

Proof of Proposition 3.4.11. We recall that $H_1 \subseteq [t]\setminus[t/4]$ is a labeled directed subgraph containing v_1, such that it is an admissible realization of the neighborhood $U_{\le k_t^-}(v_1)$ of the minimally-k_t^--connected vertex v_1 (recall Definition 3.4.2); in particular, $H_1 \setminus \{v_1\} \subseteq [t/2]\setminus[t/4]$. We also recall that, for all $u \in H_1^o := H_1 \setminus \partial H_1$ and $j \in [m]$, the j-th edge of u is connected to a well specified vertex in H_1, denoted by $\theta_j^{H_1}(u)$. Analogous considerations apply to H_2.

We have to bound the probability

$$\mathbb{P}\left(U_{\leq k_t^-}(v_1) = H_1,\ U_{\leq k_t^-}(v_2) = H_2,\ \pi \subseteq \mathrm{PA}_t,\ B_t\right), \tag{3.4.33}$$

where $\pi = (\pi_0, \pi_1, \dots, \pi_k) \in [t]^{k+1}$ is a given sequence of vertices with $\pi_0 \in \partial H_1$ and $\pi_k \in \partial H_2$. The event in (3.4.33) is not factorizable, because the degrees of the vertices in the path π are not specified, hence it is not easy to evaluate its probability. To get a factorizable event, we need to give more information. For a vertex $v \in [t]$, define its *incoming neighborhood* $\mathcal{N}(v)$ by

$$\mathcal{N}(v) := \{(u, j) \in [t] \times [m] :\ u \xrightarrow{j} v\}. \tag{3.4.34}$$

The key observation is that *the knowledge of $\mathcal{N}(v)$ determines the degree $D_v(s)$ at any time $s \leq t$* (for instance, at time t we simply have $D_v(t) = |\mathcal{N}(v)| + m$).

We are going to fix the incoming neighborhoods $\mathcal{N}(\pi_1) = K_1$, …, $\mathcal{N}(\pi_{k-1}) = K_{k-1}$ of all vertices in the path π, except the extreme ones π_0 and π_k (note that $\mathcal{N}(\pi_0)$ and $\mathcal{N}(\pi_k)$ reduce to single points in H_1^o and H_2^o, respectively, because $\pi_0 \in \partial H_1$ and $\pi_k \in \partial H_2$). We emphasize that such incoming neighborhoods allow us to determine whether $\pi = (\pi_0, \dots, \pi_k)$ is a path in PA_t. Recalling the definition of the event B_t in (3.4.28), we restrict to

$$|K_i| \leq \sqrt{t}, \qquad \text{for } i \in [k-1], \tag{3.4.35}$$

and simply drop B_t from (3.4.33). We will then prove the following relation: for all $v_1, v_2, H_1, H_2, \pi = (\pi_0, \dots, \pi_k)$, and for all $K_1, \dots, K_{k-1}$ satisfying (3.4.35), we have

$$\begin{aligned}\mathbb{P}\Big(U_{\leq k_t^-}(v_1) = H_1,\ U_{\leq k_t^-}(v_2) = H_2,\ \{\mathcal{N}(\pi_1) = K_1,\ \dots, \mathcal{N}(\pi_{k-1}) = K_{k-1}\}\Big)& \\ \leq q\,\mathbb{P}\Big(U_{\leq k_t^-}(v_1) = H_1,\ U_{\leq k_t^-}(v_2) = H_2\Big)& \\ \times\,\mathbb{P}\Big(\mathcal{N}(\pi_1) = K_1,\ \dots, \mathcal{N}(\pi_{k-1}) = K_{k-1}\Big).&\end{aligned} \tag{3.4.36}$$

Our goal (3.4.31) follows by summing this relation over all $K_1, \dots, K_{k-1}$ for which $\pi \subseteq \mathrm{PA}_t$.

The first line of (3.4.36) is the probability of a factorizable event. In fact, setting for short

$$R \ :=\ \big(H_1^o \times [m]\big)\ \cup\ \big(H_2^o \times [m]\big)\ \cup\ K_1\ \cup\ \ldots\ \cup\ K_{k-1}\,,$$

the event in the first line of (3.4.36) is the intersection of the following four events (see (3.4.8)):

$$\bigcap_{u \in H_1^o} \bigcap_{j=1}^{m} \{u \xrightarrow{j} \theta_j^{H_1}(u)\}\,, \qquad \bigcap_{u \in H_2^o} \bigcap_{j=1}^{m} \{u \xrightarrow{j} \theta_j^{H_2}(u)\}\,, \qquad \bigcap_{i=1}^{k-1} \bigcap_{(u,j) \in K_i} \{u \xrightarrow{j} \pi_i\},$$

$$\bigcap_{(u,j) \in [t] \times [m] \setminus R} \{u \overset{j}{\not\to} (H_1 \cup H_2 \cup \pi^o)\}\,,$$

where we set $\pi^o := \pi \setminus \{\pi_0, \pi_k\} = (\pi_1, \dots, \pi_{k-1})$. Generalizing (3.4.9)-(3.4.10), we can rewrite the first line of (3.4.36) as follows, recalling (3.1.10):

$$\begin{aligned}
&\mathbb{P}\left(U_{\le k_t^-}(v_1) = H_1,\, U_{\le k_t^-}(v_2) = H_2,\, \{\mathcal{N}(\pi_1) = K_1, \dots, \mathcal{N}(\pi_{k-1}) = K_{k-1}\}\right) \\
&= \left\{ \prod_{u \in H_1^o} \prod_{j=1}^{m} \frac{m+\delta}{c_{u,j}} \right\} \left\{ \prod_{u \in H_2^o} \prod_{j=1}^{m} \frac{m+\delta}{c_{u,j}} \right\} \left\{ \prod_{i=1}^{k-1} \prod_{(u,j) \in K_i} \frac{D_{\pi_i}(u, j-1) + \delta}{c_{u,j}} \right\} \\
&\quad \times \left\{ \prod_{(u,j) \in [t] \times [m] \setminus R} \left(1 - \frac{D_{H_1 \cup H_2 \cup \pi^o}(u, j-1) + |(H_1 \cup H_2 \cup \pi^o) \cap [u-1]|\delta}{c_{u,j}}\right) \right\}.
\end{aligned} \tag{3.4.37}$$

We stress that $D_{\pi_i}(u, j-1)$ is *non-random*, because it is determined by K_i. Analogous considerations apply to $D_{H_1 \cup H_2 \cup \pi^o}(u, j-1)$. We have thus obtained a factorizable event.

Next we evaluate the second and third lines of (3.4.36). Looking back at (3.4.16) and (3.4.17), we have

$$\begin{aligned}
&\mathbb{P}\Big(U_{\le k_t^-}(v_1) = H_1,\, U_{\le k_t^-}(v_2) = H_2\Big) \\
&= \left\{ \prod_{u \in H_1^o} \prod_{j=1}^{m} \frac{m+\delta}{c_{u,j}} \right\} \left\{ \prod_{u \in H_2^o} \prod_{j=1}^{m} \frac{m+\delta}{c_{u,j}} \right\} \\
&\quad \times \left\{ \prod_{(u,j) \in [t] \times [m] \setminus (H_1^o \cup H_2^o) \times [m]} \left(1 - \frac{D_{H_1 \cup H_2}(u, j-1) + |(H_1 \cup H_2) \cap [u-1]|\delta}{c_{u,j}}\right) \right\}.
\end{aligned} \tag{3.4.38}$$

On the other hand,

$$\begin{aligned}
&\mathbb{P}\Big(\mathcal{N}(\pi_1) = K_1, \dots, \mathcal{N}(\pi_{k-1}) = K_{k-1}\Big) \\
&\qquad = \left\{ \prod_{i=1}^{k-1} \prod_{(u,j) \in K_i} \frac{D_{\pi_i}(u, j-1) + \delta}{c_{u,j}} \right\} \\
&\qquad \quad \times \left\{ \prod_{(u,j) \in [t] \times [m] \setminus K_1 \cup \dots \cup K_{k-1}} \left(1 - \frac{D_{\pi^o}(u, j-1) + |\pi^o \cap [u-1]|\delta}{c_{u,j}}\right) \right\}.
\end{aligned} \tag{3.4.39}$$

Using the bound $(1 - (a+b)) \le (1-a)(1-b)$ in the second line of (3.4.37), and comparing with (3.4.38)-(3.4.39), we only need to take into account the missing terms in the product in the last lines. This shows that relation (3.4.36) holds if one sets $q = C_1\, C_2$ therein, where

$$C_1 := \left\{ \prod_{(u,j) \in K_1 \cup \dots \cup K_{k-1}} \left(1 - \frac{D_{H_1 \cup H_2}(u, j-1) + |(H_1 \cup H_2) \cap [u-1]|\delta}{c_{u,j}}\right) \right\}^{-1},$$

$$C_2 := \left\{ \prod_{(u,j)\in(H_1^o\cup H_2^o)\times[m]} \left(1 - \frac{D_{\pi^o}(u,j-1) + |\pi^o \cap [u-1]|\delta}{c_{u,j}}\right)\right\}^{-1}.$$

To complete the proof, it is enough to give uniform upper bounds on C_1 and C_2, that does not depend on H_1, H_2, π. We start with C_1. In the product we may assume $u > t/4$, because the terms with $u \le t/4$ are identically one, since $H_1, H_2 \subseteq [t] \setminus [t/4]$. Moreover, for $u > t/4$ we have $c_{u,j} \ge t(2m+\delta)/4 \ge mt/4$ by (3.1.11) and $\delta > -m$. Since $D_{H_1\cup H_2}(u, j-1) \le 2(m+1)i_k$, using $1 - x \ge \mathrm{e}^{-2x}$ for x small and recalling that $\delta < 0$, it follows that

$$C_1^{-1} \ge \prod_{(u,j)\in K_1 \cup \ldots \cup K_{k-1}} \left(1 - \frac{2(m+1)i_k}{\frac{m}{4}t}\right) \ge \mathrm{e}^{-\frac{8(m+1)}{tm}|K_{[k-1]}|i_k},$$

where $K_{[k-1]} = K_1 \cup \ldots \cup K_{k-1}$. Since i_k is given by (3.4.4), for $k = k_t^-$ as in (3.4.12) we have $i_k = \frac{m}{m-1} m^{k_t^-}(1 + o(1)) \le 2(\log t)^{1-\varepsilon}$. Recalling also (3.4.35) and bounding $m + 1 \le 2m$, we obtain

$$C_1 \le \mathrm{e}^{\frac{8(m+1)}{tm}|K_{[k-1]}|i_k} \le \mathrm{e}^{16k\, i_k/\sqrt{t}} = \mathrm{e}^{O(\log t/\sqrt{t})} = 1 + o(1)\,.$$

For C_2, since $D_{\pi^o}(u, j-1) \le D_{\pi^o}(t) = |K_{[k-1]}| \le k\sqrt{t}$, again by (3.4.35), we get

$$C_2^{-1} \ge \prod_{(u,j)\in(H_1^o\cup H_2^o)\times[m]} \left(1 - \frac{k\sqrt{t}}{\frac{m}{4}t}\right) \ge \mathrm{e}^{-\frac{8}{m}\frac{k}{\sqrt{t}}|H_1^o\cup H_2^o|m} \ge \mathrm{e}^{-16\,k\,i_k/\sqrt{t}} = 1 - o(1)\,.$$

It follows that C_1C_2 is bounded from above by some constant q. This completes the proof. □

3.4.5. Proof of Theorem 3.1.6

Dereich, Mönch and Mörters [57] have already proved the upper bound. For the lower bound we use Proposition 3.4.9. In fact, for $\bar{k}_t$ as in (3.4.21),

$$\mathbb{P}\left(H_t \le 2\bar{k}_t\right) = \sum_{v_1,v_2\in[t]} \mathbb{P}\left(V_1 = v_1, V_2 = v_2, \mathrm{dist}(v_1,v_2) \le 2\bar{k}_t\right).$$

If v_1 and v_2 are both larger or equal than $g_0 = \lceil \frac{t}{(\log t)^2} \rceil$, then we can apply Proposition 3.4.9. The probability that $V_1 < g_0$ or $V_2 < g_0$ is

$$\mathbb{P}\left(\{V_1 < g_0\} \cup \{V_2 < g_0\}\right) \le 2g_0/t = o(1),$$

hence we get

$$\frac{1}{t^2} \sum_{v_1,v_2\in[t]\setminus[g_0]} \mathbb{P}\left(\mathrm{dist}(v_1,v_2) \le 2\bar{k}_t\right) + o(1) \le \frac{(t-g_0)^2}{t^2}\frac{p}{(\log t)^2} + o(1) = o(1),$$

and this completes the proof of Theorem 3.1.6. □

3.5. Upper bound for CM

In this section we prove Statements 3.2.5 and 3.2.6 for the configuration model. By the discussion in Section 3.2.2, this completes the proof of the upper bound in Theorem 3.1.3, because the proof of Statement 3.2.4 is already known in the literature, as explained below Statement 3.2.4.

Throughout this section, the assumptions of Theorem 3.1.3 apply. In particular, we work on a configuration model CM_n, with $\tau \in (2,3)$ and $d_{\min} \geq 3$.

3.5.1. Proof of Statement 3.2.5

We first recall what Core_n is, and define the k-exploration graph. Recall from (3.2.8) that, for CM_n, Core_n is defined as

$$\mathrm{Core}_n = \{i \in [n] \text{ such that } d_i > (\log n)^\sigma\}\,,$$

where $\sigma > 1/(3-\tau)$. Since the degrees d_i are fixed in the configuration model, Core_n is a deterministic subset.

For any $v \in [n]$, we recall that $U_{\leq k}(v) \subseteq [n]$ denotes the subgraph of CM_n consisting of the vertices at distance at most k from v. We next consider the *k-exploration graph* $\widehat{U}_{\leq k}(v)$ as a modification of $U_{\leq k}(v)$, where we only explore $d_{\min}$ half-edges of the starting vertex v, and only $d_{\min}-1$ for the following vertices:

Definition 3.5.1 (k-exploration graph in CM_n)**.** *The* k-exploration graph *of a vertex v is the subgraph $\widehat{U}_{\leq k}(v)$ built iteratively as follows:*

- ▷ *Starting from $\widehat{U}_{\leq 0}(v) = \{v\}$, we consider the first $d_{\min}$ half-edges of v and we pair them, one by one, to a uniformly chosen unpaired half-edge (see Remark 3.3.1), to obtain $\widehat{U}_{\leq 1}(v)$.*
- ▷ *Assume that we have built $\widehat{U}_{\leq \ell}(v)$, for $\ell \geq 1$, and set $\widehat{U}_{=\ell}(v) := \widehat{U}_{\leq \ell}(v) \setminus \widehat{U}_{\leq (\ell-1)}(v)$. For each vertex in $\widehat{U}_{=\ell}(v)$, we consider the first $d_{\min}-1$* unpaired *half-edges and we pair them, one by one, to a uniformly chosen unpaired half-edge, to obtain $\widehat{U}_{\leq (\ell+1)}(v)$. (Note that, by construction, each vertex in $\widehat{U}_{=\ell}(v)$ has* at least *one already paired half-edge.)*

Definition 3.5.2 (Collision)**.** *In the process of building the k-exploration graph $\widehat{U}_{\leq k}(v)$, we say that there is a* collision *when a half-edge is paired to a vertex already included in the k-exploration graph.*

We now prove Statement 3.2.5. Let us fix $\varepsilon > 0$ and set

$$k_n^+ = (1+\varepsilon)\,\frac{\log\log n}{\log(d_{\min}-1)}. \tag{3.5.1}$$

Proposition 3.5.3 (At most one collision)**.** *Under the assumption of Theorem 3.1.3, the following holds with high probability: the k_n^+-exploration graph of* every *vertex either intersects* Core_n*, or it has at most one collision.*

Proof. Let us fix a vertex $v \in [n]$. We are going to estimate the probability

$$q_n(v) := \mathbb{P}\Big(\text{there are at least 2 collisions in } \widehat{U}_{\le k_n^+}(v) \text{ and } \widehat{U}_{\le k_n^+}(v) \cap \mathrm{Core}_n = \varnothing\Big)\,.$$

If we show that $\sup_{v\in[n]} q_n(v) = o(1/n)$, then it follows that $\sum_{v\in[n]} q_n(v) = o(1)$, completing the proof.

Starting from the vertex v, we pair successively one half-edge after the other, as described in Definition 3.5.1 (recall also Remark 3.3.1). In order to build $\widehat{U}_{\le k_n^+}(v)$, we need to make a number of pairings, denoted by $\mathcal{N}$, which is *random*, because collisions may occur. In fact, when there are no collisions, $\mathcal{N}$ is deterministic and takes its maximal value given by $i_{k_n^+}$ in (3.3.1), therefore

$$\mathcal{N} \le i_{k_n^+} \le \frac{d_{\min}}{d_{\min}-2}(d_{\min}-1)^{k_n^+} \le 3\,(\log n)^{1+\varepsilon}\,.$$

Introducing the event $C_i :=$ "there is a collision when pairing the i-th half-edge", we can write

$$\begin{aligned} q_n(v) &\le \mathbb{E}\Bigg[\sum_{1\le i<j\le \mathcal{N}} \mathbb{1}_{\{C_i,\, C_j,\, \widehat{U}_{\le k_n^+}(v)\cap \mathrm{Core}_n = \varnothing\}}\Bigg] \\ &= \sum_{1\le i<j\le 3(\log n)^{1+\varepsilon}} \mathbb{P}\big(C_i,\, C_j,\, j\le \mathcal{N},\, \widehat{U}_{\le k_n^+}(v)\cap \mathrm{Core}_n = \varnothing\big)\,. \end{aligned} \tag{3.5.2}$$

Let E_ℓ be the event that the first ℓ half-edges are paired to vertices with degree $\le (\log n)^\sigma$ (i.e., the graph obtained after pairing the first ℓ half-edges is disjoint from Core_n). Then

$$\begin{aligned} \mathbb{P}\big(C_i,\, C_j,\, j\le \mathcal{N},\, \widehat{U}_{\le k_n^+}(v)\cap \mathrm{Core}_n = \varnothing\big) &\le \mathbb{P}\big(C_i,\, C_j,\, E_{j-1}\big) \\ &= \mathbb{P}(E_{i-1})\,\mathbb{P}(C_i \,|\, E_{i-1})\,\mathbb{P}(C_j \,|\, C_i,\, E_{j-1})\,. \end{aligned} \tag{3.5.3}$$

On the event E_{i-1}, before pairing the i-th half-edge, the graph is composed by at most $i-1$ vertices, each with degree at most $(\log n)^\sigma$, hence, for $i \le 3(\log n)^{1+\varepsilon}$,

$$\mathbb{P}(C_i \,|\, E_{i-1}) \le \frac{(i-1)(\log n)^\sigma}{\ell_n - 2i + 1} \le \frac{3(\log n)^{1+\varepsilon}(\log n)^\sigma}{\ell_n - 6(\log n)^{1+\varepsilon}} \le c\,\frac{(\log n)^{\sigma+1+\varepsilon}}{n}\,,$$

for some $c \in (0,\infty)$, thanks to $\ell_n = n\mu(1+o(1))$ (recall (3.3.4)). The same arguments show that

$$\mathbb{P}(C_j \,|\, C_i,\, E_{j-1}) \le c\,\frac{(\log n)^{\sigma+1+\varepsilon}}{n}\,.$$

Looking back at (3.5.2)-(3.5.3), we obtain

$$\sup_{v\in[n]} q_n(v) \le \sum_{1\le i<j\le 3(\log n)^{1+\varepsilon}} c^2\,\frac{(\log n)^{2(\sigma+1+\varepsilon)}}{n^2} \le 9\,c^2\,\frac{(\log n)^{2\sigma+4(1+\varepsilon)}}{n^2} = o\Big(\frac{1}{n}\Big)\,,$$

which completes the proof. □

Corollary 3.5.4 (Large boundaries)**.** *Under the assumptions of Theorem 3.1.3 and on the event $\widehat{U}_{\le k_n^+}(v)\cap \mathrm{Core}_n = \varnothing$, with high probability, the boundary $\widehat{U}_{=k_n^+}(v)$ of the k_n^+-exploration graph of any vertex $v\in[n]$ contains at least $(d_{\min}-2)(d_{\min}-1)^{k_n^+-1} \ge \frac{1}{2}(\log n)^{1+\varepsilon}$ vertices, each one with at least two unpaired half-edges.*

Proof. By Proposition 3.5.3, with high probability, every k_n^+-exploration graph has at most one collision before hitting Core_n. The worst case is when the collision happens immediately, i.e. a half-edge incident to v is paired to another half-edge incident to v: in this case, removing both half-edges, the k_n^+-exploration graph becomes a tree with $(d_{\min}-2)(d_{\min}-1)^{k_n^+-1}$ vertices on its boundary, each of which has at least $(d_{\min}-1)\ge 2$ yet unpaired half-edges. Since $(d_{\min}-2)/(d_{\min}-1)\ge\frac{1}{2}$ for $d_{\min}\ge 3$, and moreover $(d_{\min}-1)^{k_n^+}=(\log n)^{1+\varepsilon}$ by (3.5.1), we obtain the claimed bound.

If the collision happens at a later stage, i.e. for a half-edge incident to a vertex different from the starting vertex v, then we just remove the branch from v to that vertex, getting a tree with $(d_{\min}-1)(d_{\min}-1)^{k_n^+-1}$ vertices on its boundary. The conclusion follows. □

Together, Proposition 3.5.3 and Corollary 3.5.4 prove Statement 3.2.5. □

3.5.2. Proof of Statement 3.2.6

Consider the k_n^+-exploration graph $\widehat{U}=\widehat{U}_{\le k_n^+}(v)$ of a fixed vertex $v\in[n]$, as in Definition 3.5.1, and let $x_1,\dots,x_N$ be the (random) vertices on its boundary. We stress that, by Corollary 3.5.4, with high probability $N\ge\frac{1}{2}(\log n)^{1+\varepsilon}$. Set

$$h_n = \big\lceil B\log\log\log n + C\big\rceil\,, \tag{3.5.4}$$

where B, C are fixed constants, to be determined later on.

Henceforth we fix a realization H of $\widehat{U}=\widehat{U}_{\le k_n^+}(v)$ and we work *conditionally* on the event $\{\widehat{U}=H\}$. By Remark 3.3.1, we can complete the construction of the configuration model CM_n by pairing uniformly all the yet unpaired half-edges. We do this as follows: for each vertex $x_1,\dots,x_N$ on the boundary of $\widehat{U}$, we explore its neighborhood, looking for *fresh* vertices with higher and higher degree, up to distance h_n (we call a vertex *fresh* if it is connected to the graph for the first time, hence it only has one paired half-edge). We now describe this procedure in detail:

Definition 3.5.5 (Exploration procedure)**.** *Let $x_1,\dots,x_N$ denote the vertices on the boundary of a k_n^+-exploration graph $\widehat{U}=\widehat{U}_{\le k_n^+}(v)$. We start the exploration procedure from x_1.*

- ▷ *Step 1. We set* $v_0^{(1)} := x_1$ *and we pair all its unpaired half-edges.* Among the fresh vertices *to which* $v_0^{(1)}$ *has been connected, we call* v_1 *the one with maximal degree.*
- ▷ *When there are no fresh vertices at some step, the procedure for* x_1 *stops.*
- ▷ *Step 2. Assuming we have built* $v_1^{(1)}$*, we pair all its unpaired half-edges: among the fresh connected vertices, we denote by* $v_2^{(1)}$ *the vertex with maximal degree.*
- ▷ *We continue in this way for (at most)* h_n *steps, defining* $v_j^{(1)}$ *for* $0 \le j \le h_n$ *(recall* (3.5.4)*).*

After finishing the procedure for x_1*, we perform the same procedure for* $x_2, x_3, \dots, x_N$*, defining the vertices* $v_0^{(i)}, v_1^{(i)}, \dots, v_{h_n}^{(i)}$ *starting from* $v_0^{(i)} = x_i$*.*

Definition 3.5.6 (Success)**.** *Let* $x_1, \dots, x_N$ *be the vertices on the boundary of a* k_n^+*-exploration graph* $\widehat{U} = \widehat{U}_{\le k_n^+}(v)$*. We define the event* $S_{x_i} :=$ *"*x_i *is a* success*" by*

$$S_{x_i} := \big\{\{v_0^{(i)}, v_1^{(i)}, \dots, v_{h_n}^{(i)}\} \cap \mathrm{Core}_n \neq \varnothing\big\} = \big\{d_{v_j^{(i)}} > (\log n)^\sigma \text{ for some } 0 \le j \le h_n\big\}\,.$$

Here is the key result, proved below:

Proposition 3.5.7 (Hitting the core quickly)**.** *There exists a constant* $\eta > 0$ *such that, for every* $n \in \mathbb{N}$ *and for every realization* H *of* $\widehat{U}$*,*

$$\mathbb{P}\big(S_{x_1} \,\big|\, \widehat{U} = H\big) \ge \eta, \tag{3.5.5}$$

and, for each $i = 2, \dots, N$*,*

$$\mathbb{P}\big(S_{x_i} \,\big|\, \widehat{U} = H\,,\, S_{x_1}^c\,,\, \dots\,,\, S_{x_{i-1}}^c\big) \ge \eta. \tag{3.5.6}$$

This directly leads to the proof of Statement 3.2.6, as the following corollary shows:

Corollary 3.5.8 (Distance between periphery and Core_n)**.** *Under the hypotheses of Theorem 3.1.3, with high probability, the distance of every vertex in the graph from* Core_n *is at most*

$$(1+\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)} + o\left(\log\log n\right). \tag{3.5.7}$$

Proof. By Corollary 3.5.4, with high probability, every vertex $v \in [n]$ either is at distance at most k_n^+ from Core_n, or has a k_n^+-exploration graph $\widehat{U} = \widehat{U}_{\le k_n^+}(v)$ with at least $N \ge \frac{1}{2}(\log n)^{1+\varepsilon}$ vertices on its boundary. It suffices to consider the latter case. Conditionally on $\widehat{U} = H$, the probability that none of these vertices is a success can be bounded by Proposition 3.5.7:

$$\begin{aligned}\mathbb{P}\big(S_{x_1}^c \cap \dots \cap S_{x_N}^c \,\big|\, \widehat{U} = H\big) &= \mathbb{P}\big(S_{x_1}^c \,\big|\, \widehat{U} = H\big) \prod_{j=2}^{N} \mathbb{P}\big(S_{x_j}^c \,\big|\, \widehat{U} = H\,,\, S_{x_1}^c\,,\, \dots\,,\, S_{x_{j-1}}^c\big) \\ &\le (1-\eta)^N \le (1-\eta)^{\frac{1}{2}(\log n)^{1+\varepsilon}} = o(1/n)\,.\end{aligned} \tag{3.5.8}$$

This is uniform over H, hence the probability that no vertex is a success, without conditioning, is still $o(1/n)$. It follows that, with high probability, every $v \in [n]$ has at least one successful vertex on the boundary of its k_n^+-exploration graph. This means that the distance of every vertex $v \in [n]$ from Core_n is at most $k_n^+ + h_n = k_n^+ + o(\log\log n)$, by (3.5.4). Recalling (3.5.1), we have completed the proof of Corollary 3.5.8 and thus of Statement 3.2.6. □

To prove Proposition 3.5.7, we need the following technical (but simple) result:

Lemma 3.5.9 (High-degree fresh vertices)**.** *Consider the process of building a configuration model* CM_n *as described in Remark 3.3.1. Let* $\mathcal{G}_l$ *be the random graph obtained after* l *pairings of half-edges and let* V_l *be the random vertex incident to the half-edge to which the* l*-th half-edge is paired. For all* $l, n \in \mathbb{N}$ *and* $z \in [0, \infty)$ *such that*

$$l \leq \frac{n}{4}(1 - F_{\boldsymbol{d},n}(z)), \tag{3.5.9}$$

the following holds:

$$\mathbb{P}\left(d_{V_{l+1}} > z\,,\, V_{l+1} \notin \mathcal{G}_l \,\middle|\, \mathcal{G}_l\right) \geq z[1 - F_{\boldsymbol{d},n}(z)]\frac{n}{2\ell_n}. \tag{3.5.10}$$

In particular, when Conditions 3.1.1 and 3.1.2 hold, for every $\zeta > 0$ *there are* $c > 0$, $n_0 < \infty$ *such that* $\forall\, n \geq n_0\,,\ 0 \leq z \leq n^{1/3}\,,\ l \leq n^{1/3}$,

$$\mathbb{P}\left(d_{V_{l+1}} > z\,,\, V_{l+1} \notin \mathcal{G}_l \,\middle|\, \mathcal{G}_l\right) \geq \frac{c}{z^{\tau-2+\zeta}}\,. \tag{3.5.11}$$

Proof. By definition of CM_n, the $(l+1)$-st half-edge is paired to a uniformly chosen half-edge among the $\ell_n - 2l - 1$ that are not yet paired. Consequently

$$\mathbb{P}\left(d_{V_{l+1}} > z\,,\, V_{l+1} \notin \mathcal{G}_l \,\middle|\, \mathcal{G}_l\right) = \frac{1}{\ell_n - 2l - 1}\sum_{v \notin \mathcal{G}_l} d_v\, \mathbb{1}_{\{d_v > z\}}.$$

Since $|\mathcal{G}_l| \leq 2l \leq \frac{n}{2}(1 - F_{\boldsymbol{d},n}(z))$ by (3.5.9), we obtain

$$\frac{1}{\ell_n - 2l - 1}\sum_{v \notin \mathcal{G}_l} d_v\, \mathbb{1}_{\{d_v > z\}} \geq \frac{z}{\ell_n}\big(n(1 - F_{\boldsymbol{d},n}(z)) - |\mathcal{G}_l|\big) \geq z(1 - F_{\boldsymbol{d},n}(z))\frac{n}{2\ell_n},$$

which proves (3.5.10).

Assuming Conditions 3.1.1 and 3.1.2, we have $\ell_n = \mu n(1 + o(1))$, with $\mu \in (0, \infty)$, see (3.3.4), and there are $c_1 > 0$ and $\alpha > 1/2$ such that $1 - F_{\boldsymbol{d},n}(z) \geq c_1\, z^{-(\tau-1)}$ for $0 \leq z \leq n^{\alpha}$. Consequently, for $0 \leq z \leq n^{1/3}$, the right hand side of (3.5.9) is at least $\frac{n}{4}\frac{c_1}{n^{(\tau-1)/3}}$. Note that $(\tau-1)/3 < 2/3$ (because $\tau < 3$), hence we can choose n_0 so that $\frac{n}{4}\frac{c_1}{n^{(\tau-1)/3}} \geq n^{1/3}$ for all $n \geq n_0$. This directly leads to (3.5.11). □

With Lemma 3.5.9 in hand, we are able to prove Proposition 3.5.7:

Proof of Proposition 3.5.7. We fix $v \in [n]$ and a realization H of $\widehat{U} = \widehat{U}_{\le k_n^+}(v)$. We abbreviate

$$\mathbb{P}^*(\,\cdot\,) := \mathbb{P}(\,\cdot \,|\, \widehat{U} = H)\,. \tag{3.5.12}$$

The vertices on the boundary of $\widehat{U}$ are denoted by $x_1, \dots, x_N$. We start proving (3.5.5), hence we focus on x_1 and we define $v_0^{(1)}, v_1^{(1)}, \dots, v_{h_n}^{(1)}$ as in Definition 3.5.5, with $v_0^{(1)} = x_1$.

We first fix some parameters. Since $2 < \tau < 3$, we can choose $\zeta, \gamma > 0$ small enough so that

$$\xi := 1 - \mathrm{e}^{\gamma}(\tau - 2 + \zeta) > 0\,. \tag{3.5.13}$$

Next we define a sequence $(g_\ell)_{\ell \in \mathbb{N}_0}$ that grows *doubly exponentially* fast:

$$g_\ell := 2^{\mathrm{e}^{\gamma \ell}} = \exp\big((\log 2)\exp(\gamma\,\ell)\big)\,. \tag{3.5.14}$$

Then we fix $B = 1/\gamma$ and $C = \log(\sigma/\log 2)$ in (3.5.4), where σ is the same constant as in Core_n, see (3.2.8). With these choices, we have

$$g_{h_n} = \mathrm{e}^{\sigma \mathrm{e}^{\lceil \log\log\log n\rceil}} > \mathrm{e}^{\sigma \log\log n} = (\log n)^\sigma\,, \quad \text{while} \quad g_{h_n - 1} < (\log n)^\sigma\,. \tag{3.5.15}$$

Roughly speaking, the idea is to show that, with positive probability, one has $d_{v_j^{(1)}} > g_j$. As a consequence, $d_{v_{h_n}^{(1)}} > g_{h_n} \ge (\log n)^\sigma$, that is $v_{h_n}^{(1)}$ belongs to Core_n and x_1 is a success. The situation is actually more involved, since we can only show that $d_{v_j^{(1)}} > g_j$ before reaching Core_n.

Let us make the above intuition precise. Recalling (3.5.12), let us set

$$H_{-1} := \varnothing\,, \qquad H_0 := H\,, \qquad H_k := H \cup \{v_1^{(1)}, \dots, v_k^{(1)}\} \quad \text{for } 1 \le k \le h_n\,.$$

Then we introduce the events

$$T_\ell := \bigcup_{k=0}^{\ell} \big\{d_{v_k^{(1)}} > (\log n)^\sigma\big\}\,, \qquad W_\ell := \bigcap_{k=0}^{\ell} \big\{d_{v_k^{(1)}} > g_k\,,\ v_k^{(1)} \notin H_{k-1}\big\}\,. \tag{3.5.16}$$

In words, the event T_ℓ means that one of the vertices $v_0^{(1)}, \dots, v_\ell^{(1)}$ has already reached Core_n, while the event W_ℓ means that the degrees of vertices $v_0^{(1)}, \dots, v_\ell^{(1)}$ grow at least like $g_0, \dots, g_\ell$ and, furthermore, each v_k is a fresh vertex (this is actually already implied by Definition 3.5.5, otherwise v_k would not even be defined). We finally set

$$E_0 := W_0\,, \qquad E_j := T_{j-1} \cup W_j \quad \text{for } 1 \le j \le h_n\,.$$

Note that T_{h_n} coincides with $S_{x_1} =$ "x_1 is a success". Also note that $W_{h_n} \subseteq \{d_{v_{h_n}^{(1)}} > (\log n)^\sigma\}$, because $d_{v_{h_n}^{(1)}} > g_{h_n} > (\log n)^\sigma$ by (3.5.15), hence

$$E_{h_n} = T_{h_n - 1} \cup W_{h_n} \subseteq T_{h_n - 1} \cup \{d_{v_{h_n}^{(1)}} > (\log n)^\sigma\} = T_{h_n} = S_{x_1}\,.$$

Consequently, if we prove that $\mathbb{P}^*(E_{h_n}) \geq \eta$, then our goal $\mathbb{P}^*(S_{x_1}) \geq \eta$ follows (recall (3.5.5)).

The reason for working with the events E_j is that their probabilities can be controlled by an induction argument. Recalling (3.5.12), we can write

$$\begin{aligned}\mathbb{P}^*(E_{j+1}) &= \mathbb{P}^*(T_j) + \mathbb{P}^*(T_j^c \cap W_{j+1}) \\ &= \mathbb{P}^*(T_j) + \mathbb{P}\big(d_{v_{j+1}^{(1)}} > g_{j+1}\,,\, v_{j+1}^{(1)} \notin H_j \,\big|\, \{\widehat{U} = H\} \cap T_j^c \cap W_j\big)\, \mathbb{P}^*(T_j^c \cap W_j)\,.\end{aligned}$$

The key point is the following estimate on the conditional probability, proved below:

$$\mathbb{P}\big(d_{v_{j+1}^{(1)}} > g_{j+1}\,,\, v_{j+1}^{(1)} \notin H_j \,\big|\, \{\widehat{U} = H\} \cap T_j^c \cap W_j\big) \;\geq\; 1 - \mathrm{e}^{-c(g_j)^{\xi}/2} =: 1 - \varepsilon_j\,, \quad (3.5.17)$$

with $\xi > 0$ is defined in (3.5.13) and $c > 0$ is the constant appearing in relation (3.5.11). This yields

$$\begin{aligned}\mathbb{P}^*(E_{j+1}) \;&\geq\; \mathbb{P}^*(T_j) + (1-\varepsilon_j)\,\mathbb{P}^*(T_j^c \cap W_j) \;\geq\; (1-\varepsilon_j)\big(\mathbb{P}^*(T_j) + \mathbb{P}^*(T_j^c \cap W_j)\big) \\ &=\; (1-\varepsilon_j)\mathbb{P}^*(T_j \cup W_j) \;\geq\; (1-\varepsilon_j)\mathbb{P}^*(T_{j-1} \cup W_j) \\ &=\; (1-\varepsilon_j)\mathbb{P}^*(E_j)\,,\end{aligned}$$

which leads us to

$$\mathbb{P}^*(E_{h_n}) \geq \mathbb{P}^*(E_0) \prod_{j=0}^{h_n-1} (1-\varepsilon_j) \geq \mathbb{P}^*(E_0) \prod_{j=0}^{\infty} (1-\varepsilon_j) =: \eta\,.$$

Since $\sum_{j\geq 0} \varepsilon_j < \infty$ and $\varepsilon_j < 1$ for every $j \geq 0$, by (3.5.17) and (3.5.14), the infinite product is strictly positive. Also note that $\mathbb{P}^*(E_0) = \mathbb{P}^*(d_{v_0^{(1)}} \geq 2) = 1$, because $g_0 = 2$ and $d_{v_0^{(1)}} \geq d_{\min} \geq 3$. Then $\eta > 0$, as required.

It remains to prove (3.5.17). To lighten notation, we rewrite the left hand side of (3.5.17) as

$$\begin{aligned} &q_{j+1} := \mathbb{P}\big(d_{v_{j+1}^{(1)}} > g_{j+1}\,,\, v_{j+1}^{(1)} \notin H_j \,\big|\, D_j\big)\,, \\ &\text{where} \quad D_j := \{\widehat{U} = H\} \cap T_j^c \cap W_j\,. \end{aligned} \quad (3.5.18)$$

Note that, on the event $D_j \subseteq W_j$, vertex $v_j^{(1)}$ is fresh (i.e., it is connected to the graph for the first time), hence it has $m = d_{v_j^{(1)}} - 1$ unpaired half-edges. These are paired uniformly, connecting $v_j^{(1)}$ to (not necessarily distinct) vertices $w^{(1)}, \ldots, w^{(m)}$. Let us introduce for $1 \leq \ell \leq m$ the event

$$C_\ell := \bigcap_{k=1}^{\ell} \big\{d_{w^{(k)}} > g_{j+1}\,,\, w^{(k)} \notin H_j\big\}^c\,. \quad (3.5.19)$$

By Definition 3.5.5, $v_{j+1}^{(1)}$ is the *fresh* vertex with maximal degree among them, hence

$$\big\{d_{v_{j+1}^{(1)}} > g_{j+1}\,,\, v_{j+1}^{(1)} \notin H_j\big\}^c = C_m\,.$$

Since $m = d_{v_j^{(1)}} - 1 > g_j - 1$ on $W_j \subseteq D_j$, the left hand side of (3.5.17) can be estimated by

$$\begin{aligned} q_{j+1} &= 1 - \mathbb{P}\big(C_m \,\big|\, D_j\big) \geq 1 - \prod_{k=1}^{g_j-1} \mathbb{P}\big(C_k \,\big|\, D_j \cap C_{k-1}\big) \\ &= 1 - \prod_{k=1}^{g_j-1} \Big(1 - \mathbb{P}\big(d_{w^{(k)}} > g_{j+1}\,,\, w^{(k)} \notin H_j \,\big|\, D_j \cap C_{k-1}\big)\Big)\,. \end{aligned} \tag{3.5.20}$$

We claim that we can apply relation (3.5.11) from Lemma 3.5.9 to each of the probabilities in the last line of (3.5.20). To justify this claim, we need to look at the conditioning event $D_j \cap C_{k-1}$, recalling (3.5.19), (3.5.18) and (3.5.16). In order to produce it, we have to do the following:

- ▷ First we build the k_n^+-exploration graph $\widehat{U}_{\leq k_n^+}(v) = H$, which requires to pair at most $O((d_{\min}-1)^{k_n^+}) = O((\log n)^{1+\varepsilon})$ half-edges (recall Definition 3.5.1);
- ▷ Next, starting from the boundary vertex x_1, we generate the fresh vertices $v_0^{(1)}$, $\ldots$, $v_j^{(1)}$ *all outside* Core_n, because we are on the event T_j^c, and this requires to pair a number of half-edges which is at most

$$(\log n)^\sigma j \leq (\log n)^\sigma h_n = O((\log n)^{\sigma+1});$$

- ▷ Finally, in order to generate $w^{(1)}, \ldots, w^{(k-1)}$, we pair exactly $k-1$ half-edges, and note that $k - 1 \leq g_j - 1 \leq g_{h_n} - 1 = O((\log n)^\sigma)$ (always because $v_j \notin \mathrm{Core}_n$).

It follows that the conditioning event $D_j \cap C_{k-1}$ is in the σ-algebra generated by $\mathcal{G}_l$ for $l \leq O((\log n)^{1+\sigma+\varepsilon})$ (we use the notation of Lemma 3.5.9). In particular, $l \leq n^{1/3}$. Also note that $z = g_{j+1} \leq g_{h_n} = O((\log n)^\sigma)$, see (3.5.15), hence also $z \leq n^{1/3}$. Applying (3.5.11), we get

$$\begin{aligned} q_{j+1} &\geq 1 - \left(1 - \frac{c}{(g_{j+1})^{\tau-2+\zeta}}\right)^{g_j-1} \geq 1 - \exp\left(-c\frac{g_j-1}{(g_{j+1})^{\tau-2+\zeta}}\right) \\ &\geq 1 - \exp\left(-\frac{c}{2}\,\frac{g_j}{(g_{j+1})^{\tau-2+\zeta}}\right) \end{aligned} \tag{3.5.21}$$

because $1 - x \leq \mathrm{e}^{-x}$ and $n - 1 \geq n/2$ for all $n \geq 2$ (note that $g_j \geq g_0 = 2$). Since $g_{j+1} = (g_j)^{\mathrm{e}^\gamma}$, by (3.5.14), we finally arrive at

$$q_{j+1} \geq 1 - \exp\left(-\frac{c}{2}\,(g_j)^{1-\mathrm{e}^\gamma(\tau-2+\zeta)}\right) = 1 - \mathrm{e}^{-c\,(g_j)^\xi/2}\,, \tag{3.5.22}$$

which is precisely (3.5.17). This completes the proof of (3.5.5).

In order to prove (3.5.6), we proceed in the same way: for any fixed $2 \le i \le N$, we start from the modification of (3.5.12) given by $\mathbb{P}^*(\,\cdot\,) := \mathbb{P}(\,\cdot \mid \widehat{U} = H\,,\, S^c_{x_1}\,,\,\ldots\,,\, S^c_{x_{i-1}})$ and we follow the same proof, working with the vertices $v_1^{(i)}$, …, $v_{h_n}^{(i)}$ instead of $v_1^{(1)}, \ldots, v_{h_n}^{(1)}$ (recall Definition 3.5.5). We leave the details to the reader. □

3.6. Upper bound for PAM

In this section we prove Statements 3.2.5 and 3.2.6 for the preferential attachment model. By the discussion in Section 3.2.2, this completes the proof of the upper bound in Theorem 3.1.5, because the proof of Statement 3.2.4 is already known in the literature, as explained below Statement 3.2.4.

3.6.1. Proof of Statement 3.2.5

Recall the definition of Core_t in (3.2.8). It is crucial that in Core_t, we let $D_v(t/2)$ be large. We again continue to define what a k-exploration graph and its collisions are, but this time for the preferential attachment model:

Definition 3.6.1 (k-exploration graph)**.** *Let $(\mathrm{PA}_t)_{t\ge1}$ be a preferential attachment model. For $v \in [t]$, we call the* k-exploration graph *of v to be the subgraph of PA_t, where we consider the m edges originally incident to v, and the m edges originally incident to any other vertex that is connected to v in this procedure, up to distance k from v.*

Definition 3.6.2 (Collision)**.** *Let $(\mathrm{PA}_t)_{t\ge1}$ be a preferential attachment model with $m \ge 2$, and let v be a vertex. We say that we have a* collision *in the k-exploration graph of v when one of the m edges of a vertex in the k-exploration graph of v is connected to a vertex that is already in the k-exploration graph of v.*

Now we want to show that every k-exploration graph has at most a finite number of collisions before hitting the Core_t, as we did for the configuration model. The first step is to use [59, Lemma 3.9]:

Lemma 3.6.3 (Early vertices have large degree)**.** *Fix $m \ge 1$. There exists $a > 0$ such that*

$$\mathbb{P}\Big(\min_{i\le t^a} D_i(t) \ge (\log t)^\sigma\Big) \longrightarrow 1$$

for some $\sigma > 1/(3-\tau)$. As consequence, $[t^a] \subseteq \mathrm{Core}_t$ with high probability.

In agreement with (3.2.10) (see also (3.4.12)), we set

$$k_t^+ = (1+\varepsilon)\frac{\log\log t}{\log m}. \tag{3.6.1}$$

We want to prove that the exploration graph $\widehat{U}_{\le k_t^+}(v)$ has at most a finite number of collisions before hitting Core_t, similarly to the case of CM_n, now for PA_t. As it is possible to see from (3.2.8), $\mathrm{Core}_t \subseteq [t/2]$, i.e., is a subset defined in PA_t when the

graph has size $t/2$. As a consequence, we do not know the degree of vertices in $[t/2]$ when the graph has size t. However, in [59, Appendix A.4] the authors prove that at time t all the vertices $t/2+1,\dots,t$ have degree smaller than $(\log t)^\sigma$.

We continue by giving a bound on the degree of vertices that are not in Core_t. For vertices $i \in [t/2] \setminus \mathrm{Core}_t$ we know that $D_i(t/2) < (\log t)^\sigma$, see (3.2.8), but in principle their degree $D_i(t)$ at time t could be quite high. We need to prove that this happens with very small probability. Precisely, we prove that, for some $B > 0$,

$$\mathbb{P}\left(\max_{i\in[t/2]\setminus\mathrm{Core}_t} D_i(t) \geq (1+B)(\log t)^\sigma\right) = o(1). \tag{3.6.2}$$

This inequality implies that when a degree is at most $(\log t)^\sigma$ at time $t/2$, then it is unlikely to grow by $B(\log t)^\sigma$ between time $t/2$ and t. This provides a bound on the cardinality of incoming neighborhoods that we can use in the definition of the exploration processes that we will rely on, in order to avoid Core_t. We prove (3.6.2) in the following lemma that is an adaptation of the proof of [59, Lemma A.4]. Its proof is deferred to Appendix 3.8:

Lemma 3.6.4 (Old vertex not in Core_t)**.** *There exists $B \in (0,\infty)$ such that, for every $i \in [t/2]$,*

$$\mathbb{P}\left(D_i(t) \geq (1+B)(\log t)^\sigma \mid D_i(t/2) < (\log t)^\sigma\right) = o(1/t).$$

We can now get to the core of the proof of Statement 3.2.5, that is we show that there are few collisions before reaching Core_t:

Lemma 3.6.5 (Few collisions before hitting the core)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. Fix $a \in (0,1)$ and $l \in \mathbb{N}$ such that $l > 1/a$. With k_t^+ as in* (3.6.1)*, the probability that there exists a vertex $v \in [t]$ such that its k_t^+-exploration graph has at least l collisions before hitting $\mathrm{Core}_t \cup [t^a]$ is $o(1)$.*

Next we give a lower bound on the number of vertices on the boundary of a k_n^+-exploration graph. First of all, for any fixed $a \in (0,1)$, we notice that the probability of existence of a vertex in $[t]\setminus[t^a]$, that has only self loops is $o(1)$. Indeed, the probability that a vertex s has only self-loops is $O(\frac{1}{s^m})$. Thus, the probability that there exists a vertex in $[t] \setminus [t^a]$ that has only self-loops is bounded above by

$$\sum_{s>t^a} O(\frac{1}{s^m}) = O(t^{-a(m-1)}) = o(1),$$

since we assume that $m \geq 2$. We can thus assume that no vertex in $[t] \setminus [t^a]$ has only self-loops. This leads us formulate the following Lemma, whose proof is given in Section 3.8:

Lemma 3.6.6 (Lower bound on boundary vertices)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. For $a \in (0,1)$, consider a vertex $v \in [t] \setminus (\mathrm{Core}_t \cup [t^a])$ and its k-exploration graph. If there are at most l collisions in the k-exploration graph, and no vertex in $[t] \setminus [t^a]$ has only self loops, then there exists a constant*

$s = s(m, l) > 0$ such that the number of vertices in the boundary of the k-exploration graph is at least $s(m, l)m^k$.

Together, Lemmas 3.6.3, 3.6.5 and 3.6.6 complete the proof of Statement 3.2.5.

The rest of this section is devoted to the proof of Lemma 3.6.5. We first need to introduce some notation, in order to be able to express the probability of collisions. We do this in the next subsection.

Ulam-Harris notation for regular trees

Recall the definition of the Ulam-Harris set given in Definition 2.1.3. In this section, we need a subset of Ulam-Harris set that can be used to describe *regular trees*, i.e., trees where all vertices have the same degree. Define

$$W_\ell := [m]^\ell\,, \qquad W_{\le k} := \bigcup_{\ell=0}^{k} W_\ell\,,$$

where $W_0 := \varnothing$. We use $W_{\le k}$ as a universal set to label any regular tree of depth k, where each vertex has m children.

Given $y \in W_\ell$ and $z \in W_m$, we denote by $(y, z) \in W_{\ell+m}$ the concatenation of y and z. Given $x, y \in W_{\le k}$, we write $y \succeq x$ if y is a descendant of x, that is $y = (x, z)$ for some $z \in W_{\le k}$.

Given a finite number of points $z_1, \dots, z_m \in W_{\le k}$, abbreviate $\vec{z}_m = (z_1, \dots, z_m)$, and define $W_{\le k}^{(\vec{z}_m)}$ to be the tree obtained from $W_{\le k}$ by cutting the branches starting from any of the z_i's (including the z_i's themselves):

$$W_{\le k}^{(\vec{z}_k)} := \left\{x \in W_{\le k} :\ x \not\succeq z_1,\ \dots, x \not\succeq z_m\right\}. \tag{3.6.3}$$

Remark 3.6.7 (Total order)**.** The set $W_{\le k}$ comes with a natural total order relation, called *shortlex order*, in which shorter words precede longer ones, and words with equal length are ordered lexicographically. More precisely, given $x \in W_\ell$ and $y \in W_m$, we say that x precedes y if either $\ell < m$, or if $\ell = m$ and $x_i \le y_i$ for all $1 \le i \le \ell$. We stress that this is a *total* order relation, unlike the descendant relation $\succeq$ which is only a partial order. (Of course, if $y \succeq x$, then x precedes y, but not vice versa).

Collisions

We recall that, given $z \in [t]$ and $j \in [m]$, the j-th half-edge starting from vertex z in PA_t is attached to a random vertex, denoted by $\xi_{z,j}$. We can use the set $W_{\le k}$ to label the exploration graph $\widehat{U}_{\le k}(v)$, as follows:

$$\widehat{U}_{\le k}(v) = \left\{V_z\right\}_{z \in W_{\le k}}, \tag{3.6.4}$$

where $V_\varnothing = v$ and, iteratively, $V_z = \xi_{V_x, j}$ for $z = (x, j)$ with $x \in W_{\le k-1}$ and $j \in [m]$.

The first vertex generating a *collision* is V_{Z_1}, where the random index $Z_1 \in W_{\leq k}$ is given by

$$Z_1 := \min\left\{z \in W_{\leq k} :\ V_z = V_y \text{ for some } y \text{ which precedes } z\right\},$$

where "min" refers to the total order relation on $W_{\leq k}$ as defined in Remark 3.6.7.

Now comes a tedious observation. Since $V_{Z_1} = V_y$ for some y which precedes Z_1, by definition of Z_1, then all descendants of Z_1 will coincide with the corresponding descendants of y, that is $V_{(Z_1,r)} = V_{(y,r)}$ for all r. In order not to over count collisions, in defining the second collision index Z_2, we avoid exploring the descendants of index Z_1, that is we only look at indices in $W_{\leq k}^{(Z_1)}$, see (3.6.3). The second vertex representing a (true) collision is then V_{Z_2}, where we define

$$Z_2 := \min\left\{z \in W_{\leq k}^{(Z_1)} \colon z \text{ follows } Z_1, \text{ i.e., } \ V_z = V_y \text{ for some } y \text{ which precedes } z\right\},$$

Iteratively, we define

$$Z_{i+1} := \min\left\{z \in W_{\leq k}^{(\vec{Z}_i)} \colon z \text{ follows } Z_i, \text{ i.e., } V_z = V_y \text{ for some } y \text{ which precedes } z\right\},$$

so that V_{Z_i} is the i-th vertex that represents a collision. The procedure stops when there are no more collisions. Denoting by $\mathcal{C}$ the (random) number of collisions, we have a family

$$Z_1,\ Z_2,\ \ldots,\ Z_{\mathcal{C}}$$

of random elements of $W_{\leq k}$, such that $(V_{Z_i})_{1\leq i\leq \mathcal{C}}$ are the vertices generating the collisions.

Proof of Lemma 3.6.5

Recalling (3.6.4) and (3.6.3), given arbitrarily $z_1, \ldots, z_l \in W_{\leq k}$, we define

$$\widehat{U}_{\leq k}^{(\vec{z}_l)}(v) = \left\{V_z\right\}_{z \in W_{\leq k}^{(\vec{z}_l)}}, \tag{3.6.5}$$

that is, we consider a subset of the full exploration graph $\widehat{U}_{\leq k}(v)$, consisting of vertices V_z whose indexes $z \in W_{\leq k}$ are not descendants of $z_1, \ldots, z_l$. The basic observation is that

$$\widehat{U}_{\leq k}(v) = \widehat{U}_{\leq k}^{(\vec{z}_l)}(v) \qquad \text{on the event} \quad \left\{\mathcal{C} = l\,,\ Z_1 = z_1, \ldots, Z_l = z_l\right\}. \tag{3.6.6}$$

In words, this means that to recover the full exploration graph $\widehat{U}_{\leq k}(v)$, it is irrelevant to look at vertices V_z for z that is a descendant of a collision index $z_1, \ldots, z_l$.

We will bound the probability that there are l collisions before reaching $\mathrm{Core}_t \cup [t^a]$, occurring at specified indices $z_1, \ldots, z_l \in W_{\leq k}$, for $k = k_t^+$ as in (3.6.1), as follows:

$$\mathbb{P}\big(\mathcal{C} = l\,,\ Z_1 = z_1, \dots, Z_l = z_l, \widehat{U}_{\le k}(v) \cap (\mathrm{Core}_t \cup [t^a]) = \varnothing\big) \ \le\ \alpha(t)^l\,, \tag{3.6.7}$$

where, for the constant B given by Lemma 3.6.4, we define

$$\alpha(t) = \frac{4(1+B)}{m}\,\frac{(\log t)^{\sigma+1+\varepsilon}}{t^a}\,. \tag{3.6.8}$$

Summing (3.6.7) over $z_1, \dots, z_l \in W_{\le k}$ we get

$$\mathbb{P}(\mathcal{C} = l,\ \widehat{U}_{\le k}(v) \cap (\mathrm{Core}_t \cup [t^a]) = \varnothing) \le \alpha(t)^l\, |W_{\le k}|^l\,.$$

Since, for $k = k_t^+$ as in (3.6.1), we can bound

$$|W_{\le k}| = \frac{m^{k+1}-1}{m-1} \le 2\, m^k \le 2\,(\log t)^{1+\varepsilon}\,, \tag{3.6.9}$$

the probability of having at least l collisions, before reaching $\mathrm{Core}_t \cup [t^a]$, is

$$O(\alpha(t)^l (\log t)^{2l}) = o(1/t),$$

because $l > 1/a$ by assumption. This completes the proof of Lemma 3.6.5. It only remains to show that (3.6.7) holds true.

Proof of (3.6.7)**: case** $l = 1$

We start proving (3.6.7) for one collision. By (3.6.6), we can replace $\widehat{U}_{\le k}(v)$ by $\widehat{U}^{(z_1)}_{\le k}(v)$ in the left hand side of (3.6.7), i.e., we have to prove that

$$\mathbb{P}(\mathcal{C} = 1\,,\ Z_1 = z_1,\ \widehat{U}^{(z_1)}_{\le k}(v) \cap (\mathrm{Core}_t \cup [t^a]) = \varnothing) \ \le\ \alpha(t)\,. \tag{3.6.10}$$

Since v, k and z_1 are fixed, let us abbreviate, and recalling (3.6.5),

$$\mathcal{W} := W^{(z_1)}_{\le k}(v)\,, \qquad \widehat{U} := \widehat{U}^{(z_1)}_{\le k}(v) = \big\{V_z\big\}_{z\in\mathcal{W}}\,. \tag{3.6.11}$$

Note that V_{z_1} is the only collision precisely when $\widehat{U}$ *is a tree and* $V_{z_1} \in \widehat{U}$. Then (3.6.10) becomes

$$\mathbb{P}(\widehat{U} \text{ is a tree}\,,\ V_{z_1} \in \widehat{U}\,,\ \widehat{U} \cap (\mathrm{Core}_t \cup [t^a]) = \varnothing) \le \alpha(t)\,. \tag{3.6.12}$$

We will actually prove a stronger statement: for any fixed *deterministic* labeled di-

rected tree $H \subseteq [t]$ and for any $y \in H$,

$$\mathbb{P}(\widehat{U} = H\,,\; V_{z_1} = y\,,\; H \cap (\mathrm{Core}_t \cup [t^a]) = \varnothing) \\ \leq \frac{\alpha(t)}{2(\log t)^{1+\varepsilon}}\, \mathbb{P}\big(\widehat{U} = H\,,\; V_{z_1} \notin H\big)\,. \tag{3.6.13}$$

This yields (3.6.12) by summing over $y \in H$ —note that $|H| \leq |W_{\leq k}| \leq 2(\log t)^{1+\varepsilon}$ by (3.6.9)— and then summing over all possible realizations of H.

It remains to prove (3.6.13). We again use the notion of a *factorizable event*, as in the proof of the lower bound. Since the events in (3.6.13) are not factorizable, we will specify the incoming neighborhood $\mathcal{N}(y)$ (recall (3.4.34)) of all $y \in H$. More precisely, by labeling the vertices of H, see (3.6.11), as

$$H = \{v_s\}_{s \in \mathcal{W}} \qquad \text{and} \qquad y = v_{\bar{s}}\,, \qquad \text{for some } \bar{s} \in \mathcal{W}\,, \tag{3.6.14}$$

we can consider the events $\{\mathcal{N}(v_s) = N_{v_s}\}$ where N_{v_s} are (deterministic) disjoint subsets of $[t] \times [m]$. We say that the subsets $(N_{v_s})_{s \in \mathcal{W}}$ are *compatible* with the tree H when $(v_{s'}, j) \in N_{v_s}$ whenever $s = (s', j)$ with $s, s' \in \mathcal{W}$, $j \in [m]$. Then we can write

$$\{\widehat{U} = H\} = \bigcup_{\text{compatible } (N_{v_s})_{s \in \mathcal{W}}} \{\mathcal{N}(v_s) = N_{v_s} \text{ for every } s \in \mathcal{W}\}\,. \tag{3.6.15}$$

Since the degree of vertex v_s equals $D_t(v_s) = m + |N_{v_s}|$, we can ensure that $H \cap (\mathrm{Core}_t \cup [t^a]) = \varnothing$ by restricting the union in (3.6.15) to those N_{v_s} satisfying the constraints

$$v_s > t^a \qquad \text{and} \qquad |N_{v_s}| \leq (1+B)(\log t)^\sigma - m\,, \qquad \forall s \in \mathcal{W}\,. \tag{3.6.16}$$

Finally, if we write

$$z_1 = (x, j) \qquad \text{for some} \quad x \in \mathcal{W}\,,\; j \in [m]\,,$$

then, since $V_{z_1} = \xi_{V_x, j}$, the event $\{V_{z_1} = v_{\bar{s}}\}$ amounts to require that[2]

$$(v_x, j) \in N_{v_{\bar{s}}}\,. \tag{3.6.17}$$

Let us summarize where we now stand: When we fix a family of $(N_{v_s})_{s \in \mathcal{W}}$ that is compatible and satisfies the constraints (3.6.16) and (3.6.17), in order to prove (3.6.13) it is enough to show that

$$\mathbb{P}(\mathcal{N}(v_s) = N_{v_s} \text{ for every } s \in \mathcal{W}) \\ \leq \frac{\alpha(t)}{2(\log t)^{1+\varepsilon}}\, \mathbb{P}(\mathcal{N}(v_s) = N_{v_s} \text{ for every } s \in \mathcal{W} \setminus \{\bar{s}\},\; \mathcal{N}(v_{\bar{s}}) = N_{v_{\bar{s}}} \setminus \{(v_x, j)\})\,. \tag{3.6.18}$$

[2]Incidentally, we observe that the constraint (3.6.17) is not included in the requirement that $(N_{v_s})_{s \in \mathcal{W}}$ are compatible, because $z_1 = (x, j) \notin \mathcal{W}$ by definition (3.6.11) of $\mathcal{W}$.

Let us set

$$N := \bigcup_{s\in\mathcal{W}} N_{v_s} \subseteq [t]\times[m]\,. \tag{3.6.19}$$

The probability on the left-hand side of (3.6.18) can be factorized, using conditional expectations and the tower property, as a product of two kinds of terms:

▷ For every edge $(u,r)\in N$ —say $(u,r)\in N_{v_s}$, with $s\in\mathcal{W}$— we have the term

$$\frac{D_{v_s}(u,r-1)+\delta}{c_{u,r}} \tag{3.6.20}$$

corresponding to the fact that the edge needs to be connected to v_s;

▷ On the other hand, for every edge $(u,r)\notin N$, we have the term

$$1-\frac{D_H(u,r-1)+|H\cap[u-1]|\delta}{c_{u,r}}, \tag{3.6.21}$$

corresponding to the fact that the edge may not connect to any vertex in H.

We emphasize that all the degrees $D_\cdot(\cdot,\cdot)$ appearing in (3.6.20) and (3.6.21) are *deterministic*, since they are fully determined by the realizations of the incoming neighborhoods $(N_{v_s})_{s\in\mathcal{W}}$.

We can obtain the right-hand side in (3.6.18) by replacing some terms in the product.

▷ Among the edges $(u,r)\in N$, whose contribution is (3.6.20), we have the one that creates the collision, namely (v_x,j). If we want this edge to be connected *outside* H, as in the right-hand side in (3.6.18), we need to divide the left hand side of (3.6.18) by

$$\left(\frac{D_{v_{\bar s}}(v_x,j-1)+\delta}{c_{v_x,j}}\right)\left(1-\frac{D_H(v_x,j-1)+|H\cap[v_x-1]|\delta}{c_{v_x,j}}\right)^{-1}. \tag{3.6.22}$$

We also have to replace some other terms corresponding to edges $(u,r)\in N_{v_{\bar s}}$, because the degree of vertex $v_{\bar s}$ is decreased by one after connecting (v_x,j) outside H. More precisely, for every edge $(u,r)\in N_{v_{\bar s}}$ that is younger than (v_x,j), that is $(u,r)>(v_x,j)$, we can reduce the degree of $v_{\bar s}$ by one by dividing the left-hand side of (3.6.18) by

$$\prod_{(u,r)\in N_{v_{\bar s}},\ (u,r)>(v_x,j)} \frac{D_{v_{\bar s}}(u,r-1)+\delta}{D_{v_{\bar s}}(u,r-1)-1+\delta} = \frac{D_{v_{\bar s}}(t)+\delta}{D_{v_{\bar s}}(v_x,j-1)+\delta}\,. \tag{3.6.23}$$

Finally, the contribution of the edges $(u,r)\in N_{v_s}$ for $s\neq\bar s$ is unchanged.

▷ For every edge $(u,r)\notin N$, the probability that such edge is not attached to H, after we reconnect the edge (v_x,j), becomes larger, since the degree of H is reduced by one.

It follows that the inequality (3.6.18) holds with $\alpha(t)/(2(\log t)^{1+\varepsilon})$ replaced by β, defined by

$$\begin{aligned}\beta &= \left(\frac{D_{v_{\bar{s}}}(v_x, j-1)+\delta}{c_{v_x,j}}\right) \\ &\quad\times\left(1-\frac{D_H(v_x,j-1)+|H\cap[v_x-1]|\delta}{c_{v_x,j}}\right)^{-1}\frac{D_{v_{\bar{s}}}(t)+\delta}{D_{v_{\bar{s}}}(v_x,j-1)+\delta} \\ &= \left(\frac{D_t(v_{\bar{s}})+\delta}{c_{v_x,j}}\right)\left(1-\frac{D_{v_x,j-1}(H)+|H\cap[v_x-1]|\delta}{c_{v_x,j}}\right)^{-1} \\ &\le \left(\frac{D_t(v_{\bar{s}})}{c_{v_x,j}}\right)\left(1-\frac{D_{v_x,j-1}(H)}{c_{v_x,j}}\right)^{-1} =: \beta'\,,\end{aligned} \tag{3.6.24}$$

because $\delta \le 0$. We only need to show that $\beta' \le \alpha(t)/(2(\log t)^{1+\varepsilon})$. Since $c_{v,j} \ge m(v-1)$, the first relation in (3.6.16) yields

$$c_{v_x,j} \ge t^a.$$

Hence, since $D_{v_{\bar{s}}}(t) \le (1+B)(\log t)^\sigma$ by the second relation in (3.6.16), we can bound

$$\left(\frac{D_{v_{\bar{s}}}(t)}{c_{v_x,j}}\right) \le \frac{(1+B)(\log t)^\sigma}{mt^a}\,.$$

Likewise, since $D_H(t) \le |H|(1+B)(\log t)^\sigma$, for $k = k_t^+$ we get, by (3.6.9),

$$\left(1-\frac{D_H(v_x,j-1)}{c_{v_x,j}}\right)^{-1} \le \left(1-\frac{2(\log t)^{1+\varepsilon}(1+B)(\log t)^\sigma}{t^a}\right)^{-1} \le 2\,,$$

where the last inequality holds for t large enough. Recalling (3.6.8),

$$\beta' \le 2\frac{(1+B)(\log t)^\sigma}{mt^a} = \frac{\alpha(t)}{2(\log t)^{1+\varepsilon}}\,.$$

This completes the proof of (3.6.18), and hence of (3.6.7), in the case where $l = 1$. □

Proof of (3.6.7)**: general case** $l \ge 2$

The proof for the general case is very similar to that for $l = 1$, so we only highlight the (minor) changes.

In analogy with (3.6.10), we can replace $\widehat{U}_{\le k}(v)$ by $\widehat{U}^{(\vec{z}_l)}_{\le k}(v)$ in the left-hand side of (3.6.7), thanks to (3.6.6). Then, as in (3.6.11), we write

$$\mathcal{W} := W^{(\vec{z}_l)}_{\le k}(v)\,, \qquad \widehat{U} := \widehat{U}^{(\vec{z}_l)}_{\le k}(v) = \left\{V_z\right\}_{z\in\mathcal{W}}.$$

The extension of (3.6.13) becomes that for any fixed *deterministic* labeled directed tree

$H \subseteq [t]$ and for all $y_1, \dots, y_l \in H$,

$$\begin{aligned}&\mathbb{P}(\widehat{U} = H\,,\, V_{z_1} = y_1\,,\, \dots, V_{z_l} = y_l\,,\, H \cap (\mathrm{Core}_t \cup [t^a]) = \varnothing)\\ &\quad\le \left(\frac{\alpha(t)}{2(\log t)^{1+\varepsilon}}\right)^l \mathbb{P}\big(\widehat{U} = H\,,\, V_{z_1} \notin H\,,\, V_{z_2} \notin H\,,\, \dots\,,\, V_{z_l} \notin H\big)\,.\end{aligned}$$

As in (3.6.14), we can write

$$H = \{v_s\}_{s\in\mathcal{W}} \qquad \text{and} \qquad y_1 = v_{\bar{s}_1}\,,\, \dots\,,\, y_l = v_{\bar{s}_l} \qquad \text{for some } \bar{s}_1\,,\, \dots\,,\, \bar{s}_l \in \mathcal{W}\,.$$

To obtain a factorizable event, we must specify the incoming neighborhoods $\mathcal{N}_{v_s} = N_{v_s}$ for all $s \in \mathcal{W}$, which must be compatible with H and satisfy the constraint (3.6.16). If we write

$$z_1 = (x_1, j_1)\,,\, \dots\,,\, z_l = (x_l, j_l)\,, \qquad \text{for some} \quad x_1, \dots, x_l \in \mathcal{W},\, j_1, \dots, j_l \in [m]\,,$$

then we also impose the constraint that obviously generalizes (3.6.17), namely

$$(v_{x_1}, j_1) \in N_{v_{\bar{s}_1}}\,,\, \dots\,,\, (v_{x_l}, j_l) \in N_{v_{\bar{s}_l}}\,.$$

The analogue of (3.6.18) then becomes

$$\begin{aligned}&\mathbb{P}(\mathcal{N}(v_s) = N_{v_s} \text{ for every } s \in \mathcal{W}) \qquad (3.6.25)\\ &\quad\le \left(\frac{\alpha(t)}{2(\log t)^{1+\varepsilon}}\right)^l \mathbb{P}\big(\mathcal{N}(v_s) = N_{v_s} \text{ for every } s \in \mathcal{W} \setminus \{\bar{s}_1, \dots, \bar{s}_l\},\\ &\qquad\qquad \mathcal{N}(v_{\bar{s}_i}) = N_{v_{\bar{s}_i}} \setminus \{(v_{x_i}, j_i)\} \text{ for every } i = 1, \dots, l\big)\,.\end{aligned}$$

When we define N as in (3.6.19), the probability in the left-hand side of (3.6.25) can be factorized in a product of terms of two different types, which are given precisely by (3.6.20) and (3.6.21). In order to obtain the probability in the right-hand side of (3.6.25), we have to divide the left-hand side by a product of factors analogous to (3.6.22) and (3.6.23). More precisely, (3.6.22) becomes

$$\prod_{i=1}^{l} \left(\frac{D_{v_{\bar{s}_i}}(v_{x_i}, j_i - 1) + \delta}{c_{v_{x_i}, j_i}}\right) \left(1 - \frac{D_H(v_{x_i}, j_i - 1) + |H \cap [v_{x_i} - 1]|\delta}{c_{v_{x_i}, j_i}}\right)^{-1},$$

while (3.6.23) becomes

$$\prod_{i=1}^{l} \frac{D_{v_{\bar{s}_i}}(t) + \delta}{D_{v_{\bar{s}_i}}(v_{x_i}, j_i - 1) + \delta}\,.$$

We define β accordingly, namely we take the product for $i = 1, \dots, l$ of (3.6.24) with

$x, j, \bar{s}$ replaced respectively by $x_i, j_i, \bar{s}_i$. Then it is easy to show that

$$\beta \leq \left(\frac{\alpha(t)}{2(\log t)^{1+\varepsilon}} \right)^l ,$$

arguing as in the case $l = 1$. This completes the proof of (3.6.25). □

3.6.2. Proof of Statement 3.2.6

The next step is to prove that the boundaries of the k_t^+-exploration graphs are at most at distance

$$h_t = \lceil B \log\log\log t + C \rceil \tag{3.6.26}$$

from Core_t, where B, C are constants to be chosen later on. Similarly to the proof in Section 3.5.2, we consider a k_t^+-exploration graph, and we enumerate the vertices on the boundary as $x_1, \ldots, x_N$, where $N \geq s(m, l)m^{k_t^+}$ from Lemma 3.6.6 and l is chosen as in Lemma 3.6.5. We next define what it means to have a success:

Definition 3.6.8 (Success)**.** *Consider the vertices $x_1, \ldots, x_N$ on the boundary of a k_t^+-exploration graph. We say that x_i is a* success *when the distance between x_i and Core_t is at most $2h_t$.*

The next lemma is similar to Lemma 3.5.7 (but only deals with vertices in $[t/2]$):

Lemma 3.6.9 (Probability of success)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. Consider $v \in [t/2] \setminus \mathrm{Core}_t$ and its k_t^+-exploration graph. Then there exists a constant $\eta > 0$ such that*

$$\mathbb{P}\left(S_{x_1} \mid \mathrm{PA}_{t/2}\right) \geq \eta, \tag{3.6.27}$$

and for all $j = 2, \ldots, N$,

$$\mathbb{P}\left(S_{x_1} \mid \mathrm{PA}_{t/2}, S^c_{x_1}, \ldots, S^c_{x_{j-1}}\right) \geq \eta. \tag{3.6.28}$$

The aim is to define a sequence of vertices $w_0, \ldots, w_h$ that connects a vertex x_i on the boundary with Core_t. In order to do this, we need some preliminary results. We start with the crucial definition of a t-connector:

Definition 3.6.10 (t-connector)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$. Consider two subsets $A, B \subseteq [t/2]$, with $A \cap B = \varnothing$. We say that a vertex $j \in [t] \setminus [t/2]$ is a* t-connector for A and B *if at least one of the edges incident to j is attached to a vertex in A and at least one is attached to a vertex in B.*

The notion of t-connector is useful, because, unlike in the configuration model, in the preferential attachment model typically two high-degree vertices are not directly connected. From the definition of the preferential attachment model, it is clear that the older vertices have with high probability large degree, and the younger vertices

have lower degree. When we add a new vertex, this is typically attached to vertices with large degrees. This means that, with high probability, two vertices with high degree can be connected by a young vertex, which is the t-connector.

A further important reason for the usefulness of t-connectors is that we have effectively *decoupled* the preferential attachment model at time $t/2$ and what happens in between times $t/2$ and t. When the sets A and B are appropriately chosen, then each vertex will be a t-connector with reasonable probability, and the events that distinct vertices are t-connectors are close to being independent. Thus, we can use comparisons to binomial random variables to investigate the existence of t-connectors. In order to make this work, we need to identify the structure of $\mathrm{PA}_{t/2}$ and show that it has sufficiently many vertices of large degree, and we need to show that t-connectors are likely to exist. We start with the latter.

In more detail, we will use t-connectors to generate the sequence of vertices $w_1, \dots, w_h$ between the boundary of a k_n^+-exploration graph and the Core_t, in the sense that we use a t-connector to link the vertex w_i to the vertex w_{i+1}. (This is why we define a vertex x_i to be a success if its distance from Core_t is at most $2h_t$, instead of h_t.) We rely on a result implying the existence of t-connectors between sets of high total degree:

Lemma 3.6.11 (Existence of t-connectors)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. There exists a constant $\mu > 0$ such that, for every $A \subseteq [t/2]$, and $i \in [t/2] \setminus A$,*

$$\mathbb{P}\Big(\nexists j \in [t] \setminus [t/2] \colon j \text{ is a } t\text{-connector for } i \text{ and } A \mid \mathrm{PA}_{t/2}\Big) \leq \exp\left(-\frac{\mu D_A(t/2) D_i(t/2)}{t}\right),$$

where $D_A(t/2) = \sum_{v\in A} D_v(t/2)$ is the total degree of A at time $t/2$.

Proof. The proof of this lemma is present in the proof of [59, Proposition 3.2]. □

Remark 3.6.12. Notice that this bound depends on the fact that the number of possible t-connectors is of order t.

A last preliminary result that we need is a technical one, which plays the role of Lemma 3.5.9 for the configuration model and shows that at time $t/2$ there are sufficiently many vertices of high degree, uniformly over a wide range of what 'large' could mean:

Lemma 3.6.13 (Tail of degree distribution)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model, with $m \geq 2$ and $\delta \in (-m, 0)$. Then, for all $\zeta > 0$ there exists a constant $c = c(\zeta)$ such that, for all $1 \leq x \leq (\log t)^q$, for any $q > 0$, and uniformly in t,*

$$P_{\geq x}(t) = \frac{1}{t} \sum_{v\in[t]} \mathbb{1}_{\{D_v(t)\geq x\}} \geq c x^{-(\tau-1+\zeta)}. \tag{3.6.29}$$

Proof. The degree distribution sequence $(p_k^{(m)})_{k\in\mathbb{N}}$ in (1.3.1) satisfies a power law with exponent $\tau \in (2, 3)$. As a consequence, for all $\zeta > 0$ there exists a constant $\bar{c} = \bar{c}(\zeta)$

such that

$$p^{(m)}_{\geq x} := \sum_{k \geq x} p^{(m)}_k \geq \bar{c} x^{-(\tau-1+\zeta)}. \tag{3.6.30}$$

We now use a concentration result on the empirical degree distribution (for details, see [85, Theorem 8.2]), which assures us that there exists a second constant $C > 0$ such that, with high probability, for every $x \in \mathbb{N}$,

$$|P_{\geq x} - p_{\geq x}| \leq C\sqrt{\frac{\log t}{t}}.$$

Fix now $\zeta > 0$, then from this last bound we can immediately write, for a suitable constant $\bar{c}$ as in (3.6.30),

$$P_{\geq x} \geq p_{\geq x} - C\sqrt{\frac{\log t}{t}} \geq \bar{c} x^{-(\tau-1+\zeta)} - C\sqrt{\frac{\log t}{t}} \geq \frac{\bar{c}}{2} x^{-(\tau-1+\zeta)},$$

if and only if

$$C\sqrt{\frac{\log t}{t}} = o\left(x^{-(\tau-1+\zeta)}\right).$$

This is clearly true for $x \leq (\log t)^q$, for any positive q. Taking $c = \bar{c}/2$ completes the proof. □

With the above tools, we are now ready to complete the proof of Lemma 3.6.9:

Proof of Lemma 3.6.9. As in the proof of Proposition 3.5.7, we define the super-exponentially growing sequence g_ℓ as in (3.5.14), where $\gamma > 0$ is chosen small enough, as well as $\zeta > 0$, so that (3.5.13) holds. The constants B and C in the definition (3.6.26) of h_t are fixed as prescribed below (3.5.14).

We will define a sequence of vertices $w_0, \dots, w_h$ such that, for $i = 1, \dots, h$, $D_{w_i}(t) \geq g_i$ and w_{i-1} is connected to w_i. For this, we define, for $i = 1, \dots, h-1$,

$$H_i = \left\{u \in [t] \colon D_u(t/2) \geq g_i\right\} \subseteq [t/2],$$

so that we aim for $w_i \in H_i$.

We define the vertices recursively, and start with $w_0 = x_1$. Then, we consider t-connectors between w_0 and H_1, and denote by w_1 the vertex in H_1 with minimal degree among the ones that are connected to w_0 by a t-connector. Recursively, consider t-connectors between w_i and H_{i+1}, and denote by w_{i+1} the vertex in H_{i+1} with minimal degree among the ones that are connected to w_i by a t-connector. Recall (3.5.15) to see that $g_{h_t} \geq (\log t)^\sigma$, where h_t is defined in (3.6.26). The distance between w_0 and Core_t is at most $2h_t = 2\lceil B \log\log\log t + C\rceil$. If we denote the event that there exists a t connector between w_{i-1} and H_i by $\{w_{i-1} \sim H_i\}$, then we will bound from below

$$\mathbb{P}(S_{x_1} \mid \mathrm{PA}_{t/2}) \geq \mathbb{E}\Big[\prod_{i=1}^{h_t} \mathbb{1}_{\{w_{i-1} \sim H_i\}} \mid \mathrm{PA}_{t/2}\Big].$$

In Lemma 3.6.11, the bound on the probability that a vertex $j \in [t] \setminus [t/2]$ is a t-connector between two subsets of $[t]$ is independent of the fact that the other vertices are t-connectors or not. This means that, with $\mathcal{F}_i$ the σ-field generated by the path formed by $w_0, \dots, w_i$ and their respective t-connectors,

$$\mathbb{E}\Big[\mathbb{1}_{\{w_{i-1}\sim H_i\}} \mid \mathrm{PA}_{t/2}, \mathcal{F}_{i-1}\Big] \geq 1 - \mathrm{e}^{-\mu D_{w_{i-1}}(t/2) D_{H_i}(t/2)/t},$$

where $D_{H_i}(t) = \sum_{u\in H_i} D_u(t/2)$. This means that

$$\mathbb{E}\Big[\prod_{i=1}^{h_t} \mathbb{1}_{\{w_{i-1}\sim H_i\}} \mid \mathrm{PA}_{t/2}\Big] \geq \prod_{i=1}^{h_t} \Big(1 - \mathrm{e}^{-\mu D_{w_{i-1}}(t/2) D_{H_i}(t/2)/t}\Big).$$

We have to bound every term in the product. Using Lemma 3.6.13, for $i = 1$,

$$1 - \mathrm{e}^{-\mu D_{w_0}(t/2) D_{H_1}(t/2)/t} \geq 1 - \mathrm{e}^{-\mu D_{w_0}(t/2) g_1 P_{\geq g_1}(t/2)},$$

while, for $i = 2, \dots, h-1$

$$1 - \mathrm{e}^{-\mu D_{w_{i-1}}(t/2) D_{H_i}(t/2)/t} \geq 1 - \mathrm{e}^{-\mu g_{i-1} g_i P_{\geq g_i}(t/2)}.$$

Applying (3.6.29) and recalling (3.5.21)–(3.5.22), the result is

$$\begin{aligned}\mathbb{P}(S_{x_1} \mid \mathrm{PA}_t) &\geq \Big(1 - \mathrm{e}^{-\mu D_{w_0}(t/2) g_1 P_{\geq g_1}(t/2)}\Big) \prod_{i=2}^{h_t} \Big(1 - \mathrm{e}^{-\mu g_{i-1} g_i P_{\geq g_i}(t/2)}\Big) \\ &\geq \Big(1 - \mathrm{e}^{-\mu m g_1 P_{\geq g_1}(t/2)}\Big) \prod_{i=2}^{\infty} \Big(1 - \mathrm{e}^{-\tilde{c}\,(g_i)^\xi}\Big),\end{aligned}$$

for some constant $\tilde{c}$. Since $h_t = \lceil B \log\log\log t + C \rceil$, and

$$P_{\geq g_1}(t/2) \to \sum_{k\geq g_1} p_k > 0$$

with high probability as $t \to \infty$, we can find a constant η such that

$$\Big(1 - \mathrm{e}^{-\eta m g_1 P_{\geq g_1}(t/2)}\Big) \prod_{i=2}^{h_t} \Big(1 - \mathrm{e}^{-\tilde{c}\,(g_i)^\xi}\Big) > \eta > 0,$$

which proves (3.6.27).

To prove (3.6.28), we observe that all the lower bounds that we have used on the probability of existence of t-connectors only depend on the existence of sufficiently many potential t-connectors. Thus, it suffices to prove that, on the event $S_{x_1}^c \cap \cdots \cap S_{x_{j-1}}^c$, we have not used too many vertices as t-connectors. On this event, we have used at most $h_t \cdot (j-1)$ vertices as t-connectors, which is $o(t)$. Thus, this means that, when we bound the probability of S_{x_j}, we still have $t - h_t \cdot (j-1)$ possible t-connectors,

where j is at most $(\log t)^{1+\varepsilon}$. Thus, with the same notation as before,

$$\mathbb{E}\Big[\mathbb{1}_{\{w_{i-1}\sim H_i\}} \mid \mathrm{PA}_{t/2}, S^c_{x_1}, \dots, S^c_{x_{j-1}}\Big] \geq 1 - \mathrm{e}^{-\mu D_{w_{i-1}}(t/2) D_{H_i}(t/2)/t},$$

so that we can proceed as we did for S_{x_1}. We omit further details. □

We are now ready to identify the distance between the vertices outside the core and the core:

Proposition 3.6.14 (Distance between periphery and Core_t)**.** *Let $(\mathrm{PA}_t)_{t\geq 1}$ be a preferential attachment model with $m \geq 2$ and $\delta \in (-m, 0)$. Then, with high probability and for all $v \in [t] \setminus \mathrm{Core}_t$,*

$$\mathrm{dist}_{\mathrm{PA}_t}(v, \mathrm{Core}_t) \leq k_t^+ + 2h_t.$$

Proof. We start by analyzing $v \in [t/2]$. By Lemma 3.6.3, with high probability there exists $a \in (0, 1]$ such that $[t^a] \subseteq \mathrm{Core}_t$. Consider $l > 1/a$, and fix a vertex $v \in [t/2]$. Then, by Lemma 3.6.5 and with high probability, the k_t^+-exploration graph starting from v has at most l collisions before hitting Core_t. By Lemma 3.6.6 and with high probability, the number of vertices on the boundary of the k_t^+-exploration graph is at least $N = s(m, l)(\log t)^{1+\varepsilon}$. It remains to bound the probability that none of the N vertices on the boundary is a success, meaning that it does not reach Core_t in at most $2h_t = 2\lceil B \log\log t + C\rceil$ steps.

By Lemma 3.6.9,

$$\mathbb{P}(S^c_{x_1} \cap \cdots \cap S^c_{x_N} \mid \mathrm{PA}_{t/2}) \leq (1-\eta)^N = o(1/t),$$

thanks to the bound $N \geq s(m, l)(\log t)^{1+\epsilon}$. This means that the probability that there exists a vertex $v \in [t/2]$ such that its k_n^+-exploration graph is at distance more than $A \log\log\log t$ from Core_t is $o(1)$. This proves the statement for all $v \in [t/2]$.

Next, consider a vertex $v \in [t] \setminus [t/2]$. Lemma 3.6.5 implies that the probability that there exists a vertex $v \in [t] \setminus [t/2]$ such that its k_t^+-exploration graph contains more than one collision before hitting $\mathrm{Core}_t \cup [t/2]$ is $o(1)$. As before, the number of vertices on the boundary of a k_t^+-exploration graph starting at $v \in [t] \setminus [t/2]$ is at least $N \geq s(m, 1)m^{k_n^+} = s(m, 1)(\log t)^{1+\varepsilon}$. We denote these vertices by $x_1, \dots, x_N$. We aim to show that, with high probability,

$$\Delta_N = \sum_{i=1}^{N} \mathbb{1}_{(x_i \in [t/2])} \geq N/4.$$

For every $i = 1, \dots, N$, there exists a unique vertex y_i such that y_i is in the k_t^+-exploration graph and it is attached to x_i. Obviously, if $y_i \in [t/2]$ then also $x_i \in [t/2]$, since x_i has to be older than y_i. If $y_i \notin [t/2]$, then

$$\mathbb{P}\left(x_i \in [t/2] \mid \mathrm{PA}_{y_i - 1}\right) = \mathbb{P}\left(y_i \to [t/2] \mid \mathrm{PA}_{y_i - 1}\right) \geq \frac{1}{2},$$

and this bound does not depend on the attaching of the edges of the other vertices $\{y_j : j \neq i\}$. This means that we obtain the stochastic domination

$$\Delta_N \geq \sum_{i=1}^{N} \mathbb{1}_{(x_i \in [t/2])} \succeq \mathrm{Bin}\big(N, \frac{1}{2}\big),$$

where we write that $X \succeq Y$ when the random variable X is stochastically larger than Y. By concentration properties of the binomial, $\mathrm{Bin}\big(N, \frac{1}{2}\big) \geq N/4$ with probability at least

$$1 - \mathrm{e}^{-N/4} = 1 - \mathrm{e}^{-s(m,1)(\log t)^{1+\varepsilon}/4} = 1 - o(1/t).$$

Thus, the probability that none of the vertices on the boundary intersected with $[t/2]$ is a success is bounded by

$$\mathbb{P}\big(S^c_{x_1} \cap \cdots \cap S^c_{x_{\Delta_N}} \mid \mathrm{PA}_{t/2}\big) \leq (1-\eta)^{N/4} + o(1/t) = o(1/t).$$

We conclude that the probability that there exists a vertex in $[t] \setminus [t/2]$ such that it is at distance more than $k_t^+ + 2h_t$ from Core_t is $o(1)$. □

This completes the proof of Statement 3.2.6, and thus of Theorem 3.1.5. □

3.7. Bound on distances between periphery vertices in PAM: proof of Proposition 3.4.9

We prove Proposition 3.4.9. As mentioned in Section 3.4.3, the proof of (3.4.22) is an adaptation of an argument in [57, Section 4.1]. The final aim is to prove that (3.4.26) is $o(1)$.

First, we start with a technical lemma. Let us fix $R \in (0, \infty)$ and define

$$p(n,l) = R(n \wedge l)^{-\gamma}(n \vee l)^{\gamma-1}. \tag{3.7.1}$$

Our interest is for $\gamma = m/(2m+\delta) \in (1/2, 1)$, so that $\gamma \in (1/2, 1)$ (because $\delta \in (-m, 0)$).

Lemma 3.7.1. *Let $\gamma \in (1/2, 1)$ and suppose that $2 \leq g \leq t$, $\alpha, \beta \geq 0$ and $q\colon [t] \to [0, \infty)$ satisfy*

$$q(n) \leq \mathbb{1}_{\{n \geq g\}} \left(\alpha n^{-\gamma} + \beta n^{\gamma-1}\right)$$

for all $n \in [t]$. Then there exists a constant $c = c(R, \gamma) > 1$ such that, for all $l \in [t]$,

$$\sum_{n=1}^{t} q(n)p(n,l) \leq c\big(\alpha \log(t/g) + \beta t^{2\gamma-1}\big) l^{-\gamma} + c\mathbb{1}_{\{l>g\}}\big(\alpha g^{1-2\gamma} + \beta \log(t/g)\big) l^{\gamma-1}.$$

Proof. We split

$$\sum_{n=1}^{t} q(n)p(n,l) = \sum_{n=g\vee l}^{t} q(n)p(n,l) + \mathbb{1}_{\{l>g\}} \sum_{n=g}^{l-1} q(n)p(n,l),$$

because $q(n) = 0$ when $n < g$. Therefore,

$$\begin{aligned}\sum_{n=g\vee l}^{t} q(n)p(n,l) &= \sum_{n=g\vee l}^{t} q(n)R(n\wedge l)^{-\gamma}(n\vee l)^{\gamma-1} \\ &\leq \sum_{n=g\vee l}^{t} \left(\alpha n^{-\gamma} + \beta n^{\gamma-1}\right) Rn^{\gamma-1}l^{-\gamma},\end{aligned}$$

because the sum is over $n \geq g \vee l \geq l$. For the other term, since we may assume that $l > g$,

$$\mathbb{1}_{\{l>g\}} \sum_{n=g}^{l-1} q(n)p(n,l) \leq \mathbb{1}_{\{l>g\}} \sum_{n=g}^{l-1} \left(\alpha n^{-\gamma} + \beta n^{\gamma-1}\right) Rn^{-\gamma}l^{\gamma-1}.$$

This means that $\sum_{n=1}^{t} q(n)p(n,l)$ is bounded above by

$$\begin{aligned}&R\Big[\alpha \sum_{n=g\vee l}^{t} n^{-1} + \beta \sum_{n=g\vee l}^{t} n^{2\gamma-2}\Big]l^{-\gamma} + \mathbb{1}_{\{l>g\}}R\Big[\alpha\sum_{n=g}^{l-1} n^{-2\gamma} + \beta\sum_{n=g}^{l-1} n^{-1}\Big]l^{\gamma-1} \\ &\leq c_1\left[\alpha\log(t/g) + \beta t^{2\gamma-1}\right]l^{-\gamma} + \mathbb{1}_{\{l>g\}}c_2\left[\alpha g^{1-2\gamma} + \beta\log(t/g)\right]l^{\gamma-1},\end{aligned}$$

where we have used that $\gamma > 1/2$. We take $c = \max(c_1, c_2)$ to obtain the statement. □

We now define recursively the sequences $(\alpha_k)_{k\in\mathbb{N}}$, $(\beta_k)_{k\in\mathbb{N}}$ and $(g_k)_{k\in\mathbb{N}}$, for which we will prove the bound (3.4.27). This will allow us to control (3.4.26).

Definition 3.7.2. *We define*

$$g_0 = \left\lceil \frac{t}{(\log t)^2} \right\rceil, \qquad \alpha_1 = R\,(g_0)^{\gamma-1}, \qquad \beta_1 = R\,(g_0)^{-\gamma},$$

and recursively, for $k \geq 1$:

(1) g_k is the smallest integer such that

$$\frac{1}{1-\gamma}\alpha_k g_k^{1-\gamma} \geq \frac{6}{\pi^2 k^2 (\log t)^2}; \tag{3.7.2}$$

(2)

$$\alpha_{k+1} = c\big(\alpha_k \log(t/g_k) + \beta_k t^{2\gamma-1}\big); \tag{3.7.3}$$

(3)

$$\beta_{k+1} = c\big(\alpha_k g_k^{1-2\gamma} + \beta_k \log(t/g_k)\big), \tag{3.7.4}$$

where $c = c(R,\gamma) > 1$ is the same constant appearing in Lemma 3.7.1.

One can check that $k \mapsto g_k$ is non-increasing, while $k \mapsto \alpha_k, \beta_k$ are non-decreasing.

We recall that $f_{k,t}(x,l)$ was introduced in (3.4.25), with $p(z,w)$ defined in (3.4.23) (where we set $R = Cm$, to match with (3.7.1)). As a consequence, the following recursive relation is satisfied:

$$\forall k \geq 1: \qquad f_{k+1,t}(x,w) = \sum_{z=g_k}^{t} f_{k,t}(x,z)p(z,w), \tag{3.7.5}$$

where $p(z,w)$ is given in (3.7.1). The following lemma derives recursive bounds on $f_{k,t}$.

Lemma 3.7.3 (Recursive bound on $f_{k,t}$)**.** *For the sequences in Definition 3.7.2, for every $l \in [t]$ and $k \in \mathbb{N}$,*

$$f_{k,t}(x,l) \leq \alpha_k l^{-\gamma} + \mathbb{1}_{\{l > g_{k-1}\}}\beta_k l^{\gamma-1}. \tag{3.7.6}$$

Proof. We prove (3.7.6) by induction on k. For $k = 1$, using $\alpha_1 = Rg_0^{\gamma-1}$ and $\beta_1 = Rg_0^{-\gamma}$,

$$\begin{aligned} f_{1,t}(x,l) = p(x,l)\mathbb{1}_{\{x \geq g_0\}} &\leq R(g_0)^{\gamma-1}l^{-\gamma} + \mathbb{1}_{\{l>g_0\}}R(g_0)^{-\gamma}l^{\gamma-1} \\ &= \alpha_1 l^{-\gamma} + \mathbb{1}_{\{l>g_0\}}\beta_1 l^{\gamma-1}, \end{aligned}$$

as required. This initiates the induction hypothesis. We now proceed with the induction: suppose that g_{k-1}, α_k and β_k are such that

$$f_{k,t}(x,l) \leq \alpha_k l^{-\gamma} + \mathbb{1}_{(l>g_{k-1})}\beta_k l^{\gamma-1}.$$

We use the recursive property of $f_{k,t}$ in (3.7.5). We apply Lemma 3.7.1, with $g = g_k$ and $q(n) = f_{k,t}(x,n)\mathbb{1}_{\{n \geq g_k\}}$, so, by Definition 3.7.2,

$$\begin{aligned} f_{k+1,t}(x,l) &\leq c\Big[\alpha_k \log(t/g_k) + \beta_k t^{2\gamma-1}\Big]l^{-\gamma} + c\mathbb{1}_{\{l>g_k\}}\Big[\alpha_k g_k^{1-2\gamma} + \beta_k \log(t/g_k)\Big]l^{\gamma-1} \\ &= \alpha_{k+1}l^{-\gamma} + \mathbb{1}_{\{l>g_k\}}\beta_{k+1}l^{\gamma-1}. \end{aligned}$$

This advances the induction hypothesis, and thus completes the proof. □

In order to proceed, we define $\eta_k = t/g_k$. We aim to derive a bound on the growth of η_k.

Lemma 3.7.4 (Recursive relation of η_k)**.** *Let $\eta_k = t/g_k$ be defined as above. Then there exists a constant $C > 0$ such that*

$$\eta_{k+2}^{1-\gamma} \leq C\left[\eta_k^{\gamma} + \eta_{k+1}^{1-\gamma}\log\eta_{k+1}\right]. \tag{3.7.7}$$

Proof. By definition of g_k in (3.7.2),

$$\eta_{k+2}^{1-\gamma} = t^{1-\gamma} g_{k+2}^{\gamma-1} \leq t^{1-\gamma} \frac{1}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \alpha_{k+2}.$$

By definition of α_k in (3.7.3),

$$\begin{aligned} &t^{1-\gamma} \frac{1}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \alpha_{k+2} \\ &\quad = t^{1-\gamma} \frac{c}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \left[\alpha_{k+1} \log \eta_{k+1} + \beta_{k+1} t^{2\gamma-1} \right]. \end{aligned} \tag{3.7.8}$$

By definition of g_k, relation (3.7.2) holds with the opposite inequality if we replace g_k by $g_k - 1$ in the left hand side. This, with $k+1$ instead of k, yields

$$\alpha_{k+1} \leq \frac{6(1-\gamma)}{\pi^2 (k+1)^2 (\log t)^2} (g_{k+1} - 1)^{\gamma-1}. \tag{3.7.9}$$

Since $\alpha_{k+1} \geq 2$, we must have that $g_{k+1} \geq 2$, so that

$$(g_{k+1} - 1)^{\gamma-1} \leq 2^{1-\gamma} g_{k+1}^{\gamma-1}. \tag{3.7.10}$$

We conclude that

$$\begin{aligned} &t^{1-\gamma} \frac{c}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \alpha_{k+1} \log \eta_{k+1} \\ &\quad \leq t^{1-\gamma} \frac{c 2^{1-\gamma}}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \frac{6(1-\gamma)}{\pi^2 (k+1)^2 (\log t)^2} g_{k+1}^{\gamma-1} \log \eta_{k+1} \\ &\quad = c 2^{1-\gamma} \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log \eta_{k+1}. \end{aligned}$$

We now have to bound the remaining term in (3.7.8), which equals

$$t^{1-\gamma} \frac{c}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \beta_{k+1} t^{2\gamma-1} = \frac{c}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \beta_{k+1} t^{\gamma}.$$

We use the definition of β_k in (3.7.4) to write

$$\frac{c}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 \beta_{k+1} t^{\gamma} = \frac{c}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 t^{\gamma} c \left[\alpha_k g_k^{1-2\gamma} + \beta_k \log \eta_k \right],$$

and again use the fact that $\alpha_k \leq 2^{1-\gamma} \frac{6(1-\gamma)}{\pi^2 k^2 (\log t)^2} g_k^{\gamma-1}$, so that

$$\frac{c}{1-\gamma} \frac{\pi^2 (\log t)^2}{6} (k+2)^2 t^{\gamma} c \alpha_k g_k^{1-2\gamma}$$

$$\leq \frac{c2^{1-\gamma}}{1-\gamma}\frac{\pi^2(\log t)^2}{6}(k+2)^2t^\gamma c\frac{6(1-\gamma)}{\pi^2k^2(\log t)^2}g_k^{\gamma-1}g_k^{1-2\gamma} = c^2 2^{1-\gamma}\frac{(k+2)^2}{k^2}\eta_k^\gamma.$$

By Definition 3.7.2, we have $c\beta_k t^{2\gamma-1} \leq \alpha_{k+1}$, so that, using (3.7.9) and (3.7.10),

$$\begin{aligned}
&\frac{c}{1-\gamma}\frac{\pi^2(\log t)^2}{6}(k+2)^2t^\gamma c\beta_k\log\eta_k \\
&\quad\leq \frac{c}{1-\gamma}\frac{\pi^2(\log t)^2}{6}(k+2)^2\alpha_{k+1}t^{1-\gamma}\log\eta_k \leq c2^{1-\gamma}\frac{(k+2)^2}{(k+1)^2}g_{k+1}^{\gamma-1}t^{1-\gamma}\log\eta_k \\
&\quad= c2^{1-\gamma}\frac{(k+2)^2}{(k+1)^2}\eta_{k+1}^{1-\gamma}\log\eta_k.
\end{aligned}$$

Since $k \mapsto \eta_k$ is increasing,

$$c2^{1-\gamma}\frac{(k+2)^2}{(k+1)^2}\eta_{k+1}^{1-\gamma}\log\eta_k \leq c2^{1-\gamma}\frac{(k+2)^2}{(k+1)^2}\eta_{k+1}^{1-\gamma}\log\eta_{k+1}.$$

Putting together all the bounds and taking a different constant $C = C(\gamma)$, we obtain (3.7.7). □

We can now obtain a useful bound on the growth of η_k.

Lemma 3.7.5 (Inductive bound on η_k)**.** *Let $(\eta_k)_{k\in\mathbb{N}}$ be given by $\eta_k = t/g_k$ and let $\kappa = \gamma/(1-\gamma) \in (1,\infty)$. Then, there exists a constant $B \geq 2$ such that*

$$\eta_k \leq \exp\bigl(B(\log\log t)\kappa^{k/2}\bigr).$$

for any $k = O(\log\log t)$.

Proof. We prove the lemma by induction on k, and start by initializing the induction. For $k = 0$,

$$\eta_0 = t/g_0 = \frac{t}{\left\lceil \frac{t}{(\log t)^2}\right\rceil} \leq (\log t)^2 = \mathrm{e}^{2\log\log t} \leq \mathrm{e}^{B\log\log t},$$

for any $B \geq 2$, which initializes the induction.

We next suppose that the statement is true for $l = 1,\dots,k-1$, and will prove it for k. Using that

$$(z+w)^{\frac{1}{1-\gamma}} \leq 2^{\frac{1}{1-\gamma}}\left(z^{\frac{1}{1-\gamma}} + w^{\frac{1}{1-\gamma}}\right),$$

we can write by Lemma 3.7.4, for a different constant C,

$$\eta_k \leq C\bigl[\eta_{k-2}^{\frac{\gamma}{1-\gamma}} + \eta_{k-1}(\log\eta_{k-1})^{\frac{1}{1-\gamma}}\bigr] = C\bigl[\eta_{k-2}^{\kappa} + \eta_{k-1}(\log\eta_{k-1})^{\frac{1}{1-\gamma}}\bigr].$$

Using this inequality, we can write

$$\eta_{k-2} \leq C\bigl[\eta_{k-4}^{\kappa} + \eta_{k-3}(\log\eta_{k-3})^{\frac{1}{1-\gamma}}\bigr],$$

so that, by $(z+w)^\kappa \le 2^\kappa(z^\kappa + w^\kappa)$,

$$\eta_k \le C(2C)^\kappa \big[\eta_{k-4}^{\kappa^2} + \eta_{k-3}^{\kappa}(\log \eta_{k-3})^{\frac{\kappa}{1-\gamma}}\big] + C\eta_{k-1}(\log \eta_{k-1})^{\frac{1}{1-\gamma}}.$$

Renaming $2C$ as C for simplicity, and iterating these bounds, we obtain

$$\eta_k \le C^{\sum_{l=0}^{k/2}\kappa^l}\eta_0^{\kappa^{k/2}} + \sum_{i=1}^{k/2} C^{\sum_{l=0}^{i-1}\kappa^l}\eta_{k-2i+1}^{\kappa^{i-1}}(\log \eta_{k-2i+1})^{\frac{\kappa^{i-1}}{1-\gamma}}. \tag{3.7.11}$$

For the first term in (3.7.11), we use the precise expression for η_0 to obtain

$$\begin{aligned} C^{\sum_{l=0}^{k/2}\kappa^l}\eta_0^{\kappa^{k/2}} &\le C^{\sum_{l=0}^{k/2}\kappa^l}\exp\Big(2(\log\log t)\kappa^{k/2}\Big) \\ &\le \frac{1}{2}\exp\Big(B(\log\log t)\kappa^{k/2}\Big), \end{aligned} \tag{3.7.12}$$

for a constant $B \ge 2$ large enough.

For the second term in (3.7.11), we use the induction hypothesis to obtain

$$\begin{aligned} &\sum_{i=1}^{k/2} C^{\sum_{l=0}^{i-1}\kappa^l}\eta_{k-2i+1}^{\kappa^{i-1}}(\log \eta_{k-2i+1})^{\frac{\kappa^{i-1}}{1-\gamma}} \\ &\quad\le \sum_{i=1}^{k/2} C^{\sum_{l=0}^{i-1}\kappa^l}\exp\Big(B(\log\log t)\kappa^{(k-1)/2}\Big)\Big[B(\log\log t)\kappa^{(k-2i+1)/2}\Big]^{\frac{\kappa^{i-1}}{1-\gamma}}. \end{aligned} \tag{3.7.13}$$

We can write

$$\begin{aligned} &\exp\Big(B(\log\log t)\kappa^{(k-1)/2}\Big) \\ &\qquad = \exp\Big(B(\log\log t)\kappa^{k/2}\Big)\exp\Big(B(\log\log t)\kappa^{k/2}\big(\sqrt{1/\kappa}-1\big)\Big). \end{aligned}$$

Since $\sqrt{1/\kappa}-1<0$, for $k = O(\log\log t)$ we can take B large enough such that

$$\begin{aligned} &\sum_{i=1}^{k/2} C^{\sum_{l=0}^{i-1}\kappa^l}\exp\Big(B(\log\log t)\kappa^{k/2}\big(\sqrt{1/\kappa}-1\big)\Big) \\ &\qquad\qquad \times\Big[B(\log\log t)\kappa^{(k-2i+1)/2}\Big]^{\frac{\kappa^{i-1}}{1-\gamma}} < \frac{1}{2}. \end{aligned} \tag{3.7.14}$$

We can now sum the bounds in (3.7.12) and (3.7.13)–(3.7.14) to obtain

$$\eta_k \le \Big(\frac{1}{2}+\frac{1}{2}\Big)\exp\Big(B(\log\log t)\kappa^{k/2}\Big),$$

as required. This completes the proof of Lemma 3.7.5. □

Now we are ready to complete the proof of Proposition 3.4.9:

Proof of Proposition 3.4.9. Recall the definition of $\bar{k}_t$ in (3.4.21). By (3.4.26),

$$\mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t}(x,y) \le \bar{k}_t) \le \sum_{k=1}^{\bar{k}_t} \sum_{l=1}^{g_k-1} f_{k,t}(x,l) + \sum_{k=1}^{\bar{k}_t} \sum_{l=1}^{g_k-1} f_{k,t}(y,l) + \sum_{k=1}^{2\bar{k}_t} \sum_{l=g_{\lfloor k/2 \rfloor}}^{t} f_{\lfloor k/2 \rfloor,t}(x,l) f_{\lceil k/2 \rceil,t}(y,l). \tag{3.7.15}$$

We start with the first two sums, which are equal except that x is replaced by y in the second. We use (3.7.6), together with the fact that $l \le g_k - 1$, to obtain

$$\sum_{k=1}^{\bar{k}_t} \sum_{l=1}^{g_k-1} f_{k,t}(x,l) \le \sum_{k=1}^{\bar{k}_t} \alpha_k \sum_{l=1}^{g_k-1} l^{-\gamma}.$$

Since $\gamma \in (1/2, 1)$, there exists a constant b such that

$$\sum_{k=1}^{\bar{k}_t} \alpha_k \sum_{l=1}^{g_k-1} l^{-\gamma} \le b \sum_{k=1}^{\bar{k}_t} \alpha_k (g_k - 1)^{1-\gamma}.$$

We use the definition of g_k to bound

$$b \sum_{k=1}^{\bar{k}_t} \alpha_k (g_k - 1)^{1-\gamma} < b \sum_{k=1}^{\bar{k}_t} \frac{6}{\pi^2 k^2 (\log t)^2} \le \frac{b}{(\log t)^2},$$

as required. The term with y replacing x is identical.

We next consider now the third sum in (3.7.15), we again use the bound in (3.7.6) as well as the fact that $k \mapsto g_k$ is non-increasing, while $k \mapsto \alpha_k, \beta_k$ are non-decreasing, to obtain

$$\begin{aligned}
&\sum_{k=1}^{2\bar{k}_t} \sum_{l=g_{\lfloor k/2 \rfloor}}^{t} (\alpha_{\lfloor k/2 \rfloor} l^{-\gamma} + \beta_{\lfloor k/2 \rfloor} l^{\gamma-1})(\alpha_{\lceil k/2 \rceil} l^{-\gamma} + \beta_{\lceil k/2 \rceil} l^{\gamma-1}) \qquad (3.7.16)\\
&\le \sum_{k=1}^{2\bar{k}_t} \sum_{l=g_{\lceil k/2 \rceil}}^{t} (\alpha_{\lceil k/2 \rceil} l^{-\gamma} + \beta_{\lceil k/2 \rceil} l^{\gamma-1})^2 \le 2 \sum_{k=1}^{2\bar{k}_t} \sum_{l=g_{\lceil k/2 \rceil}}^{t} \left(\alpha^2_{\lceil k/2 \rceil} l^{-2\gamma} + \beta^2_{\lceil k/2 \rceil} l^{2\gamma-2}\right)\\
&= 2 \sum_{k=1}^{2\bar{k}_t} \sum_{l=g_{\lceil k/2 \rceil}}^{t} \alpha^2_{\lceil k/2 \rceil} l^{-2\gamma} + 2 \sum_{k=1}^{2\bar{k}_t} \sum_{l=g_{\lceil k/2 \rceil}}^{t} \beta^2_{\lceil k/2 \rceil} l^{2\gamma-2}.
\end{aligned}$$

This leads to two terms that we bound one by one. For the first term in (3.7.16), we

can write

$$2\sum_{k=1}^{2\bar{k}_t}\sum_{l=g_{\lceil k/2\rceil}}^{t}\alpha^2_{\lceil k/2\rceil}l^{-2\gamma}\le 2b'\sum_{k=1}^{2\bar{k}_t}\alpha^2_{\lceil k/2\rceil}g^{1-2\gamma}_{\lceil k/2\rceil}=2b'\sum_{k=1}^{2\bar{k}_t}\alpha^2_{\lceil k/2\rceil}g^{2-2\gamma}_{\lceil k/2\rceil}\eta_{\lceil k/2\rceil}\frac{1}{t}.$$

By definition of g_k,

$$\alpha_{\lceil k/2\rceil}g^{1-\gamma}_{\lceil k/2\rceil}\le 2^{1-\gamma}\alpha_{\lceil k/2\rceil}(g_{\lceil k/2\rceil}-1)^{1-\gamma}\le\frac{6(1-\gamma)}{\pi^2(k/2)^2(\log t)^2}.$$

Therefore,

$$\begin{aligned}\frac{2b'}{t}\eta_{\bar{k}_t}\sum_{k=1}^{2\bar{k}_t}\alpha^2_{\lceil k/2\rceil}g^{2-2\gamma}_{\lceil k/2\rceil}&\le\frac{2b'2^{2(1-\gamma)}}{t}\eta_{\bar{k}_t}\sum_{k=1}^{2\bar{k}_t}\Big(\frac{24(1-\gamma)}{\pi^2k^2(\log t)^2}\Big)^2\\&\le\frac{C}{t}\eta_{\bar{k}_t}\frac{1}{(\log t)^4}=o\left((\log t)^{-4}\right),\end{aligned}\tag{3.7.17}$$

since $\eta_{\bar{k}_t}=o(t)$ by the definition of $\bar{k}_t$ in (3.4.21) and Lemma 3.7.5: in fact, $\gamma=m/(2m+\delta)$ and consequently $\kappa=\gamma/(1-\gamma)=(1+\delta/m)^{-1}$, that is $\kappa=1/(\tau-2)$ (recall that $\tau=3+\delta/m$).

For the second term in (3.7.16), we use that $2-2\gamma\in(0,1)$ to compute

$$2\sum_{k=1}^{2\bar{k}_t}\sum_{l=g_{\lceil k/2\rceil}}^{t}\beta^2_{\lceil k/2\rceil}l^{2\gamma-2}\le 2b''\sum_{k=1}^{2\bar{k}_t}\beta^2_{\lceil k/2\rceil}t^{2\gamma-1}.$$

By definition of α_k, we have $\beta_k\le\alpha_{k+1}t^{1-2\gamma}$, which means that

$$\sum_{r=1}^{\bar{k}_t}\beta^2_rt^{2\gamma-1}\le\sum_{r=1}^{\bar{k}_t}\alpha^2_{r+1}t^{2-4\gamma}t^{2\gamma-1}=\sum_{r=1}^{\bar{k}_t}\alpha^2_{r+1}t^{1-2\gamma}\le\frac{1}{t}\eta^{2-2\gamma}_{\bar{k}_t}\sum_{r=1}^{\bar{k}_t}\alpha^2_{r+1}g^{2-2\gamma}_{r+1},$$

which is $o((\log t)^{-4})$ as in (3.7.17). We conclude that there exists a constant p such that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t}(x,y)\le 2\bar{k}_t)\le p(\log t)^{-2},$$

as required. This completes the proof of Proposition 3.4.9. □

3.8. Technical bounds on exploration trees in PAM: Proof of Lemmas 3.6.4 and 3.6.6

3.8.1. Proof of Lemma 3.6.4

We adapt the proof of [59, Lemma A.4]. We design a Pólya urn experiment to bound the probability that a fixed vertex $i\in[t/2]\setminus\mathrm{Core}_t$ accumulates too many

edges from the vertices $t/2+1,\dots,t$.

Let us fix $i \in [t/2]$. We consider one urn containing blue balls and containing red balls. For every edge that we add to the graph from $t/2+1$ to t (which means $mt/2$ edges), we need to keep track of the number of edges attached to i. To give an upper bound, we can assume that $D_i(t/2) = (\log t)^\sigma$. Let R_k and B_k denote the number of red and blue balls after k draws, so that $R_0 = (\log t)^\sigma$ and $B_0 = m(t/2) - (\log t)^\sigma$. Thus, R_0 is the maximal degree of vertex i at time $t/2$, while B_0 is the minimal degree of all vertices unequal to i at time $t/2$. We consider two linear weight functions for the number of balls in each urn,

$$W_k^{(r)} = k + \delta, \quad \text{and} \quad W_k^{(b)} = k + \delta(t/2 - 1). \tag{3.8.1}$$

At time $k \geq 0$, let R_k and B_k denote the number of red and blue balls after k draws. Then we draw a ball colored red or blue according to the weights $W_{R_k}^{(r)}$ and $W_{B_k}^{(b)}$. Here $W_{R_k}^{(r)}$ represents the weight of the vertex i and $W_{B_k}^{(b)}$ is the weight of the rest of the graph. Naturally, $R_k + B_k = m(t/2) + k$ is deterministic, as it should be in a Polyá urn, and also $W_{R_k}^{(r)} + W_{B_k}^{(b)} = (m+\delta)t/2 + k$ is deterministic.

We consider $mt/2$ draws, and at every one, we pick a red ball with probability proportional to $W_{R_k}^{(r)}$ and a blue ball with probability proportional to $W_{B_k}^{(b)}$, respectively. We add one ball of the same color as the selected color (next to the drawn ball, which we put back). Recalling (3.1.10), one could notice that at every draw we should add $1+\delta/m$ to the weight of the blue balls since the weight of the graph (with i excluded) is always increasing due to the fact that the new edge is attached to a new vertex (that is not equal to i). However, $1+\delta/m \geq 0$ and thus ignoring this effect only increases the probability of choosing i. This suffices for our purposes, since we are only interested in an upper bound.

We denote by $(X_n)_{n=1}^{mt/2}$ a sequence of random variables, where $X_n = 1$ whenever the n-th extraction is a red ball (a new edge is attached to i). As a consequence,

$$\mathbb{P}\Big(D_i(t) \geq (1+B)(\log t)^\sigma \mid D_i(t/2) < (\log t)^\sigma\Big) \leq \mathbb{P}\Big(\sum_{n=1}^{mt/2} X_n \geq B(\log t)^\sigma\Big).$$

As the reader can check, the sequence of random variables $(X_n)_{n=1}^{mt/2}$ is exchangeable, so we can apply De Finetti's Theorem, and obtain that

$$\mathbb{P}\Big(\sum_{n=1}^{mt/2} X_n \geq B(\log t)^\sigma\Big) = \mathbb{E}\left[\mathbb{P}\left(\mathrm{Bin}(mt/2, U) \geq B(\log t)^\sigma \mid U\right)\right], \tag{3.8.2}$$

where U is a distribution on $[0,1]$. In the case of the Polyá urn with two colors and linear weights, as discussed in Section 4.4, U has a Beta distribution with parameters given by

$$\alpha_t = (\log t)^\sigma + \delta, \quad \text{and} \quad \beta_t = \frac{t}{2}(2m+\delta) - ((\log t)^\sigma + \delta).$$

We call

$$\psi(u) = \mathbb{P}\left(\mathrm{Bin}(mt/2, u) \geq B(\log t)^\sigma\right) \leq 1.$$

By the classical Chernoff bound

$$\mathbb{P}(\mathrm{Bin}(n,u) \geq k) \leq e^{-nI_u(k/n)}, \quad \text{where} \quad I_u(a) = a(\log(a/u) - 1) + u$$

is the large deviation function of a $\mathrm{Pois}(u)$ random variable, see [85, Corollary 2.20]. For $a \geq 8u$ we can bound $I_u(a) \geq a(\log 8 - 1) \geq a$, hence

$$\psi(u) \leq \mathrm{e}^{-B(\log t)^\sigma} \quad \text{whenever} \quad 4mtu \leq B(\log t)^\sigma.$$

Define

$$g(t) = \frac{B(\log t)^\sigma}{4mt}.$$

Using g and ψ in (3.8.2), we have

$$\mathbb{E}\left[\mathbb{P}\left(\mathrm{Bin}(mt/2, U) \geq B(\log t)^\sigma \mid U\right)\right] \leq \mathrm{e}^{-B(\log t)^\sigma} + \mathbb{P}\left(U > g(t)\right). \tag{3.8.3}$$

Note that $\mathrm{e}^{-B(\log t)^\sigma} = o(1/t)$ since $\sigma > 1$ and for B large enough. What remains is to show that also the second term in the right-hand side of (3.8.3) is $o(1/t)$. Since U has a Beta distribution, its density $f_U(u) = \frac{\Gamma(\alpha_t+\beta_t)}{\Gamma(\alpha_t)\Gamma(\beta_t)} u^{\alpha_t - 1}(1-u)^{\beta_t - 1}$ attains its maximum at the point $\bar{u}_t = \frac{\alpha_t - 1}{\alpha_t + \beta_t - 2}$. Since $g(t) \geq \bar{u}_t$ (which can easily be checked, for B large enough), we have

$$\mathbb{P}\left(U > g(t)\right) \leq \frac{\Gamma(\alpha_t + \beta_t)}{\Gamma(\alpha_t)\Gamma(\beta_t)} \left(1 - g(t)\right)^{\beta_t} g(t)^{\alpha_t - 1}. \tag{3.8.4}$$

We next bound each of these terms separately. Firstly, asymptotically as $t \to \infty$ and since $\delta < 0$,

$$\frac{\Gamma(\alpha_t + \beta_t)}{\Gamma(\beta_t)} \leq (\alpha_t + \beta_t)^{\alpha_t} = (mt(1 + \delta/2))^{\alpha_t} \leq (mt)^{\alpha_t}. \tag{3.8.5}$$

Secondly,

$$g(t)^{\alpha_t} = \left(\frac{B(\log t)^\sigma}{4mt}\right)^{\alpha_t} = (B/4)^{\alpha_t} \left(\frac{(\log t)^\sigma}{mt}\right)^{\alpha_t}. \tag{3.8.6}$$

Thirdly, since $1 - x \leq \mathrm{e}^{-x}$,

$$\left(1 - g(t)\right)^{\beta_t} = \left(1 - \frac{B(\log t)^\sigma}{4mt}\right)^{\beta_t} \leq \exp\left(-cB(\log t)^\sigma\right), \tag{3.8.7}$$

for some $c \in (0, \infty)$. Finally, by Stirling's formula $\Gamma(\alpha_t) \geq (\alpha_t/\mathrm{e})^{\alpha_t}$, so that

$$\Gamma(\alpha_t)^{-1} \leq (\mathrm{e}/\alpha_t)^{\alpha_t}. \tag{3.8.8}$$

Substituting (3.8.5), (3.8.6), (3.8.7) and (3.8.8) into (3.8.4), we obtain

$$\mathbb{P}\left(U > g(t)\right) \leq \frac{\exp\left(-cB(\log t)^{\sigma}\right)}{g(t)} \Big(\frac{B\mathrm{e}(\log t)^{\sigma}}{4\alpha_t}\Big)^{\alpha_t} \leq t \exp\left(-cB(\log t)^{\sigma}/2\right) = o(1/t),$$

for B large enough and using that $(\log t)^{\sigma}/\alpha_t \leq 1 - \delta/\alpha_t = 1 + O(1/\alpha_t)$. This completes the proof of the lemma. □

3.8.2. Proof of Lemma 3.6.6

For $i = 0, \ldots, k$, we denote by N_i the number of vertices in the k-exploration graph at distance i from v; the set of such vertices will be called "level i". Clearly, $N_0 = 1$ because the only vertex at level 0 is v. Plainly, if there are no collisions between level $i-1$ and level i, then $N_i = mN_{i-1}$. The number of collisions l_i between level $i-1$ and level i is then given by

$$l_i := mN_{i-1} - N_i\,, \tag{3.8.9}$$

and the total number of collisions is, by assumption,

$$\bar{l}_k := l_1 + \ldots + l_k \leq l\,.$$

The assumption that no vertex has only self-loops implies that $N_i \geq 1$ for every i, because it ensures that the youngest vertex at level $i-1$ has at least one "descendant" at level i. Therefore

$$l_i \leq mN_{i-1} - 1\,. \tag{3.8.10}$$

For later purposes, it is convenient to start with $N_0 \geq 1$ vertices (even though we are eventually interested in the case $N_0 = 1$). Rewriting (3.8.9) as $N_i = mN_{i-1} - l_i$, a simple iteration yields

$$N_k = m^k N_0 - m^{k-1} l_1 - m^{k-2} l_2 + \ldots - m l_{k-1} - l_k\,. \tag{3.8.11}$$

This yields $N_k \geq m^k N_0 - m^{k-1}(l_1 + l_2 + \ldots + l_k)$, that is

$$N_k \geq (mN_0 - \bar{l}_k) m^{k-1}\,. \tag{3.8.12}$$

This lower bound is only useful if $\bar{l}_k < mN_0$, otherwise the right hand side is negative. To deal with the complementary case, we now show by induction the following useful bound:

$$N_k \geq m^{-\frac{\bar{l}_k - (mN_0 - 1)}{m-1}}\, m^{k-2} \qquad \text{when} \quad \bar{l}_k \geq mN_0 - 1\,. \tag{3.8.13}$$

The case $k = 1$ is easy: since $\bar{l}_1 = l_1 \leq mN_0 - 1$ by (3.8.10), we only need to consider the extreme case $\bar{l}_1 = mN_0 - 1$, when (3.8.13) reduces to $N_1 \geq \frac{1}{m}$, which holds since $N_1 \geq 1$. Next we fix $k \geq 1$ and our goal is to prove (3.8.13) for $k+1$, assuming that it holds for k.

The crucial observation is that, iterating relation $N_i = mN_{i-1} - l_i$ (that is (3.8.9))

from $i = k+1$ until $i = 2$, we get an analogue of (3.8.11), that is

$$N_{k+1} = m^k N_1 - m^{k-1} l_2 - m^{k-2} l_3 + \ldots - m l_k - l_{k+1}\,.$$

This means that N_{k+1} *coincides with* N_k *in* (3.8.11) *where we replace* N_0 *by* N_1 *and* $(l_1, \ldots, l_k)$ *by* $(l_2, \ldots, l_{k+1})$ *(therefore* $\bar{l}_k$ *is replaced by* $\bar{l}_{k+1} - l_1$*).* As a consequence, by the inductive assumption, we can apply (3.8.13) which yields

$$N_{k+1} \geq m^{-\frac{(\bar{l}_{k+1} - \bar{l}_1) - (mN_1 - 1)}{m-1}}\, m^{k-2} \qquad \text{when} \quad (\bar{l}_{k+1} - \bar{l}_1) \geq mN_1 - 1\,. \tag{3.8.14}$$

For later use, we note that, analogously, relation (3.8.12) gives

$$N_{k+1} \geq (mN_1 - (\bar{l}_{k+1} - l_1))m^{k-1}\,. \tag{3.8.15}$$

By (3.8.9) and (3.8.10), which yield $N_1 = mN_0 - l_1$ and $l_1 \leq mN_0 - 1$, we can write

$$(\bar{l}_{k+1} - l_1) - (mN_1 - 1) = \bar{l}_{k+1} - m^2 N_0 + (m-1)l_1 + 1 \leq \bar{l}_{k+1} - (mN_0 - 1) - (m-1),$$

which plugged into (3.8.14) gives

$$N_{k+1} \geq m^{-\frac{\bar{l}_{k+1} - (mN_0 - 1)}{m-1}}\, m^{k-1} \qquad \text{when} \quad (\bar{l}_{k+1} - \bar{l}_1) \geq mN_1 - 1\,. \tag{3.8.16}$$

This is precisely the analogue of (3.8.13) for $k+1$, which is our goal, except for the "wrong" restriction $(\bar{l}_{k+1} - \bar{l}_1) \geq mN_1$, instead of $\bar{l}_{k+1} \geq mN_0$. We are thus left with showing that the inequality in (3.8.16) still holds if $(\bar{l}_{k+1} - \bar{l}_1) < mN_1 - 1$ and $\bar{l}_{k+1} \geq mN_0 - 1$, but this is easy, because these two conditions, together with (3.8.15), imply

$$N_{k+1} \geq (mN_1 - (\bar{l}_{k+1} - l_1))m^{k-1} \geq m^{k-1} \geq m^{-\frac{\bar{l}_{k+1} - (mN_0 - 1)}{m-1}}\, m^{k-1}\,.$$

We are ready to conclude. The bounds (3.8.12) and (3.8.13), for $N_0 = 1$ and $l_k \leq l$, yield

$$N_k \geq m^{-\frac{l}{m-1}}\, m^{k-1} = s(m,l)\, m^k\,, \qquad \text{where} \qquad s(m,l) := m^{-1-\frac{l}{m-1}}\,,$$

which is what we wanted to prove. □

4

Treelike property and local weak limit of PAMs

Content and structure of the chapter

In this chapter, we formalize the heuristic idea that PAMs are locally treelike, using the notion of local weak convergence. We move from the analysis made by Berger et al. [21], which was stated for $\tau \geq 3$, and we extend it to $\tau > 2$. In particular, we use the interpretation of PAMs as *Pólya urn scheme*, where the number of balls in each urn represents the degree of a vertex. We extend this argument to many different versions of PAMs, in particular looking at those models that for $m \geq 2$ are defined from the tree setting ($m = 1$) through a collapsing procedure.

The chapter is structured as follows: In Section 4.1 we introduce local weak convergence, giving the rigorous construction. In Section 4.2 we state the main results of the chapter, namely the local weak convergence of general PAMs (Theorem 4.2.1) and the specialized statement about PAMs obtained through collapsing (Theorem 4.2.2). In Section 4.3 we list the different versions of PAMs that we consider in this chapter. In Section 4.4 we present the general theory of Pólya urn schemes, showing how we can use this to represent the PAM model (e) of Section 4.3, and in Section 4.6 we extend this construction to other PAMs. Section 4.7 contains the sketch of the proof of the original result in [21], that is useful in order to understand the extension to Theorem 4.2.1. In Section 4.9 we prove our main results. The content of this chapter is based on a work in preparation [71].

4.1. Preliminaries: local weak convergence

We present the definition of LWC for undirected graphs. As mentioned in Section 1.7, we will extend this construction to directed graphs in Chapter 6. In the present section, we only present the construction for undirected graphs.

We start by defining what a rooted graph is:

Definition 4.1.1 (Rooted graph)**.** *Let G be a locally finite graph with vertex set $V(G)$ (finite or countable), and edge set $E(G)$. Fix a vertex $\varnothing \in G$ and call it the* root*. The pair $(G, \varnothing)$ is called a* rooted graph*.*

We are not interested in the labeling of the vertices, but only in the graph structure. For this, we define isomorphisms between rooted graphs as follows:

Definition 4.1.2 (Isomorphism)**.** *An isomorphism between two rooted graphs $(G, \varnothing)$ and $(G', \varnothing')$ is a bijection $\gamma : V(G) \to V(G')$ such that*

(1) $(j, i) \in E(G)$ if and only if $(\gamma(j), \gamma(i)) \in E(G)$;
(2) $\gamma(\varnothing) = \varnothing'$.

We write $(G, \varnothing) \cong (G', \varnothing')$ to denote that $(G, \varnothing)$ and $(G', \varnothing')$ are isomorphic rooted graphs.

Denote the space of all rooted graphs (up to isomorphisms) by $\mathcal{G}_\star$. Formally, $\mathcal{G}_\star$ is the quotient space of the set of all locally-finite rooted graphs with respect to the equivalence relation given by isomorphisms.

For a rooted graph $(G, \varnothing) \in \mathcal{G}_\star$, we let $U_{\leq k}(\varnothing)$ denote the subgraph of G of all vertices at graph distance at most k away from $\varnothing$. Formally, this means that $U_{\leq k}(\varnothing) = (V(U_{\leq k}(\varnothing)), E(U_{\leq k}(\varnothing)))$, where

$$V(U_{\leq k}(\varnothing)) = \{i \colon d_G(i, \varnothing) \leq k\}, \quad E(U_{\leq k}(\varnothing)) = \{\{j, i\} \colon j, i \in V(U_{\leq k}(\varnothing))\}.$$

We call $U_{\leq k}(\varnothing)$ the k-neighborhood around $\varnothing$. We use this notion to define the distance between two rooted graphs:

Definition 4.1.3 (Local distance)**.** *The function $d_{loc}((G, \varnothing), (G', \varnothing')) = 1/(1+\kappa)$, where*

$$\kappa = \inf_{k \geq 1} \{U_{\leq k}(\varnothing) \ncong U_{\leq k}(\varnothing')\},$$

is called the local distance *on the space of rooted graphs $\mathcal{G}_\star$.*

It is possible to prove that d_{loc} is an actual distance on the space of rooted graphs. In particular, the space $(\mathcal{G}_\star, d_{loc})$ is a Polish space (see [68, Appendix A] for the proof for an equivalent definition of a distance). The function d_{loc} measures how distant two rooted graphs are *from the point of view of the root*. In many graphs though, there is no vertex that can be naturally chosen as a root, for instance in configuration models or Erdős-Rényi random graph. For this reason, it is useful to choose the root *at random*. Define, for any graph G,

$$\mathcal{P}(G) = \frac{1}{n} \sum_{i \in [n]} \delta_{(G,i)}. \tag{4.1.1}$$

Given a graph G of size n, $\mathcal{P}(G)$ is a probability measure that assigns the root uniformly at random among the n vertices. When we consider a sequence of graphs $(G_n)_{n\in\mathbb{N}}$, we denote $\mathcal{P}(G_n)$ simply by $\mathcal{P}_n$. With this notion, we are ready to define LWC for undirected deterministic graphs:

Definition 4.1.4 (Local weak convergence)**.** *Consider a deterministic sequence of locally finite graphs $(G_n)_{n\in\mathbb{N}}$. We say that $(G_n)_{n\in\mathbb{N}}$* converges in the local weak sense *to a (possibly) random element $(G,\varnothing)$ of $\mathcal{G}_\star$ with law $\mathcal{P}$, if, for any bounded continuous (with respect to the topology induced by d_{loc}) function $f:\mathcal{G}_\star\to\mathbb{R}$,*

$$\mathbb{E}_{\mathcal{P}_n}[f]\longrightarrow\mathbb{E}_{\mathcal{P}}[f],$$

where $\mathbb{E}_{\mathcal{P}_n}$ and $\mathbb{E}_{\mathcal{P}}$ denote the expectation with respect to $\mathcal{P}_n$ and $\mathcal{P}$, respectively.

In particular, this means that probabilities of open sets in the topology converge. Fix (H,y) finite, then

$$\begin{aligned} B_R(H,y) &= \{(G,\varnothing)\in\mathcal{G}_\star : d_{loc}((H,y),(G,\varnothing))\le R\} \\ &= \left\{(G,\varnothing)\in\mathcal{G}_\star : U_{\le\lfloor 1/R\rfloor}(\varnothing)\cong(H,y)\right\}. \end{aligned} \tag{4.1.2}$$

Elements in this open ball are determined by the neighborhood of the root up to distance $\lfloor 1/R\rfloor$. As a consequence, the probability $\mathcal{P}_n$ of the ball $B_R(h,y)$ is given by

$$\mathcal{P}_n(B_R(h,y))=\frac{1}{n}\sum_{i\in[n]}\mathbb{1}\left\{U_{\le\lfloor 1/R\rfloor}(i)\cong(H,y)\right\}.$$

This implies that it suffices to consider the local structure of the neighborhood of a typical vertex to obtain the probability $\mathcal{P}_n$ of any open ball. We now state a criterion for a sequence of deterministic graphs to converge in the LW sense as in Definition 4.1.4:

Theorem 4.1.5 (Criterion for local weak convergence)**.** *Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of graphs. Then G_n converges in the local weak sense to $(G,\varnothing)$ with law $\mathcal{P}$ when, for every finite rooted graph (H,y),*

$$\mathcal{P}_n(H)=\frac{1}{n}\sum_{i\in[n]}\mathbb{1}\left\{U_{\le k}(i)\cong(H,y)\right\}\longrightarrow\mathcal{P}\left(U_{\le k}(\varnothing)\cong(H,y)\right). \tag{4.1.3}$$

The proof Theorem 4.1.5 can be found in [86, Section 1.4]. Notice that for $(H,y)\in\mathcal{G}_\star$, the functions $\mathbb{1}\{U_{\le k}(\varnothing)\cong(H,y)\}$ are continuous with respect to the local weak topology and uniquely identify the limit.

So far we have considered sequences of deterministic graphs. Whenever we consider a *random graph* G_n, we have two sources of randomness. First, we have the randomness of the choice of the root, and then the randomness of the graph itself. For this reason, it is necessary to specify the randomness we take expectation with respect to, giving rise to different ways of convergence. We specify this in the following definition:

Definition 4.1.6 (Local weak convergence)**.** *Consider a sequence of random graphs $(G_n)_{n\in\mathbb{N}}$, and a probability $\mathcal{P}$ on $\mathcal{G}_\star$. Denote by $\mathcal{P}_n$ the probability associated to G_n as in* (4.1.1)*.*

▷ *We say that* G_n converges in distribution in the local weak sense *to $\mathcal{P}$ if, for any bounded continuous function $f : \mathcal{G}_\star \to \mathbb{R}$,*

$$\mathbb{E}\left[\mathbb{E}_{\mathcal{P}_n}\left[f\right]\right] \longrightarrow \mathbb{E}_{\mathcal{P}}\left[f\right]; \tag{4.1.4}$$

▷ *We say that* G_n converges in probability in the local weak sense *to $\mathcal{P}$ if, for any bounded continuous function $f : \mathcal{G}_\star \to \mathbb{R}$,*

$$\mathbb{E}_{\mathcal{P}_n}\left[f\right] \xrightarrow{\mathbb{P}} \mathbb{E}_{\mathcal{P}}\left[f\right]; \tag{4.1.5}$$

▷ *We say that* G_n converges almost surely in the local weak sense *to $\mathcal{P}$ if, for any bounded continuous function $f : \mathcal{G}_\star \to \mathbb{R}$,*

$$\mathbb{E}_{\mathcal{P}_n}\left[f\right] \overset{\mathbb{P}-a.s.}{\longrightarrow} \mathbb{E}_{\mathcal{P}}\left[f\right]. \tag{4.1.6}$$

Notice that the left-hand term in (4.1.5) is a random variable, while the right-hand side is deterministic. In fact, (4.1.5) implies (4.1.4), but the opposite is not true. Similarly, (4.1.6) implies (4.1.5).

Similarly to Theorem 4.1.5, we can give a criterion for the convergence of a sequence of random graphs:

Theorem 4.1.7 (Criterion for local weak convergence of random graphs)**.** *Consider a sequence of random graphs $(G_n)_{n\in\mathbb{N}}$, and a distribution $\mathcal{P}$ on $\mathcal{G}_\star$. Then, as $n \to \infty$,*

(1) G_n converges in distribution *in the LW sense to $\mathcal{P}$ if, for every fixed $k \in \mathbb{N}$ and finite directed marked rooted graph (H, y),*

$$\mathbb{E}\left[\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\left\{U_{\leq k}(i) \cong (H, y)\right\}\right] \longrightarrow \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y)\right)$$

(2) G_n converges in probability *in the LW sense to $\mathcal{P}$ if, for every fixed $k \in \mathbb{N}$ and finite directed marked rooted graph (H, y),*

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\left\{U_{\leq k}(i) \cong (H, y)\right\} \xrightarrow{\mathbb{P}} \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y)\right);$$

(3) G_n converges almost surely *in the LW sense to $\mathcal{P}$ if for every fixed $k \in \mathbb{N}$ and finite directed marked rooted graph (H, y),*

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\left\{U_{\leq k}(i) \cong (H, y)\right\} \overset{\mathbb{P}-a.s.}{\longrightarrow} \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y)\right),$$

where $\mathbb{P}$ denotes the law of the random sequence $(G_n)_{n\in\mathbb{N}}$.

Proof. The proof of Theorem 4.1.7 follows immediately from Theorem 4.1.5. □

4.2. Convergence of PAMs: main result

The main results we prove in the present chapter are the two theorems stated in this section. The first Theorem is about the convergence of all the different versions of PAMs defined in Section 4.3:

Theorem 4.2.1 (Universality of LW limit)**.** *PAMs (a)-(g) defined in Section 4.3 converge in probability in the LW sense to the Pólya point tree for all $\delta > -m$.*

For models (d), (f) and (g) the result is proven in [21] Models (d), (f) and (g) are defined by a mixtures of PA function depending on the degree ($f(k) = k$) and a uniform choice of attachment (an edge is attached to a vertex chosen uniformly at random). For a more precise description, we refer to Section 4.3 (in particular, see (4.3.4)). This mechanism is equivalent to considering $\delta \geq 0$.

The original proof in [21] makes extensive use of Pólya urn schemes (see Section 4.4). In simple words, a Pólya urn schemes consists of an urn with balls of two colors (red and blue). At every time step, we draw a ball, and we put it back in the urn with an additional ball of the same color as the one we have chosen. Notice that *the probability of choosing a ball of a color is propostional to the number of balls of that color*. This is similar to PAMs, where vertices are chosen proportionally to their degree. We extend this construction using *affine urn schemes*, a generalized construction where we choose balls according to an affine function of the number of balls in the urn (similarly to affine PA functions). We define a new class of random graphs called *unit graphs*, that link affine urn schemes with PAMs with affine PA function. In this way, we are able to extend the result in [21] to values all values $\delta > -m$.

The last model defined in Section 4.3, called model (h), does not converge in the LW sense to the Pólya point tree. As shown by Dereich and Mőrters [54, 55, 56], model (h) converges to an inhomogeneous multi-type branching processes called *idealized random tree*. This random tree has similar features as the Pólya point tree, but it is not the same rooted tree. In particular, this implies that Model (h) belongs to a different universality class as the other versions of PAM in Section 4.3.

The second theorem of this chapter consists in the LWC of Model (a) in Section 4.3. This particular version of PAM for $m \geq 2$ is defined in terms of the model for $m = 1$ through the *collapsing procedure* defined in Chapter 2. More precisely, for every $t \in \mathbb{N}$, $\mathrm{PA}_t^{(a)}(m, \delta)$ is defined by $\mathrm{PA}_{mt}^{(a)}(1, \delta/m)$, where vertices are grouped together by groups of m.

We can formulate a general statement:

Theorem 4.2.2 (Collapsing PAMs)**.** *Let $(\mathrm{PA}_t(m, \delta, G_0))_{t\geq 1}$ be a PAM with parameters $m \geq 1$ and $\delta > -m$, with initial graph G_0. Assume that $\mathrm{PA}_t(m, \delta, G_0)$ can be defined by $\mathrm{PA}_{mt}(1, \delta/m, \bar{G}_0)$, for some initial graph $\bar{G}_0$, through the collapsing procedure. Then, if*

$\mathrm{PA}_t(1, \delta/m, \bar{G}_0)$ *converges locally weakly to the Pólya point tree with parameters 1 and* δ/m*, then* $\mathrm{PA}_t(m, \delta, G_0)$ *converges locally weakly to Pólya point tree with parameters* m *and* δ*.*

Theorem 4.2.2 shows that the properties of PAMs defined through collapsing are determined by the underlying tree model. In particular, the initial graph is irrelevant for the convergence result. In other words, for PAMs with $m \geq 2$ defined through collapsing, a sufficient condition for the LW convergence to the Pólya point tree is the convergence of the corresponding model with $m = 1$ to the Pólya point tree (with different parameters).

4.3. Variations of PAMs and Pólya point tree

There exists several definitions of PAMs, each one defined by slightly different attachment probabilities. Here we make a list of what we consider.

Model (a). Model (a) is taken from [85, Model (a)]. This model is considered in Chapter 3 when we investigate the diameter of PAMs. For the precise definition, we refer to Definition 3.1.4. Here we just recall that the model for $m \geq 2$ is defined through collapsing, and every edge in the graph is allowed to be a self-loop. Multiple edges are allowed.

Model (b). Model (b) is taken from [85, Model (b)]. Model (b) for $m = 1$ is a tree without self-loops. In particular, we start with with $\mathrm{PA}_2^{(b)}(1, \delta)$ consisting of two vertices with two edges between them. Then, the attachment probabilities for $t \geq 3$ are given by

$$\mathbb{P}\left(t \to i \mid \mathrm{PA}_{t-1}^{(b)}(1, \delta)\right) = \frac{D_i(t-1) + \delta}{(t-1)(2+\delta)} \quad \text{for } i \in [t-1]. \tag{4.3.1}$$

When $m \geq 2$, we obtain a graph where edges are added sequentially with intermediate degree updates, but the first edge of every new vertex is not allowed to form a self-loop. This means that $\mathrm{PA}_1^{(m,\delta)}(b)$ consists of a single vertex with m self loops. Then, for $t \geq 2$,

$$\mathbb{P}\left(t \xrightarrow{j} i \mid \mathrm{PA}_{t-1}^{(m,\delta)}(b)\right) = \frac{D_i(t-1, j-1) + \delta}{(t-1)(2m+\delta) + (j-1)(2+\delta/m)}, \tag{4.3.2}$$

where we point out that $D_t(t-1, j-1) = 0$ if $j = 1$, and for $j \geq 2$, $D_t(t-1, j-1)$ is equal to j plus the number of self-loops of vertex t after the $(j-1)$-st edge has been attached.

Model (b'). Model (b') is a modification of model (b). We modify the starting graph by assuming that $\mathrm{PA}_2^{(1,\delta)}(b)$ consists of two vertices with only one edge. Then, model (b') is exactly the discrete-time version of a continuous-time branching process tree (individual 1 does not have a self-loop, where vertex 1 in model (b) has one). When $m \geq 2$, $\mathrm{PA}_t^{(m,\delta)}(b')$ is a *collapsed branching process* of size t, as introduced in [70]. The reason why (b) and (b') are different lies in the normalization constants in (4.3.1) and

(4.3.2), since we start with one edge less in $\mathrm{PA}_2^{(1,\delta)}(b')$. The normalization constants are now

$$c'_{t,j} = \begin{cases} 2(t-2)+(t-1)\delta & \text{if } m=1 \text{ and } t \geq 3, \\ (t-1)(2m+\delta)-2+(j-1)(2+\delta/m) & \text{if } m \geq 2,\, t \geq 3 \text{ and } j \in [m]. \end{cases} \tag{4.3.3}$$

Notice that the difference between the constants (4.3.3) and the normalization constants in (4.3.1) and (4.3.2) is -2, given by the fact that we start with one edge less.

Model (c). Model (c) is taken from [85, Model (c)]. We again let the graph at time 2 consist of two vertices with two edges between them. We fix $\alpha \in [0,1]$. Then, we first draw a Bernoulli random variable I_{t+1} with success probability $1-\alpha$. The random variables $(I_t)_{t\geq 1}$ are independent. When $I_{t+1}=0$, then we attach the $(t+1)$st edge to a uniform vertex in $[t]$. When $I_{t+1}=1$, then we attach the $(t+1)$st edge to vertex $i \in [t]$ with probability $D_i(t)/(2t)$. No self-loops are allowed when $m=1$. When $m \geq 2$ the graph is again defined by collapsing. As a consequence, only the first edge of every vertex is not allowed to be a self loop. We have

$$\begin{aligned} \mathbb{P}\left(t \to i \mid \mathrm{PA}_{t-1}^{(c)}(1,\delta)\right) &= (1-\alpha)\frac{1}{t-1} + \alpha\frac{D_i(t-1)}{2(t-1)} \\ &= \frac{D_i(t-1)+2(1-\alpha)/\alpha}{2(t-1)/\alpha} = \frac{D_i(t-1)+\hat{\delta}}{2(t-1)/\alpha}, r \end{aligned} \tag{4.3.4}$$

where $\hat{\delta} = 2(1-\alpha)/\alpha > 0$.

Model (d). Model (d) is taken from [21, Model 3]. In the tree case, it coincides with model (c). For $m \geq 2$, edges are again assigned interpolating uniformly and degree-biased choice with intermediate updating of degrees, but no self loops are allowed, so attachment probabilities are similar to (4.3.4). In this model, the graph case *is not defined by the tree case through collapsing*.

Model (e). Model (e) is a modification of model (d), made to extend the "no-self-loops" regime to negative values of δ. This model is called the *sequential model* in [21]. Definition 4.3.1 in the form below is given in [86], where the graph is defined for any $\delta > -m$. We give a proper definition of this model, since we use it as the starting point of our analysis:

Definition 4.3.1 (Model (e))**.** *Fix $m \geq 1$, $\delta > -m$. Then $(\mathrm{PA}_t^{(e)}(m,\delta))_{t\geq 2}$ is a sequence of random graphs defined as follows:*

- ▷ *for $t=2$, $\mathrm{PA}_2^{(e)}(m,\delta)$ consists of two vertices with m edges between them;*
- ▷ *for $t \geq 3$, $\mathrm{PA}_t^{(e)}(m,\delta)$ is constructed recursively as follows: conditioning on the graph at time $t-1$, we add a vertex t to the graph, with m new edges. Edges start from vertex t*

and, for $j = 1, \dots, m$, they are attached sequentially to vertices $E_{t,1}, \dots, E_{t,m}$ chosen with the following probabilities: for $j \in [m]$,

$$\mathbb{P}\left(E_{t,j} = i \mid \mathrm{PA}^{(e)}_{t-1,j-1}(m,\delta)\right) = \frac{D_i(t-1,j-1)+\delta}{2m(t-2)+(j-1)+(t-1)\delta}. \tag{4.3.5}$$

In (4.3.5), *$D_i(t-1)$ denotes the degree of i in $\mathrm{PA}^{(e)}_{t-1}(m,\delta)$, while $D_i(t-1,j-1)$ denotes the degree of vertex i after the first $j-1$ edges of vertex t have been attached. Here we assume that $\mathrm{PA}^{(e)}_{t-1,-1} = \mathrm{PA}^{(e)}_{t-1}$.*

To keep notation light, we write PA_t instead of $\mathrm{PA}^{(e)}_t(m,\delta)$. The normalization constants in (4.3.5) depends on the fact that, when $t-1$ vertices are present and $j-1$ edges have been attached, the total degree of the first $t-1$ vertices in $\mathrm{PA}_{t-1,j-1}$ is $2m(t-2)+(j-1)$. The term $(t-1)\delta$ comes from the fact that we can attach to $t-1$ vertices. We do not allow self-loops, but we allow multiple edges.

Model (f). Model (f) is taken from [21, Independent model]. In this model, the graph starts as two vertices with m edges between each other. Fix $\alpha \in [0,1]$. Then, at every step $t \geq 3$, a new vertex is added to the graph with m edges. Each edge $j \in [m]$ is attached to a vertex $i \in [t-1]$ as follows: with probability α, i is chosen uniformly in $[t-1]$, while with probability $1-\alpha$ it is chosen according to

$$\mathbb{P}\left(E_{t,j} = i \mid \mathrm{PA}^{(f)}_{t-1}(m,\delta)\right) = \frac{D_i(t-1)}{Z(t-1)},$$

where $Z(t-1)$ is the normalization constant. In particular, the m edges are attached *(conditionally) independently* of each other, *without intermediate degree updates*, in contrast for instance with Model (a) and Model (e). In this case, no self-loops are allowed.

Model (g). Model (g) is taken from [21, Conditional model], and it is defined as Model (f), with the difference that, for every $t \in \mathbb{N}$, the m edges of vertex t are conditioned to be attached to *distinct vertices*, thus generating a simple graph, so no self-loops and multiple edges are allowed. Notice that to be consistent, we start with a complete graph on $m+1$ vertices (so that every vertex has degree m).

Model (h). Model (h) is investigated in a series of papers by Dereich and Mőrters [54, 55, 56], where the authors call it the *PAM with conditionally independent edges*. In this model, we consider PA functions $f : \mathbb{N} \to (0,\infty)$ that satisfy that $f(k+1)-f(k) < 1$ for every $k \geq 0$, with the additional assumption that $f(0) \leq 1$. Then, every edge comes into the graph and it is connected to vertex $i \in [t-1]$ for every $i \in [t-1]$, *conditionally independently* of the other vertices, with probability

$$\mathbb{P}\left(E_{t,j} = i \mid \mathrm{PA}^{(h)}_{t-1}(m,\delta)\right) = \frac{f(D_i(t-1))}{t-1}. \tag{4.3.6}$$

Thus self-loops and multiple edges are not allowed in this model. In particular, this model is sometimes called *Bernoulli* PAM since the number of original edges incident to a vertex, instead being a fixed number m, is the sum of (conditionally) independent Bernoulli random variables. In the present paper, we consider $f(k) = ck + \delta$, for $c \in (0, 1)$ and $\delta < 1$.

4.3.1. The Pólya point tree

The *Pólya point tree* (PPT) is an *infinite multi-type rooted random tree* constructed as follows. Let $m \geq 1$, $\delta > -m$ be given and define

- $\chi = (m + \delta)/(2m + \delta)$, $\phi = (1 - \chi)/\chi$;
- Γ_{in} denote a Gamma distribution with parameters $m + \delta$ and 1;
- Γ'_{in} denote a Gamma distribution with parameters $m + \delta + 1$ and 1.

Vertices in the graph have five characteristics:

- a *label* i in the Ulam-Harris set $\mathcal{N}$ (recall Definition 2.1.3);
- a *position* $x \in [0, 1]$;
- a positive number γ called *strength*;
- a type: *younger* denoted by Y and *older* denoted by O, with the exception of the root $\varnothing$ that does not have a type;
- an integer m^{out} that is either m if the vertex is of type O or $m - 1$ if the vertex is of type Y;

Then, the Pólya point graph is constructed as follows:

(1) Assign to $\varnothing$ a position $x_\varnothing = U^\chi$, where U is a uniform random variable on $[0, 1]$;
(2) Assign to $\varnothing$ an out-degree $m_\varnothing^{out} = m$ and strength $\gamma_\varnothing \sim \Gamma_{in}$;
(3) Sample $U_1, \ldots, U_m$, independent of the rest, uniform random variables on $[0, x_\varnothing]$; to vertices $\varnothing 1, \ldots, \varnothing m$ assign the positions $U_1, \ldots, U_m$ and type O;
(4) To vertices $\varnothing(m+1), \ldots, \varnothing(m+d_\varnothing^{(\mathrm{in})})$ assign positions $x_{\varnothing(m+1)}, \ldots, x_{\varnothing(m+d_\varnothing^{(\mathrm{in})})}$, that are the points given by the Poisson point process on $[x_\varnothing, 1]$ defined by the intensity
$$\rho_\varnothing(x) = \gamma_\varnothing \frac{\phi x^{\phi-1}}{x_\varnothing^\phi},$$
where $d_\varnothing^{in}$ is the total number of points of of this process;
(5) Label $\varnothing$ as explored and vertices $\varnothing 1, \ldots, \varnothing(m + d_\varnothing^{(\mathrm{in})})$ as unexplored.

Then, recursively over the elements in the set of unexplored vertices:

(1) Let i denote the current unexplored vertex;
(2) Assign to i two values:

$$m_i^{(\mathrm{out})} \sim \begin{cases} m-1 & \text{if } x_i \text{ is of type Y,} \\ m & \text{if } x_i \text{ is of type O,} \end{cases} \tag{4.3.7}$$

$$\gamma_i \sim \begin{cases} \Gamma_{in} & \text{if} \quad x_i \quad \text{is of type Y,} \\ \Gamma'_{in} & \text{if} \quad x_i \quad \text{is of type O;} \end{cases} \tag{4.3.8}$$

(3) Sample $m_i^{(\text{out})}$ points $x_{i1}, \dots, x_{im_i^{(\text{out})}}$ independently from all the previous steps and from each other, uniformly on $[0, x_i]$. Assign to $i1, \dots, im_i^{(\text{out})}$ type O and set them unexplored;

(4) Let $x_{i(m_i^{(\text{out})}+1)}, \dots, x_{i(m_i^{(\text{out})}+d_i^{(\text{in})})}$ be the random $d_i^{(\text{in})}$ points given by an independent Poisson process on $[x_i, 1]$ with intensity

$$\rho_i(x) = \gamma_i \frac{\phi x^{\phi-1}}{x_i^{\phi}}. \tag{4.3.9}$$

Assign to $i(m_i^{(\text{out})}+1), \dots, i(m_i^{(\text{out})}+d_i^{(\text{in})})$ type Y and set them unexplored;

(5) Draw an edge between i and each one of the vertices $i1, \dots, i(m_i^{(\text{out})}+d_i^{(\text{in})})$;

(6) Set i as explored.

This inhomogeneous random tree is the LW limit of the sequential PAM in Definition 4.3.1, as stated in the following theorem:

Theorem 4.3.2 (LWC - Sequential model)**.** *Consider a* PA_t *with parameters* m *and* δ *as in Definition 4.3.1. Then,* $(\text{PA}_t)_{t\in\mathbb{N}}$ converges locally weakly in probability *to the Pólya point tree with the same parameters.*

The original proof of Theorem 4.3.2 is given for $\delta \geq 0$ in [21, Theorem 3], while [86, Chapter 4] contains a sketch of the proof for all $\delta > -m$, by indicating how the Pólya urn description (see Section 4.4) can be extended to all values of δ. In Section 4.7 we explain the structure of the proof, highlighting the main steps.

4.4. Affine urn schemes

In this section, we present the theory of Pólya urn schemes with affine weight functions. We follow the structure given in [86, Section 4.1].

4.4.1. Two-urn schemes

An urn scheme consists of an urn, with blue balls and red balls. At every time step, we draw a ball from the urn and we replace it with two balls of the same color of the one drawn. We start with $B_0 = b_0$ blue balls and $R_0 = r_0$ red balls.

The so-called *weighted urn process* is defined as follows: consider two weight functions:

$$W_b(k) = a_b + k, \quad \text{and} \quad W_r(k) = a_r + k. \tag{4.4.1}$$

Conditionally on the number of blue balls B_n and red balls R_n, at time $n+1$ the probability of drawing a blue ball (and then adding an additional one) is equal to

$$\frac{W_b(B_n)}{W_b(B_n) + W_r(R_n)}.$$

Then, the general result on the Pólya urn scheme $((B_n, R_n))_{n\in\mathbb{N}}$ is the following:

Theorem 4.4.1 (Affine urn scheme)**.** *Let $((B_n, R_n))_{n\in\mathbb{N}}$ be a Pólya urn scheme with weight functions as in* (4.4.1)*. Then, as $n \to \infty$,*

$$\frac{B_n}{B_n + R_n} \xrightarrow{\mathbb{P}-a.s.} \psi, \tag{4.4.2}$$

where ψ has a Beta distribution with parameters $B_0 + a_b$ and $R_0 + a_r$, and, for every $n \in \mathbb{N}$,

$$\mathbb{P}(B_n = B_0 + k) = \mathbb{E}\left[\mathbb{P}\left(\mathrm{Bin}(n, \psi) = k \middle| \psi\right)\right]. \tag{4.4.3}$$

Theorem 4.4.1 can be found in [86, Theorem 4.2], and the proof is based on De Finetti's theorem on exchangeable random variables. Sometimes we call the random variable ψ the *intensity* or *strength* of the blue balls in the urn. Conditionally on ψ, the number of blue balls in the urn at step n is then a Binomial distribution with n attempts and probability of success ψ. Notice that the parameters of the distribution ψ are completely determined by the initial weights of the process, i.e., $B_0 + a_b$ and $R_0 + a_r$.

We can equivalently see the urn process as in Theorem 4.4.1 as two different urns, one containing only blue balls and the other only red balls, and we choose an urn proportionally to the number of balls in the urns. In this case, the result is the same, but we can say that ψ is the strength of the blue balls urn and $1 - \psi$ is the strength of the red balls urn.

4.4.2. More than one urn

The previous one urn experiment can be generalized to a number $t \geq 3$ of urns. With two urns, we already mentioned that a one-urn experiment can be seen as a two-urns experiment, by assuming that every urn contains balls of the same color, and we pick the urns according to the total weight of the balls inside each urn.

For $t \geq 3$, we assume to have a single urn with t colors, or equivalently, t urns where every urn contains balls of a different color. Denote the number of balls of each color at time n by $(C_i(n))_{i\in[t]}^{n\in\mathbb{N}}$, where $i \in [t]$ denotes the color, and $n \in \mathbb{N}$ denotes the number of extractions already made. Similarly to the single-urn case, we pick a ball of a color (or equivalently, an urn) with probability proportionally to $W_i(n) = C_i(n)+a_i$, and we put back the picked ball with another one of the same color. We assume that $i \in [t]$ starts with k_i balls in the urn.

We start by observing that the number of balls of color 1 and the number of balls of the other colors can be seen as a two-colors urn experiment, thus by Theorem 4.4.1 we have

$$\frac{C_1(n)}{n} \xrightarrow{\mathbb{P}-a.s.} \psi_1, \tag{4.4.4}$$

where ψ_1 is a Beta random variable with parameters $a_1 + k_1$ and $\sum_{j=2}^{t} a_j + k_j$. In particular, ψ_1 gives the asymptotic fraction of balls with color 1. This means that the fraction $1 - \psi_1$ is made by balls of the remaining $t - 1$ colors.

When one of the colors $2, \dots, t$ is drawn, we have that color 2 and colors $3, \dots, t$ can again be seen as a two-colors urn experiment, so

$$\frac{C_2(n)}{n} \xrightarrow{\mathbb{P}-a.s.} \psi_2(1-\psi_1), \tag{4.4.5}$$

where ψ_2 is a Beta random variable independent of ψ_1, with parameters $a_2 + k_2$ and $\sum_{j=3}^{t} a_j + k_j$. We can see the right-hand side of (4.4.5) as follows: conditionally on picking one of the colors $2, \dots, t$, the restriction of the urn experiment to this color evolve as a two-urn experiment. Thus, the fraction of color 2 balls converges to ψ_2, but by (4.4.4) we know that ψ_1 fraction of the total is made by color 1 balls, thus the fraction of balls made by color 2 balls converges to $\psi_2(1-\psi_1)$ of the total. In general, the fraction of balls of color $i \in [t]$ satisfies

$$\frac{C_i(n)}{n} \xrightarrow{\mathbb{P}-a.s.} \psi_i \prod_{j=1}^{i-1}(1-\psi_j), \tag{4.4.6}$$

where $(\psi_j)_{j\in[i]}$ are independent Beta random variables, and ψ_j has parameters $a_j + k_j$ and $\sum_{l=j+1}^{t} a_l + k_l$.

4.4.3. Model (e) as a Pólya urn graph

The sequential model in Definition 4.3.1 can be interpreted as an experiment with t urns, where the number of balls in each urn represents the degree of a vertex in the graph. First, we introduce a random graph model:

Definition 4.4.2 (Pólya urn graph)**.** *Fix $m \geq 1$ and $\delta > -m$. Let $t \in \mathbb{N}$ be the size of the graph. Let $\psi_1 = 1$, and consider $\psi_2, \dots, \psi_t$ independent random variables, where*

$$\psi_k \stackrel{d}{=} \mathrm{Beta}\left(m+\delta, m(2k-3) + (k-1)\delta\right). \tag{4.4.7}$$

Define

$$\varphi_j^{(t)} = \psi_j \prod_{i=j+1}^{t}(1-\psi_i), \quad S_k^{(t)} = \sum_{j=1}^{k} \varphi_j, \quad I_k^{(t)} = [S_{k-1}^{(t)}, S_k^{(t)}). \tag{4.4.8}$$

Conditionally on $\psi_1, \dots, \psi_t$, let $(U_{k,j})_{k=2,\dots,t}^{j=1,\dots,m}$ be independent random variables, with $U_{k,j}^{(t)}$ uniformly distributed on $[0, S_{k-1}^{(t)}]$. Then, the corresponding Pólya urn graph PU *is the graph of size t where, for $u < v$, the number of edges between u and v is equal to the number of variables $U_{v,j}^{(t)}$ in I_u, for $j = 1, \dots, m$ (multiple edges are allowed).*

The Beta distributions in Definition 4.4.2 come from the Pólya urn interpretation of the sequential model, using urns with affine weight functions, as introduced above. Notice the similarity between (4.4.6) and the expressions of $(\varphi_j^{(t)})_{j\in[t]}$. In particular, since $\sum_{j=1}^{t} \varphi_j^{(t)} = 1$, $\varphi_j^{(t)}$ represents the asymptotic fraction of edges received by vertex j among the first t vertices, when the size of the graph diverges.

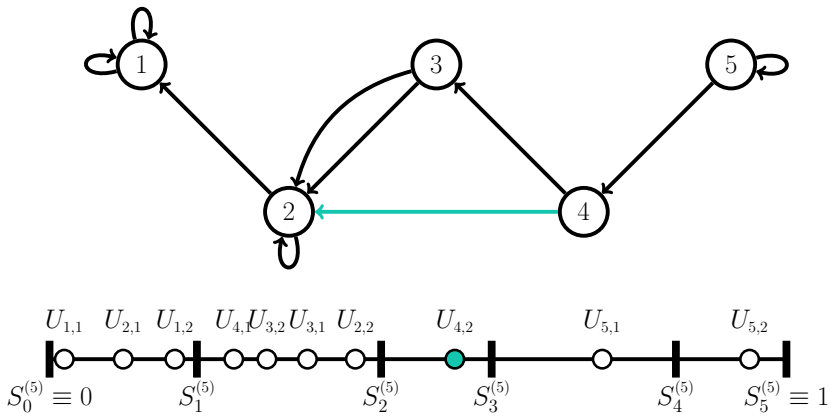


FIGURE 4.1: Example of Pólya urn graph as in Definition 4.4.2 of size 5 where $m = 2$. At every uniform random variable corresponds an edge in the graph, as highlighted in green in the figure. The relative age i/t of vertex $i \in [t]$ is represented as the position $S_i^{(t)}$, or, equivalently, as the position of the interval $I_i^{(t)}$ in the interval $[0, 1]$. The more the point $S_i^{(t)}$ is close to zero, the older the vertex i is among the t vertices of the graph.

The Pólya urn graph is an equivalent formulation of the sequential model, as proven in the following theorem:

Theorem 4.4.3 (Equivalence of formulations)**.** *Fix $m \geq 2$ and $\delta > -m$. For every $t \in \mathbb{N}$, $\mathrm{PA}_t^{(e)}(m, \delta)$ and $\mathrm{PU}_t(m, \delta)$ have the same distribution.*

The proof can be found in [21, Theorem 2.1] for $\delta \geq 0$, while [86, Theorem 4.8] gives the sketch of the proof for general δ. We give the formal proof for general values of δ in Section 4.5. The advantage of the formulation in Definition 4.3.1 is that there are no self-loops, thus the corresponding Pólya urn experiment is easier to define. We explain how self-loops affect the definition of the Pólya urn graph in Section 4.6.1.

4.5. Equivalence of Pólya urn graph and model (e): proof of Theorem 4.4.3

In this section, we give a rigorous proof of Theorem 4.4.3, following the sketch given in [86, Theorem 4.8]. The proof is divided in two steps: First, we prove a preliminary lemma on consistency of Pólya urn experiments in model (e), then we give the general proof of the result.

We recall the notation introduced in Chapter 2 that we used to denote individuals during collapsing. For $m \geq 1$, we write $(k, j) = m(k-1) + j$, for $k \in \mathbb{N}$ and $j \in [m]$. We now can start state a preliminary lemma:

Lemma 4.5.1 (Consistency of Pólya urns in model (e))**.** *Fix $m \geq 1$ and $\delta > -m$, and let $(\mathrm{PA}_t^{(e)}(m,\delta))_{t\in\mathbb{N}}$ be the corresponding PAM model (e). For $k \geq 2$, denote by $D_k(t)$ and $D_{<k}(t)$ the degree of vertex k and the total degree of $[k-1]$ respectively. Define $\theta_0 = km$ and, for $n > m+1$:*

$$\theta_n := \{(h,j) > (k,m) \colon (h,j) \text{ is the } n\text{th edge attached to } [k]\}, \tag{4.5.1}$$

i.e., $(\theta_n)_{n\geq mk}$ is the sequence of the indices of the edges in model (e) that are attached to the first $[k]$ vertices. Then, $(D_k(\theta_n), D_{<k}(\theta_n))_{n\geq km}$ is a two-urn Pólya urn experiment in the sense of Section 4.4. As a consequence,

$$\frac{D_k(\theta_n)}{D_k(\theta_n) + D_{<k}(\theta_n)} \xrightarrow{a.s.} \psi_k,$$

where ψ_k is a Beta random variable with parameter $m+\delta$ and $m(2k-3)+(k-1)\delta$.

Proof. Fix $k \geq 2$. We consider the evolution of the degree of vertex k, $D_k(t)$, and the degree of the union of the first $k-1$ vertices $D_{<k}(t)$. We start at time $i = k$, that coincides with $\theta_0 = (k,m)$ (recall (4.5.1)). At this time, we have just added the m original edges of vertex k to the graph, so $D_k(\theta_0) = m$ and $D_{<k}(\theta_0) = m(2k-3)$.

For every edge (t,j) such that $\theta_0 < (t,j) < \theta_1$, the degree of the first k does not change by definition of the sequence $(\theta_n)_{n\in\mathbb{N}}$. At time $(t,j) = \theta_1$, the edge (t,j) is attached to one of the first k vertices. In particular,

$$\mathbb{P}(t \xrightarrow{j} [k] \mid \mathrm{PA}_{t,j-1}^{(e)}(m,\delta)) = \frac{D_k(t,j-1) + D_{<k}(t,j-1) + k\delta}{m(2t-2) + t\delta + j - 1}.$$

As a consequence, conditionally on $\{t \xrightarrow{j} [k]\}$ and $\mathrm{PA}_{t,j-1}^{(e)}(m,\delta)$,

$$\begin{aligned}\mathbb{P}(t \xrightarrow{j} k \mid \mathrm{PA}_{t,j-1}^{(e)}(m,\delta),\ t \xrightarrow{j} [k]) &= \frac{\mathbb{P}(t \xrightarrow{j} k,\ t \xrightarrow{j} [k] \mid \mathrm{PA}_{t,j-1}^{(e)}(m,\delta))}{\mathbb{P}(t \xrightarrow{j} [k] \mid \mathrm{PA}_{t,j-1}^{(e)}(m,\delta))} \\ &= \frac{D_k(t,j-1) + \delta}{D_k(t,j-1) + D_{<k}(t,j-1) + k\delta},\end{aligned} \tag{4.5.2}$$

and similarly

$$\mathbb{P}(t \xrightarrow{j} [k-1] \mid \mathrm{PA}_{t,j-1}^{(e)}(m,\delta),\ t \xrightarrow{j} [k]) = \frac{D_{<k}(t,j-1)}{D_k(t,j-1) + D_{<k}(t,j-1)}. \tag{4.5.3}$$

In particular, we obtain (4.5.2) and (4.5.3) since the normalization constants in the attachment probabilities are the same for all the events $\{t \xrightarrow{j} [k-1]\}$, $\{t \xrightarrow{j} k\}$ and $\{t \xrightarrow{j} [k]\}$.

Notice that the right-hand side of (4.5.2) and (4.5.3) are depending only on $D_k(\theta_0)$ and $D_{<k}(\theta_0)$. In the same way we can prove that (4.5.2) and (4.5.3) hold for θ_n, con-

ditionally on $D_k(\theta_{n-1})$ and $D_{<k}(\theta_{n-1})$. In particular,

$$\begin{aligned}\mathbb{P}(D_k(\theta_n) = D_k(\theta_{n-1}) + 1 \mid D_k(\theta_{n-1}),\ D_{<k}(\theta_{n-1})) \\ = \frac{D_{<k}(\theta_{n-1}) + \delta}{D_k(t, j-1) + D_{<k}(t, j-1) + k\delta}.\end{aligned} \tag{4.5.4}$$

Equation (4.5.4) represents the evolution of a two-urns Pólya urn experiment, where the two urns have initial weight $m + \delta$ and $m(2k-3) + (k-1)\delta$. This completes the proof. □

We use Lemma 4.5.1 to prove Theorem 4.4.3:

Proof of Theorem 4.4.3. Fix $t \geq 2$. We want to prove that $\mathrm{PA}_t^{(e)}(m, \delta)$ and $\mathrm{PU}_t(m, \delta)$ have the same distribution. Let $(X_n)_{n\in\mathbb{N}}$ be the vertex receiving the nth edge of the graph. In particular, we have $m(t-1)$ edges in the graph of size t (recall that we start with G_0 being a graph with two vertices and m edges between them).

To keep notation light, we denote $\mathbb{P}_t(\cdot) = \mathbb{P}(\cdot \mid \psi_1, \dots, \psi_t)$. For $n = 1, \dots, m$, $X_1 \equiv \cdots \equiv X_m = 1$, since vertex 1 deterministically receives all the m edges of vertex 2. For $n = m+1, \dots, 2m$, $X_{m+1}, \dots, X_{2m}$ can either be equal to 1 or 2. By Lemma 4.5.1, the evolution of the degrees of vertex 1 and 2 can be described as a two-urn Pólya urn model, with intensities $1 - \psi_2$ and ψ_2. In this case, ψ_2 is a Beta random variable with parameters $m + \delta$ and $m + \delta$.

When $2m + 1 \leq n \leq 3m$, X_n can take three values. In this case we have

$$\mathbb{P}_t(X_n \in [2]) = 1 - \psi_3, \qquad \text{and} \quad \mathbb{P}_t(X_n = 3) = \psi_3, \tag{4.5.5}$$

where $1 - \psi_3$ and ψ_3 are the intensities of the two-urns experiment corresponding to the evolution of the degree of vertex 3 and the union of vertices $\{1, 2\}$. In particular, ψ_3 is a Beta random variable with parameters $m + \delta$ and $3m + 2\delta$. Conditionally on $X_n < 3$, by Lemma 4.5.1 the degrees of vertex 1 and 2 again evolve as a two-urns model. This implies that

$$\mathbb{P}_t(X_n = 1 \mid X_n \in [2]) = 1 - \psi_2 \qquad \text{and} \quad \mathbb{P}_t(X_t = 2 \mid X_t \in [2]) = \psi_2. \tag{4.5.6}$$

Using (4.5.6) in (4.5.5), we obtain

$$\begin{aligned}\mathbb{P}_t(X_n = 1) &= (1 - \psi_3)(1 - \psi_2), \\ \mathbb{P}_t(X_n = 2) &= \psi_2(1 - \psi_3), \\ \mathbb{P}_t(X_n = 3) &= \psi_3,\end{aligned} \tag{4.5.7}$$

as required. Recursively, we can extend (4.5.7) to all $k \in [t]$. In fact, we can look at the indices $m(k-1)+1, \dots, mk$, i.e., the indices such that $X_{m(k-1)+1}, \dots, X_{mk}$ are the vertices selected by the m edges of vertex $k+1$ (remember that vertex 1 has no edges,

thus the total number of edges in the graph will be $m(t-1)$). Then, by Lemma 4.5.1,

$$\mathbb{P}_t(X_n \in [k-1]) = 1 - \psi_k, \qquad \text{and} \quad \mathbb{P}_t(X_n = k) = \psi_k. \tag{4.5.8}$$

Then, for every $h \in [k-1]$, similarly to (4.5.6),

$$\begin{aligned}
&\mathbb{P}_t(X_n \in [h] \mid X_n \in [k-1]) \\
&\qquad = \prod_{\ell=k-1}^{h} \mathbb{P}_t(X_n \in [h] \mid X_n \in [\ell]) \mathbb{P}_t(X_n \in [\ell] \mid X_n \in [k-1]) \\
&\qquad = \psi_k \prod_{\ell=k-1}^{h} (1 - \psi_\ell).
\end{aligned}$$

This completes the proof. □

Remark 4.5.2 (Relation between models (d) and (e))**.** Theorem 4.4.3 was originally stated in [21, Theorem 2] for model (d). The main difference between model (e) and model (d), lies in the interpretation of the attachment mechanisms, and the corresponding Pólya urn experiment that represents the graph. Model (d) is defined by a mixture of a uniform PA function (vertices are chosen uniformly among existing ones) and an affine PA function where $\delta = 0$. This mechanism allows is equivalent to consider an affine PA function with $\hat{\delta} = 2(1-\alpha)/\alpha > 0$ (recall (4.3.4) and the description given in Section 4.3). Model (e) instead is defined using urns with weight functions as in 4.4. In this setting, we can consider affine weights functions for any value $\delta > -m$. Notice that the initial weight of an urn is then $m + \delta > 0$. This allows us to directly extend the results in [21] stated for model (d) to model (e).

4.6. Generalization of Pólya urn graphs

4.6.1. Different urn evolution

In this section, we show the universality of the Pólya urn construction of PAMs. In particular, we show that all the PAMs defined in Section 4.3 can *almost* be defined as Pólya urn graphs. Here we use the word "almost" to indicate that all the small differences in the definitions of the models in Section 4.3 make the Pólya urn graph construction not immediate as in Definition 4.4.2, but a close definition is possible.

There are two elements that generate differences between the models in Section 4.3: the presence of *self-loops*, the *initial state of the models*, and whether *edges are updated intermediately*. As it turns out, the sequential model in Definition 4.3.1, with no self-loops and as a starting graph a single vertex with no edges turns out to be the model for which the corresponding Pólya urn graph is the easiest to define.

From Definition 4.4.2, it is possible to see that the parameters of the sequence $(\psi_k)_{k\in[t]}$ depend on *the initial number of balls in the urns* when the graph has k vertices.

More precisely, the random variable ψ_k has parameters a_k and b_k given by

$$a_k = m + \delta, \qquad b_k = m(2k-3) + (k-1)\delta. \tag{4.6.1}$$

Notice that the attachment probabilities as in (4.3.5) for $t = k+1$, and $j = 1$, are exactly given by $a_k/(a_k+b_k)$ for $i = k$ and $b_k/(a_k+b_k)$ for the *union* of the vertices $[k-1]$.

In general, all the versions of PAMs in Section 4.3 can be defined using sequences of Beta random variables $(\psi_k)_{k\in\mathbb{N}}$, where the parameters (4.6.1) satisfy

$$a_k = m + \delta, \qquad b_k = (2m+\delta)k + c, \tag{4.6.2}$$

for some constant c that does not depend on k. In particular, c depends on the initial state of the graph and on the presence of self-loops.

In Definition 4.4.2, for $k \in \mathbb{N}$ and $j \in [m]$, the random variable $U_{k,j}^{(t)}$ is uniform on $[0, S_{k-1}^{(t)}]$. In particular, $U_{k,j} \in I_v^{(t)}$, for some $v < k$, thus creating an edge between vertex k and vertex v. Notice that self-loops are not allowed. We can define the urn graphs in a different way, assuming that $U_{k,j}$ is uniform on $[0, S_k^{(t)}]$, thus allowing for self-loops. This complicates the Pólya urn scheme, since now the edges of vertex k can increase the degree of k itself. More precisely, the intensity ψ_k has parameters a_k and b_k as in (4.6.1), where b_k is the total weight of the first $k-1$ vertices when we start attaching the edges of the vertex $k+1$.

For this, it is sufficient to distinguish between the initial weight of an urn *before* or *after* starting attaching edges. Consider $m = 1$, $k \in \mathbb{N}$, and assume that we consider model (a) in Section 4.3. Then, the intensity ψ_k obeys a Beta distribution with parameters

$$a_k = 1 + \delta, \qquad b_k = (2+\delta)(k-1),$$

where now a_k and b_k represent the initial weight of vertex k and the first $k-1$ vertices respectively *before attaching the edge of vertex* k. In this case, the random variable $U_{k,1}$ is uniform on $[0, S_k^{(t)}]$. More precisely, model (a) allows for self-loops, thus U_k is uniform on $[0, S_k^{(t)}]$. In particular, the parameters of ψ_k are $1+\delta$ and $(2+\delta)(k-1)$, that are exactly the tonal weight of k and $[k-1]$ *before* attaching the edge of k. In model (e) this does not happen, and the edge of k is surely attached to $[k-1]$. As a consequence, the parameters of ψ_k are $1+\delta$ and $(2k-3)+(k-1)\delta$, that are the total weight of k and $[k-1]$ *after* attaching the edge of vertex k.

So far, we have considered $m = 1$. When $m \geq 2$, this setting can be further complicated by the differences between edges. For example, model (b) in Section 4.3 does allow for self-loops, with the exception of the first edge of every vertex (thus forcing the graph to be connected). In this case, we have a mixed situation, since, for every vertex, the first edge cannot be a self-loop (similar to model (e)), but the others are allowed to be self-loops (similarly to model (a)).

The consequence of these problems is that not every PAM defined in Section 4.3 can be directly defined as a Pólya urn graph. Some models can be defined as a urn graph for every $m \geq 1$, while others can be defined as a Pólya urn graph only for

$m = 1$, followed by a collapsing procedure.

4.6.2. Unit graphs

In this section, we define a general framework that allows to describe all versions of PAMs, with the exception of model (h), with a similar formulation to the urn graph in Section 4.4.2. First of all, we define a new class of random graphs:

Definition 4.6.1 (Unit graph)**.** *Fix $m \geq 1$, and define the following:*

- ▷ *Let $(U_{k,j})_{k\in\mathbb{N}}^{j\in[m]}$ be a sequence of i.i.d. random variables, where $U_{1,1}$ is a uniform random variable in $[0,1]$;*
- ▷ *Let $(\psi_{k,j})_{k\in\mathbb{N}}^{j\in[m]}$ be a sequence of random variables with support in $[0,1]$, such that $\mathbb{P}(\psi_{k,j} = 1) = 0$ for all $k \geq 1$ and $j \in [m]$, but $\psi_{1,1} \equiv 1$;*
- ▷ *Let $t \in \mathbb{N}$ be the size of the graph.*

Then, for every $k \in [t]$, define $\mathcal{S}_0^{(t)} = 0$, $\mathcal{S}_{t,m}^{(t)} = 1$, and, for $j \in [m]$,

$$\mathcal{S}_{k,j}^{(t)} = \psi_{k,j} \prod_{l=j+1}^{m} (1 - \psi_{k,l}) \prod_{h=k+1}^{t} \prod_{l=1}^{m} (1 - \psi_{h,l}).$$

Denote $S_k^{(t)} = \mathcal{S}_{k,m}^{(t)}$. Define the intervals

$$\mathcal{I}_k^{(t)} = [S_{k-1}^{(t)}, S_k^{(t)}).$$

Then, we call UG_t the unit graph *of size t defined as follows: The jth edge of vertex k is attached to vertex $u \in [k]$ if and only if*

$$U_{k,j}\mathcal{S}_{k,j-1}^{(t)} \in \mathcal{I}_u^{(t)}, \tag{4.6.3}$$

and we label the graph NSL (no self-loops). The unit graph UG_t is labeled SL (self-loops) if the condition (4.6.3) *is replaced by*

$$U_{k,j}\mathcal{S}_{k,j}^{(t)} \in \mathcal{I}_u^{(t)}. \tag{4.6.4}$$

When $m = 1$, we call this model unit tree*, and we denote it by UT_t.*

The labels SL and NSL are purely technical. We are aware that this notation can be confusing. In fact, in a NSL graph an edge can still generate a self-loop. The labels SL and NSL coincide with the possibility that an edge can create a self-loop only when $m = 1$. When $m \geq 2$, the difference between (4.6.3) and (4.6.4) is that in the latter, the random variable $U_{v,j}$ is rescaled on $\mathcal{S}_{k,j}^{(t)}$, while in (4.6.3) it is rescaled on $\mathcal{S}_{k,j-1}^{(t)}$. The meaning of this notation will be clear when we adapt it to PAMs.

Notice that in a unit graph, the starting graph is a single vertex with no self-loops. Intuitively, we can start from any fixed graph G_0 on $t_0 \leq t$ vertices, and use (4.6.3) and (4.6.4) to define the edges of the remaining $t - t_0$ vertices. In this case, the random

variables associated to the first t_0 vertices can be arbitrarily determined, or kept as random variables. In general, in PAMs the random values associated to the starting graph are assumed to be deterministic. The initial graph does not influence the LWC result, since it involves a finite number of vertices t_0 that does not grow with t.

Observe that the graph satisfies a generalized recursive property: it is possible to define UG_t such that $\mathrm{UG}_{t-1} \subseteq \mathrm{UG}_t$ as a subgraph. In fact, for any $k \in \mathbb{N}$ and $j \in [m]$,

$$\mathcal{S}_{k,j}^{(t)} = \begin{cases} \mathcal{S}_{k,j}^{(t-1)} \displaystyle\prod_{j=1}^{m} (1-\psi_{t,j}) & \text{for } k < t, \\ \psi_{t,j} \displaystyle\prod_{l=j+1}^{m} (1-\psi_{t,l}) & \text{for } k = t \text{ and } j < m, \\ \psi_{t,m} = 1 & \text{for } k = t \text{ and } j = m. \end{cases} \tag{4.6.5}$$

In simple words, (4.6.5) states that we can take the values $(\mathcal{S}_{k,j}^{(t-1)})_{k\in[t-1]}^{j\in[m]}$ and rescale them to obtain the values $(\mathcal{S}_{k,j}^{(t)})_{k\in[t]}^{j\in[m]}$. The first point is always 0, and the last one is always 1, so that the graph is always related to the complete interval $[0,1]$.

Notice that in this way $\mathrm{UG}_{t-1} \subseteq \mathrm{UG}_t$, or, in other words, we can recursively construct UG_t from UG_{t-1}, where both have the distribution of a unit graph of size t and $t-1$ respectively. In particular, *the law of* UG_{t-1} *is independent of the vertex* t *and its attributes*. We call this fact the *recursive property* of unit graphs. In particular, the interval $\mathcal{I}_k^{(t)}$ that determines whether an edge is attached to k is *determined by its endpoints*, namely $S_{k-1}^{(t)}$ and $S_k^{(t)}$. The intermediate points $\mathcal{S}_{k,1}^{(t)}, \dots, \mathcal{S}_{k,m-1}^{(t)}$ are not relevant for the probability of receiving an edge. These intermediate points allow the random variables $U_{k,1}, \dots, U_{k,m}$ to be rescaled on slightly different intervals, thus allowing for different distribution of the m edges of vertex k.

We remark the similarity of the notation between the endpoints of the intervals in Definition 4.6.1 of a unit graph and Definition 4.4.2 of a Pólya urn graph. In fact, unit graphs are a way to generalize the Pólya urn graph, in a way to include all possible versions of PAMs. First of all, we start showing that the Pólya urn graph in Definition 4.4.2 is a unit graph:

Lemma 4.6.2 (Pólya-unit graph)**.** *Fix $m \geq 1$ and $\delta > -m$. Then, the Pólya urn graph in Definition 4.4.2 is a unit graph in the sense of Definition 4.6.1 with the following parameters:*

- ▷ *for every $k \geq 2$, $\psi_{k,1}$ is a Beta random variable with parameters a_k and b_k as in* (4.6.1)*, and for $j \geq 2$, $\psi_{k,j} \equiv 0$;*
- ▷ *$(\psi_{k,1})_{k\geq 2}$ are independent of each other;*
- ▷ *for every $k \geq 2$ and $j \in [m]$, the jth edge of k is labeled NSL.*

The proof of Lemma 4.6.2 is immediate, since Definitions 4.4.2 and 4.6.1 coincide for the parameters given in the statement. In particular, for every $k \in \mathbb{N}$, since we set

$\psi_{k,j} \equiv 0$ for $j \geq 2$, we have that

$$S_{k,1}^{(t)} \equiv S_{k,2}^{(t)} \equiv \cdots \equiv S_{k,m}^{(t)} = \psi_{k,1} \prod_{h=k+1}^{t} (1 - \psi_{h,1}), \tag{4.6.6}$$

where the expression in (4.6.6) equals the definition of $S_k^{(t)}$ given in (4.4.8).

The nice property of the unit graph class is that we can describe all versions of PAMs as in Section 4.3 as a unit graph, just specifying the distributions of the random variables $(\psi_{k,j})_{k\in\mathbb{N}}^{j\in[m]}$ and the choice of the label NSL or SL.

4.6.3. The case $m = 1$

We start by considering the tree models, i.e., $m = 1$, thus $\delta > -1$. This setting turns out to be the easiest one, since all edges are allowed to be self-loops or not, so there are no mixed situations as mentioned above. As it turns out, many of the models defined in Section 4.3 coincide in the tree setting. In fact:

- ▷ Models (b) and (b') evolve with the same rules. For $m = 1$ the only difference is given by the initial graph;
- ▷ Models (e) and model (b') evolve with the same rules for $m = 1$;
- ▷ Models (c) and (d) coincide with model (e) with the restriction that $\delta \geq 0$;
- ▷ Model (a) allows for every edge to be a self-loop.

For models (b')-(c)-(d)-(e) there is nothing to prove, since they coincide and, by Lemma 4.6.2, we know that the corresponding Pólya urn graph is a unit graph. We just need to prove that models (a) and (b) are unit graphs:

Lemma 4.6.3 (Unit graphs - Models (a) and (b))**.** *For $m = 1$ and $\delta > -1$, PAMs models (a) and (b) are both unit graphs, where*

(a) $\psi_{k,1}$ is a Beta random variable with parameters $1 + \delta$ and $(2 + \delta)(k - 1)$, and $\psi_{k,j} \equiv 0$ for every $j \geq 2$. The graph is labeled SL;

(b) $\psi_{k,1}$ is a Beta random variable with parameters $1 + \delta$ and $(2 + \delta)(k - 1) + 1$, and $\psi_{k,j} \equiv 0$ for every $j \geq 2$. The graph is labeled NSL.

Proof. The proof consists in showing that models (a) and (b) can be described as Pólya urn experiments. For model (b) the proof is immediate, since model (b) is equal to model (b') with an extra edge between the first two vertices. Replacing the sequence $(\psi_k)_{k\in\mathbb{N}}$ of model (b') with $(\psi'_k)_{k\in\mathbb{N}}$, where ψ'_k is a Beta random variable with parameters $1 + \delta$ and $(2 + \delta)(k - 1) + 1$, completes the proof.

In model (a) we do allow for self-loops, thus we have to rescale the uniform random variables that determine the edges to different intervals and look at the weight of every vertex *before* attaching the edges, as mentioned in Section 4.6.1.

For $k \geq 2$, we have that the edge of vertex k is attached to k with probability proportional to $1+\delta$ and to one of the first $k-1$ with probability proportional to $(2+$

$\delta)(k-1)$. Thus, choosing $(\psi_k)_{k\geq 2}$ where ψ_k is a Beta random variable with parameters $1+\delta$ and $(2+\delta)(k-1)$, and labeling all edges as SL, the proof is complete. □

4.6.4. The case $m \geq 2$

When we consider $m \geq 2$, we can distinguish two classes of models: the *collapsed class*, i.e., the class of models where the case $m \geq 2$ can be defined through collapsing from the tree model, and the *non-collapsed class*, where the model with $m \geq 2$ cannot be directly obtained as a collapsed version of the case $m = 1$. For instance, model (a) belongs to the first class, model (e) belongs to the second.

We now present the procedure to collapse unit graphs:

Definition 4.6.4 (Collapsing unit graphs)**.** *Let $(\psi_k)_{k\in\mathbb{N}}$ and $(U_k)_{k\in\mathbb{N}}$ be the two sequences defining a unit tree (with the corresponding graph label SL/NSL). Then, fix $m \geq 2$. Define the two sequences, for $k \in \mathbb{N}$ and $j \in [m]$,*

$$\psi'_{k,j} = \psi_{m(k-1)+j}, \qquad U'_{k,j} = U_{m(k-1)+j}.$$

Then, the collapsed unit graph *is the unit graph defined by the sequences $(\psi'_{k,j})_{k\in\mathbb{N}}^{j\in[m]}$ and $(U'_{k,j})_{k\in\mathbb{N}}^{j\in[m]}$, where the graph label is the same as the corresponding unit tree.*

We can use the collapsed unit graph in Definition 4.6.4 to describe PAMs that, when $m \geq 2$, are defined by collapsing from the tree setting. In particular, since the label is maintained, the attachment probabilities of every edge in a collapsed unit graph is the same as the one in the collapsed PAM. In fact, if the graph is labeled as SL, then $\psi_{k,j}$ has parameters $a_{k,j}$ and $b_{k,j}$ that are equal to the weights of the vertex itself and older vertices respectively, thus giving the same distribution.

This explains why the class of unit graphs has been introduced. The main difference between collapsed and non-collapsed unit graphs is the fact that, in a collapsed graph, an interval $\mathcal{I}_k^{(t)}$ is associated to a vertex $k \in \mathbb{N}$ that determines which edges are attached to k, but we need *internal points* to maintain the law of the graph. In other words, the fact that a vertex in the collapsed unit graph is obtained by m different vertices is reflected by the fact that the corresponding interval can be divided in m sub-intervals.

4.7. LWC of model (e): structure of the proof of Theorem 4.3.2

The proof of Theorem 4.3.2 is based on a coupling argument. In other words, the proof consists in constructing a coupling between the neighborhood of a uniform vertex in PU_t and the neighborhood of the root $\varnothing$ of a Pólya point tree.

The LWC in distribution is proven by constructing the coupling for the neighborhood of a *single* uniformly chosen vertex V_t. For the LWC in probability, the coupling is constructed between *two neighborhoods of two uniformly chosen vertices* V_t^1 and V_t^2 and two Pólya point trees *simultaneously*. By the second moment method the results then follow.

In this section, we state several lemmas and propositions that are used in proof of Theorem 4.3.2. We claim then that the proof of convergence for the versions of PAMs defined in Section 4.3 holds, assuming that the preliminary lemmas in this section hold even for those different definitions of PAMs.

All the proofs of the results stated in this section are proven in [21], where these argument were first stated for $\delta \geq 0$, and in [86, Chapter 4], for the precise PAM version of Definition 4.3.1, that holds for any $\delta > -m$. [86, Chapter 4] contains only a sketch of the proof. In this chapter, the structure of the proof of the convergence of model (e) is presented to explain what are the tehcnical ingredients that are necessary. Once the Pólya urn graph representing model (e) is defined (recall Remark 4.5.2), the proof is identical to the proof for model (d) in [21].

4.7.1. Properties of Pólya urn graph

The formulation of the sequential model as the Pólya urn graph in Definition 4.3.1 has immediate consequences on the evaluation of the probability of attachments. In fact, conditionally on the sequence of intensities $(\psi_k)_{k\in[t]}$, it follows that, for $k \in [t]$, the probability that the jth edge of vertex k is attached to v is equal to

$$\mathbb{P}\left(U_{k,j} \in I_v \mid \psi_1, \dots, \psi_t\right) = \psi_v \frac{S_v^{(t)}}{S_{h-1}^{(t)}} = \psi_v \prod_{h=v+1}^{k-1} (1-\psi_h). \tag{4.7.1}$$

In addition, we recall that the random variables $(U_{k,j})_{k\in[t]}^{j\in[m]}$ are independent of each other. The proof of (4.7.1) follows immediately from the definition of the Pólya urn graph and the fact that we can easily prove by induction on $k \in [t]$ that

$$S_k^{(t)} = \prod_{h=k+1}^{t} (1-\psi_h). \tag{4.7.2}$$

(4.7.1) gives an alternative expression for the probability of existence of an edge in the graph. This formulation comes from the fact that in a Pólya urn experiment the number of balls in an urn evolves as a binomial random variable with probability of success ψ. Notice that the probability in (4.7.1) is independent of the random variables $\psi_k, \dots, \psi_t$.

A direct consequence of (4.7.1) is that, once we average over the uniform random variables $(U_{k,j})_{k\in[t]}^{j\in[m]}$, the properties of the graph are determined by the sequence of intensities $(\psi_k)_{k\in[t]}$.

We now state two preliminary results involving $(\psi_k)_{k\in[t]}$. The first one is about the concentration of the positions $(S_k^{(t)})_{k\in[t]}$ around deterministic values:

Lemma 4.7.1 (Positions concentrate in PU_t)**.** *Consider a Pólya urn graph as in Definition 4.4.2. Let $\chi = (m+\delta)/(2m+\delta)$. Then, for every $\varepsilon, \omega > 0$ there exists $N_0 \in \mathbb{N}$ such that,*

for every $t \geq N_0$,

$$\mathbb{P}\Big(\bigcap_{i=N_0}^{t} \Big\{ \Big| S_i^{(t)} - \Big(\frac{i}{t}\Big)^{\chi} \Big| \leq \omega \Big(\frac{i}{t}\Big)^{\chi} \Big\} \Big) \geq 1 - \varepsilon, \tag{4.7.3}$$

and, for t *large enough,*

$$\mathbb{P}\left(\max_{i\in[t]} \Big| S_i^{(t)} - \Big(\frac{i}{t}\Big)^{\chi} \Big| \geq \omega \right) \leq \varepsilon. \tag{4.7.4}$$

As a consequence, as $t \to \infty$,

$$\max_{i\in[t]} \Big| S_i^{(t)} - \Big(\frac{i}{t}\Big)^{\chi} \Big| \xrightarrow{\mathbb{P}} 0. \tag{4.7.5}$$

Lemma 4.7.1 is originally formulated in [21, Lemma 3.2]. Lemma 4.7.1 implies that $S_k^{(t)}$ is roughly $(k/t)^{\chi}$. This can give intuition behind the intensities of the Poisson processes in the definition of the Pólya point graph. Heuristically, if we identify a vertex k with its position $S_k^{(t)}$. When we have more and more vertices as $t \to \infty$, the positions of vertices become closer and closer, thus obtaining the intensities in Definition 4.3.1.

The second preliminary result on the sequence of intensities of the Pólya urn scheme is about the convergence in distribution of such sequence, and a coupling between $(\psi_k)_{k\in\mathbb{N}}$ and a sequence of Gamma random variables:

Lemma 4.7.2 (Beta-Gamma coupling)**.** *Consider the sequence* $(\psi_k)_{k\in\mathbb{N}}$ *as in Definition 4.3.1. Then, as* $k \to \infty$, $k\psi_k \xrightarrow{d} \Gamma$, *where* Γ *has a Gamma distribution with parameters* $m + \delta$ *and* $2m + \delta$.

More precisely, take $h_k(x)$ *such that* $\mathbb{P}(\psi_k \leq h_k(x)) = \mathbb{P}(\chi_k \leq x)$, *where* χ_k *has a Gamma distribution with parameters* $m + \delta$ *and 1 (so that* $\Gamma \stackrel{d}{=} \chi_k/(2m + \delta)$ *for all* $k \in \mathbb{N}$*). For every* $\varepsilon > 0$ *there exists* $K = K_\varepsilon \geq 1$ *sufficiently large such that, for all* $k \geq K_\varepsilon$ *and* $x \leq (\log k)^2$,

$$\frac{1-\varepsilon}{k(2m+\delta)} x \leq h_k(x) \leq \frac{1+\varepsilon}{k(2m+\delta)} x. \tag{4.7.6}$$

Further, with probability at least $1 - \varepsilon$, $\chi_k \leq (\log k)^2$ *for all* $k \geq K_\varepsilon$.

Lemma 4.7.2 is formulated and proved in [21, Lemma 3.3]. As mentioned, the proof of Theorem 4.3.2 is based on a coupling argument. Lemma 4.7.2 is then crucial, since it gives a link between the intensities in the Pólya urn graph and the strengths in the Pólya point tree, that are Gamma random variables.

Both Lemma 4.7.1 and Lemma 4.7.2 are based on the fact that, for every $k \in \mathbb{N}$, the two parameters a_k and b_k of the Beta random variable ψ_k are such that $a_k \equiv a = m+\delta$ and $b_k = m(2k-3)-(k-1)\delta) = (2m+\delta)k-3m+\delta$, i.e., $a_k \equiv m+\delta$ is constant and b_k is affine in k. Notice that in Lemma 4.7.2, the limiting distribution Γ has parameters a and $2m + \delta$, that is the slope of b_k.

4.7.2. Coupling argument

We state the coupling argument that proves Theorem 4.3.2 in the following proposition:

Proposition 4.7.3 (Coupling k-neighborhoods)**.** *Fix $\varepsilon > 0$ and $k \in \mathbb{N}$. Let V_t be a uniformly chosen vertex in $[t]$. Let $\{v_i : i \in \mathcal{U}\}$ be the vertices that are at most distance k from V_t, ordered such that the edge $\{v_i, v_{ij}\}$ was created before the edge $\{v_i, v_{i(j+1)}\}$.*

Then, there exists $\nu \in (0,1)$ and $K = K(\varepsilon) < \infty$ such that, the k-neighborhood $U_{\leq k}(V_t)$ in PA_t, together with $(S^{(t)}_{v_i})_{|i|\leq k}$, and the k-neighborhood $\widehat{U}_{\leq k}(\varnothing)$ of a Pólya point tree, together with its positions $(x_i)_{|i|\leq k}$, can be coupled such that, with probability at least $1-\varepsilon$,

(1) $U_{\leq k}(V_t) \cong \widehat{U}_{\leq k}(\varnothing)$;
(2) $|\widehat{U}_{\leq k}(\varnothing)| \leq K$;
(3) $|x_i - S^{(t)}_{v_i}| \leq \nu$ *for all* $i \in \widehat{U}_{\leq k}(\varnothing)$;
(4) $(v_i)_{i\in\widehat{U}_{\leq k}(\varnothing)}$ *are all distinct and* $v_i \geq \nu t$ *for all* $v_i \in \widehat{U}_{\leq k}(\varnothing)$;
(5) $\gamma_i \leq K$ *for all* $i \in \widehat{U}_{\leq k}(\varnothing)$.

Proposition 4.7.3 proves the convergence in distribution of the sequential PAM to the Pólya point tree. More precisely, statement (1) in Proposition 4.7.3 is sufficient for the proof of the LWC. In fact, we can construct a coupling in a way that, with high probability, the two neighborhoods in $U_{\leq k}(V_t)$ in the PAM and $\widehat{U}_{\leq k}(\varnothing)$ in the Pólya point tree are isomorphic.

Proposition 4.7.3 gives more information than just the topological structure of the neighborhood of a uniformly chosen vertex in PAM. For example, statement (3) assures us that the positions $(x_i)_{i\in\widehat{U}_{\leq k}(\varnothing)}$ of the vertices in the Pólya point tree are close to the points $(S_{v_i})_{i\in\widehat{U}_{\leq k}(\varnothing)}$ that defines the intervals corresponding to the vertices in $U_{\leq k}(V_t)$.

Before explaining the structure of the proof, we state a regularity lemma:

Lemma 4.7.4 (Regularity of PPT)**.** *Fix $k \geq 1$ and $\varepsilon > 0$. Then there exist constants $\eta > 0$ and $K < \infty$ such that, with probability at least $1-\varepsilon$,*

- $|\widehat{U}_{\leq k}(\varnothing)| \leq K$;
- $x_i \geq \eta$ *for all* $i \in \widehat{U}_{\leq k}(\varnothing)$;
- $\gamma_i \leq K$, $\rho_i(\cdot) \leq K$ *for all* $i \in \widehat{U}_{\leq k}(\varnothing)$;
- $\min_{i,i'\in\widehat{U}_{\leq k}(\varnothing)} |x_i - x_{i'}| \geq \eta$.

Lemma 4.7.4 states that, with high probability, the k-neighborhood of the root $\varnothing$ is composed by finitely many vertices K. This constant K also bounds all the strengths of the vertices in $\widehat{U}_{\leq k}(\varnothing)$. The positions of every vertex in $\widehat{U}_{\leq k}(\varnothing)$ are separated by a constant $\eta > 0$, and are at least η.

As the reader can notice, the properties of the PPT listed in Lemma 4.7.4 are similar to the statements listed in Proposition 4.7.3. The idea of the coupling is to use the random variables that are assigned to the vertices of the PPT, such as the positions and the strengths, to define the k-neighborhood in the Pólya urn graph.

We now explain the coupling for the 1-neighborhood, this is enough to give the intuition behind the technicalities of the proof of Proposition 4.7.3. In addition, the 1-neighborhood contains all the delicate points that are necessary to be checked to extend the LWC result to the other PAM definitions in Section 4.3.

4.7.3. The 1-neighborhood

Consider the position of the root $x_\varnothing = U_\varnothing^\chi$ of the PPT, where $U_\varnothing$ is uniform on $[0,1]$, and the positions $x_{\varnothing 1}, \dots, x_{\varnothing m + d_\varnothing^{(\mathrm{in})}}$ of the $m + d_\varnothing^{(\mathrm{in})}$ children of the root. We choose the vertex $V_t = \lceil U_\varnothing t \rceil$, that is indeed a uniformly chosen vertex in $[t]$.

We start exploring the m neighbors of $\varnothing$ of type O, that correspond to the m vertices found by exploring the m original edges of V_t. For $i \in [m]$, we take $v_{\varnothing i}$ such that

$$S^{(t)}_{v_{\varnothing i}-1} \le \frac{x_{\varnothing i}}{x_\varnothing} S^{(t)}_{v_\varnothing} \le S^{(t)}_{v_{\varnothing i}}. \tag{4.7.7}$$

Notice that the indices $v_{\varnothing 1}, \dots, v_{\varnothing m}$ in (4.7.7) might not be distinct. By definition, the sequence $x_{\varnothing 1}/x_\varnothing, \dots, x_{\varnothing m}/x_\varnothing$ is a collection of m i.i.d. random variables with uniform distribution on $[0,1]$.

Given $\varepsilon > 0$ we can take $\eta > 0$ and $K > 0$ such that Lemma 4.7.4 holds for $k = 1$. By Lemma 4.7.1, for t large enough,

$$|S^{(t)}_{v_{\varnothing j}} - x_{\varnothing j}| \le \eta,$$

for all $j = 0, \dots, m$, where $\varnothing 0 = \varnothing$, with probability larger than $1 - 2\varepsilon$. As a consequence, $v_{\varnothing i} \in t[x_{\varnothing i} - \eta, x_{\varnothing i} + \eta]$.

In particular, since we have applied Lemma 4.7.4, all the indices $v_{\varnothing 0}, \dots, v_{\varnothing m}$ are distinct, since the positions in the PPT are all at distance at least η. From the PAM perspective, this implies that with high probability, the m edges of a uniformly chosen vertex V_t do not create self-loops.

We have now explored the neighbors of type O of the root $\varnothing$. We still have to explore the remaining $d_\varnothing^{(\mathrm{in})}$ neighbors of type Y. By (4.7.1) we know that, conditionally on $\psi_1, \dots, \psi_t$, every vertex $k > v_\varnothing$ can be connected to $v_\varnothing$ with m edges, independently of each other. Let $X_{k,j}$ denote the vertex that is chosen by the jth edge of k. Then all the events $(\{X_{k,j} = v_\varnothing\})_{k > v_\varnothing}^{j \in [m]}$, conditioning on $(\psi_k)_{k \in [t]}$ and $v_\varnothing$, are independent of each other with probability of success given by

$$P_{k \to v_\varnothing} = \psi_{v_\varnothing} \frac{S^{(t)}_{v_\varnothing}}{S^{(t)}_{k-1}}. \tag{4.7.8}$$

We can then define, for $y \in (U_\varnothing, 1]$,

$$W_{v_\varnothing}(y) = \sum_{v=v_\varnothing+1}^{\lceil yt \rceil} \sum_{j=1}^{m} \mathbb{1}\{X_{k,j} = v_\varnothing\}. \tag{4.7.9}$$

The process in (4.7.9) represents the evolution of the degree of vertex $v_\varnothing$ from time $v_\varnothing + 1$ to time $\lceil yt \rceil$. Equation (4.7.9) shows that the degree of a vertex can be written as a sum of Bernoulli random variables that, conditionally on $(\psi_k)_{k\in[t]}$ and $v_\varnothing$, are independent of each other with probability of success given by (4.7.8).

The following step of the proof is to show that, conditionally on $U_\varnothing$, the process $(W_{v_\varnothing}(y))_{y\in(U_\varnothing,1]}$ converges as $t \to \infty$ to an inhomogeneous Poisson process on $[U_\varnothing, 1]$, whose intensity is the function $\rho_\varnothing(\cdot)$ in the definition of PPT. As mentioned, since the degree of vertex $v_\varnothing$ is written as a sum of conditionally independent Bernoulli random variables, it is sufficient to prove that, as $t \to \infty$, the probabilities of success given in (4.7.8) converge to the intensity $\rho_\varnothing(\cdot)$.

For this, we define a process $(\hat{W}_{v_\varnothing}(y))_{y\in(U_\varnothing,1]}$ as sum of independent Bernoulli random variables, with different probabilities of success. We use Lemma 4.7.2, so with probability larger than $1 - \varepsilon$ and for every $k \geq K(\varepsilon)$, $\psi_k = h_k(\chi_k)$, where $(\chi_k)_{k\in\mathbb{N}}$ is a sequence of Gamma random variables with parameters 1 and $m + \delta$. Then, Lemmas 4.7.1 and 4.7.2 give that

$$(1-\omega)\hat{P}_{k\to v_\varnothing} \leq P_{k\to v_\varnothing} \leq (1+\omega)\hat{P}_{k\to v_\varnothing}, \tag{4.7.10}$$

where

$$\hat{P}_{k\to v_\varnothing} = \frac{\chi_{v_\varnothing}}{(2m+\delta)t} \frac{t}{v_\varnothing} \Big(\frac{v_\varnothing}{k}\Big)^{\chi}. \tag{4.7.11}$$

We can then define

$$\hat{W}_{v_\varnothing}(y)) = \sum_{v=v_\varnothing+1}^{\lceil yt \rceil} \sum_{j=1}^{m} \hat{X}_{k,j},$$

where $(\hat{X}_{k,j})_{k>v_\varnothing}^{j\in[m]}$ is a sequence of conditionally independent Bernoulli random variables with probability of success given by (4.7.11). We can couple $(\hat{W}_{v_\varnothing}(y))_{y\in(U_\varnothing,1]}$ to a Poisson random variable with parameter

$$\frac{\chi_{v_\varnothing}}{(2m+\delta)U_\varnothing} \int_0^y \Big(\frac{U_\varnothing}{s}\Big)^{\chi} ds. \tag{4.7.12}$$

Denoting $\phi_{v_\varnothing} = \Gamma_\varnothing$ as in Section 4.3.1, it is possible to prove that $\hat{W}_{v_\varnothing}(x^\gamma)$ converges in distribution to a Poisson point process on $[U_\varnothing, 1]$ with intensity $\Gamma_\varnothing \phi x^{\phi-1}/U_\varnothing^{\chi} = \rho_\varnothing(x)$ as in the definition of PPT.

The fact that all the vertices in the 1-neighborhood are distinct (and their positions are close to the ones in the PPT as in statement (2) in Proposition 4.7.3) follows from the concentration result in Lemma 4.7.1 and the regularity Lemma 4.7.4 on the PPT.

4.7.4. Difference with larger radius

We have just described the proof of Proposition 4.7.3 for the 1-neighborhood. The proof for $k \geq 2$ is based on induction over the radius k. Assuming that Proposition 4.7.3 holds for $k-1$, for every vertex v_i, for $i \in \mathcal{U}$ at distance k from V_t the induction hypothesis is that:

(1) the m (or $m-1$) unexplored edges of v_i identify m (or $m-1$) distinct vertices, whose position is almost uniform on $[0, S_{v_i}^{(t)}]$;
(2) the younger vertices found exploring edges that are attached to v_i are all distinct, and their positions have distribution close to a Poisson process with intensity $\rho_i(\cdot)$ as in Section 4.3.1.

It is necessary though to be careful once we explore radius larger than 1. First of all, given a vertex v_i, this vertex can be of two types: O if the ancestor v_l of v_i in the tree is such that $v_l > v_i$, otherwise v_i is of type Y. This influences the number of edges we have to explore starting from v_i: in fact, if v_i is of type Y, we can explore only $m-1$ of its original edges, since one edge has already been used to find v_i itself (recall that PA_t is defined as an undirected graph).

Also, if v_i is of type O, this implies that we have to condition on the fact that we know v_i has at least one edge attached to it, thus increasing the probability of v_i receiving more edges (recall the rich-get-richer effect).

Consider then a vertex v_i at distance k from $v_\varnothing$. Denote by m_i the original m or $m-1$ unexplored original edges of v_i. These m_i edges are determined by m_i uniform random variables on $[0, S_{v_i}^{(t)}]$. Since we have already explored the neighborhood of $v_\varnothing$ up to distance $k-1$, we have to condition on the fact that the m_i edges of v_i cannot be attached to vertices already found in the exploration process, since otherwise v_i would not be at distance k from $v_\varnothing$.

We have a similar problem for the younger neighbors of v_i. In fact, similarly to (4.7.9), it seems natural to define the process, for $y \in (x_{v_i}, 1]$,

$$W_{v_i}(y) = \sum_{v=v_i+1}^{\lceil yt \rceil} \sum_{j=1}^{m} \mathbb{1}\{X_{k,j} = v_i\}, \tag{4.7.13}$$

but this is not feasible. The difference between (4.7.9) and (4.7.13) is that, while in (4.7.9) the sum can be taken over all indices from $v_\varnothing$ to $\lceil yt \rceil$, in (4.7.13) this is not possible. In fact, we might have used some of the edges of the vertices $v_i+1, \dots, \lceil yt \rceil$ to construct the $(k-1)$-neighborhood.

Here, implication (4) of Proposition 4.7.3 comes in handy. In fact, we know that the $(k-1)$-neighborhood of V_t is finite, that the strengths of its vertices are bounded by the same constant, and that all vertices have separate positions. In other words, we can define the process in (4.7.13) by ignoring at most a finite number of indices, that are the ones used in the exploration up to radius $k-1$.

With these observations, it is easy to understand that the process $W_{v_i}(y)$ converges to a Poisson process on $[x_i^X, 1]$, similarly to the radius 1 case, since we are ignoring

only a finite number of terms in (4.7.13), and, as $t \to \infty$, this is a negligible effect. The intensity of this Poisson process is obtained by similar calculations as (4.7.12) and it is given by the intensity $\rho_i(\cdot)$ as in Section 4.3.1.

Remark 4.7.5 (Two neighborhoods)**.** So far we have discussed about the LWC in distribution. For the convergence *in probability*, the proof is based on the second moment method. Instead of looking at the probability that the k-neighborhood of V_t is distributed as a PPT, it is sufficient to prove that the k-neighborhoods of *two uniformly chosen vertices* V_t^1 and V_t^2 are distributed as two *independent* copies of the PPT. The proof consists of a coupling argument similar to Proposition 4.7.3. It is not hard to see that, since Proposition 4.7.3 is based on Lemmas 4.7.1 and 4.7.2 (regarding the sequence $(\psi_k)_{k\in[t]}$) and Lemma 4.7.4 (regarding the Pólya point tree), and the fact that, at every step, we condition on a finite number of edges, constructing one or two neighborhoods simultaneously is not very different.

4.8. Properties of the Pólya urn graph: proof of Lemma 4.7.1 and Lemma 4.7.2

4.8.1. Proof of Lemma 4.7.2

We start with the first part of the statement. Fix $x \geq 0$. We compute that

$$\begin{aligned}\mathbb{P}(k\psi_k \leq x) &= \frac{\Gamma(m(2k-2)+k\delta)}{\Gamma(m+\delta)\Gamma(m(2k-3)+(k-1)\delta)} \\ &\qquad \times \int_0^{x/k} u^{m+\delta-1}(1-u)^{(m(2k-3)+(k-1)\delta)-1}du \\ &= (1+o(1))\frac{k^{m+\delta}k^{-(m+\delta)}}{\Gamma(m+\delta)}\int_0^x u^{m+\delta-1}(1-u/k)^{(m(2k-3)+(k-1)\delta)-1}du.\end{aligned}$$

For every $u > 0$, as $k \to \infty$, $(1-u/k)^{(m(2k-3)+(k-1)\delta)-1} \to \mathrm{e}^{-u(2m+\delta)}$, so that dominated convergence implies that

$$\mathbb{P}(k\psi_k \leq x) \to \int_0^x \frac{u^{m+\delta-1}}{\Gamma(m+\delta)}\mathrm{e}^{-u(2m+\delta)}du,$$

as required. This proves the convergence in distribution as in the statement. The second part of the proof follows verbatim from [21, Lemma 3.4]. □

4.8.2. Proof of Lemma 4.7.1

Fix $\omega, \varepsilon > 0$, and let $\bar{\omega} = |\log(1+\omega)|$. We use the formulation (4.7.2) to bound the distance between $S_i^{(t)}$ and $(i/t)^\chi$. For every $i \in [t]$, $S_i^{(t)}$ is the product of $t-i-1$ independent terms that are functions of independent Beta distributions. It is easy to

show by induction that, for $i \in [t-1]$,

$$S_i^{(t)} = \prod_{k=i+1}^{t} (1-\psi_k) = \exp\left(\sum_{k=i+1}^{t} \log(1-\psi_k) \right), \tag{4.8.1}$$

while $S_t^{(t)} \equiv 1$. We can just look at the argument of the exponential in (4.8.1), ignoring $S_t^{(t)}$. Notice that

$$\mathrm{Var}(\log(1-\psi_k)) \le \mathbb{E}\left[\log^2(1-\psi_k)\right] \le \mathbb{E}\left[\frac{\psi_k^2}{(1-\psi_k)^2}\right], \tag{4.8.2}$$

By (4.8.2) and Kolmogorov's inequality we can bound

$$\mathbb{P}\Big(\max_{i\in[t-1]} \Big| \sum_{k=i+1}^{t} \log(1-\psi_k) - \mathbb{E}\Big[\sum_{k=i+1}^{t} \log(1-\psi_k) \Big] \Big| \ge \bar{\omega}/2 \Big) \tag{4.8.3}$$
$$\le \frac{4}{\bar{\omega}^2} \sum_{i=2}^{t} \mathbb{E}\left[\frac{\psi_i^2}{(1-\psi_i)^2}\right].$$

Equation (4.8.3) shows that the maximum of the fluctuations of the argument in (4.8.1) can be bounded by the variances of the singles terms. By properties of the Beta distribution, and recalling that $\psi_i \sim \beta(m+\delta, 2m(i-2)+m+(i-1)\delta)$, for $i = 1, \dots, t$, we can write, for $i \ge 2$,

$$\mathbb{E}\left[\frac{\psi_i^2}{(1-\psi_i)^2}\right] \tag{4.8.4}$$
$$= \frac{(m+\delta)(m+1+\delta)}{(2m(i-2)+m+(i-1)\delta-2)(2m(i-2)+m+(i-1)\delta-1)} = \mathcal{O}(i^{-2}).$$

Equation (4.8.4) assures us that the sum on the right-hand side of (4.8.3) is finite as $t \to \infty$. Therefore, we can fix $N_1(\bar{\omega}) \in \mathbb{N}$ such that $\sum_{i=N_1}^{\infty} \le (\varepsilon\bar{\omega}^2)/4$. As a consequence, bounding the sum in (4.8.3) by the tail of the series, for $t \ge N_1$, we have that

$$\mathbb{P}\Big(\max_{i=N_1,\dots,t-1} \Big| \sum_{k=i+1}^{t} \log(1-\psi_k) - \mathbb{E}\Big[\sum_{k=i+1}^{t} \log(1-\psi_k) \Big] \Big| \ge \bar{\omega}/2 \Big) \tag{4.8.5}$$
$$\le \frac{4}{\bar{\omega}^2} \sum_{i=N_1}^{\infty} \mathbb{E}\left[\frac{\psi_i^2}{(1-\psi_i)^2}\right] \le \varepsilon.$$

Now we want to show that the expectations of such terms are converging to the sequence of the expectations of $\psi_1, \dots, \psi_t$. Using the fact that, for $x \in (0,1)$,

$$|\log(1-x) - x| \le (x^2)/(1-x),$$

we write, for $N_1 \le i \le t$,

$$\left| \mathbb{E}\left[\sum_{k=i+1}^{t} \log(1-\psi_k) \right] - \sum_{k=i+1}^{t} \mathbb{E}[\psi_k] \right| \le \sum_{k=i+1}^{t} \mathbb{E}\left[\frac{\psi_k^2}{1-\psi_k} \right]. \tag{4.8.6}$$

By the properties of Beta distribution,

$$\begin{aligned} \mathbb{E}[\psi_k] &= \frac{m+\delta}{2m(k-1)+i\delta} = \frac{m+\delta}{2m+\delta}\frac{1}{k}(1+\mathcal{O}(k^{-1})) = \frac{\chi}{k}(1+\mathcal{O}(k^{-1})); \\ \mathbb{E}\left[\frac{\psi_k^2}{1-\psi_k}\right] &= \frac{(m+\delta)(m+1+\delta)}{(2m(k-1)+i\delta)(2m(k-2)+m+(k-1)\delta-2)} = \mathcal{O}(i^{-2}); \end{aligned} \tag{4.8.7}$$

Now, fix $N_2 \in \mathbb{N}$ such that, for every $t \ge N_2$, we have that $\sum_{t=N_2}^{\infty} \mathbb{E}\left[\frac{\psi_i^2}{1-\psi_i}\right] < \bar{\omega}/2$. Using the bounds in (4.8.7) in (4.8.6) we have that, for every $N_2 \le i \le t$, that

$$\begin{aligned} \left| \mathbb{E}\left[\sum_{k=i+1}^{t} \log(1-\psi_k) \right] - \sum_{k=i+1}^{t} \mathbb{E}[\psi_k] \right| &\le \sum_{k=i+1}^{t} \frac{\chi}{k}(1+\mathcal{O}(k^{-1})) + \bar{\omega}/2 \\ &= \chi \log(i/t) + \mathcal{O}(i^{-1}) + \bar{\omega}/2. \end{aligned}$$

As a consequence, for $t \ge N_2$,

$$\max_{i=N_2,\dots,t} \left| \mathbb{E}\left[\sum_{k=i+1}^{t} \log(1-\psi_k) \right] - \chi \log(i/t) \right| \le \bar{\omega}/2. \tag{4.8.8}$$

Let $N_0 = \max\{N_1, N_2\}$. By (4.8.8) and (4.8.5), then we have that for $t \ge N_0$

$$\mathbb{P}\left(\max_{i=N_0,\dots,t} \left| \sum_{k=i+1}^{t} \log(1-\psi_k) - \chi \log(i/t) \right| \ge \bar{\omega} \right) \le \varepsilon. \tag{4.8.9}$$

Now, recalling that $\log(S_i^{(t)}) = \sum_{k=i+1}^{t} \log(1-\psi_k)$, we can rewrite that, for every $i = N_0, \dots, t$, we have

$$-\bar{\omega}+\chi\log(i/t) \le \log S_i^{(t)} \le \bar{\omega}+\chi\log(i/t) \iff \mathrm{e}^{-\bar{\omega}}\left(\frac{i}{t}\right)^{\chi} \le S_i^{(t)} \le \mathrm{e}^{\bar{\omega}}\left(\frac{i}{t}\right)^{\chi}. \tag{4.8.10}$$

Recall then we have defined $\bar{\omega}$ as $|\log(1+\omega)|$. This means that the right hand condition in (4.8.10) becomes

$$\frac{1}{1+\omega}\left(\frac{i}{t}\right)^{\chi} \le S_i^{(t)} \le (1+\omega)\left(\frac{i}{t}\right)^{\chi}.$$

Since $(i/t)^{\chi} \le 1$ and $1/(1+\omega) \ge 1-\omega$, we have that, for every $i = N_0, \dots, t$,

$$(1-\omega)\left(\frac{i}{t}\right)^{\chi} \le S_i^{(t)} \le (1+\omega)\left(\frac{i}{t}\right)^{\chi}. \tag{4.8.11}$$

Combining (4.8.9) and (4.8.11) we obtain that

$$\mathbb{P}\Big(\bigcap_{i=N_0}^{t} \Big\{ \Big| S_i^{(t)} - \Big(\frac{i}{t}\Big)^{\chi} \Big| \leq \omega \Big(\frac{i}{t}\Big)^{\chi} \Big\} \Big) \geq 1-\varepsilon,$$

which proves (4.7.3). To prove (4.7.4), we observe that, for fixed $\omega/4 > 0$ and $\varepsilon > 0$, by (4.7.3) $S_{N_0}^{(t)}(\omega/4, \varepsilon)$ converges to 0 as t grows. Since $S_i^{(t)} \leq S_{N_0}^{(t)}$ for every $i \leq N_0$, we can take t large enough such that

$$\max_{i=1,\dots,N_0} \Big| S_i^{(t)} - \Big(\frac{i}{t}\Big)^{\chi} \Big| \leq S_{N_0}^{(t)} \Big(\frac{N_0}{t}\Big)^{\chi} \leq \frac{\omega}{2} + \Big(\frac{N_0}{t}\Big)^{\chi} \leq \omega. \tag{4.8.12}$$

This completes the proof. □

4.9. Extension of LWC result: proof of Theorem 4.2.1

In this section, we prove Theorem 4.2.1. We divide the proof into two parts: first we prove Theorem 4.2.1 when $m = 1$. Then we extend the proof to $m \geq 2$, focusing in particular on model (a), which is the standard collapsed model. The proof for model (a) is easily adapted to models (b) and (b').

4.9.1. Sufficient conditions for convergence

We presented the structure of the proof of convergence of model (e) as presented in [21] and [86, Chapter 4], highlighting the main steps. In particular, it is easy to check that for any unit graph the convergence holds when the necessary ingredients, such as Lemmas 4.7.1 and 4.7.2 hold:

Proposition 4.9.1 (Convergence conditions for unit graphs)**.** *Fix $m \geq 1$ and $\delta > -m$. Then a unit graph $(\mathrm{UG}_t(m,\delta))_{t\in\mathbb{N}}$ converges locally weakly in probability to the PPT if:*

(1) the sequence $(\psi_{k,m})_{k\in\mathbb{N}}$ consists of Beta random variables with parameters a_k and b_k, where $a_k = m+\delta$, and $b_k = (2m+\delta)k + c$, for some model-dependent constant c;
(2) Lemma 4.7.1 holds;
(3) Lemma 4.7.2 holds;
(4) a bound of the form given in (4.7.10) *and* (4.7.11) *holds for the attachment probabilities.*

Proof. The proof of this proposition is immediate. We observe that Lemmas 4.7.1 and 4.7.2 do not depend on the structure of a unit graph, but only on the parameters of the sequence $(\psi_{k,m})_{k\in\mathbb{N}}$. In particular, the intermediate points $\mathcal{S}_{k,1}^{(t)}, \dots, \mathcal{S}_{k,m-1}^{(t)}$ are not relevant, since the attachment probabilities depend on the boundaries of the intervals $(\mathcal{I}_k^{(t)})_{k\in\mathbb{N}}$. Notice that the value $\chi = (1+\delta)/(2+\delta)$ is exactly the scaling constant for which Lemma 4.7.1 holds. For a more precise argument, we refer to [21, Lemma 3.2] and [86, Chapter 4].

Statement (4) is a consequence of (1)-(3) and the fact that the degree of a vertex in a unit graph can be written as the sum of conditionally independent Bernoulli random variables, as in (4.7.9). □

Proposition 4.9.1 proves Theorem 4.2.1 for all models in Section 4.3 for $m \geq 1$ and any value of $\delta > -m$. Proposition 4.9.1 can be specialized to collapsed unit graphs, as stated in the following proposition:

Proposition 4.9.2 (Convergence conditions for collapsed unit graphs)**.** *Fix $m \geq 2$ and $\delta > -m$. Then a collapsed unit graph $(\mathrm{UG}_t(m, \delta))_{t\in\mathbb{N}}$ converges locally weakly in probability to the PPT if the corresponding unit tree $(\mathrm{UT}_{mt}(1, \delta/m))_{t\in\mathbb{N}}$ satisfies conditions (1)-(4) in Proposition 4.9.1.*

Proposition 4.9.2 proves Theorem 4.2.2. In particular, Proposition 4.9.2 shows that for collapsed PAMs, the convergence for $m \geq 2$ is determined by the convergence of the tree setting. This fact underlines the relevance of the tree setting in PAMs.

4.9.2. Proof of Proposition 4.9.2

To prove Proposition 4.9.2 it is sufficient to show that, in the case of a collapsed unit graph, conditions (1)-(4) are satisfied. Condition (1) is trivial, since it follows directly from the definition of a unit graph. Condition (2) and (3) are general results on sequences of Beta random variables, thus they are still true for the collapsed unit graph.

What is left to prove is condition (4), i.e., the bound on the conditional attachment probabilities, similarly to (4.7.10) and (4.7.11). First, we need the analytic expression for the conditional attachment probabilities:

Lemma 4.9.3 (Conditional attachment probability)**.** *Consider $u, v \in [t]$, with $u \leq v$ and $j \in [m]$. Then, in a collapsed unit graph, conditionally on $(\psi_{(k,j)})_{k\in[t]}^{j\in[m]}$:*

$$\mathbb{P}\left(v \xrightarrow{j} u \mid (\psi_{(k,j)})_{k\in[t]}^{j\in[m]}\right) = \begin{cases} \dfrac{\mathcal{S}^{(t)}_{(u,m)}}{\mathcal{S}^{(t)}_{(v,j-1)}}\left(1 - \displaystyle\prod_{i=1}^{m}(1-\psi_{(u,i)})\right) & \text{for } u < v, \\ 1 - \displaystyle\prod_{i=1}^{j}(1-\psi_{(u,i)}) & \text{for } u = v. \end{cases} \tag{4.9.1}$$

Proof. We prove the result assuming that the graph has label SL. This does not change the result, since this implies the presence of an extra term in (4.9.1) that is not present in the case of NSL graphs. This extra term comes from the different rescaling of the uniform random variables that determine edges in the graph (recall (4.6.3) and (4.6.4) in the definition of unit graphs).

We start proving (4.9.1) for $u < v$. We can write

$$\{v \xrightarrow{j} u\} = \left\{U_{(v,j)}\mathcal{S}^{(t)}_{(v,j-1)} \in \mathcal{I}^{(t)}_{(u,1)} \cup \cdots \cup \mathcal{I}^{(t)}_{(u,m)}\right\},$$

as a consequence, conditioning on $(\psi_{(n,j)})_{n\in[t]}^{j\in[m]}$,

$$\mathbb{P}\left(v \xrightarrow{j} u \mid (\psi_{(n,j)})_{n\in[t]}^{j\in[m]}\right) = \frac{|\mathcal{I}^{(t)}_{(u,1)} \cup \cdots \cup \mathcal{I}^{(t)}_{(u,m)}|}{\mathcal{S}^{(t)}_{(v,j)}} = \frac{1}{\mathcal{S}^{(t)}_{(v,j)}} \sum_{i=1}^{m} \mathcal{S}^{(t)}_{(u,i)} - \mathcal{S}^{(t)}_{(u,i-1)} = \frac{\mathcal{S}^{(t)}_{(u,m)} - \mathcal{S}^{(t)}_{(u-1,m)}}{\mathcal{S}^{(t)}_{(v,j)}}.$$

By the fact that $\mathcal{S}^{(t)}_{(k,j)} = \prod_{h=m(k-1)+j+1}^{t}(1-\psi_h)$,

$$\begin{aligned}\frac{\mathcal{S}^{(t)}_{(u,m)} - \mathcal{S}^{(t)}_{(u-1,m)}}{\mathcal{S}^{(t)}_{(v,j)}} &= \left(\prod_{h=(u,m)+1}^{(v,j)}(1-\psi_h)\right)\left(1-\prod_{i=1}^{m}(1-\psi_{(u,i)})\right) \\ &= \frac{\mathcal{S}^{(t)}_{(u,m)}}{\mathcal{S}^{(t)}_{(v,j)}}\left(1-\prod_{i=1}^{m}(1-\psi_{(u,i)})\right),\end{aligned} \tag{4.9.2}$$

and (4.9.2) coincides with (4.9.1) for $u < v$. When $u = v$, we have a self-loop. In this case, the fraction in the left-hand side of (4.9.2) is equal to

$$\frac{\mathcal{S}^{(t)}_{(u,j)} - \mathcal{S}^{(t)}_{(u-1,m)}}{\mathcal{S}^{(t)}_{(u,j)}} = \frac{\mathcal{S}^{(t)}_{(u,j)}}{\mathcal{S}^{(t)}_{(u,j)}}\left(1-\prod_{i=1}^{j}(1-\psi_{(u,i)})\right) = 1-\prod_{i=1}^{j}(1-\psi_{(u,i)}), \tag{4.9.3}$$

which coincides with (4.9.1) for $u = v$. Notice that (4.9.3) reduces to $\psi_{(u,1)}$ whenever $u = v$ and $j = 1$. □

With Lemma 4.9.3 in hand, we can prove the bound on the conditional attachment probabilities, thus showing that condition (4) in Proposition 4.9.1 holds for collapsed unit graphs:

Lemma 4.9.4 (Bound on attachment probabilities)**.** *Fix $m \geq 2$ and $\delta > -m$. Let $\chi = (m+\delta)/(2m+\delta)$. For $t \in \mathbb{N}$, consider $v \in [t]$. Then, for every $\varepsilon > 0$ there exists $\omega > 0$ such that, with probability larger than $1-\varepsilon$, for every $k \geq v$,*

$$(1-\omega)\frac{\hat{\chi}_v}{(2m+\delta)v}\left(\frac{v}{k}\right)^{\chi} \leq \mathbb{P}\left(k \to v \mid (\psi_{h,j})_{h\in[t]}^{j\in[m]}\right) \leq (1+\omega)\frac{\hat{\chi}_v}{(2m+\delta)v}\left(\frac{v}{k}\right)^{\chi},$$

where $\hat{\chi}_v$ is a Gamma random variable with parameters $m+\delta$ and 1.

Proof. Fix $t \in \mathbb{N}$, and consider the unit tree $\mathrm{UT}^{(t)}_{mt}(1, \delta/m)$. Fix $v \in [t]$. In $\mathrm{UT}^{(t)}_{mt}(1,\delta/m)$, the vertex v corresponds to m distinct elements $(v,1),\ldots,(v,m)$. For $k \geq v$, the probability that k attaches its jth edge to v is given by Lemma 4.9.3, so

$$\mathbb{P}\left(k \xrightarrow{j} v \mid (\psi_{h,j})_{h\in[t]}^{j\in[m]}\right) = \frac{\mathcal{S}^{(mt)}_{(v_\varnothing,m)} - \mathcal{S}^{(mt)}_{(v-1,m)}}{\mathcal{S}^{(mt)}_{(k,j)}} = \frac{\mathcal{S}^{(mt)}_{(v,m)}}{\mathcal{S}^{(mt)}_{(k,j)}}\left(1-\prod_{i=1}^{m}(1-\psi_{(v,i)})\right). \tag{4.9.4}$$

Now, the fraction in the last term of (4.9.4) is related to the unit tree, since the random variables $\mathcal{S}^{(mt)}_{(v,m)}$ and $\mathcal{S}^{(mt)}_{(k,j-1)}$ are defined from a tree model with parameters 1 and δ/m.

Fix $\varepsilon > 0$. Since we assume that Lemma 4.7.1 holds for the unit tree, we have that, with probability larger than $1-\varepsilon$, for some $\omega > 0$,

$$\Big|\frac{\mathcal{S}^{(mt)}_{(v,m)}}{\mathcal{S}^{(mt)}_{(k,j)}} - \Big(\frac{mv}{m(k-1)+j}\Big)^{\chi}\Big| \leq \omega\Big(\frac{mv}{m(k-1)+j}\Big)^{\chi}, \tag{4.9.5}$$

where the value χ comes from the tree model, so

$$\chi = \frac{1+\delta/m}{2+\delta/m} = \frac{m+\delta}{2m+\delta}.$$

Notice that in (4.9.5), we can approximate $(mv)/(m(k-1)+j)$ by v/k. We are left with the product inside the expectation in (4.9.4). Here we notice that we can rewrite such a product as

$$1 - \prod_{i=1}^{m}(1-\psi_{(v,i)}) = \sum_{i=1}^{m}\psi_{(v,i)} + E_m(v). \tag{4.9.6}$$

We start considering the sum in the right-hand side of (4.9.6). By the assumptions, we can apply Lemma 4.7.2 to the unit tree. As a consequence, for $i \in [m]$,

$$[m(v)+(i-1)]\psi_{(v,i)} \approx \frac{\chi_{(v,i)}}{2+\delta/m},$$

where $(\chi_h)_{h\in\mathbb{N}}$ is a sequence of independent Gamma random variables with parameters $1+\delta/m$ and 1. We can then rewrite the sum in (4.9.6) as

$$\begin{aligned}\sum_{i=1}^{m}\psi_{(v,i)} &\approx \frac{1}{2+\delta/m}\sum_{i=1}^{m}\frac{\chi_{(v,i)}}{m(v)+(i-1)} \\ &= \frac{1}{2m+\delta}\sum_{i=1}^{m}\frac{\chi_{(v,i)}}{v+(i-1)/m} \approx \frac{1}{(2m+\delta)v}\sum_{i=1}^{m}\chi_{(v,i)} \stackrel{d}{=} \frac{\hat{\chi}_v}{(2m+\delta)v},\end{aligned} \tag{4.9.7}$$

where $\hat{\chi}_v$ has a Gamma distribution with parameters $\sum_{i=1}^{m} 1+\delta/m = m+\delta$ and 1, since the random variables $\chi_{(v,1)},\ldots,\chi_{(v,m)}$ are independent.

The term $E_m(v)$ in (4.9.6) is given by the terms in the product in (4.9.4) where we have at least the product of 2 (not necessarily distinct) $\psi_{(v,i_1)}\psi_{(v,i_2)}$. Using the same arguments as in (4.9.7), we can easily show that $m(v) = \mathcal{O}(1/v^2)$. □

5 Subgraphs in PAMs

Content and structure of the chapter

In this chapter, we investigate the number of subgraphs in PAMs. Due to the dynamic nature of PAMs, we focus our attention on the expected number of occurrences of *ordered* subgraphs, i.e., subgraphs that can be constructed by adding vertices sequentially, as in PAMs. We associate to every ordered subgraph H an optimization problem, that allows us to find the most likely configuration H, thus identifying the *ages* of the vertices that create H. As a consequence, we are ble to obtain the scaling of the expected number of occurrences of H as a function of the size of the graph. This approach has limitations, since we are not able to compute precise constants. To show why this is the case, we investigate triangles further, identifying the precise constant.

The chapter is structured as follows: In Section 5.1 we introduce the optimization problem and we state the main results of the chapter. In Section 5.2 we prove a key lemma, based on Pólya urns as in Chapter 4, on the probability of a finite set of edges being present in the graph. Section 5.3 contains the proof of Theorem 5.1.2, the main result on scaling. In Section 5.4 we prove the main result on triangles (Theorem 5.1.5), highlighting the main technical difficulties that we encounter while studying more complex subgraphs. The content of this chapter is based on a joint work with Stegehuis [74].

5.1. Main results

In this section, we present our results on the number of directed subgraphs in the preferential attachment model. In this chapter we consider model (e) as defined

in Definition 4.3.1.

As mentioned in Section 1.5, we count occurrences of *labeled subgraphs* (recall Figure 1.11). We first define subgraphs in more detail. Let $H = (V_H, E_H)$ be a connected, directed graph. Let $\pi : V_H \mapsto 1, \dots, |V_H|$ be a one-to-one mapping of the vertices of H to $1, \dots, |V_H|$. In the PAM, vertices arrive one by one. We let π correspond to the order in which the vertices in H have appeared in the PAM, that is $\pi(i) < \pi(j)$ when vertex i was created before vertex j. Thus, the pair (H, π) is a directed graph, together with a prescription of the order in which the vertices of H have arrived. We call the pair (H, π) an ordered subgraph.

In PAMs, it is only possible for an older vertex to connect to a younger vertex but not the other way around. This puts constraints on the types of subgraphs that can be formed. We call the ordered subgraphs that can be formed in the PAM attainable. The following definition describes all attainable subgraphs:

Definition 5.1.1 (Attainable subgraphs)**.** *Let (H, π) be an ordered subgraph with adjacency matrix $A_\pi(H)$, where the rows and columns of the adjacency matrix are permuted by π. We say that (H, π) is attainable if $A_\pi(H)$ defines a* directed acyclic graph*, where all out-degrees are less than or equal to m.*

We now investigate how many of these subgraphs are typically present in the PAM. As we have discussed in Section 1.5, to an ordered subgraph (H, π), we associate the optimization problem

$$B(H,\pi) = \max_{s=0,1,\dots,k} -s + \sum_{i=s+1}^{k} \left[\frac{\tau-2}{\tau-1} (d_H^{(\text{in})}(\pi^{-1}(i)) - d_H^{(\text{out})}(\pi^{-1}(i))) - d_H^{(\text{in})}(\pi^{-1}(i)) \right]$$
$$:= \max_{s=0,1,\dots,k} -s + \sum_{i=s+1}^{k} \beta(\pi^{-1}(i)), \tag{5.1.1}$$

where $d_H^{(\text{out})}$ and $d_H^{(\text{in})}$ denote the in- and the out-degree in the subgraph H respectively. Let $N_t(H, \pi)$ denote the number of times the connected graph H with ordering π occurs as a subgraph of a PAM of size t. The following theorem studies the scaling of the expected number of directed subgraphs in the PAM, and relates it to the optimization problem (5.1.1):

Theorem 5.1.2. *Let H be a directed subgraph on k vertices with ordering π such that (H, π) is attainable and such that there are r different optimizers to* (5.1.1)*. Then, there exist $0 < C_1 \leq C_2 < \infty$ such that*

$$C_1 \leq \lim_{t\to\infty} \frac{\mathbb{E}\left[N_t(H,\pi)\right]}{t^{k+B(H,\pi)} \log^{r-1}(t)} \leq C_2. \tag{5.1.2}$$

Theorem 5.1.2 gives the asymptotic scaling of the number of subgraphs where the order in which the vertices appeared in the PAM is known. The total number $N_t(H)$ of copies of H for any ordering can then easily be obtained from Theorem 5.1.2:

Corollary 5.1.3. *Let H be a directed subgraph on k vertices with $\Pi \neq \varnothing$ the set of orderings π such that (H, π) is attainable. Let*

$$B(H) = \max_{\pi \in \Pi} B(H, \pi), \tag{5.1.3}$$

and let r^* be the largest number of different optimizers to (5.1.1) *among all $\pi \in \Pi$ that maximize* (5.1.3)*. Then, there exist $0 < C_1 \leq C_2 < \infty$ such that*

$$C_1 \leq \lim_{t \to \infty} \frac{\mathbb{E}\left[N_t(H)\right]}{t^{k+B(H)} \log^{r^*-1}(t)} \leq C_2. \tag{5.1.4}$$

Note that from Corollary 5.1.3 it is also possible to obtain the undirected number of subgraphs in a PAM, by summing the number of all possible directed subgraphs that create some undirected subgraph when the directions of the edges are removed.

Advantages of the optimization problem. We already gave an heuristic interpretation of the optimization problem in (5.1.1) in Section 1.5. Briefly, we have that that the probability that an attainable subgraph is present on vertices with indices $u_1 < u_2 < \cdots < u_k$ scales as

$$\prod_{i \in [k]} u_i^{\beta(i)}, \tag{5.1.5}$$

with $\beta(i)$ as in (5.1.1). The coefficients $\beta(1), \ldots, \beta(k)$ depend on τ and the structure if H. Looking at the order of magnitude of the indices of the vertices, namely $u_i \propto t^{\alpha(i)}$ for some $\alpha(i) \in [0, 1]$, the optimization problem in (5.1.1) is linear in $\alpha(1), \ldots, \alpha(k)$. In particular we identify the most likely configuration of ages (and equivalently, of degrees) of the realization of (H, π). When the optimum is not unique, several maximizers contribute equally to the number of subgraphs. In other words, when the optimizer is not unique, some if the indices are free to vary from $\mathcal{O}(1)$ to $\mathcal{O}(t)$, thus introducing the extra logarithmic factors in (5.1.2).

Fluctuations of the number of subgraphs. In Theorem 5.1.2 we investigate the expected number of subgraphs, which explains the average number of subgraphs over many PAM realizations. Another interesting question is what the distribution of the number of subgraphs in a PAM realization behaves like. In this chapter, we mainly focus on the expected value of the number of subgraphs, but here we argue that the limiting distribution of the rescaled number of subgraphs may be quite different for different subgraphs.

In Chapter 4 we discussed a Pólya urn interpretation of PAMs. In particular, we have used model (e) as example of PAM that can be directly interpreted in terms of Pólya urn experiment with multiple urns. In this section, we make use of the sequence of random independent Beta random variables $(\psi_v)_{v \in [t]}$ associated to the vertices of the PAM, as in Definition 4.3.1. Once we condition on $\psi_1, \ldots, \psi_t$, the edge statuses of the graph are independent of each other. Furthermore, the degree of a vertex v depends on the index v and ψ_v. The higher ψ_v is, the higher $D_v(t)$ is. Thus,

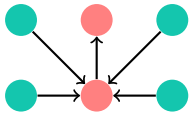


FIGURE 5.1: The order of magnitude of this subgraph containing two merged copies of the subgraph of Figure 1.13q is $t^{\frac{4}{\tau-1}}$, so that the condition in Proposition 5.1.4 is not satisfied for the subgraph in Figure 1.13q.

in this chapter, we can interpret ψ_v as a *hidden weight* associated to the vertex v.

Using this representation of the PAM we can view the PAM as a random graph model with two sources of randomness: the randomness of the ψ-variables, and then the randomness of the conditionally independent edge statuses determined by the ψ-variables. Therefore, we can define two levels of concentration for the number of ordered subgraphs $N_t(H,\pi)$. Denote $\mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)] := \mathbb{E}[N_t(H,\pi) \mid \psi_1,\dots,\psi_t]$. Furthermore, let $N_{\boldsymbol{\psi}_t}(H,\pi)$ denote the number of ordered subgraphs conditionally on $\boldsymbol{\psi}_t = \psi_1,\dots,\psi_t$. Then, the ordered subgraph (H,π) can be in the following three classes of subgraphs:

- ▷ *Concentrated*: $N_{\boldsymbol{\psi}_t}(H,\pi)$ is concentrated around its conditional expectation, i.e., as $t\to\infty$,
$$\frac{N_{\boldsymbol{\psi}_t}(H,\pi)}{\mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)]} \xrightarrow{\mathbb{P}} 1, \tag{5.1.6}$$
and as $t\to\infty$,
$$\frac{\mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)]}{\mathbb{E}[N_t(H,\pi)]} \xrightarrow{\mathbb{P}} 1. \tag{5.1.7}$$
- ▷ *Conditionally concentrated:* condition (5.1.6) holds, and as $t\to\infty$
$$\frac{N_t(H,\pi)}{\mathbb{E}[N_t(H,\pi)]} \xrightarrow{d} X \tag{5.1.8}$$
for some random variable X.
- ▷ *Non-concentrated:* condition (5.1.6) does not hold.

For example, it is easy to see that the number of subgraphs as shown in Figure 1.12d satisfies $N(H)/t \xrightarrow{\mathbb{P}} m(m-1)/2$, so that it is a subgraph that belongs to the class of concentrated subgraphs. Below we argue that the triangle belongs to the class of conditionally concentrated subgraphs. We now give a criterion for the conditional convergence in (5.1.6) in the following proposition:

Proposition 5.1.4 (Criterion for conditional convergence)**.** *Consider a subgraph (H,π) such that $\mathbb{E}[N_t(H,\pi)]\to\infty$ as $t\to\infty$. Denote by $\hat{\mathcal{H}}$ the set of all possible subgraphs*

composed by two distinct copies of (H, π) *with at least one edge in common. Then, as* $t \to \infty$,

$$\sum_{\hat{H} \in \hat{\mathcal{H}}} \mathbb{E}[N_t(\hat{H})] = o\Big(\mathbb{E}[N_t(H,\pi)]^2\Big) \qquad \Longrightarrow \qquad \frac{N_{\psi_t}(H,\pi)}{\mathbb{E}_{\psi_t}[N_t(H,\pi)]} \xrightarrow{\mathbb{P}} 1. \tag{5.1.9}$$

Proposition 5.1.4 gives a simple criterion for conditional convergence for a subgraph (H, π), and it is proved in Section 5.5. The condition in (5.1.9) is simple to evaluate in practice. We denote the subgraphs consisting of two overlapping copies of (H, π) sharing at least one edge by $\hat{H}_1, \dots, \hat{H}_r$. To identify the order of magnitude of $\mathbb{E}[\hat{H}_i]$, we apply Corollary 5.1.3 to $\hat{H}_i$ or, in other words, we apply Theorem 5.1.2 to all possible orderings $\hat{\pi}$ of $\hat{H}_i$. Once we have all orders of magnitude of $(\hat{H}_i, \hat{\pi})$ for all orderings $\hat{\pi}$, and for all $\hat{H}_i$, it is immediate to check whether hypothesis of Proposition 5.1.4 is satisfied.

There are subgraphs where the condition in Proposition 5.1.4 does not hold. For example, merging two copies of the subgraph of Figure 1.13q as in Figure 5.1 violates the condition in Proposition 5.1.4. We show in Section 5.5 that this subgraph is in the class of non-concentrated subgraphs.

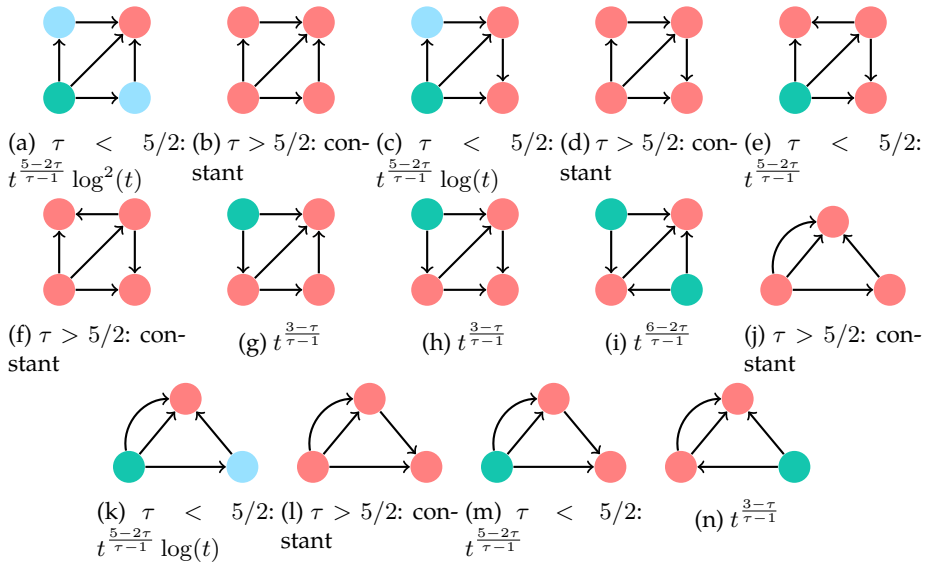


FIGURE 5.2: Order of magnitude of $N_t(\hat{H})$ for all merged triangles on 3 and 4 vertices and for $\tau \in (2, 3)$. Vertices with degree proportional to a constant are light green, vertices with free degrees are light blue, and vertices of degree proportional to $t^{1/(\tau-1)}$ are red.

5.1.1. Exact constants: triangles

Theorem 5.1.2 allows us to identify the order of magnitude of the expected number of subgraphs in the PAM. In particular, for a subgraph H with ordering π, it assures the existence of two constants $0 < C_1 \leq C_2 < \infty$ as in (5.1.2). A more detailed analysis is necessary to prove a stronger result than Theorem 5.1.2 of the type

$$\lim_{t\to\infty} \frac{\mathbb{E}\left[N_t(H,\pi)\right]}{t^{k+B(H,\pi)}\log^{r-1}(t)} = C,$$

for some constant $0 < C < \infty$. In other words, given an ordered subgraph (H,π), we want to identify the constant $C > 0$ such that

$$\mathbb{E}\left[N_t(H,\pi)\right] = Ct^{k+B(H,\pi)}\log^{r-1}(t)(1+o(1)). \tag{5.1.10}$$

We prove (5.1.10) for triangles to show the difficulties in the evaluation of the precise constant C for general subgraphs. The following theorem provides the detailed scaling of the expected number of triangles:

Theorem 5.1.5 (Phase transition for the number of triangles)**.** *Let $m \geq 2$ and $\delta > -m$ be parameters for $(\mathrm{PA}_t)_{t\geq 1}$. Denote the number of labeled triangles in PA_t by $\triangle_t$. Then, as $t \to \infty$,*

(1) if $\tau > 3$, then

$$\mathbb{E}[\triangle_t] = \frac{m^2(m-1)(m+\delta)(m+\delta+1)}{\delta^2(2m+\delta)}\log(t)(1+o(1));$$

(2) if $\tau = 3$, then

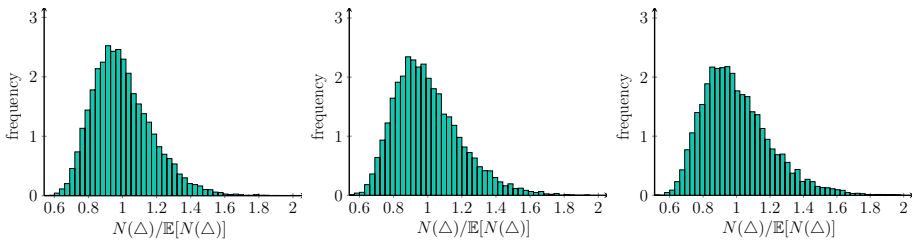
$$\mathbb{E}[\triangle_t] = \frac{m(m-1)(m+1)}{48}\log^3(t)(1+o(1));$$

(3) if $\tau \in (2,3)$, then

$$\mathbb{E}[\triangle_t] = \frac{m^2(m-1)(m+\delta)(m+\delta+1)}{\delta^2(2m+\delta)}t^{(3-\tau)/(\tau-1)}\log(t)(1+o(1)).$$

Theorem 5.1.5 in the case $\delta = 0$ coincides with [31, Theorem 14]. For $\delta > 0$ we retrieve the result in [61, Proposition 4.3], noticing that the additive constant β in the attachment probabilities in the Móri model considered in [61] coincides with (4.3.1) for $\beta = \delta/m$.

The proof of Theorem 5.1.5 in Section 5.4 shows that to identify the constant in (5.1.10) we need to evaluate the precise expectations involving the attachment probabilities of edges. The equivalent formulation of the PAM given in Section 4.4.3 simplifies the calculations, but it is still necessary to evaluate rather complicated expectations involving products of several terms as in (4.7.1). For a more detailed discussion, we refer to Remark 5.4.1.



(a) $t = 10^5$ (b) $t = 10^6$ (c) $t = 5 \cdot 10^6$

FIGURE 5.3: Density approximation of the number of triangles in 10^4 realizations of the preferential attachment model with $\tau = 5/2$ and various values of t.

The distribution of the number of triangles. Theorem 5.1.5 shows the behavior of the expected number of triangles. The distribution of the number of triangles across various PAM realizations is another object of interest. We prove the following result for the number of triangles $\triangle_t$:

Corollary 5.1.6 (Conditional concentration of triangles)**.** *For $\tau \in (2, 3)$, the number of triangles $\triangle_t$ is conditionally concentrated in the sense of* (5.1.6)*.*

Corollary 5.1.6 is a direct consequence of Proposition 5.1.4, and the atlas of the order of magnitudes of all possible realizations of the subgraphs consisting of two triangles sharing one or two edges, presented in Figure 5.2. Figure 5.3 shows a density approximation of the number of triangles obtained by simulations. These figures suggest that the rescaled number of triangles converges to a random limit, since the width of the density plots does not decrease in t. Thus, while the number of triangles concentrates conditionally, it does not seem to converge to a constant when taking the random ψ-variables into account. This would put the triangle subgraph in the class of conditionally concentrated subgraphs. Proving this and identifying the limiting random variable of the number of triangles is an interesting open question.

5.2. The probability of a subgraph being present

In this section, we prove the main ingredient for the proof of Theorem 5.1.2, the probability of a subgraph being present on a given set of vertices. The most difficult part of evaluating the probability of a subgraph H being present in PA_t is that the PAM is constructed recursively.

We consider triangles as an example. We write the event of a *labeled* triangle being present by $\{u \overset{j_1}{\leftarrow} v,\ u \overset{j_2}{\leftarrow} w,\ v \overset{j_3}{\leftarrow} w\}$, where $\{u \overset{j}{\leftarrow} v\}$ denotes the event that the jth edge of vertex v is attached to vertex u. Notice that in this way we express precisely

which edges we consider in the triangle construction. Then,

$$\begin{aligned}\mathbb{P}\Big(u \overset{j_1}{\leftarrow} v,\ u \overset{j_2}{\leftarrow} w,\ v \overset{j_3}{\leftarrow} w\Big) \\ &= \mathbb{E}\Big[\mathbb{P}\Big(u \overset{j_1}{\leftarrow} v,\ u \overset{j_2}{\leftarrow} w,\ v \overset{j_3}{\leftarrow} w \mid \mathrm{PA}_{t-1,j_3-1}\Big)\Big] \\ &= \mathbb{E}\Big[\mathbb{1}\{u \overset{j_1}{\leftarrow} v,\ u \overset{j_2}{\leftarrow} w\}\frac{D_v(w-1,j_3-1)+\delta}{2m(w-2)+(j_3-1)+(w-1)\delta}\Big].\end{aligned} \tag{5.2.1}$$

In (5.2.1), the indicator function $\mathbb{1}\{u \overset{j_1}{\leftarrow} v,\ u \overset{j_2}{\leftarrow} w\}$ and $D_v(w-1,j_3-1)$ are not independent, therefore evaluating the expectation on the right-hand side of (5.2.1) is not easy. A possible solution for the evaluation of the expectation in (5.2.1) is to rescale $D_v(w-1,j_3-1)$ with an appropriate constant to obtain a martingale, and then recursively use the conditional expectation. For a detailed explanation of this, we refer to [33, 158] and [85, Section 8.3]. This method is hardly tractable due to the complexity of the constants appearing (see Remark 5.4.1 for a more detailed explanation).

In Chapter 3 we have introduced the notion of *factorizable events* (see Definition 3.4.4). This approach is hardly tractable as the martingale approach just mentioned, since in this case we have to keep track of the evolution of the degree of *all vertices in the subgraph H at every time*.

We use a different approach to evaluate of the expectation in (5.2.1) using the interpretation of the PAM as a Pólya urn graph, focusing mainly on the *the age (the indices) of the vertices*, and not on precise constants. We give a lower and upper bound on the probability of having a finite number of edges present in the graph, as formulated in the following lemma:

Lemma 5.2.1 (Probability of finite set of labeled edges)**.** *Fix $\ell \in \mathbb{N}$. For vertices $\boldsymbol{u}_\ell = (u_1,\dots,u_\ell) \in [t]^\ell$ and $\boldsymbol{v}_\ell = (v_1,\dots,v_\ell) \in [t]^\ell$ and edge labels $\boldsymbol{j}_\ell = (j_1,\dots,j_\ell) \in [m]^\ell$, consider the corresponding finite set of ℓ distinct labeled edges $M_\ell(\boldsymbol{u}_\ell,\boldsymbol{v}_\ell,\boldsymbol{j}_\ell)$. Assume that the subgraph defined by the set $M_\ell(\boldsymbol{u}_\ell,\boldsymbol{v}_\ell,\boldsymbol{j}_\ell)$ is attainable in the sense of Definition 5.1.1. Define $\chi = (m+\delta)/(2m+\delta)$. Then the following holds:*

(1) There exist two constants $c_1, c_2 > 0$ such that, for t large enough,

$$c_1 \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi} \le \mathbb{P}\Big(M_\ell(\boldsymbol{u}_\ell,\boldsymbol{v}_\ell,\boldsymbol{j}_\ell) \subseteq E(\mathrm{PA}_t)\Big) \le c_2 \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi}. \tag{5.2.2}$$

(2) Define the set

$$J(\boldsymbol{u}_\ell,\boldsymbol{v}_\ell) = \big\{\boldsymbol{j}_\ell \in [m]^\ell \colon M_\ell(\boldsymbol{u}_\ell,\boldsymbol{v}_\ell,\boldsymbol{j}_\ell) \subseteq E(\mathrm{PA}_t)\big\}. \tag{5.2.3}$$

Then, there exist two constants $\hat{c}_1, \hat{c}_2 > 0$ such that, for t large enough,

$$\hat{c}_1 \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi} \leq \mathbb{E}[|J(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell)|] \leq \hat{c}_2 \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi}. \tag{5.2.4}$$

Formula (5.2.2) in the above lemma bounds the probability that a subgraph is present on vertices $\boldsymbol{u}_\ell$ and $\boldsymbol{v}_\ell$ such that the j_ith edge from u_i connects to v_i. Notice that (5.2.2) is independent of the precise edge labels $(j_1, \ldots, j_\ell)$. To be able to count all subgraphs, and not only subgraphs where the edge labels have been specified, (5.2.4) bounds the expected number of times a specific subgraph is present on vertices $\boldsymbol{u}_\ell$ and $\boldsymbol{v}_\ell$. This number is given exactly by the elements in the set $J(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell)$ as in (5.2.3). Note that the expectation in (5.2.4) may be larger than one, due to the fact that the PAM is a multigraph.

Lemma 5.2.1 gives a bound on the probability of the presence of $\ell \in \mathbb{N}$ distinct edges in the graph as a function of the indices $(u_1, v_1), \ldots, (u_\ell, v_\ell)$ of the endpoints of the ℓ edges. Due to the properties of PAM, the index of a vertex gives an indication of its degree, due to the old-get-richer effect. Lemma 5.2.1 is a stronger result than [59, Corollary 2.3], which gives an upper bound of the form in (5.2.2) only for self-avoiding paths.

The proof of Lemma 5.2.1 is based on the interpretation of the PAM in Definition 4.3.1 as an urn experiment as discussed in Chapter 4. We refer to Section 4.4 for a detailed discussion of Pólya urn models and the definition of the Pólya urn graph (Definition 4.4.2).

Proof of Lemma 5.2.1. We start with the proof of (5.2.2). Fix $\boldsymbol{u}_\ell, \boldsymbol{v}_\ell, \boldsymbol{j}_\ell$. In the proof, we denote $M_\ell(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell, \boldsymbol{j}_\ell)$ simply by M_ℓ to keep notation light. We use Theorem 4.4.3, that shows that the Pólya urn graph PU_t and PA_t have the same distribution and evaluate $\mathbb{P}(M_\ell \subseteq E(\mathrm{PU}_t))$. We consider ℓ distinct labeled edges, so we can use (4.7.1) to write

$$\mathbb{P}(M_\ell \subseteq E(\mathrm{PU}_t) \mid \psi_1, \ldots, \psi_t) = \prod_{l=1}^{\ell} \psi_{u_l} \frac{S_{u_l}}{S_{v_l - 1}}. \tag{5.2.5}$$

Now fix $\varepsilon > 0$. Define $\mathcal{E}_t^\varepsilon := \{\max_{i \in [t]} \left|S_i - \left(\frac{i}{t}\right)^\chi\right| \leq \varepsilon\}$. By (4.7.5), and the fact that the product of the random variables in (5.2.5) is bounded by 1,

$$\mathbb{E}\left[\prod_{l=1}^{\ell} \psi_{u_l} \frac{S_{u_l}}{S_{v_l-1}}\right] = \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_t^\varepsilon} \prod_{l=1}^{\ell} \psi_{u_l} \frac{S_{u_l}}{S_{v_l-1}}\right] + o(1). \tag{5.2.6}$$

On the event $\mathcal{E}_t^\varepsilon$, we have, for every $l \in [\ell]$,

$$(1-\varepsilon)\left(\frac{u_l}{v_l}\right)^\chi \leq \frac{S_{u_l}}{S_{v_l-1}} \leq (1+\varepsilon)\left(\frac{u_l}{v_l}\right)^\chi, \tag{5.2.7}$$

where in (5.2.7) we have replaced $v_l - 1$ with v_l with a negligible error. Notice that

since v_l is always the source of the edge, this implies $v_l \geq 2$, therefore this is allowed. Using (5.2.7) in (5.2.6) we obtain

$$(1-\varepsilon)^{\ell} \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l}\right)^{\chi} \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_t^{\varepsilon}} \prod_{l=1}^{\ell} \psi_{u_l}\right] \leq \mathbb{P}\left(M_\ell \subseteq E(\mathrm{PU}_t)\right) \leq (1+\varepsilon)^{\ell} \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l}\right)^{\chi} \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_t^{\varepsilon}} \prod_{l=1}^{\ell} \psi_{u_l}\right]. \tag{5.2.8}$$

Even though $\psi_1 \ldots, \psi_t$ depend on $\mathcal{E}_t^{\varepsilon}$, we can ignore $\mathbb{1}_{\mathcal{E}_t^{\varepsilon}}$. In fact, since the random variables $\psi_{u_1}, \ldots, \psi_{u_\ell}$ are bounded by 1, we can write

$$\left|\mathbb{E}\left[\prod_{l=1}^{\ell} \psi_{u_l}\right] - \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_t^{\varepsilon}} \prod_{l=1}^{\ell} \psi_{u_l}\right]\right| = \mathbb{E}\left[\mathbb{1}_{(\mathcal{E}_t^{\varepsilon})^c} \prod_{l=1}^{\ell} \psi_{u_l}\right] \leq 1 - \mathbb{P}(\mathcal{E}_t^{\varepsilon}) = o(1), \tag{5.2.9}$$

since $\mathbb{P}(\mathcal{E}_t^{\varepsilon}) = 1 - o(1)$ by Lemma 4.7.1. Notice that the bound in (5.2.9) depends on ε through the event $\mathcal{E}_t^{\varepsilon}$, but not on the choice of M_ℓ. As a consequence, for some constant c_1, c_2 and t large enough, from (5.2.8) and (5.2.9) we obtain that

$$c_1 \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l}\right)^{\chi} \mathbb{E}\left[\prod_{l=1}^{\ell} \psi_{u_l}\right] \leq \mathbb{P}\left(M_\ell \subseteq E(\mathrm{PU}_t)\right) \leq c_2 \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l}\right)^{\chi} \mathbb{E}\left[\prod_{l=1}^{\ell} \psi_{u_l}\right]. \tag{5.2.10}$$

What remains is to evaluate the expectation in (5.2.10). We have assumed to have ℓ distinct edges, however that does not imply that the vertices $u_1, v_1, \ldots, u_\ell, v_\ell$ are distinct. The expectation in (5.2.10) depends only on the receiving vertices of the ℓ edges, namely $u_1, \ldots, u_\ell$.

Let $\bar{u}_1, \ldots, \bar{u}_k$ denote the $k \leq \ell$ distinct elements that appear among $u_1, \ldots, u_\ell$. For $h \in [k]$, the vertex $\bar{u}_h$ appears in the product inside the expectation in (5.2.10) with multiplicity $d_h^{(\mathrm{in})}$, which is the degree of vertex $\bar{u}_k$ in the subgraph defined by M_ℓ. As a consequence, we can write

$$\mathbb{E}\left[\prod_{l=1}^{\ell} \psi_{u_l}\right] = \mathbb{E}\left[\prod_{h=1}^{k} \psi_{\bar{u}_h}^{d_h^{(\mathrm{in})}}\right] = \prod_{h=1}^{k} \mathbb{E}\left[\psi_{\bar{u}_h}^{d_h^{(\mathrm{in})}}\right], \tag{5.2.11}$$

where in (5.2.11) we have used the fact that $\psi_1, \ldots, \psi_t$ are all independent. Notice that $\mathbb{E}[\psi_1^d] = 1$ for all $d \geq 0$, since $\psi_1 \equiv 1$. Therefore, if $\bar{u}_h = 1$ for some $h \in [k]$, $\mathbb{E}[\psi_{\bar{u}_h}^d] = 1$ and the terms depending on the first vertex contribute to the expectation in (5.2.11) by a constant.

For the terms where $\bar{u}_h \geq 2$, recall that, if $X(\alpha, \beta)$ is a Beta random variable, then, for any integer $d \in \mathbb{N}$,

$$\mathbb{E}[X(\alpha,\beta)^d] = \frac{\alpha(\alpha+1)\cdots(\alpha+d-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+d-1)}.$$

Since $\psi_{\bar{u}_h}$ is Beta distributed with parameters $m+\delta$ and $2(\bar{u}_h-3)+(\bar{u}_h-1)\delta$,

$$\begin{aligned}\mathbb{E}\left[\psi_{\bar{u}_h}^{d_h^{(\mathrm{in})}}\right] &= \frac{(m+\delta)\cdots(m+\delta+d_h^{(\mathrm{in})}-1)}{[m(2\bar{u}_h-2)+\bar{u}_h\delta]\cdots[m(2\bar{u}_h-2)+\bar{u}_h\delta+d_h^{(\mathrm{in})}-1]} \\ &= \bar{u}_h^{-d_h^{(\mathrm{in})}}\frac{(m+\delta)\cdots(m+\delta+d_h^{(\mathrm{in})}-1)}{[2m+\delta-(2m)/\bar{u}_h]\cdots[2m+\delta+(d_h^{(\mathrm{in})}-1-2m)/\bar{u}_h]}.\end{aligned} \tag{5.2.12}$$

Notice that if $\bar{u}_h \geq 2$, uniformly in t and the precise choice of the ℓ edges,

$$(m+\delta)^{-\ell} \leq \left([2m+\delta-(2m)/\bar{u}]\cdots[2m+\delta+(d_h^{(\mathrm{in})}-1-2m)/\bar{u}]\right)^{-1} \leq (2m+\delta+\ell)^{-\ell}.$$

As a consequence, we can find two different constants c_1, c_2 (for simplicity we keep the same name) such that

$$c_1\prod_{h=1}^{k}\bar{u}_h^{-d_h^{(\mathrm{in})}} \leq \prod_{h=1}^{k}\mathbb{E}\left[\psi_{\bar{u}_h}^{d_h^{(\mathrm{in})}}\right] \leq c_2\prod_{h=1}^{k}\bar{u}_h^{-d_h^{(\mathrm{in})}}. \tag{5.2.13}$$

We now use (5.2.13) in (5.2.10) to obtain

$$c_1\prod_{l=1}^{\ell}\left(\frac{u_l}{v_l}\right)^{\chi}\prod_{h=1}^{k}\bar{u}_h^{-d_h^{(\mathrm{in})}} \leq \mathbb{P}\left(M_\ell \subseteq E(\mathrm{PU}_t)\right) \leq c_2\prod_{l=1}^{\ell}\left(\frac{u_l}{v_l}\right)^{\chi}\prod_{h=1}^{k}\bar{u}_h^{-d_h^{(\mathrm{in})}}. \tag{5.2.14}$$

Since $d_h^{(\mathrm{in})}$ is the multiplicity of vertex $\bar{u}_h$ as receiving vertex, we can write

$$\prod_{h=1}^{k}\bar{u}_h^{-d_h^{(\mathrm{in})}} = \prod_{l=1}^{\ell}u_l^{-1}.$$

Combining this with (5.2.14) completes the proof of (5.2.2).

The proof of (5.2.4) follows immediately from (5.2.2) and the definition of the set $J(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell)$ in (5.2.3). In fact, we can write

$$\mathbb{E}[|J(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell)|] = \sum_{\boldsymbol{j}_\ell \in [m]^\ell}\mathbb{P}\left(M_\ell(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell, \boldsymbol{j}_\ell) \subseteq E(\mathrm{PA}_t)\right).$$

Recall that $\mathbb{P}\left(M_\ell(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell, \boldsymbol{j}_\ell) \subseteq E(\mathrm{PA}_t)\right)$ is *independent of the labels* $\boldsymbol{j}_\ell$. For a fixed set of source and target vertices $\boldsymbol{u}_\ell$ and $\boldsymbol{v}_\ell$, there is only a finite combination of labels $\boldsymbol{j}_\ell$ such that the subgraph defined by $M_\ell(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell, \boldsymbol{j}_\ell)$ is attainable in the sense of Definition 5.1.1. In fact, the number of such labels $\boldsymbol{j}_\ell$ is larger than one (since the corresponding subgraph is attainable), and less than m^ℓ (the total number of elements of $[m]^\ell$). As a consequence, taking $\hat{c}_1 = c_1$ and $\hat{c}_2 = c_2 m^\ell$ proves (5.2.4). □

5.3. Scaling of expectation: proof of Theorem 5.1.2

To prove Theorem 5.1.2, we write the expected number of subgraphs as multiple integrals. W.l.o.g. we assume throughout this section that π is the identity permutation, so that the vertices of H are labeled as $1, \dots, k$, and therefore drop the dependence of the quantities on π. We first prove a lemma that states that two integrals that will be important in proving Theorem 5.1.2 are finite:

Lemma 5.3.1. *Let H be a subgraph such that the optimum of* (5.1.1) *is attained by $s_1, \dots, s_r$. Then,*

$$A_1(H) := \int_1^\infty u_1^{\beta(1)} \int_{u_1}^\infty u_2^{\beta(2)} \cdots \int_{u_{s-1}}^\infty u_{s_1}^{\beta(s_1)} \mathrm{d}u_{s_1} \cdots \mathrm{d}u_1 < \infty, \tag{5.3.1}$$

$$A_2(H) := \int_0^1 u_k^{\beta(k)} \int_0^{u_k} u_{k-1}^{\beta(k-1)} \cdots \int_0^{u_{s_r+1}} u_{s_r+1}^{\beta(s_r+1)} \mathrm{d}u_{s_r+1} \cdots \mathrm{d}u_k < \infty. \tag{5.3.2}$$

Proof. The first integral is finite as long as

$$z + \sum_{i=s_1-z}^{s_1} \beta(i) < 0 \tag{5.3.3}$$

for all $z \in [s_1]$. Suppose that (5.3.3) does not hold for some $z^* \in [s_1]$. Then, the difference between the contribution to (5.1.1) for $\tilde{s} = s_1 - z^*$ and s_1 is

$$-(s_1 - z^*) + \sum_{i=s_1-z^*}^{k} \beta(i) + s_1 - \sum_{i=s_1}^{k} \beta(i) = z^* + \sum_{i=s_1-z^*}^{s_1} \beta(i) \geq 0,$$

which would imply that $s_1 - z^*$ is also an optimizer of (5.1.1), which is in contradiction with s_1 being the smallest optimum. Thus, (5.3.3) holds for all $r \in [s]$ and $A_1(H) < \infty$.

The second integral is finite as long as

$$z - s_r + \sum_{i=s_r+1}^{z} \beta(i) > 0$$

for all $z \in \{s_r+1, \dots, k\}$. Suppose that this does not hold for some $z^* \in \{s_r+1, \dots, k\}$. Set $\tilde{s} = z^* > s_r$. Then, the difference between the contribution to (5.1.1) for $\tilde{s} = z^*$ and s_r is

$$-z^* + s_r - \sum_{i=s_r+1}^{z^*} \beta(i) \geq 0,$$

which is a contradiction with s_r being the largest optimizer. Therefore, $A_2(H) < \infty$, and the proof is complete. □

We now use this lemma to prove Theorem 5.1.2:

Proof of Theorem 5.1.2. Again, we assume that π is the identity mapping, so that we may drop all dependencies on π. Suppose that the optimal solution to (5.1.1) is attained by $s_1, s_2, \ldots, s_r$ for some $r \geq 1$. Let the ℓ edges of H be denoted by (u_l, v_l) for $l \in [\ell]$. Let $N_t(H, i_1, \ldots, i_k)$ denote the number of times subgraph H is present on vertices $i_1, \ldots, i_k$. We then use Lemma 5.2.1, which proves that, for some $0 < C < \infty$,

$$\begin{aligned}\mathbb{E}\left[N_t(H)\right] &= \sum_{i_1<\cdots<i_k\in[t]} \mathbb{E}\left[N_t(H, i_1, \ldots, i_k)\right] \\ &\leq C \sum_{i_1<\cdots<i_k\in[t]} \prod_{l=1}^{\ell} i_{u_l}^{\chi-1} i_{v_l}^{-\chi} = C \sum_{i_1<\cdots<i_k\in[t]} \prod_{q=1}^{k} i_q^{\beta(q)}.\end{aligned} \tag{5.3.4}$$

We then bound the sums by integrals as

$$\begin{aligned}\mathbb{E}\left[N_t(H)\right] &\leq \tilde{C} \int_1^t u_1^{\beta(1)} \cdots \int_{u_{k-1}}^t u_k^{\beta(k)} \mathrm{d}u_k \cdots \mathrm{d}u_1 \\ &\leq \tilde{C} \int_1^\infty u_1^{\beta(1)} \cdots \int_{u_s-1}^\infty u_{s_1}^{\beta(s_1)} \mathrm{d}u_{s_1} \ldots \mathrm{d}u_1 \\ &\quad \times \int_1^t u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^\infty u_{s_1+2}^{\beta(s_1+2)} \cdots \int_{u_{s_2-1}}^t u_{s_2}^{\beta(s_2)} \mathrm{d}u_{s_2} \ldots \mathrm{d}u_{s_1+1} \\ &\quad \times \int_1^t u_{s_2+1}^{\beta(s_2+1)} \int_{u_{s_2+1}}^\infty u_{s_2+2}^{\beta(s_2+2)} \cdots \int_{u_{s_3-1}}^t u_{s_3}^{\beta(s_3)} \mathrm{d}u_{s_3} \ldots \mathrm{d}u_{s_2+1} \times \cdots \\ &\quad \times \int_1^t u_{s_{r-1}+1}^{\beta(s_{r-1}+1)} \int_{u_{s_{r-1}+1}}^\infty u_{s_{r-1}+2}^{\beta(s_{r-1}+2)} \cdots \int_{u_{s_r-1}}^t u_{s_r}^{\beta(s_r)} \mathrm{d}u_{s_r} \ldots \mathrm{d}u_{s_{r-1}+1} \\ &\quad \times \int_0^t u_{s_r+1}^{\beta(s_r+1)} \int_{s_r+1}^t u_{s_r+2}^{\beta(s_r+2)} \cdots \int_{u_{k-1}}^t u_k^{\beta(k)} \mathrm{d}u_k \cdots \mathrm{d}u_{s_r+1},\end{aligned} \tag{5.3.5}$$

for some $0 < \tilde{C} < \infty$. The first set of integrals is finite by Lemma 5.3.1 and independent of t. For the last set of integrals, we obtain

$$\begin{aligned}&\int_0^t u_{s_r+1}^{\beta(s_r+1)} \int_{u_{s_r+1}}^t u_{s_r+2}^{\beta(s+2)} \cdots \int_{u_{k-1}}^t u_k^{\beta(k)} \mathrm{d}u_k \cdots \mathrm{d}u_{s_r+1} \\ &= t^{k-s_r+\sum_{i=s_r+1}^k \beta(i)} \int_0^1 w_{s_r+1}^{\beta(s_r+1)} \int_{w_{s_r+1}}^1 w_{s_r+2}^{\beta(s_r+2)} \cdots \int_{w_{k-1}}^1 w_k^{\beta(k)} \mathrm{d}w_k \cdots \mathrm{d}w_{s_r+1} \\ &= K t^{k+B(H)},\end{aligned} \tag{5.3.6}$$

for some $0 < K < \infty$, where we have used the change of variables $w = u/t$ and Lemma 5.3.1. For $r = 1$, this finishes the proof, because then the middle integrals in (5.3.5) are empty. We now investigate the behavior of the middle sets of integrals

for $r > 1$. Because the optimum to (5.1.1) is attained for s_1 as well as s_2,

$$-s_1 + \sum_{i=s_1+1}^{k} \beta(i) + s_2 - \sum_{i=s_2+1}^{k} \beta(i) = s_2 - s_1 + \sum_{i=s_1+1}^{s_2} \beta(i) = 0. \tag{5.3.7}$$

Therefore, when $s_2 = s_1 + 1$, the second set of integrals in (5.3.5) equals

$$\int_1^t u_{s_1}^{-1} \mathrm{d}u_{s_1} = \log(t).$$

Now suppose that $s_1 < s_2 + 1$. Then, any $\tilde{s} \in [s_1 + 1, s_2 - 1]$ is a non-optimal solution to (5.1.1), and therefore

$$-s_2 + \sum_{i=s_2+1}^{k} \beta(i) + \tilde{s} - \sum_{i=\tilde{s}+1}^{k} \beta(i) = \tilde{s} - s_2 - \sum_{i=\tilde{s}+1}^{s_2} \beta(i) > 0,$$

or

$$\sum_{i=\tilde{s}+1}^{s_2} \beta(i) < s_2 - \tilde{s}. \tag{5.3.8}$$

This implies that

$$\begin{aligned} &\int_1^t u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^{\infty} u_{s_1+2}^{\beta(s_1+2)} \cdots \int_{u_{s_2-1}}^{t} u_{s_2}^{\beta(s_2)} \mathrm{d}u_{s_2} \dots \mathrm{d}u_{s_1+1} \\ &\quad = \tilde{K} \int_1^t u_{s_1+1}^{\sum_{i=s_1+1}^{s_2} \beta(i) + s_2 - s_1 - 1} \mathrm{d}u_{s_1+1} \\ &\quad = \tilde{K} \int_1^t u_{s_1+1}^{-1} \mathrm{d}u_{s_1+1} = \tilde{K} \log(t), \end{aligned} \tag{5.3.9}$$

for some $0 < C < \infty$. A similar reasoning holds for the other integrals, so that combining (5.3.5), (5.3.6) and (5.3.9) yields

$$\lim_{t \to \infty} \frac{\mathbb{E}\left[N_t(H)\right]}{t^{k+B(H)} \log^{r-1}(t)} \leq C_2, \tag{5.3.10}$$

for some $0 < C_2 < \infty$.

We now proceed to prove a lower bound on the expected number of subgraphs. Again, by Lemma 5.2.1 and lower bounding the sums by integrals as in (5.3.4), we obtain that, for some $0 < C < \infty$

$$\mathbb{E}\left[N_t(H)\right] \geq C \int_1^t u_1^{\beta(1)} \cdots \int_{u_{k-1}}^{t} u_k^{\beta(k)} \mathrm{d}u_k \cdots \mathrm{d}u_1. \tag{5.3.11}$$

Fix $\varepsilon > 0$. We investigate the contribution where vertices $1, \dots, s_1$ have indices in $[1, 1/\varepsilon]$, vertices $s_1 + 1, \dots, s_2$ have indices in $[1/\varepsilon, \varepsilon t^{1/r}]$, vertices $s_2 + 1, \dots, s_3$ have

indices in $[t^{1/r}, \varepsilon t^{2/r}]$ and so on, and vertices $s_r+1, \ldots, s_k$ have indices in $[\varepsilon t, t]$. Thus, we bound

$$\begin{aligned}
&\mathbb{E}\left[N_t(H)\right] \\
&\geq C \int_1^{1/\varepsilon} u_1^{\beta(1)} \int_{u_1}^{1/\varepsilon} u_2^{\beta(2)} \cdots \int_{u_{s_1-1}}^{1/\varepsilon} u_{s_1}^{\beta(s)} \mathrm{d}u_{s_1} \ldots \mathrm{d}u_1 \\
&\quad \times \int_{1/\varepsilon}^{\varepsilon t^{1/r}} u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^{u_{s_1+1}/\varepsilon} u_{s_1+2}^{\beta(s_1+2)} \cdots \int_{u_{s_2-1}}^{u_{s_2-1}/\varepsilon} u_{s_2}^{\beta(s_2)} \mathrm{d}u_{s_2} \ldots \mathrm{d}u_{s_1+1} \\
&\quad \times \int_{t^{1/r}}^{\varepsilon t^{2/r}} u_{s_2+1}^{\beta(s_2+1)} \int_{u_{s_2+1}}^{u_{s_2+1}/\varepsilon} u_{s_2+2}^{\beta(s_2+2)} \cdots \int_{u_{s_3-1}}^{u_{s_3-1}/\varepsilon} u_{s_3}^{\beta(s_3)} \mathrm{d}u_{s_3} \ldots \mathrm{d}u_{s_2+1} \qquad (5.3.12) \\
&\quad \times \int_{t^{(r-2)/r}}^{\varepsilon t^{(r-1)/r}} u_{s_{r-1}+1}^{\beta(s_{r-1}+1)} \int_{u_{s_{r-1}+1}}^{u_{s_{r-1}+1}/\varepsilon} u_{s_{r-1}+2}^{\beta(s_{r-1}+2)} \cdots \int_{u_{s_r-1}}^{u_{s_r-1}/\varepsilon} u_{s_r}^{\beta(s_r)} \mathrm{d}u_{s_r} \ldots \mathrm{d}u_{s_{r-1}+1} \\
&\quad \times \int_{\varepsilon t}^{t} u_{s_r+1}^{\beta(s_r+1)} \int_{u_{s_r+1}}^{t} u_{s_r+2}^{\beta(s_r+2)} \cdots \int_{u_{k-1}}^{t} u_k^{\beta(k)} \mathrm{d}u_k \ldots \mathrm{d}u_{s_r+1}
\end{aligned}$$

The first set of integrals equals $A_1(H)$ plus terms that vanish as ε becomes small by Lemma 5.3.1. For the last set of integrals, we use the change of variables $w = u/t$ to obtain

$$\begin{aligned}
&\int_{\varepsilon t}^{t} u_{s_r+1}^{\beta(s_r+1)} \int_{u_{s_r+1}}^{t} u_{s_r+2}^{\beta(s_r+2)} \cdots \int_{u_{k-1}}^{t} u_k^{\beta(k)} \mathrm{d}u_k \ldots \mathrm{d}u_{s_r+1} \qquad (5.3.13) \\
&= t^{k-s_r+\sum_{i=s_r+1}^{k}\beta(i)} \int_{\varepsilon}^{1} w_{s_r+1}^{\beta(s_r+1)} \int_{w_{s_r+1}}^{1} w_{s_r+2}^{\beta(s_r+2)} \cdots \int_{w_{k-1}}^{1} w_k^{\beta(k)} \mathrm{d}w_k \ldots \mathrm{d}w_{s_r+1} \\
&= t^{k+B(H)}(A_2(H) - h_1(\varepsilon)),
\end{aligned}$$

for some function $h_1(\varepsilon)$. By Lemma 5.3.1 $h_1(\varepsilon)$ satisfies $\lim_{\varepsilon\to 0} h_1(\varepsilon) = 0$. Again, if $r = 1$, the middle sets of integrals in (5.3.12) are empty, so we are done.

We now investigate the second set of integrals in (5.3.12) for $r > 1$. Using the substitution $w_{s_1+1} = u_{s_1+1}$ and $w_i = u_i/u_{i-1}$ for $i > s_1 + 1$, we obtain

$$\begin{aligned}
&\int_{1/\varepsilon}^{\varepsilon t^{1/r}} u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^{u_{s_1+1}/\varepsilon} u_{s_1+2}^{\beta(s_1+2)} \cdots \int_{u_{s_2-1}}^{u_{s_2-1}/\varepsilon} u_{s_2}^{\beta(s_2)} \mathrm{d}u_{s_2} \ldots \mathrm{d}u_{s_1+1} x \\
&= \int_{1/\varepsilon}^{\varepsilon t^{1/r}} w_{s_1+1}^{s_2-s_1-1+\sum_{i=s_1+1}^{s_2}\beta(i)} \mathrm{d}w_{s_1+1} \qquad (5.3.14) \\
&\quad \times \int_1^{1/\varepsilon} w_{s_1+2}^{s_2-s_1-2+\sum_{i=s_1+2}^{s_2}\beta(i)} \mathrm{d}w_{s_2+1} \cdots \int_1^{1/\varepsilon} w_{s_2}^{\beta(s_2)} \mathrm{d}w_{s_2}.
\end{aligned}$$

The first integral equals by (5.3.7)

$$\int_{1/\varepsilon}^{\varepsilon t^{1/r}} w_{s_1+1}^{-1} \mathrm{d}w_{s_1+1} = \frac{1}{r}\log(t) + \log(\varepsilon^2).$$

The integrand in all other integrals in (5.3.14) equals $w_i^{\gamma_i}$ for some $\gamma_i < -1$ by (5.3.8). Therefore, these integrals equal a constant plus a function of ε that vanishes as ε becomes small so that

$$\begin{aligned}&\int_{1/\varepsilon}^{\varepsilon t^{1/r}} u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^{u_{s_1+1}/\varepsilon} u_{s_1+2}^{\beta(s_1+2)} \cdots \int_{u_{s_2-1}}^{u_{s_2-1}/\varepsilon} u_{s_2}^{\beta(s_2)} \mathrm{d}u_{s_2} \ldots \mathrm{d}u_{s_1+1} \\ &= \left(\frac{1}{r}\log(t) + \log(\varepsilon^2)\right)(K + h_2(\varepsilon)), \end{aligned} \tag{5.3.15}$$

for some $0 < K < \infty$ and some $h_2(\varepsilon)$ such that $\lim_{\varepsilon\to 0} h_2(\varepsilon) = 0$. The other integrals in (5.3.12) can be estimated similarly.

Combining (5.3.12), (5.3.13) and (5.3.15) we obtain

$$\lim_{t\to\infty} \frac{\mathbb{E}\left[N_t(H)\right]}{t^{k+B(H)}\log^{r-1}(t)} \geq C_1 + h(\varepsilon),$$

for some constant $0 < C_1 < \infty$ and some function $h(\varepsilon)$ such that $\lim_{\varepsilon\to 0} h(\varepsilon) = 0$. Taking the limit for $\varepsilon \downarrow 0$ then proves the theorem. □

5.4. Exact constant for triangles: proof of Theorem 5.1.5

Fix $m \geq 2$ and $\delta > -m$. The first step of the proof consists of showing that

$$\begin{aligned}\mathbb{E}[\triangle_t] =& \frac{\tau-2}{\tau-1}\frac{m^2(m-1)(m+\delta)(m+\delta+1)}{(2m+\delta)^2} \\ &\times \sum_{u=1}^{t-2}[(u-(2m)/(2m+\delta))(u-(2m-1)/(2m+\delta))]^{-1} \\ &\times \frac{\Gamma(u+2-(2m)/(2m+\delta))}{\Gamma(u+2-(3m+\delta)/(2m+\delta))}\frac{\Gamma(u+2-(2m-1)/(2m+\delta))}{\Gamma(u+2-(3m+\delta-1)/(2m+\delta))} \\ &\times \sum_{v=u+1}^{t-1}(v-(3m+\delta-1)/(2m+\delta))^{-1} \\ &\times \sum_{w=v+1}^{t}\frac{\Gamma(w-(3m+\delta)/(2m+\delta))}{\Gamma(w-(2m)/(2m+\delta))}\frac{\Gamma(w-(3m+\delta-1)/(2m+\delta))}{\Gamma(w-(2m-1)/(2m+\delta))}.\end{aligned} \tag{5.4.1}$$

We can write

$$\triangle_t := \sum_{u=1}^{t-2} \sum_{v=u+1}^{t-1} \sum_{w=v+1}^{t} \sum_{j_1 \in [m]} \sum_{j_2, j_3 \in [m]} \mathbb{1}\{u \overset{j_1}{\leftarrow} v,\ u \overset{j_2}{\leftarrow} w,\ v \overset{j_3}{\leftarrow} w\}. \tag{5.4.2}$$

Since there are $m^2(m-1)$ possible choices for the edges j_1, j_2, j_3,

$$\mathbb{E}[\triangle_t] = m^2(m-1) \sum_{u=1}^{t-2} \sum_{v=u+1}^{t-1} \sum_{w=v+1}^{t} \mathbb{E}\left[\psi_u \frac{S_u}{S_{v-1}} \psi_u \frac{S_u}{S_{w-1}} \psi_v \frac{S_v}{S_{w-1}}\right]. \tag{5.4.3}$$

Recalling (4.7.1), we can write every term in the sum in (5.4.3) as

$$\mathbb{E}\left[\left(\psi_u \prod_{h=u+1}^{v-1} (1-\psi_h)\right)\left(\psi_u \prod_{k=u+1}^{w-1} (1-\psi_k)\right)\left(\psi_v \prod_{l=v+1}^{w-1} (1-\psi_l)\right)\right]. \tag{5.4.4}$$

Since the random variables $\psi_1, \dots, \psi_t$ are independent, we can factorize the expectation to obtain

$$\begin{aligned} &\mathbb{E}[\psi_u^2]\mathbb{E}[\psi_v(1-\psi_v)] \prod_{k=u+1, k\neq v}^{w-1} \mathbb{E}[(1-\psi_k)^2] \\ &\qquad = \mathbb{E}[\psi_u^2] \frac{\mathbb{E}[\psi_v(1-\psi_v)]}{\mathbb{E}[(1-\psi_v)^2]} \prod_{k=u+1}^{w-1} \mathbb{E}[(1-\psi_k)^2]. \end{aligned} \tag{5.4.5}$$

Recall that, for a Beta random variable $X(\alpha, \beta)$, we have

$$\begin{aligned} \mathbb{E}[X] &= \frac{\alpha}{\alpha+\beta}, \\ \mathbb{E}[X(1-X)] &= \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}, \\ \mathbb{E}[X^2] &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}, \end{aligned} \tag{5.4.6}$$

and $1-X(\alpha,\beta)$ is distributed as $X(\beta,\alpha)$. Using (5.4.6), we can rewrite (5.4.5) in terms of the parameters of $\psi_1, \dots, \psi_t$. Since ψ_k has parameters $\alpha = m+\delta$ and $\beta = \beta_k = m(2k-3) + (k-1)\delta$, the first term in (5.4.5) can be written as

$$\begin{aligned} &\mathbb{E}[\psi_u^2] \frac{(m+\delta)(m+\delta+1)}{(m(2u-2)+u\delta)(m(2u-2)+u\delta+1)} \\ &\qquad = \frac{(m+\delta)(m+\delta+1)}{(2m+\delta)^2} \left[(u-2m/2m+\delta)(u-(2m-1)/(2m+\delta))\right]^{-1}. \end{aligned} \tag{5.4.7}$$

For the second term, we have

$$\frac{\mathbb{E}[\psi_v(1-\psi_v)]}{\mathbb{E}[(1-\psi_v)^2]} = \frac{m+\delta}{m(2v-3)+(v-1)\delta} = \frac{\tau-2}{\tau-1}(v-(3m+\delta-1)/(2m+\delta))^{-1}. \quad (5.4.8)$$

The last product in (5.4.5), for $k = u+1, \dots, w-1$ results in

$$\begin{aligned}\mathbb{E}[(1-\psi_k)^2] &= \frac{(m(2k-3)+(k-1)\delta)(m(2k-3)+(k-1)\delta+1)}{(m(2k-2)+k\delta)(m(2k-2)+k\delta+1)} \\ &= \frac{k-(3m+\delta)/(2m+\delta)}{k-2m/(2m+\delta)}\frac{k-(3m+\delta-1)/(2m+\delta)}{k-(2m-1)/(2m+\delta)}.\end{aligned} \quad (5.4.9)$$

Using the recursive property $\Gamma(a+1) = a\Gamma(a)$ of the Gamma function,

$$\begin{aligned}&\prod_{k=u+1}^{w-1}\mathbb{E}[(1-\psi_k)^2] \quad (5.4.10)\\ &= \frac{\Gamma(u+2-(2m)/(2m+\delta))}{\Gamma(u+2-(3m+\delta)/(2m+\delta))}\frac{\Gamma(u+2-(2m-1)/(2m+\delta))}{\Gamma(u+2-(3m+\delta-1)/(2m+\delta))} \\ &\quad\times\frac{\Gamma(w-(3m+\delta)/(2m+\delta))}{\Gamma(w-(2m)/(2m+\delta))}\frac{\Gamma(w-(3m+\delta-1)/(2m+\delta))}{\Gamma(w-(2m-1)/(2m+\delta))}.\end{aligned}$$

Equation (5.4.3) follows by combining (5.4.5), (5.4.7), (5.4.8), (5.4.9) and (5.4.10).

The last step of the proof is to evaluate the sum in (5.4.3), and combining the result with the multiplicative constant in front in (5.4.3). By Stirling's formula

$$\frac{\Gamma(x+a)}{\Gamma(x+b)} = x^{a-b}(1+O(1/x)).$$

As a consequence, recalling that $\chi = (m+\delta)/(2m+\delta)$, the sum in (5.4.3) can be written as

$$\sum_{u=1}^{t-2} u^{2\chi-2}(1+O(1/u))\sum_{v=u+1}^{t-1} v^{-1}(1+O(1/v))\sum_{w=v+1}^{t} w^{-2\chi}(1+O(1/w)). \quad (5.4.11)$$

We can approximate the sum in (5.4.11) with the corresponding integral using Euler-Maclaurin formula, thus obtaining

$$\int_1^t u^{2\chi-2}du\int_u^t v^{-1}dv\int_v^t w^{-2\chi}dw. \quad (5.4.12)$$

As $t\to\infty$, the order of magnitude of the integral in (5.4.12) is predicted by Theorem 5.1.2. If we evaluate the integral, then we obtain that the coefficient of the dominant term in (5.4.12) is $(2m+\delta)^2/\delta^2$ for $\tau > 2, \tau \neq 3$, and $1/6$ for $\tau = 3$.

Putting together these coefficients with the constant in front of the sum in (5.4.1) completes the proof of Theorem 5.1.5. □

Remark 5.4.1 (Constant for general subgraphs)**.** In the proof of Theorem 5.1.5, the hardest step is to prove (5.4.2), i.e., to find the expectation of the indicator functions in (5.4.1). This is the reason why for a general ordered subgraph (H, π) on k vertices it is hard to find the explicit constant as in (5.1.10). In fact, as we have done to move from (5.4.3) to (5.4.4), it is necessary to identify precisely, for every $v \in [t]$, how many times the terms ψ_v and $(1 - \psi_v)$ appear in the product inside the expectations in (5.4.3). This makes the evaluation of such terms complicated.

Typically, as it shown in (5.4.5), (5.4.7), (5.4.8), (5.4.9) and (5.4.10), the product of the constants obtained by evaluating the probability of an ordered subgraph (H, π) being present can be written as ratios of Gamma functions. The same constants can be found using the martingale approach as in [33, 158] and [85, Section 8.3], even though in this case constants are obtained through a recursive use of conditional expectation.

We remark that our method and the martingale method are equivalent. We focused on the Pólya urn interpretation of the graph since it highlights the dependence of the presence of edges on the *age* of vertices, that is directly related to the order of magnitude of degrees.

5.5. Conditional convergence: proof of Proposition 5.1.4

In the previous sections, we have considered the order of magnitude of the expectation of the number of occurrences of ordered subgraphs in the PAM. In other words, for an ordered subgraph (H, ψ) we are able to identify the order of magnitude $h(t)$ of the expected number of occurrences $N_t(H, \pi)$, so that $\mathbb{E}[N_t(H, \pi)] = \Theta(h(t))$. We now show how these orders of magnitude of the expected number of subgraphs determines the conditional convergence given in (5.1.6).

5.5.1. Bound with overlapping subgraphs

The Pólya urn graph in Definition 4.4.2 consists of a function of uniform random variables $(U_{v,j})_{v\in[t]}^{j\in[m]}$ and an independent sequence of Beta random variables $(\psi_v)_{v\in[t]}$. We can interpret the sequence $(\psi_v)_{v\in[t]}$ as a sequence of *intensities* associated to the vertices, where a higher intensity corresponds to a higher probability of receiving a connection. The sequence $(U_{v,j})_{v\in[t]}^{j\in[m]}$ determines the attachment of edges. In particular, conditionally on the sequence $(\psi_v)_{v\in[t]}$, every edge is present *independently* (but with different probabilities).

For $t \in \mathbb{N}$, denote $\mathbb{P}_{\boldsymbol{\psi}_t}(\cdot) = \mathbb{P}(\,\cdot\,|\psi_1, \dots, \psi_t)$, and similarly $\mathbb{E}_{\boldsymbol{\psi}_t}[\cdot] = \mathbb{E}[\,\cdot\,|\psi_1, \dots, \psi_t]$. Furthermore, let $N_{\boldsymbol{\psi}_t}(H, \pi)$ denote the number of times subgraph (H, π) appears conditionally on the ψ-variables. We now apply a conditional second moment method to $N_{\boldsymbol{\psi}_t}(H, \pi)$. We use the notation introduced in Section 5.2, so that every possible realization of H in PAM corresponds to a finite set of edges $M_\ell(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell, \boldsymbol{j}_\ell)$, where ℓ is the number of edges in H such that $v_h \xrightarrow{j_h} u_h$, i.e., u_h is the receiving vertex, and j_h is the label of the edge. For simplicity, we denote the set $M_\ell(\boldsymbol{u}_\ell, \boldsymbol{v}_\ell, \boldsymbol{j}_\ell)$ by M. For ease of notation, we assume that π is the identity map and drop the dependence on π. We prove the following result:

Lemma 5.5.1 (Bound on conditional variance)**.** *Consider an ordered subgraph (H, π). Then, $\mathbb{P}$-a.s.,*

$$\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi)) \leq \mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)] + \sum_{\hat{H}\in\hat{\mathcal{H}}} \mathbb{E}_{\boldsymbol{\psi}_t}[N_t(\hat{H})],$$

where $\hat{\mathcal{H}}$ denotes the set of all possible attainable subgraphs $\hat{H}$ that are obtained by merging two copies of (H, π) such that they share at least one edge.

Lemma 5.5.1 gives a bound on the conditional variance in terms of the conditional probabilities of observing two overlapping of the subgraph (H, π) at the same time. Notice that we require these copies to overlap at at least one edge, which is different than requiring that they are disjoint (they can share one or more vertices but no edges).

Proof of Lemma 5.5.1. We prove the bound in Lemma 5.5.1 by evaluating the conditional second moment of $N_t(H)$ as

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)^2] &= \mathbb{E}_{\boldsymbol{\psi}_t}\Big[\sum_{M,M'} \mathbb{1}_{\{M\subseteq E(\mathrm{PA}_t)\}}\mathbb{1}_{\{M'\subseteq E(\mathrm{PA}_t)\}}\Big] \\ &= \sum_{M,M'} \mathbb{P}_{\boldsymbol{\psi}_t}\Big(M\subseteq E(\mathrm{PA}_t),\ M'\subseteq E(\mathrm{PA}_t)\Big),\end{aligned}$$

where M and M' are two sets of edges corresponding to two possible realizations of the subgraph (H, π). Notice that M and M' are not necessarily distinct. We then have to evaluate the conditional probability of having both the sets M and M' simultaneously present in the graph. As a consequence, we conditional variance in Lemma 5.5.1 can be written as

$$\begin{aligned}\sum_{M\neq M'} \mathbb{P}_{\boldsymbol{\psi}_t}\Big(M\subseteq E(\mathrm{PA}_t),\ M'\subseteq E(\mathrm{PA}_t)\Big) \\ - \mathbb{P}_{\boldsymbol{\psi}_t}(M\subseteq E(\mathrm{PA}_t))\mathbb{P}_{\boldsymbol{\psi}_t}(M'\subseteq E(\mathrm{PA}_t)).\end{aligned} \tag{5.5.1}$$

We define

$$\begin{aligned}\mathcal{M} := \Big\{(M,M')\colon \exists\,(u,v,j)\colon (u,v,j)\in M,\ (u,v,j)\in M', \\ M\neq M',\ (M\cup M') \text{ defines an attainable subgraph}\Big\}.\end{aligned} \tag{5.5.2}$$

We then consider two different cases, i.e., whether (M, M') is in $\mathcal{M}$ or not. If $(M, M') \notin \mathcal{M}$, then one of the three following situations occurs:

- ▷ $M\cup M'$ defines a subgraph that is not attainable (for instance, M and M' require that the same edge is attached to different vertices);
- ▷ $M \cup M'$ defines a subgraph that is attainable, M and M' are disjoint sets of labeled edges (they are allowed to share vertices);

▷ M and M' define the same attainable subgraph (so $M = M'$, thus labels of edges coincide).

When $M = M'$ we have that

$$\mathbb{P}_{\boldsymbol{\psi}_t}\Big(M \subseteq E(\mathrm{PA}_t),\ M' \subseteq E(\mathrm{PA}_t)\Big) = \mathbb{P}_{\boldsymbol{\psi}_t}(M \subseteq E(\mathrm{PA}_t)),$$

so that the corresponding contribution in the sum in (5.5.1) is

$$\mathbb{P}_{\boldsymbol{\psi}_t}(M \subseteq E(\mathrm{PA}_t)) - \mathbb{P}_{\boldsymbol{\psi}_t}(M \subseteq E(\mathrm{PA}_t))^2 \leq \mathbb{P}_{\boldsymbol{\psi}_t}(M \subseteq E(\mathrm{PA}_t)),$$

and the sum over M gives the term $\mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)]$ in the statement of Lemma 5.5.1. When $M \neq M'$ and $M \cup M'$ is attainable and their sets of edges are disjoint it follows directly from the independence of $(U_{v,j})_{v\in[t]}^{j\in[m]}$ and $(\psi_v)_{v\in[t]}$ that

$$\mathbb{P}_{\boldsymbol{\psi}_t}\Big(M \subseteq E(\mathrm{PA}_t),\ M' \subseteq E(\mathrm{PA}_t)\Big) = \mathbb{P}_{\boldsymbol{\psi}_t}(M \subseteq E(\mathrm{PA}_t))\mathbb{P}_{\boldsymbol{\psi}_t}(M' \subseteq E(\mathrm{PA}_t)).$$

Thus, in this situation the corresponding contribution is zero. When (M, M') is not attainable the corresponding contribution is negative. When $(M, M') \in \mathcal{M}$ we bound the corresponding terms in (5.5.1) by $\mathbb{P}_{\boldsymbol{\psi}_t}\Big(M \subseteq E(\mathrm{PA}_t),\ M' \subseteq E(\mathrm{PA}_t)\Big)$, thus obtaining

$$\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi)) \leq \mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)] + \sum_{(M,M')\in\mathcal{M}} \mathbb{P}_{\boldsymbol{\psi}_t}\Big(M \cup M' \subseteq E(\mathrm{PA}_t)\Big),$$

We then rewrite this as

$$\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi)) \leq \mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)] + \sum_{\hat{H}\in\hat{\mathcal{H}}} \mathbb{E}_{\boldsymbol{\psi}_t}[N_t(\hat{H})],$$

which proves the lemma. □

5.5.2. Criterion for conditional convergence

We now prove Proposition 5.1.4 using Lemma 5.5.1 and Lemma 5.5.3:

Proof of Proposition 5.1.4. It sufficient to show that for every fixed $\varepsilon > 0$,

$$\mathbb{P}\left(|N_{t,\psi}(H,\pi) - \mathbb{E}_t[N_t(H,\pi)] > \varepsilon\mathbb{E}[N_t(H,\pi)]\right) = o(1).$$

We now apply Chebychev inequality to write

$$\mathbb{P}\Big(|N_{\boldsymbol{\psi}_t}(H,\pi) - \mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)] > \varepsilon\mathbb{E}[N_t(H,\pi)]\Big) \leq \frac{\mathbb{E}\big[\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi))\big]}{\varepsilon^2\mathbb{E}[N_t(H,\pi)]^2}.$$

We can then apply Lemma 5.5.1, which yields

$$\frac{\mathbb{E}\big[\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi))\big]}{\varepsilon^2\mathbb{E}[N_t(H,\pi)]^2} \leq \frac{\mathbb{E}\Big[\mathbb{E}_{\boldsymbol{\psi}_t}[N_t(H,\pi)] + \sum_{\hat{H}\in\hat{\mathcal{H}}}\mathbb{E}_{\boldsymbol{\psi}_t}[N_t(\hat{H})]\Big]}{\varepsilon^2\mathbb{E}[N_t(H,\pi)]^2}$$
$$= \frac{\mathbb{E}[N_t(H,\pi)] + \sum_{\hat{H}\in\hat{\mathcal{H}}}\mathbb{E}[N_t(\hat{H})]}{\varepsilon^2\mathbb{E}[N_t(H,\pi)]^2} = o(1).$$

□

As an example, we consider triangles. Theorem 5.1.5 identifies the expected number of triangles, and by Theorem 5.1.2 we can show that $\mathbb{E}[\triangle_t^2] = \Theta(\mathbb{E}[\triangle_t]^2)$, so we are not able to apply the second moment method to $\triangle_t$. Figure 5.3 suggests that $\triangle_t/\mathbb{E}[\triangle_t]$ converges to a limit that is not deterministic, i.e., in (5.1.8) the limiting X is a random variable.

However, we can prove that $\triangle_t$ is conditionally concentrated, as stated in Corollary 5.1.6. The proof of Corollary 5.1.6 follows directly from Proposition 5.1.4, the fact that $\mathbb{E}[\triangle_t] = \Theta(t^{(3-\tau)/(\tau-1)}\log(t))$ as given by Theorem 5.1.5, and Figure 5.2, that contains the information on the subgraphs consisting of two triangles sharing one or two edges.

5.5.3. Non-concentrated subgraphs

We now show that for most ψ-sequences, the other direction in Proposition 5.1.4 also holds. That is, if there exists a subgraph composed of two merged copies of H such that the condition in Proposition 5.1.4 does not hold, then for most ψ-sequences, H is not conditionally concentrated.

Proposition 5.5.2. *Consider a subgraph (H,π) such that $\mathbb{E}[N_t(H,\pi)] \to \infty$ as $t\to\infty$. Suppose that there exists a subgraph $\hat{H}$, composed of two distinct copies of (H,π) with at least one edge in common such that $\mathbb{E}\left[N_t(\hat{H})\right]/\mathbb{E}[N_t(H,\pi)] \not\to 0$ as $t\to\infty$. Then, for any $\varepsilon>0$, there exists $\eta>0$ such that*

$$\mathbb{P}\left(\frac{\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi))}{\mathbb{E}[N_t(H,\pi)]^2} > \eta\right) \geq 1-\varepsilon. \tag{5.5.3}$$

To prove Proposition 5.5.2, we need Lemma 4.7.2, that gives a coupling between the sequence $(\psi_k)_{k\in\mathbb{N}}$ and a sequence of Gamma random variables. For the precise statement, we refer to Chapter 4. We now state the lemma we need to prove Proposition 5.5.2:

Lemma 5.5.3 (Maximum intensity)**.** *For every $\varepsilon>0$ there exists $\omega=\omega(\varepsilon)\in(0,1)$ such that, for every $t\in\mathbb{N}$,*

$$\mathbb{P}\Big(\max_{i\in 2,\dots,t}\psi_i < \omega\Big) \geq 1-\varepsilon.$$

Proof. Fix $\varepsilon > 0$, and consider $K(\varepsilon/2)$ as given by Lemma 4.7.2. For every $\omega \in (0,1)$, we can write

$$\mathbb{P}\Big(\max_{i\in 2,\dots,t} \psi_i < \omega\Big) = \mathbb{P}\Big(\max_{i\in 2,\dots,K} \psi_i < \omega\Big)\mathbb{P}\Big(\max_{i\in[t]\setminus[K]} \psi_i < \omega\Big), \tag{5.5.4}$$

where we have used the independence of $\psi_2, \dots, \psi_t$. If $t > K$ the second term in the right-hand side of (5.5.4) is well defined, otherwise we only have the first term. Define

$$\omega_1 = \begin{cases} \frac{(\log K)^2}{(2m+\delta)K} & \text{if } t > K, \\ 0 & \text{if } t \leq K. \end{cases}$$

Notice that, since the function $k \mapsto \frac{(\log k)^2}{(2m+\delta)k}$ is decreasing, it follows that

$$\mathbb{P}\Big(\max_{i\in[t]\setminus[K]} \psi_i < \omega_1\Big) \geq 1 - \varepsilon/2. \tag{5.5.5}$$

Define the random variable $X_K = \max_{i\in 2,\dots,K} \psi_i$, denote its distribution function by F_K and the inverse of its distribution function by F_K^{-1}. Consider $\omega_2 = F_K^{-1}(1-\varepsilon/2)$, that implies

$$\mathbb{P}\Big(\max_{i\in[K]} \psi_i < \omega_2\Big) = 1 - \varepsilon/2. \tag{5.5.6}$$

Consider then $\omega = \max\{\omega_1, \omega_2\}$. Using a(5.5.5) and (5.5.6) with ω in (5.5.4), it follows that

$$\mathbb{P}\Big(\max_{i\in 2,\dots,K} \psi_i < \omega\Big)\mathbb{P}\Big(\max_{i\in[t]\setminus[K]} \psi_i < \omega\Big) \geq (1-\varepsilon/2)^2 \geq 1 - \varepsilon,$$

which completes the proof. □

Proof of Proposition 5.5.2. We use the expression of the conditional variance of (5.5.1). We first study the term in the conditional variance corresponding to $\hat{H}$. Let $\tilde{\mathcal{M}}$ denote the set of labeled edges M, M' that together form the subgraph $\hat{H}$. Let the edges that M and M' share be denoted by M_s. Furthermore, let $\tilde{\mathcal{M}}_1$ denote the set of labeled edges M, M' that together form subgraph $\hat{H}$ that do not use vertex 1.

The, we can then write a term in the sum in (5.5.1) as

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{\psi}_t}\Big(M \subseteq E(\mathrm{PA}_t),\ M' \subseteq E(\mathrm{PA}_t)\Big) - \mathbb{P}_{\boldsymbol{\psi}_t}(M \subseteq E(\mathrm{PA}_t))\mathbb{P}_{\boldsymbol{\psi}_t}(M' \subseteq E(\mathrm{PA}_t)) \\ &= \mathbb{P}_{\psi_t}\left(M \cup M' \subseteq E(PA_t)\right)\left(1 - \mathbb{P}_{\psi_t}\left(M_s \subseteq E(PA_t)\right)\right). \end{aligned} \tag{5.5.7}$$

In fact, the difference between the first term and the second term in in (5.5.7) is given by the fact that in the second term we count the probability of the presence of edges in M_s twice, thus generating the term $\mathbb{P}_{\psi_t}\left(M_s \subseteq E(PA_t)\right)$. As a consequence, using (5.5.7) in (5.5.1), we can write

$$\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi)) = \sum_{M,M'\in\tilde{\mathcal{M}}} \mathbb{P}_{\psi_t}\left(M \cup M' \subseteq E(PA_t)\right)\left(1 - \mathbb{P}_{\psi_t}\left(M_s \subseteq E(PA_t)\right)\right)$$

$$\geq \sum_{M,M' \in \tilde{\mathcal{M}}_1} \mathbb{P}_{\psi_t}\left(M \cup M' \subseteq E(PA_t)\right)(1-\psi_{\max})$$
$$= (1-\psi_{\max})\mathbb{E}_{\psi_t}\left[N_t(\hat{H})\right]$$

where the inequality uses (4.7.1), and $\psi_{\max} = \max_{i\in 2,\ldots,t}\psi_i$. Note that here we excluded vertex 1 from the number of subgraphs with negligible error. By Lemma 5.5.3 there exists ω such that with probability at least $1-\varepsilon$, $\psi_{\max} < \omega < 1$.

By the assumption on $\hat{H}$, $\mathbb{E}\left[N_t(\hat{H})\right] \geq \tilde{C}\mathbb{E}\left[N_t(H,\pi)\right]^2$ for some $\tilde{C} > 0$. We then use that $\mathbb{E}_{\psi_t}\left[N_t(\hat{H})\right] = \mathcal{O}_{\mathbb{P}}(\mathbb{E}\left[N_t(\hat{H})\right])$. Thus, for t sufficiently large, we can bound the contribution from subgraph $\hat{H}$ to the conditional variance from below with probability at least $1-\varepsilon$ by

$$\sum_{M,M' \in \tilde{\mathcal{M}}} \mathbb{P}_{\psi_t}\left(M \cup M' \subseteq E(PA_t)\right)\left(1 - \mathbb{P}_{\psi_t}\left(M_s \subseteq E(PA_t)\right)\right) \geq C\mathbb{E}_{\psi_t}\left[N_t(H,\pi)\right]^2,$$

for some $C > 0$.

Note that the only terms that have a negative contribution to (5.5.1) are the terms where $M \cup M'$ is a non-attainable subgraph. In that situation,

$$\mathbb{P}_{\boldsymbol{\psi}_t}\Big(M \subseteq E(\mathrm{PA}_t),\ M' \subseteq E(\mathrm{PA}_t)\Big) = 0.$$

Furthermore,

$$\sum_{M,M' \in \tilde{\mathcal{M}}} \mathbb{P}_{\boldsymbol{\psi}_t}(M \subseteq E(\mathrm{PA}_t))\mathbb{P}_{\boldsymbol{\psi}_t}(M' \subseteq E(\mathrm{PA}_t) \leq \mathbb{E}_{\psi_t}\left[N_t(H,\pi)\right]^2/t^2,$$

since the two subgraphs share at least two vertices. Therefore, the negative terms in the conditional variance scale as most as $\mathbb{E}_{\psi_t}\left[N_t(H,\pi)\right]/t^2$. We therefore obtain that with probability at least $1-\varepsilon$,

$$\mathrm{Var}_{\boldsymbol{\psi}_t}(N_t(H,\pi)) \geq \eta\mathbb{E}_{\psi_t}\left[N_t(H,\pi)\right]^2,$$

for some $\eta > 0$, which proves the proposition. □

6

Local weak convergence for PageRank

Content and structure of the chapter

The motivation of this chapter is in finding general conditions for the existence of an asymptotic PageRank distribution. We prove the convergence (in distribution and/or in probability) of PageRank for a large class of models, by adapting the notion of local weak convergence (LWC) to directed graphs. Our results also shed light on the power-law hypothesis. When the limit is a branching tree, this directly implies the power-law hypothesis, based on mentioned results in the literature (see Section 1.7). When the limit is different, e.g., the tree generated by a continuous-time branching process, proving the power-law hypothesis remains an open problem. Our results imply however that it is sufficient to study PageRank on the limiting object, which hopefully is simpler since the graph-size asymptotics no longer interfere.

The chapter is structured as follows: In Section 6.1 we give the formal definition of PageRank on graphs. In Section 6.2 we state the main result. In Section 6.3 we define directed LWC. In Section 6.2.2 we give a high-level structure of the proof of Theorem 6.2.1, that we prove in Section 6.4. In Section 6.5 we explain how our results can be extended to generalized PageRanks. Section 6.6 contains the proof of the directed LWC for directed continuous-time branching processes, directed configuration models and directed PAMs, showing the generality of our argument. Proposition 6.6.8 together with Remark 6.2.2 prove a lower bound for the power-law PageRank hypothesis in PAMs. Novel results of this chapter are based on [72].

6.1. Formal definition of PageRank

Consider a finite directed (multi-)graph G of size n. We write $[n] = \{1, \dots, n\}$. Let $e_{j,i}$ be the number of directed edges from j to i. Denote the in-degree of vertex $i \in [n]$ by $d_i^{(\text{in})}$ and the out-degree by $d_i^{(\text{out})}$. Fix a parameter $c \in (0,1)$, which is called the *damping factor*, or teleportation parameter. PageRank is the unique vector $\boldsymbol{\pi}(n) = (\pi_1(n), \dots, \pi_n(n))$ that satisfies, for every $i \in [n]$,

$$\pi_i(n) = c \sum_{j \in [n]} \frac{e_{j,i}}{d_j^{(\text{out})}} \pi_j(n) + \frac{1-c}{n}. \tag{6.1.1}$$

PageRank has the natural interpretation as the invariant measure of a random walk with restarts on G. With probability c the random walk takes a simple random walk step on G, while with probability $(1-c)$ it moves to a uniformly chosen vertex. Here by simple random walk we mean the random walk that chooses, at every step, an outgoing edge from the current position uniformly at random. When $d_j^{(\text{out})} > 0$ for all $j \in [n]$, the invariant measure of this random walk is given exactly by (6.1.1). The interpretation is easily extended to the case when some vertices j have $d_j^{(\text{out})} = 0$ by introducing a random jump from such vertices; in this case the stationary distribution will be the solution of (6.1.1) renormalized to sum up to one [108].

We consider the *graph-normalized version* of PageRank, which is the vector defined as $\boldsymbol{R}(n) = n\boldsymbol{\pi}(n)$. We call both the algorithm and the vector $\boldsymbol{R}(n)$ PageRank, the meaning will always be clear from the context. The graph-normalized version of (6.1.1) is the unique solution $\boldsymbol{R}(n)$ to

$$R_i(n) = c \sum_{j \in [n]} \frac{e_{j,i}}{d_j^{(\text{out})}} R_j(n) + (1-c). \tag{6.1.2}$$

PageRank has numerous generalizations. For example, after a random jump, the random walk might not restart from a uniformly chosen vertex, but rather choose vertex i with probability b_i, where $\sum_{i=1}^n b_i = 1$. Equation (6.1.1) then becomes

$$R_i(n) = c \sum_{j \in [n]} \frac{e_{j,i}}{d_j^{(\text{out})}} R_j(n) + (1-c) b_i. \tag{6.1.3}$$

This generalized version of PageRank is sometimes called *topic-sensitive* [83] or *personalized*. We note that the term *personalized PageRank* often refers to the case when the vector $\mathbf{b} = (b_1, \dots, b_n)$ has one of its coordinates equal to 1, and the rest equal to zero, so that the random walk always restarts from the same vertex. One can generalize further, e.g., allow the probability c to be random as well. The literature [45, 101, 110, 162] usually studies the following graph-normalized equation:

$$R_i(n) = \sum_{j : e_{j,i} \geq 1} A_j R_j(n) + B_i, \quad i \in [n], \tag{6.1.4}$$

where $(A_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ are values assigned to the vertices in the graph. For simplicity of the argument, we will focus on the basic model (6.1.2) and then, in Section 6.5, extend the results to the more general model (6.1.4) with $A_j = C_j/d_j^{(\text{out})}$, where C_j's are random variables bounded by $c < 1$, and $(B_i)_{i\in[n]}$ are i.i.d. across vertices.

6.2. Main result

For any deterministic graph, PageRank is a deterministic vector. We are interested in the PageRank associated to *random* graphs. In particular, we want to investigate the asymptotic behavior of the PageRank value of a uniformly chosen vertex V_n, as the size of the graph grows. In this case we have two sources of randomness: the choice of the vertex and the randomness of the graph itself. Our main result is the following:

Theorem 6.2.1 (Existence of asymptotic PageRank distribution)**.** *Consider a sequence of directed random graphs $(G_n)_{n\in\mathbb{N}}$. Then, the following hold:*

(1) If G_n converges in distribution in the local weak sense*, then there exists a limiting distribution $R_\varnothing$, with $\mathbb{E}[R_\varnothing] \le 1$, such that*

$$R_{V_n}(n) \xrightarrow{d} R_\varnothing;$$

(2) If G_n converges in probability in the local weak sense*, then there exists a limiting distribution $R_\varnothing$, with $\mathbb{E}[R_\varnothing] \le 1$, such that, for every $r > 0$,*

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{R_i(n) > r\} \xrightarrow{\mathbb{P}} \mathbb{P}\left(R_\varnothing > r\right).$$

Theorem 6.2.1 establishes that, whenever a sequence of directed random graphs converges in the *local weak sense*, then the distribution $R_{V_n}(n)$ admits a limit in distribution, $R_\varnothing$. This limit has the interpretation of PageRank on the (possibly infinite) limiting graph. Theorem 6.2.1 can be extended to personalized PageRank defined in (6.1.4) under additional conditions on the random variables $(A_i)_{i\in\mathbb{N}}$ and $(B_i)_{i\in\mathbb{N}}$. The precise formulation, that requires more notation, is given in Theorem 6.5.1.

Remark 6.2.2 (Stochastic lower bound for PageRank)**.** Theorem 6.2.1 gives a rough lower bond on the tail of the asymptotic PageRank distribution for a graph sequence. In simple words, we can write

$$R_\varnothing \ge (1-c)\left(1 + c\sum_{i=1}^{D_\varnothing^{(\text{in})}} \frac{1}{m_i^{(\text{out})}}\right), \tag{6.2.1}$$

where $\varnothing$ is a vertex called root in the local weak limit of the graph sequence $(G_n)_{n\in\mathbb{N}}$, $D_\varnothing^{(\text{in})}$ is the graph limiting in-degree distribution, and $m_i^{(\text{out})}$ represent the out-degree in the LW limit. All the notation in (6.2.1) is introduced in Sections 6.3 and 6.4. In

particular, (6.2.1) implies that a.s. $R_\varnothing > 1-c$. Since $m^{(\text{out})}$ represents the limiting out-degree distribution, it follows that, if $(G_n)_{n\in\mathbb{N}}$ has out-degrees uniformly bounded by a constant $A < \infty$,

$$R_\varnothing \geq (1-c)\Big(1 + \frac{c}{A} D_\varnothing^{(\text{in})}\Big).$$

As a consequence, if the limiting in-degree distribution obeys a power law, then the tail of the distribution $R_\varnothing$ is bounded from below by a multiple of the tail of the in-degree. This establishes a power-law lower bound for $R_\varnothing$. This is a partial solution of the power-law hypothesis mentioned in Chapter 1.

6.2.1. Local weak convergence for directed graphs

Theorem 6.2.1 is based on local weak convergence (LWC), that we have introduced in Chapter 4. The notion of LWC in Chapter 4 is defined for undirected graphs. Since we want to investigate PageRank, that is naturally defined on directed graphs, we need to extend the definition of LWC to directed graphs.

LWC is defined by looking at neighborhoods of vertices in a graph. For undirected graphs, this is not a problem since the exploration process is well defined. In directed graphs, such exploration of neighborhoods is not uniquely defined. Indeed, in the exploration process (rigorous definition is given in Definition 6.3.3), motivated by the PageRank problem, we naturally explore directed edges *only in their opposite direction*. In other words, a directed edge (j, i) is only explored from i to j. Clearly, since edges are not explored in both directions, starting from the root we might not be able to explore all the graph. Heuristically, from the point of view of the root $\varnothing$, *only part of the graph influences the incoming neighborhood of* $\varnothing$. This is very different from the undirected case, where the exploration process continues until the entire graph is explored (when the graph is connected). We resolve this by introducing so-called marks to track the explored and not-explored out-edges in the graph. The precise definition of LW convergence in directed graphs is given in Section 6.3.

We point out that our construction is one of many possible ways to define LW convergence for directed graphs. For instance, Aldous and Steele [5] allow edge weights. This might be sufficient to define an inclusion of directed graphs in the space of undirected graphs with edge weights, and use the notion for undirected graphs to define an exploration process for directed graphs. The advantage of our construction is that it requires the minimum amount of information, sufficient to prove the convergence of PageRank, which is the main problem we aim to resolve.

Definition 6.3.9 below, together with Theorem 6.3.10, gives a criterion for the convergence of a sequence of directed random graphs, that can be presented as marked graphs by just assigning *marks equal to out-degrees*. The precise formulation requires heavy notation that we have not introduced yet, therefore we do not state it here.

The advantage of having a LW limit $(G, \varnothing)$ is that a whole family of local properties of the graph sequence *can pass to the limit*, and the limit is given by a local property of $(G, \varnothing)$ itself. More precisely, in the construction of LW convergence, one defines a *distance* between (marked directed) rooted graphs (see Definition 4.1.3). Then, any

function h from the space of rooted graphs to $\mathbb{R}$ that is *bounded and continuous* with respect to the distance function can pass to the limit, i.e., for V_n a uniformly chosen vertex in G_n,

$$\lim_{n\to\infty} \mathbb{E}\left[h(G_n, V_n)\right] = \mathbb{E}\left[h(G, \varnothing)\right].$$

This can be rather useful in understanding the asymptotic behavior of local properties of a graph sequence. As a toy example, in the undirected setting, take the function $h(G, \varnothing) = \mathbb{1}\{d_\varnothing = k\}$. It is easy to show, using Definition 4.1.3, that h is a continuous function. Assume that a sequence of graphs $G_n \to (G, \varnothing)$ locally weakly, where $(G, \varnothing)$ is random rooted graph. Then, for every $n \in \mathbb{N}$,

$$\mathbb{E}\left[f(G_n, V_n)\right] = \frac{1}{n}\sum_{i\in[n]} \mathbb{P}\left(d_i = k\right) = \mathbb{P}\left(d_{V_n} = k\right),$$

i.e., h evaluated on a random root is just the probability that a uniformly chosen vertex has degree k. As a consequence, the sequence $(G_n)_{n\in\mathbb{N}}$ has a limiting degree distribution given by

$$\lim_{n\to\infty} \mathbb{P}\left(d_{V_n} = k\right) = \mathbb{P}\left(d_\varnothing = k\right),$$

where $\varnothing$ is the root of G. Other examples of continuous functions in the undirected setting are the nearest-neighbor average degree of a uniform vertex, the finite-distance neighborhood of a uniform vertex and the average pressure per particle in the Ising model. In our directed setting, it follows that, if $G_n \to (G, \varnothing, M(G))$,

$$(m^{(\mathrm{out})}_{V_n}, d^{(\mathrm{in})}_{V_n}) \xrightarrow{d} (m^{(\mathrm{out})}_{\varnothing}, d^{(\mathrm{in})}_{\varnothing}),$$

where $M(G)$ is the set of marks of the limiting graph, $(m^{(\mathrm{out})}_{V_n}, d^{(\mathrm{in})}_{V_n})$ are the mark and the in-degree of a uniformly chosen vertex V_n, and $(m^{(\mathrm{out})}_{\varnothing}, d^{(\mathrm{in})}_{\varnothing})$ are the mark and the in-degree of the root $\varnothing$ in the limiting directed graph G. The notation $m^{(\mathrm{out})}$ hints on the relation between marks and out-degrees. When marks are assigned that are equal to the out-degree, this implies the convergence of the in- and out-degree of a uniformly chosen vertex. One of the surprises in our version of LW convergence is that in the limiting graph, the mark of the root $m^{(\mathrm{out})}_{\varnothing}$ is not necessarily equal to the out-degree of the root.

6.2.2. Application of LW convergence to PageRank

The proof of Theorem 6.2.1 is given in Section 6.4. Here we describe the structure of the proof, explaining why the LW convergence for directed graphs is useful. Schematically, the structure of our proof of Theorem 6.2.1 is presented in Figure 6.1. The implication (A), denoted by the dashed red arrow, is the one we aim to prove. We split it in three steps (a), (b), (c), denoted by the solid black arrows. We will now explain each step.

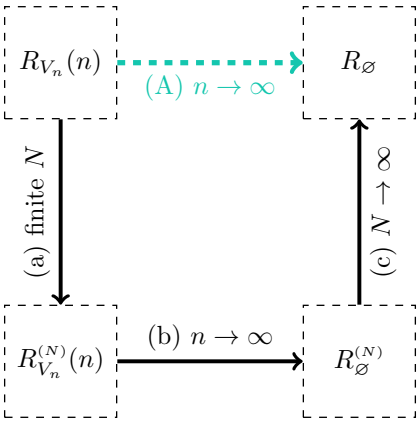


FIGURE 6.1: Structure of the proof of Theorem 6.2.1. The (A) convergence is what we are after, the convergence in distribution of $R_{V_n}(n)$ to a limiting random variable. To prove that, we need the three different steps (a), (b), (c) given by the other arrows.

Step (a): Finite approximations. It is well known [8, 13, 23, 45] that PageRank can be written as

$$R_i(n) = (1-c)\left(1 + \sum_{k=1}^{\infty} c^k \sum_{\ell \in \mathrm{path}_i(k)} \prod_{h=1}^{k} \frac{e_{\ell_h,\ell_{h+1}}}{d^{(\mathrm{out})}_{\ell_h}}\right),$$

where $\mathrm{path}_i(k)$ is the set of directed paths of k steps that end at i. In other words, $R_i(n)$ is a *weighted sum of all the directed paths that end at* i. In particular, we can write finite approximations for PageRank as

$$R^{(N)}_i(n) = (1-c)\left(1 + \sum_{k=1}^{N} c^k \sum_{\ell \in \mathrm{path}_i(k)} \prod_{h=1}^{k} \frac{e_{\ell_h,\ell_{h+1}}}{d^{(\mathrm{out})}_{\ell_h}}\right),$$

where now the sum is taken over all paths of length at most $N \in \mathbb{N}$. We use the sequence of finite approximations $\left(R^{(N)}_{V_n}(n)\right)_{n\in\mathbb{N}}$ to estimate the PageRank of a random vertex with *exponentially small precision* by its finite approximations. We prove that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(|R_{V_n}(n) - R^{(N)}_{V_n}(n)| \geq \varepsilon\right) \leq \frac{c^{N+1}}{\varepsilon}.$$

Notice that the bound is *independent of the graph size that we consider*. This bound is true for any directed graph of any size, so it does not require any assumption on the graph sequence.

Step (b): LW convergence. The finite approximations of PageRank are *continuous* with respect to the local weak topology. Indeed, by definition, the Nth approximation of PageRank depends only on the incoming neighborhood of a vertex *up to distance* N. Note that $R_i(n)$ and $R_i^{(N)}(n)$ are not bounded. However, for any $r \geq 0$, the function $\mathbb{1}\{R_{V_n}^{(N)} > r\}$ is a continuous and bounded function on marked directed rooted graphs, therefore we can pass to the limit for any $N \in \mathbb{N}$. It follows that

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbb{1}\{R_{V_n}^{(N)} > r\}\right] = \lim_{n\to\infty} \mathbb{P}\left(R_{V_n}^{(N)}(n) > r\right) = \mathbb{P}\left(R_{\varnothing}^{(N)} > r\right),$$

where in the last term $\varnothing$ is the root of the limiting random marked directed rooted graph $(G, \varnothing, M(G))$. As a consequence, every term of the sequence $(R_{V_n}^{(N)}(n))_{n\in\mathbb{N}}$ converges in distribution. Notice that similar arguments apply for Theorem 6.2.1(b).

Step (c): Finite approximations on the limiting graph. On the limiting random marked directed rooted graph $(G, \varnothing, M(G))$, the sequence $(R_{\varnothing}^{(N)})_{N\in\mathbb{N}}$ is a monotonically increasing sequence of random variables. Therefore, there exists an almost sure limiting random variable $R_{\varnothing}$. Using the fact that $(G, \varnothing, M(G))$ is a local weak limit of a sequence of random directed graphs, and $\mathbb{E}[R_{V_n}] = 1$ for every $n \geq 1$, it is possible to prove that $\mathbb{E}[R_{\varnothing}] \leq 1$, so that $\mathbb{P}(R_{\varnothing} < \infty) = 1$.

Remark 6.2.3. We emphasize that the above strategy is meant just to give the intuition behind the proof. In particular, in the proof it is necessary to be careful and specify with respect to which randomness we take expectations. In fact, when we consider local weak convergence of random graphs, we have two sources of randomness: the choice of the root and the randomness of the graphs. All these are made rigorous in Section 6.4.

6.2.3. Examples

We consider examples of directed random graphs, for which we prove LWC and find the limiting random graph. Thus, PageRank in these models converges to PageRank on the limiting graph. The following theorem makes this precise for several random graph models that have been studied in the literature. For precise definitions of the models, as well as the proof, we refer to Section 6.6.

Theorem 6.2.4 (Examples of convergence)**.** *The following models converge in the directed local weak sense:*

(1) the directed configuration model *converges* in probability*;*

(2) the continuous-time branching processes *converge* almost surely*;*

(3) the directed preferential attachment model *converges* in probability*.*

As a consequence, for these models there exists a limiting PageRank distribution, and the convergence holds as specified.

Remark 6.2.5 (Power-law lower bound)**.** The directed preferential attachment model and continuous-time branching processes both have constant out-degree. Therefore, they satisfy the condition in Remark 6.2.2. Thus, their limiting PageRank distributions are stochastically bounded from below by a multiple of the limiting in-degree distributions. The directed configuration model satisfies Remark 6.2.2 whenever the out-degree distribution has bounded support.

The proof of Theorem 6.2.4 is divided into three propositions, respectively Proposition 6.6.2 for the directed configuration model, Proposition 6.6.5 for continuous-time branching processes and Proposition 6.6.8 for the directed preferential attachment model.

6.3. Directed local weak convergence

The construction of local weak convergence for directed graphs is similar to the undirected case (see Section 4.1). It is necessary though to define an exploration process to construct the neighborhood of the root and keep track of in- and out-degrees of vertices. To keep notation as simple as possible, we use the same notation as in Section 4.1, while here we refer to directed graphs. We start giving the definition of rooted marked directed graphs:

Definition 6.3.1 (Rooted marked directed graph)**.** *Let G be a directed graph with vertex set $V(G)$ and edge set $E(G)$. Let $\varnothing \in V(G)$ be a vertex called the* root*. Assume that for every $i \in V(G)$, the in-degree $d_i^{(\mathrm{in})}$ and the out-degree $d_i^{(\mathrm{out})}$ of the vertex i are finite. Assign to every $i \in V(G)$ an integer value $m_i^{(\mathrm{out})}$ called a* mark*, such that $d_i^{(\mathrm{out})} \leq m_i^{(\mathrm{out})} < \infty$. Denote the set of marks by $M(G) = (m_i^{(\mathrm{out})})_{i \in V(G)}$. We call the triplet $(G, \varnothing, M(G))$ a* rooted marked directed graph*.*

To simplify notation in Definition 6.3.1, we will specify the marks only when nec-

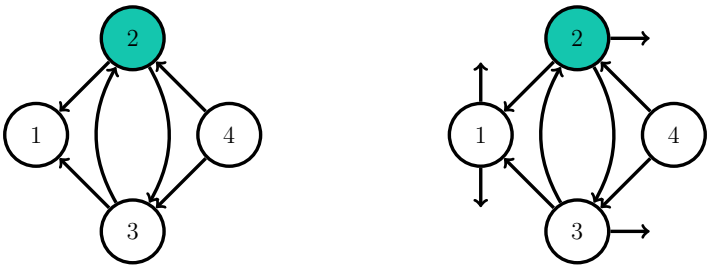


Figure 6.2: Two examples of rooted marked directed graphs. The graph on the left is considered with marks equal to the out-degree, while in the example on the right we have assigned marks larger than the out-degree. The difference between the mark and the out-degree of a vertex can be visualized as the number of arrows starting at the vertex and pointing nowhere.

essary. In simple words, a rooted marked directed graph is a locally finite directed graph where one of the vertices is marked as root, and to every vertex we assign a mark, which is larger than the out-degree of the vertex. If $m_i^{(\text{out})} = d_i^{(\text{out})}$ we keep i intact, and if $m_i^{(\text{out})} - d_i^{(\text{out})} > 0$ then we attach to i exactly $m_i^{(\text{out})} - d_i^{(\text{out})}$ outgoing arrows pointing nowhere. This is illustrated in Figure 6.2. We call a directed graph with marks, without specifying the root, a *marked graph*.

Every directed graph can be seen as a rooted marked directed graph, with marks equal to the out-degrees and a root picked from the set of vertices. In what follows, sometimes we specify the marks, and sometimes we specify the out-degree and the number of edges pointing nowhere.

As in the undirected case, we are not interested in the precise labeling of the vertices. This leads us to define the notion of isomorphism, including the presence of marks:

Definition 6.3.2 (Isomorphism of rooted marked directed graphs)**.** *Two rooted marked directed graphs $(G, \varnothing, M(G))$ and $(G', \varnothing', M(G'))$ are* isomorphic *if and only if there exists a bijection $\gamma : V(G) \to V(G')$ such that*

(1) $(i, j) \in E(G)$ if and only if $(\gamma(i), \gamma(j)) \in E(G')$;
(2) $\gamma(\varnothing) = \varnothing'$;
(3) for every $i \in V(G)$, $m_i^{(\text{out})} = m_{\gamma(i)}^{(\text{out})}$.

We write $(G, \varnothing, M(G)) \cong (G', \varnothing', M(G'))$ to denote that the two marked rooted graphs $(G, \varnothing, M(G))$ and $(G', \varnothing', M(G'))$ are isomorphic.

Notice the similarity with Definition 4.1.2. Denote the space of rooted marked directed graphs by $\mathcal{G}_\star$, which is again a quotient space with respect to the equivalence given by isomorphisms. We now define the exploration process that identifies the neighborhood of the root, see Figure 6.3 for an example.

Definition 6.3.3 (Directed root neighborhood)**.** *Consider a rooted marked directed graph $(G, \varnothing, M(G))$. Fix $k \in \mathbb{N}$. The k-*neighborhood of root $\varnothing$ *is a rooted marked directed graph $(U_{\leq k}(\varnothing), \varnothing, M(U_{\leq k}(\varnothing)))$ constructed as follows:*

▷ for $k = 0$, $U_{\leq k}(\varnothing)$ is a graph with a single vertex $\varnothing$, no edges, and mark $m_\varnothing^{(\text{out})}$;

▷ for $k > 0$, consider $\varnothing$ as active, and proceed recursively as follows, for $h = 1, \ldots, k$:

(1) for every vertex active at step $h - 1$, explore the incoming edges to the vertices in the opposite direction, finding the source of the edges;
(2) label the vertices that were active to be explored, and label the vertices just found as active, but only if they were not already found in the exploration process;
(3) for every vertex i (explored or active), assign the mark $m_i^{(\text{out})}$ to it, that is equal to the mark in the original graph $(G, \varnothing, M(G))$. In addition, draw every edge between two vertices that are already found in the exploration process;
(4) if there are no more active vertices, then stop the process.

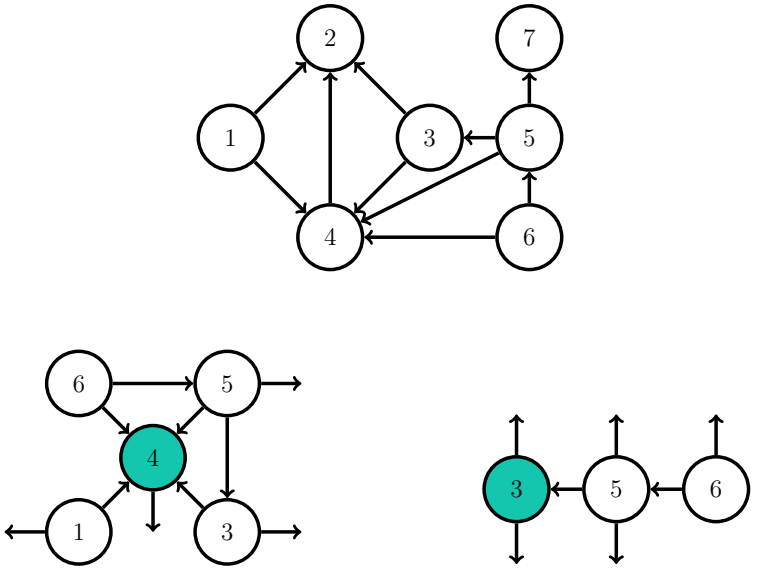


FIGURE 6.3: Example of two root neighborhoods in the same graph above, where we have assigned marks equal to the out-degrees, with a different choice of the root. The root on the left is vertex 4, and vertex 3 on the right. We explore the root neighborhood up to the maximum possible distance. Notice that the graph is only partially explored in this example.

In this way we explore the *incoming neighborhood* of the root. As stated in Definition 6.3.3, we explore edges in the opposite direction: if $(j,i) \in E(G)$ is a directed edge, then the exploration process goes from vertex i to vertex j. Notice that it is possible that we do not explore the entire graph in this process, because we do not explore edges in all directions. This is different to the undirected case, where, for k large enough, we always explore the entire graph (if connected).

We can define a local distance d_{loc} on $\mathcal{G}_\star$ as in but this time for rooted marked directed graphs, using Definitions 6.3.2 and 6.3.3. Next we define a local distance on $\mathcal{G}_*$:

Definition 6.3.4 (Local distance)**.** *For any two rooted marked directed graphs $(G,\varnothing,M(G))$ and $(G',\varnothing',M(G'))$, define*

$$d_{loc}\left((G,\varnothing,M(G)),(G',\varnothing',M(G'))\right) = \frac{1}{1+\kappa},$$

where

$$\kappa = \inf\left\{k \geq 0\colon (U_{\leq k}(\varnothing),\varnothing,M(U_{\leq n}(\varnothing))) \ncong (U_{\leq n}(\varnothing'),\varnothing',M(U_{\leq k}(\varnothing')))\right\},$$

and the inf *of the empty set is defined as* $+\infty$*. The function* d_{loc} *is called* local distance *on* $\mathcal{G}_*$.

As in the undirected setting (see Definition 4.1.3), the function d_{loc} tells us up to what distance the neighborhoods of two roots in two different rooted marked directed graphs are isomorphic. However, in the directed setting the function d_{loc} is *not* a metric on $\mathcal{G}_\star$, but it is a *pseudonorm*. We can prove the following result:

Lemma 6.3.5 (Pseudonorm)**.** *The space* $(\mathcal{G}_*, d_{loc})$ *is complete and separable.*

Proof. The function d_{loc} is positive by definition, and obviously symmetric. It is not hard to prove that it satisfies the triangle inequality. In fact, consider three elements $(G_1, \varnothing_1), (G_2, \varnothing_2), (G_3, \varnothing_3) \in \mathcal{G}_*$, where we omit for simplicity the sets of marks. Then, assume that, for $i, j \in \{1, 2, 3\}$, $i \neq j$,

$$d_{loc}((G_i, \varnothing_i), (G_j, \varnothing_j)) = \frac{1}{1 + N_{i,j}},$$

where $N_{1,2}, N_{2,3}, N_{1,3}$ are integer numbers (possibly ∞). Then we need to show that

$$\frac{1}{1 + N_{1,3}} \leq \frac{1}{1 + N_{1,2}} + \frac{1}{1 + N_{2,3}}.$$

Without loss of generality, suppose $N_{1,2} \leq N_{2,3}$. This implies

$$U_{\leq N_{1,2}}(\varnothing_1) \cong U_{\leq N_{1,2}}(\varnothing_2) \cong U_{\leq N_{1,2}}(\varnothing_3),$$

which implies that $U_{\leq N_{1,2}}(\varnothing_1) \cong U_{\leq N_{1,2}}(\varnothing_3)$, by composition of isomorphisms. As a consequence, $N_{1,3} \geq N_{1,2}$, which means

$$\begin{aligned} d_{loc}((G_1, \varnothing_1), (G_3, \varnothing_3)) &\leq \frac{1}{1 + N_{1,2}} \\ &\leq d_{loc}((G_1, \varnothing_1), (G_2, \varnothing_2)) + d_{loc}((G_2, \varnothing_2), (G_3, \varnothing_3)), \end{aligned}$$

which is the triangular inequality. Notice that this holds even when $N_{1,2} = \infty$.

We now have to prove that every Cauchy sequence in $(\mathcal{G}_*, d_{loc})$ has a limiting point. We use a modification of an argument in [68, Appendix A, Proposition 2]. Consider a Cauchy sequence $(G_n, \varnothing_n)_{n \in \mathbb{N}} \subseteq \mathcal{G}_*$. We need to prove that there exists $(G, \varnothing) \in \mathcal{G}_*$ such that

$$d_{loc}\left((G_n, \varnothing_n), (G, \varnothing)\right) \to 0.$$

In this case, since d_{loc} is a pseudometric, the limit is not unique. At first consider a subsequence $(G_{n_i}, \varnothing_{n_i})_{i \in \mathbb{N}}$ such that

$$d_{loc}\left((G_{n_i}, \varnothing_{n_i}), (G_{n_{i+1}}, \varnothing_{n_{i+1}})\right) \leq 2^{-i}. \tag{6.3.1}$$

Assume that

$$d_{loc}\left((G_{n_i}, \varnothing_{n_i}), (G_{n_{i+1}}, \varnothing_{n_{i+1}})\right)$$

is not definitively zero with respect to i, otherwise there is nothing to prove, since a limiting $(G, \varnothing)$ is given by the equivalence class of the neighborhood of the root. We can assume that $d_{loc}\left((G_{n_i}, \varnothing_{n_i}), (G_{n_{i+1}}, \varnothing_{n_{i+1}})\right) > 0$ for every $i \in \mathbb{N}$. If it is not, then we can restrict ourselves to a subsequence of $(n_i)_{i\in\mathbb{N}}$ where this is true.

For every $\delta > 0$ there exists $i_0 = i_0(\delta)$ and $N_0 = N_0(\delta)$ such that, for every $i, j \geq i_0$, $U_{\leq N_0}(\varnothing_{n_i}) \cong U_{\leq N_0}(\varnothing_{n_j})$. In fact, fix $\delta > 0$ such that $\sum_{i=i_0}^{\infty} 2^{-i} \leq \delta$. Then, by the triangle inequality, assuming $i < j$,

$$\begin{aligned} d_{loc}\left((G_{n_i}, \varnothing_{n_i}), (G_{n_j}, \varnothing_{n_j})\right) &\leq \sum_{h=i}^{j} d_{loc}\left((G_{n_h}, \varnothing_{n_h}), (G_{n_{h+1}}, \varnothing_{n_{h+1}})\right) \\ &\leq \sum_{h=i}^{j} 2^{-h} \leq \sum_{h=i_0}^{\infty} 2^{-h} \leq \delta. \end{aligned}$$

This means that there exists $N_0 > -1 + 1/\delta$ such that, for any $i, j > i_0$,

$$U_{\leq N_0}(\varnothing_{n_i}) \cong U_{\leq N_0}(\varnothing_{n_j}). \tag{6.3.2}$$

We can construct the limiting $(G, \varnothing)$ up to radius N as follows: fix $\delta > 0$ such that $N_0(\delta) > N$. Then, define $U_{\leq N}(\varnothing)$ as $U_{\leq N}(\varnothing_{n_{i_0+1}})$. Clearly, for $N' > N$, by this definition we have $U_{\leq N}(\varnothing) \subseteq U_{\leq N'}(\varnothing)$ in the subgraph sense (compatibly with the marks). In particular, we can define $(G, \varnothing)$ up to any radius N from the root.

Fix now $\varepsilon > 0$. By definition, there exists $i_0(\varepsilon)$ and $N(\varepsilon)$ such that, for every $i > i_0$, we have $U_{\leq N}(\varnothing) \cong U_{\leq N}(\varnothing_{n_i})$, which implies that $d_{loc}\left((G, \varnothing), (G_{n_i}, \varnothing_{n_i})\right) < \varepsilon$. This proves that for the subsequence $(n_i)_{i\in\mathbb{N}}$ we have a limit in $\mathcal{G}_*$.

We can finish proving that the whole sequence converges. Fix $\varepsilon > 0$. Then there exists $n_0(\varepsilon)$ such that $d_{loc}\left((G_n, \varnothing_n), (G_m, \varnothing_m)\right) < \varepsilon/2$ for every $n, m > n_0$, and there exists $i_0(\varepsilon)$ such that, for every $i > i_0$, $n_i > n_0$ and $d_{loc}\left((G, \varnothing), (G_{n_i}, \varnothing_{n_i})\right) < \varepsilon/2$. As a consequence for any $n > n_0$ and $i > i_0$, by the triangle inequality,

$$\begin{aligned} d_{loc}\left((G, \varnothing), (G_n, \varnothing_n)\right) \leq\ & d_{loc}\left((G, \varnothing), (G_{n_i}, \varnothing_{n_i})\right) \\ & + d_{loc}\left((G_n, \varnothing_n), (G_{n_i}, \varnothing_{n_i})\right) < \varepsilon. \end{aligned}$$

□

The reason that d_{loc} is not a metric is that two rooted marked directed graphs can be at distance 0 without being isomorphic. This is due to the fact that the edges can be explored only in one direction, possibly leaving parts of the graph unexplored, as mentioned above. If the *explorable parts* or *incoming neighborhoods* of two graphs from the roots are isomorphic, then the two rooted marked directed graphs are at distance zero, while these graphs still might not be isomorphic. An example is given in Figure 6.4.

Knowing that $(\mathcal{G}_*, d_{loc})$ is a complete pseudometric space, we can give the following definition:

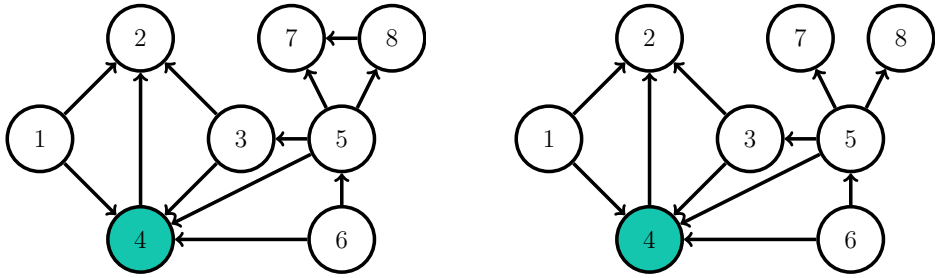


FIGURE 6.4: Example of two rooted marked directed graphs that are at distance zero, but are not isomorphic. The distance between the two graphs is zero since the explorable parts of the graphs from vertex 4 (including vertices 1–6) are isomorphic, but there exists no isomorphisms between the two graphs.

Definition 6.3.6 (Space $\tilde{\mathcal{G}}_*$)**.** *Define the equivalence relation $\sim_*$ on $\mathcal{G}_*$ as follows: two elements $(G_1, \varnothing_1, M(G_1))$ and $(G_2, \varnothing_2, M(G_2))$ are $\sim_*$-equivalent if and only if*

$$d_{loc}((G_1, \varnothing_1, M(G_1)), (G_2, \varnothing_2, M(G_2))) = 0.$$

Denote the quotient space by $\tilde{\mathcal{G}}_*$.

On the space $\tilde{\mathcal{G}}_*$, d_{loc} is a metric. As a consequence, $(\tilde{\mathcal{G}}_*, d_{loc})$ is a Polish space. We point out that by $(G, \varnothing, M(G)) \in \tilde{\mathcal{G}}_*$ we denote an equivalence class of the relation given by $\sim_*$. Denote the explorable neighborhood of the root by $U_\infty(\varnothing)$, i.e., the (possibly infinite) subgraph of a rooted marked directed graph that it is possible to explore from the root. Then

$$d_{loc}((G_1, \varnothing_1, M(G_1)), (G_2, \varnothing_2, M(G_2))) = 0 \quad \iff \quad U_\infty(\varnothing_1) \cong U_\infty(\varnothing_2).$$

Any equivalence class in $\tilde{\mathcal{G}}_*$ is composed by directed marked rooted graphs whose neighborhoods of the root are isomorphic. Heuristically, everything that it is in the part of the graph that is not explorable from the root *does not have any influence on the incoming neighborhood of the root*. This means that any function on $\tilde{\mathcal{G}}_*$ is well defined if and only if it is a function of the neighborhood of the root.

As in the undirected sense, we denote

$$\mathcal{P}(G) = \frac{1}{|V(G)|} \sum_{i \in V(G)} \delta_{(G,i,M(G))}. \tag{6.3.3}$$

When we consider a sequence of marked graphs $((G_n, M(G_n)))_{n\in\mathbb{N}}$, we denote $\mathcal{P}(G_n)$ by $\mathcal{P}_n$. From the definition, we have that $\mathcal{P}(G)$ is a probability on $\tilde{\mathcal{G}}_\star$, that assigns a uniformly chosen root to the marked directed finite graph. Notice that the mark set is fixed. In fact, the triplet $(G, i, M(G))$ is mapped to the equivalence class of the

explorable neighborhood $U_\infty(i)$ of i in G with the same set of marks.

Definition 6.3.7 (Local weak convergence - directed)**.** *Consider a sequence of marked directed graphs $(G_n, M(G_n))_{n\in\mathbb{N}}$. We say that $(G_n, M(G_n))$* converges in the directed local weak sense *to a probability $\mathcal{P}$ on $\tilde{\mathcal{G}}_*$ if, for any bounded continuous function $f : \tilde{\mathcal{G}}_* \to \mathbb{R}$,*

$$\mathbb{E}_{\mathcal{P}_n}[f] \longrightarrow \mathbb{E}_{\mathcal{P}}[f],$$

where $\mathbb{E}_{\mathcal{P}_n}$ and $\mathbb{E}_{\mathcal{P}}$ denote the expectation with respect to $\mathcal{P}_n$ and $\mathcal{P}$.

This definition is similar to Definition 4.1.4 in the undirected case. As the next step, we will give a criterion for the convergence of a sequence of marked directed graphs:

Theorem 6.3.8 (Criterion for directed LWC)**.** *Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of directed graphs. For every $n \in \mathbb{N}$, assign to G_n marks equal to the out-degrees. Then, G_n converges in the directed LW sense to $\mathcal{P}$ if, for every fixed $k \in \mathbb{N}$ and finite directed rooted marked graph $(H, y, M(H))$,*

$$\begin{aligned}&\mathcal{P}_n\left(U_{\leq k}(\varnothing_n) \cong (H, y, M(H))\right)\\ &\quad= \frac{1}{n}\sum_{i\in[n]} \mathbb{1}\left\{U_{\leq k}(i) \cong (H, y, M(H))\right\} \longrightarrow \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y, M(H))\right).\end{aligned}$$

Proof. The proof follows the same argument as in Theorem 4.1.5. For $(H, y, M(H)) \in \tilde{\mathcal{G}}_*$, the functions $\mathbb{1}\{U_{\leq k}(\varnothing) \cong (H, y, M(H))\}$ uniquely identify the explorable part of the marked directed rooted graphs we are considering, which implies we identify the equivalence class that forms the limiting element in $\tilde{\mathcal{G}}_*$. □

The reader can observe that, once the notion of exploration process and isomorphisms in the directed case are introduced, the construction of the definition of local weak convergence for directed graphs is the same as in the undirected case. With the presence of marks we are able to keep track of the out-degree of vertices, while we explore the incoming edges.

We define now the notion of convergence for random graphs. Notice that, considering marked graphs, the marks can be random as well as the graph:

Definition 6.3.9 (Directed LWC (random graphs))**.** *Consider a sequence of directed graphs with vertex marks $(G_n, M(G_n))_{n\in\mathbb{N}}$. Let $\mathcal{P}$ be a probability on $\tilde{\mathcal{G}}_*$. Then:*

(1) We say that $(G_n, M(G_n))_{n\in\mathbb{N}}$ converges in distribution in the directed LW sense *to $\mathcal{P}$ if, for any bounded continuous function $f : \tilde{\mathcal{G}}_* \to \mathbb{R}$,*

$$\mathbb{E}\left[\mathbb{E}_{\mathcal{P}_n}[f]\right] \longrightarrow \mathbb{E}_{\mathcal{P}}[f];$$

(2) We say that $(G_n, M(G_n))_{n\in\mathbb{N}}$ converges in probability in the directed LW sense *to $\mathcal{P}$ if, for any bounded continuous function $f : \tilde{\mathcal{G}}_* \to \mathbb{R}$,*

$$\mathbb{E}_{\mathcal{P}_n}[f] \xrightarrow{\mathbb{P}} \mathbb{E}_{\mathcal{P}}[f];$$

(3) We say that $(G_n, M(G_n))_{n\in\mathbb{N}}$ converges almost surely in the directed LW sense *to* $\mathcal{P}$ *if, for any bounded continuous function* $f : \tilde{\mathcal{G}}_* \to \mathbb{R}$,

$$\mathbb{E}_{\mathcal{P}_n}[f] \xrightarrow{\mathbb{P}-a.s.} \mathbb{E}_{\mathcal{P}}[f],$$

where $\mathbb{E}_{\mathcal{P}_n}$ *denotes the expectation with respect the random choice of the root,* $\mathbb{E}_{\mathcal{P}}$ *the expectation with respect to* $\mathcal{P}$ *and* $\mathbb{E}$ *the expectation with respect to the randomness of the graph sequence.*

Since we are interested in sequences of random graphs, we give the definition of LW convergence only for random graphs:

Similarly to the undirected case, we want to state a criterion for the convergence of random graphs. Again, we consider a sequence of directed random graphs $(G_n)_{n\in\mathbb{N}}$, where we assign marks equal to the out-degrees:

Theorem 6.3.10 (Criterion of convergence for random graphs)**.** *Consider a sequence of directed random graphs* $(G_n)_{n\in\mathbb{N}}$*, and assign marks equal to the out-degrees. Then,*

(1) G_n converges in distribution *in the directed LW sense to* $\mathcal{P}$ *if, for every fixed* $k \in \mathbb{N}$ *and finite directed marked rooted graph* $(H, y, M(H))$*,*

$$\mathbb{E}\left[\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{U_{\leq k}(i) \cong (H, y, M(H))\}\right] \longrightarrow \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y, M(H))\right);$$

(2) G_n converges in probability *in the directed LW sense to* $\mathcal{P}$ *if, for every fixed* $k \in \mathbb{N}$ *and finite directed marked rooted graph* $(H, y, M(H))$*,*

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{U_{\leq k}(i) \cong (H, y, M(H))\} \xrightarrow{\mathbb{P}} \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y, M(H))\right);$$

(3) G_n converges almost surely *in the directed LW sense to* $\mathcal{P}$ *if for every fixed* $k \in \mathbb{N}$ *and finite directed marked rooted graph* $(H, y, M(H))$*,*

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{U_{\leq k}(i) \cong (H, y, M(H))\} \xrightarrow{\mathbb{P}-a.s.} \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y, M(H))\right),$$

where $\mathbb{P}$ *denotes the law of the random sequence* $(G_n)_{n\in\mathbb{N}}$*.*

The proof of Theorem 6.3.10 follows from Theorem 6.3.8.

6.4. Convergence of PageRank

The main result on PageRank is Theorem 6.2.1. It states that, for a locally weakly convergent sequence of directed random graphs $(G_n)_{n\in\mathbb{N}}$, there exists a random variable $R_\varnothing$ such that the PageRank value of a uniformly chosen vertex $R_{V_n}(n)$ satisfies

$$R_{V_n}(n) \xrightarrow{d} R_\varnothing.$$

The random variable $R_\varnothing$ is defined in Proposition 6.4.3 below. Notice that, even though local weak convergence is defined in terms of *local properties* of the graph, it is sufficient for the existence of the limiting distribution for a global property such as PageRank.

The existence of $R_\varnothing$ for a sequence $(G_n)_{n\in\mathbb{N}}$ is assured by the convergence in *distribution* in the local weak sense. If $(G_n)_{n\in\mathbb{N}}$ converges in probability (or almost surely), then the fraction of vertices whose PageRank value exceeds a fixed value $r > 0$ converges in probability (or almost surely) to a deterministic value.

6.4.1. Finite approximation of PageRank

Consider a directed graph G_n, and define the matrix $\boldsymbol{Q}(n)$, where $\boldsymbol{Q}(n)_{i,j} = e_{i,j}/d_i^{(\mathrm{out})}$, for $e_{i,j}$ the number of directed edges from i to j. For $c \in (0,1]$, the PageRank vector $\boldsymbol{\pi}(n) = (\pi_1, \dots, \pi_n)$ is the unique solution of

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(n)\left[c\boldsymbol{Q}(n)\right] + \frac{1-c}{n}\mathbf{1}_n \quad \text{and} \quad \sum_{i=1}^{n}\pi_i = 1, \tag{6.4.1}$$

where $c \in (0,1)$ and $\mathbf{1}_n$ is the vector of all ones of size n. We are interested in the graph-normalized version of PageRank, so $\boldsymbol{R}(n) = n\boldsymbol{\pi}(n)$, which is just the PageRank vector rescaled with the size of the graph. The vector $\boldsymbol{R}(n)$ satisfies

$$\boldsymbol{R}(n) = \boldsymbol{R}(n)\left[c\boldsymbol{Q}(n)\right] + (1-c)\mathbf{1}_n. \tag{6.4.2}$$

Denote Id_n the identity matrix of size n. We can solve (6.4.2) to obtain the well-known expression [8, 13, 23, 45]

$$\boldsymbol{R}(n) = (1-c)\mathbf{1}_n\left[\mathrm{Id}_n - c\boldsymbol{Q}(n)\right]^{-1}. \tag{6.4.3}$$

In practice, the inversion operation on the matrix $\mathrm{Id}_n - c\boldsymbol{Q}(n)$ is inefficient, therefore, power expansion is used to approximate the matrix in (6.4.3) (see e.g. [8]), as

$$\left[\mathrm{Id}_n - c\boldsymbol{Q}(n)\right]^{-1} = \sum_{k=0}^{\infty} c^k \boldsymbol{Q}(n)^k.$$

Notice that $\boldsymbol{Q}(n)^k_{i,j} > 0$ if and only if there exists a path of length exactly k from i to j, possibly with repetition of vertices and edges. Define, for $k \in \mathbb{N}$,

$$\mathrm{path}_i(k) = \{\text{directed path } \ell = (\ell_0, \ell_1, \ell_2, \dots, \ell_k = i)\}.$$

With this notation, we can write, for $i \in [n]$,

$$R_i(n) = (1-c)\left(1 + \sum_{k=1}^{\infty} c^k \sum_{\ell\in\mathrm{path}_i(n)} \prod_{h=1}^{k} \frac{e_{\ell_h,\ell_{h+1}}}{d^+_{\ell_h}}\right), \tag{6.4.4}$$

while the Nth finite approximation of PageRank is

$$R_i^{(N)}(n) = (1-c)\left(1 + \sum_{k=0}^{N} c^k \sum_{\ell \in \mathrm{path}_i(n)} \prod_{h=1}^{k} \frac{e_{\ell_h,\ell_{h+1}}}{d^+_{\ell_h}}\right). \tag{6.4.5}$$

Heuristically, the PageRank formulation in (6.4.4) includes paths of every length, while the Nth approximation in (6.4.5) discards the paths of length $N+1$ or higher. In particular, for every $i \in [n]$, $R_i^{(N)}(n) \uparrow R_i(n)$. One can write the difference between the PageRank and its finite approximation as

$$\left|R_i(n) - R_i^{(N)}(n)\right| = (1-c)\mathbf{1}_n \sum_{k=N+1}^{\infty} (c\boldsymbol{Q}(n))_i^k. \tag{6.4.6}$$

We can prove that we can approximate the PageRank value of a randomly chosen vertex by a finite approximation with an exponentially small error, that is independent of the size of the graph:

Lemma 6.4.1 (Finite iterations)**.** *Consider a directed graph G_n and denote a uniformly chosen vertex by V_n. Then,*

$$\mathbb{E}\left[R_{V_n}(n) - R_{V_n}^{(N)}(n)\right] \leq c^{N+1},$$

where the bound is independent of n.

Proof. Consider (6.4.6) for a uniformly chosen vertex. We have

$$\mathbb{E}\left[R_{V_n}(n) - R_{V_n}^{(N)}(n)\right] = \frac{1-c}{n} \sum_{i=1}^{n} \sum_{k=N+1}^{\infty} \left[\mathbf{1}_n (c\boldsymbol{Q}(n))^k\right]_i. \tag{6.4.7}$$

We write $\boldsymbol{Q}(n)_{j,i}^k$ to denote the element (j,i) of the matrix $\boldsymbol{Q}(n)^k$. We write

$$\left[\mathbf{1}_n (c\boldsymbol{Q}(n))^k\right]_i = c^k \sum_{j=1}^{n} \boldsymbol{Q}(n)_{j,i}^k. \tag{6.4.8}$$

Substituting (6.4.8) in (6.4.7), we obtain

$$\mathbb{E}\left[R_{V_n}(n) - R_{V_n}^{(N)}(n)\right] = (1-c) \sum_{k=N+1}^{\infty} c^k \frac{1}{n} \sum_{i,j} \boldsymbol{Q}(n)_{j,i}^k.$$

Since $\boldsymbol{Q}(n)^k$ is a (sub)stochastic matrix,

$$\sum_{i=1}^{n} \boldsymbol{Q}(n)_{j,i}^k \leq 1$$

for every $j \in [n]$. It follows that

$$\mathbb{E}\left[R_{V_n}(n) - R^{(N)}_{V_n}(n)\right] \leq (1-c)\sum_{k=N+1}^{\infty} c^k \frac{1}{n}\sum_{i=1}^{n} 1 = (1-c)\sum_{k=N+1}^{\infty} c^k = c^{N+1}.$$

□

Lemma 6.4.1 means that we can approximate the PageRank value of a uniformly chosen vertex with an arbitrary precision in a finite number of iterations, that is independent of the graph size. This is the starting point of our analysis.

6.4.2. PageRank on marked directed graphs

In this section we show how the graph-normalized version of PageRank of a uniformly chosen vertex in a sequence of directed graphs $(G_n)_{n\in\mathbb{N}}$ admits a limiting distribution whenever G_n converges in the local weak sense to a distribution $\mathcal{P}$. The advantage is that such a limiting distribution is expressed in terms of functions of $\mathcal{P}$.

The first step is to write PageRank as functions of marked directed rooted graphs that are *bounded and continuous* with respect to the topology given by d_{loc}. In this way, by the definition of local weak convergence, we can pass to the limit and find the limiting distribution.

Fix $n \in \mathbb{N}$. Consider a marked rooted directed graph $(G, \varnothing, M(G)) \in \mathcal{G}_\star$ of size n. Denote as before, for $k \in \mathbb{N}$,

$$\mathrm{path}_\varnothing(k) = \{\text{directed paths } \ell = (\ell_0, \ell_1, \ell_2, \dots, \ell_k = \varnothing)\},$$

i.e., the set of directed paths in $(G, \varnothing, M(G))$ of length exactly $k+1$ whose endpoint is the root $\varnothing$. It is clear that this set is completely determined by $U_{\leq k}(\varnothing)$ in $(G, \varnothing, M(G))$.

Consider a directed marked graph $(G_n, M(G_n))$, where we consider marks equal to the out-degrees. We have that

$$\begin{aligned} R^{(N)}_{V_n}(n) &= \sum_{i\in[n]} \mathbb{1}_{\{V_n=i\}}(1-c)\left(1 + \sum_{k=1}^{N}\sum_{\pi\in\mathrm{path}_i(k)}\prod_{h=1}^{k} c\frac{e_{\pi_h,\pi_{h+1}}}{d^{(\mathrm{out})}_{\pi_h}}\right) \\ &=: R^{(N)}[(G_n, V_n, M(G_n))], \end{aligned} \tag{6.4.9}$$

where the last term in (6.4.9) is a function of a marked rooted graph, evaluated on $(G_n, V_n, M(G_n))$, with V_n a uniformly chosen root. In particular, we can see the Nth approximation of PageRank as a function of the marked rooted graph. We call the function $R^{(N)} : \tilde{\mathcal{G}}_\star \to \mathbb{R}$ the *root N-PageRank*.

Clearly, the root N-PageRank $R^{(N)}$ is a function of $U_{\leq N}(\varnothing)$ only. It depends in fact on the vertices, edges and marks that are considered when exploring the graph from the root up to distance N. Notice that, since the dependence on the marked directed rooted graph is given only by $U_{\leq k}(\varnothing)$, the function $R^{(N)}$ is well defined on any equivalence class in $\tilde{\mathcal{G}}_\star$.

In addition, the function $R^{(N)}$ is *continuous* with respect to the topology generated by d_{loc}. In fact, since $R^{(N)}$ depends only on the root neighborhood up to distance N, whenever two elements $(G, \varnothing, M(G))$ and $(G', \varnothing', M(G'))$ are at distance less than $1/(1+N)$, their roots neighborhoods are isomorphic up to distance $N+1$, which implies that $R^{(N)}[(G, \varnothing, M(G))] = R^{(N)}[(G', \varnothing', M(G'))]$.

The problem is that $R^{(N)}$ *is not bounded*, so LWC does not assure that we can pass to the limit. To resolve this, we introduce a different type of function:

Definition 6.4.2 (Root N-PageRank tail)**.** *Fix $N \in \mathbb{N}$. For $r > 0$, define $\Psi_{r,N} : \tilde{\mathcal{G}}_\star \to \{0,1\}$ by*

$$\Psi_{r,N}\left[(G, \varnothing, M(G))\right] := \mathbb{1}\left\{R^{(N)}\left[(G, \varnothing, M(G))\right] > r\right\}.$$

We call the function $\Psi_{r,N}$ the root N-PageRank tail *at r.*

The function $\Psi_{r,N}$ is clearly bounded, and it depends only on the neighborhood of the root $\varnothing$ up to distance N through the function $R^{(N)}$. This means that, for any $r > 0$, $\Psi_{r,N}$ is continuous.

Since the root N-PageRank on $\tilde{\mathcal{G}}_\star$ represents the Nth approximation of PageRank on directed graphs, it follows that

$$\mathbb{E}_{\mathcal{P}_n}[\Psi_{r,N}] = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}\left\{R_i^{(N)}(n) > r\right\},$$

i.e., $\mathbb{E}_{\mathcal{P}_n}[\Psi_{r,N}]$ is the empirical fraction of vertices in G such that the Nth approximation of PageRank exceeds r. In particular, for every $r \geq 0$, if $G_n \to \mathcal{P}$ in distribution,

$$\mathbb{P}\left(R_{V_n}^{(N)}(n) > r\right) = \mathbb{E}\left[\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\left\{R_i^{(N)}(n) > r\right\}\right] \longrightarrow \mathcal{P}\left(R_\varnothing^{(N)} \geq r\right), \tag{6.4.10}$$

while for convergence in probability (or almost surely), the limit in (6.4.10) exists in probability (or almost surely). Consider the sequence of random variables $(R_\varnothing^{(N)})_{N\in\mathbb{N}}$, where

$$R_\varnothing^{(N)} := R^{(N)}[(G, \varnothing, M(G))],$$

where $(G, \varnothing, M(G))$ is a random directed rooted graph with law $\mathcal{P}$. From (6.4.10), it follows that $R_{V_n}^{(N)}(n) \to R_\varnothing^{(N)}$ in distribution.

We have just proved that, for a sequence of directed graphs $(G_n)_{n\in\mathbb{N}}$ that converges locally weakly to $\mathcal{P}$, any finite approximation of the PageRank value of a uniformly chosen vertex converges in distribution to a limiting random variable, which is given by a function of $\mathcal{P}$.

6.4.3. The limit of finite root ranks

Assume that the sequence $(G_n)_{n\in\mathbb{N}}$ of directed graphs converges to a directed rooted marked graph $(G, \varnothing, M(G))$ with law $\mathcal{P}$. In principle, $(G, \varnothing, M(G))$ can be an infinite directed rooted marked graph. Because of this, we cannot simply take the limit as $N \to \infty$ of the sequence $(R_\varnothing^{(N)})_{N\in\mathbb{N}}$, where $\varnothing$ is the root of $(G, \varnothing, M(G))$,

because the PageRank algorithm is not defined on an infinite graph. Nevertheless, if $\mathcal{P}$ is a LW limit of some sequence of directed graphs, it admits a such limit:

Proposition 6.4.3 (Existence of limiting root rank)**.** *Let $\mathcal{P}$ be a probability on $\tilde{\mathcal{G}}_\star$. If $\mathcal{P}$ is the LW limit in distribution of a sequence of marked directed graphs $(G_n)_{n\in\mathbb{N}}$, then there exists a random variable $R_\varnothing$ with $\mathbb{E}_\mathcal{P}[R_\varnothing] \leq 1$, such that $\mathcal{P}$-a.s. $R_\varnothing^{(N)} \to R_\varnothing$. As a consequence, $\mathcal{P}(R_\varnothing < \infty) = 1$.*

Proof. Clearly, the sequence $(R_\varnothing^{(N)})_{N\in\mathbb{N}}$ is $\mathcal{P}$-a.s. increasing. Therefore, the almost sure limit $R_\varnothing = \lim_{N\to\infty} R_\varnothing^{(N)}$ exists. This is independent of the fact that $\mathcal{P}$ is a LW limit.

By LW convergence, we know that $R_{V_n}^{(N)}(n) \to R_\varnothing^{(N)}$ in distribution. For every $N \in \mathbb{N}$, by Fatou's Lemma we can bound

$$\mathbb{E}_\mathcal{P}\left[R_\varnothing^{(N)}\right] \leq \liminf_{n\in\mathbb{N}} \mathbb{E}\left[R_{V_n}^{(N)}(n)\right] \leq \liminf_{n\in\mathbb{N}} \mathbb{E}\left[R_{V_n}(n)\right] = 1,$$

where the second bound comes from the fact that any N-finite approximation of PageRank is less than the actual PageRank value, and the fact that the graph-normalized PageRank has expected value 1. Since $(R_\varnothing^{(N)})_{N\in\mathbb{N}}$ is increasing, we conclude that there exists $z \leq 1$ such that

$$\mathbb{E}_\mathcal{P}\left[R_\varnothing\right] = \lim_{N\to\infty} \mathbb{E}_\mathcal{P}\left[R_\varnothing^{(N)}\right] = z.$$

□

6.4.4. Proof of Theorem 6.2.1

We start with implication (a) of Theorem 6.2.1. We want to prove that $R_{V_n}(n)$ converges to $R_\varnothing$ in distribution. So, for every $r \geq 0$ and $\varepsilon > 0$ there exists $M(\varepsilon) \in \mathbb{N}$ such that, for every $n \geq M(\varepsilon)$,

$$\left|\mathbb{P}\left(R_{V_n}(n) > r\right) - \mathcal{P}\left(R_\varnothing > r\right)\right| \leq \varepsilon. \tag{6.4.11}$$

We can write, using the triangle inequality,

$$\begin{aligned}\left|\mathbb{P}\left(R_{V_n}(n) > r\right) - \mathcal{P}\left(R_\varnothing > r\right)\right| \leq &\left|\mathbb{P}\left(R_{V_n}(n) > r\right) - \mathbb{E}\left[\mathcal{P}_n\left(R_\varnothing^{(N)} > r\right)\right]\right| \\ &+ \left|\mathbb{E}\left[\mathcal{P}_n\left(R_\varnothing^{(N)} > r\right)\right] - \mathcal{P}\left(R_\varnothing^{(N)} > r\right)\right| \\ &+ \left|\mathcal{P}\left(R_\varnothing^{(N)} > r\right) - \mathcal{P}\left(R_\varnothing > r\right)\right|.\end{aligned} \tag{6.4.12}$$

We show that (6.4.11) holds by proving that every term in the left hand side of (6.4.12) can be bounded by $\varepsilon/3$.

By Lemma 6.4.1 we can bound the first term with c^{N+1} (independently of n). Therefore, defining $N_1 = \log_c(\varepsilon/3)$ and taking $N > N_1$, the first term is bounded by $\varepsilon/3$.

For the last term, we apply Proposition 6.4.3, so we can find $N_2 = N_2(\varepsilon) \in \mathbb{N}$ such

that, for every $N \geq N_2$,

$$\left|\mathcal{P}\left(R_{\varnothing}^{(N)} > r\right) - \mathcal{P}\left(R_{\varnothing} > r\right)\right| \leq \varepsilon/3.$$

Set $N_0(\varepsilon) = \max(N_1, N_2)$. For any $N \geq N_0$, both the first and third terms are bounded by $\varepsilon/3$. Using LW convergence in distribution, we can find $M(N_0, \varepsilon) \in \mathbb{N}$ such that, for every $n \geq M$, the second term is bounded by $\varepsilon/3$. This completes the proof of statement (a).

For statement (b), we need to show that, for every $r > 0$, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{R_i(n) > r\} \xrightarrow{\mathbb{P}} \mathcal{P}\left(R_{\varnothing} > r\right).$$

For every $N \in \mathbb{N} \cup \{\infty\}$ and $r \geq 0$, we denote the empirical fraction of vertices whose Nth approximation of PageRank in G_n exceeds r by

$$\bar{R}(n; r, N) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{R_i^{(N)}(n) > r\}.$$

If $N = \infty$, then $\bar{R}(n; r, N) = \bar{R}(n; r)$ is the empirical tail distribution of PageRank. By LW convergence in probability of $(G_n)_{n\in\mathbb{N}}$, we know that, for every $N \in \mathbb{N}$ and $r > 0$,

$$\bar{R}(n; r, N) \xrightarrow{\mathbb{P}} \mathcal{P}\left(R_{\varnothing}^{(N)} > r\right). \tag{6.4.13}$$

Fix $r > 0$, $\varepsilon > 0$. We need to show that for every $\delta > 0$ there exists $n_0(\delta) \in \mathbb{N}$ such that, for any $n \geq n_0$, $\mathbb{P}\left(\left|\bar{R}(n; r) - \mathcal{P}(R_{\varnothing} > r)\right| \geq \varepsilon\right) \leq \delta$. We can write, for N to be fixed,

$$\begin{aligned}\mathbb{P}\left(\left|\bar{R}(n; r) - \mathcal{P}(R_{\varnothing} > r)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon}\Big[&\mathbb{E}[\bar{R}(n; r) - \bar{R}(n; r, N)] \\ &+ \mathbb{E}[|\bar{R}(n; r, N) - \mathcal{P}(R_{\varnothing}^{(N)} > r)|] \\ &+ |\mathcal{P}(R_{\varnothing}^{(N)} > r) - \mathcal{P}(R_{\varnothing} > r)|\Big].\end{aligned} \tag{6.4.14}$$

Similarly to (6.4.12), we can find n and N large enough such that every term in the right-hand side of (6.4.14) is less than $\delta\varepsilon/3$.

For the first term, we apply Lemma 6.4.1, so we can find N_1 large enough such that $c^{N_1+1} \leq \delta\varepsilon/3$. For the last term, we apply Proposition 6.4.3, so we can find N_2 such that the last term is less than $\delta\varepsilon/3$.

Take $N_0 = \max\{N_1, N_2\}$. Then, by (6.4.13) and the fact that $\{\bar{R}(n; r, N)\}_{n\in\mathbb{N}}$ is uniformly integrable (since $\bar{R}(n; r, N) \leq 1$), we can find n_0 big enough such that

$$\mathbb{E}[|\bar{R}(n; r, N) - \mathcal{P}(R_{\varnothing}^{(N)} > r)|] \leq \delta\varepsilon/3$$

for all $n > n_0$, $N > N_0$. As a consequence, we conclude that, for any $n \geq n_0$,

$$\mathbb{P}\left(\left|\bar{R}(n;r) - \mathcal{P}(R_\varnothing > r)\right| \geq \varepsilon\right) \leq \delta,$$

which proves the convergence in probability.

6.4.5. Undirected graphs

Undirected graphs are in fact a special case of directed graphs, where each link is reciprocated. Theorem 6.2.1 does not make any assumption concerning link reciprocation, and thus it simply holds for undirected graphs as well. In that case, we may use the standard notion of the LWC for undirected graphs, as described in Section 4.1, and it is not hard to see that our notion of directed LW convergence reduces to this.

Let us explain why the special case of undirected graphs deserves our attention. Indeed, usually, undirected graphs are easier to analyze than directed ones. For example, the adjacency matrix of an undirected graph is symmetric, which implies many nice properties. However, PageRank is based on directed paths, and its analysis is greatly simplified when these paths do not contain cycles, with high probability.

For example, PageRank can be written as a product of three terms, one of which is the expected number of visits to i, starting from i, by a simple random walk, which terminates at each step with probability c [14]. Now notice that in undirected graphs, each edge can be traversed in both directions, hence, a path starting at i may return to i in only two steps, so the average number of visits to i will be a random variable that depends on the entire neighborhood. In contrast, e.g., in the directed configuration model, returning to i is highly unlikely. This makes PageRank in undirected graphs hard to analyze, and only few results have been obtained so far (see e.g. [12]).

Our result simultaneously covers the directed and the undirected cases because we only state the equivalence between the behavior of PageRank on a graph and on its limiting object. In this setting, the difficulties that arise in the analysis of PageRank on undirected graphs are, in fact, 'postponed' to the (undirected) limiting random graph.

6.5. Generalized PageRank

6.5.1. Universality of finite approximations

In this section we will show that Theorem 6.2.1 extends to generalized PageRank as given in (6.1.4). We will assume that $C_j \leq c < 1$, $j \in [n]$ are bounded away from one and that the vector $\mathbf{B}_n = (B_i)_{i\in[n]}$ consists of i.i.d. random variables that are independent of the graph G_n, and we let $\mathbb{E}(B_1) = 1 - c$ to keep the argument close to the basic case.

In this generalized setting, the proof of Lemma 6.4.1 goes through almost without changes. Let $\boldsymbol{A}$ be a matrix such that $A_{ij}(n) = C_j e_{ji}/D_j^{(\mathrm{out})}$. Recall that $Q_{ij}(n) =$

$e_{ji}/D_j^{(\mathrm{out})}$. Since $C_i \le c < 1$ holds for all $i \in [n]$,

$$\begin{aligned}
\mathbb{E}\left[R_{V_n}(n) - R^{(N)}_{V_n}(n)\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=N+1}^{\infty}\left[\mathbf{B}_n(\boldsymbol{A}(n))^k\right]_i\right] \\
&\le \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=N+1}^{\infty} c^k\left[\mathbf{B}_n\boldsymbol{Q}(n)^k\right]_i\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=N+1}^{\infty} c^k\sum_{j=1}^{n} B_j\boldsymbol{Q}(n)^k_{j,i}\right] \\
&= \sum_{j=1}^{n}\sum_{k=N+1}^{\infty} c^k\,(1-c)\,\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Q}(n)^k_{j,i} \le c^{N+1},
\end{aligned}$$

where in the final equality we have used the independence of B_j and the graph G_n (and thus $\boldsymbol{Q}(n)$).

Furthermore, Proposition 6.4.3 goes through without changes. The only difference is that additional randomness arises through the random $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$. Therefore, for generalized PageRank, the first and the last terms in (6.4.12) and (6.4.14) can be bounded exactly as before. This is natural because the first and the last terms approximate the PageRank in, respectively, original graph and the limiting graph, by finite iterations, and this approximation does not depend on the random $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ under quite general assumptions.

It remains to analyze the second term in (6.4.12) and (6.4.14). This is more tricky because this term bounds the difference between the finite random graph and the limiting object. Difficulties arise since $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ are associated to vertex labels in $[n]$. This information is lost in the LW limit, therefore additional assumptions are necessary to prove that the second term in (6.4.12) and (6.4.14) is small. We next discuss two possible settings how LWC can be used in the generalized PageRank setting.

6.5.2. Independent $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$

First, we assume that $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ are independent of the graph sequence $(G_n)_{n\in\mathbb{N}}$, and $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ are each i.i.d. sequences that are independent of each other. In this case, on the limiting marked rooted graph $(G, \varnothing, M(G))$ we assign to every vertex $v \in V(G)$ independent samples C_v and B_v. In this case, for $(H, y, M(H))$ a finite marked rooted graph, since $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ are independent of the graph,

$$\begin{aligned}
&\frac{1}{n}\sum_{i\in[n]}\mathbb{P}\left(R_i^{(N)}(n) > r \mid U_{\le N}(i) \cong (H, y, M(H))\right)\mathbb{P}\left(U_{\le N}(i) \cong (H, y, M(H))\right) \\
&= \mathbb{P}(\widehat{R}^{(N)}(H, y, M(H)) > r)\frac{1}{n}\sum_{i\in[n]}\mathbb{P}\left(U_{\le N}(i) \cong (H, y, M(H))\right),
\end{aligned} \tag{6.5.1}$$

where now $\mathbb{1}\{\widehat{R}^{(N)}(H,y,M(H)) > r\}$ is a function of the finite structure given by $(H,y,M(H))$, where the randomness is only given by a finite number of $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$. We note that (6.5.1) only assumes that $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ are independent of the graph sequence $(G_n)_{n\in\mathbb{N}}$. In order to be able to define the local-weak limit, though, we further need the independence and i.i.d. assumptions on $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$. Then, a similar expression to (6.5.1) holds for the limiting graph $(G,\varnothing,M(G))$. As a consequence, the second term in (6.4.12) can be written as

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathcal{P}_n\left(R^{(N)}_{\varnothing} > r\right)\right] - \mathcal{P}\left(R^{(N)}_{\varnothing} > r\right)\right| \\
&\quad= \sum_{(H,y,M(H))} \mathbb{P}(\widehat{R}^{(N)}(H,y,M(H)) > r) \\
&\qquad\times \left|\mathbb{E}[\mathcal{P}_n(U_{\le N}(\varnothing) \cong (H,y,M(H)))] - \mathcal{P}(U_{\le N}(\varnothing)(H,y,M(H)))\right| \qquad (6.5.2)\\
&\quad\le \sum_{(H,y,M(H))} \left|\mathbb{E}[\mathcal{P}_n(U_{\le N}(\varnothing) \cong (H,y,M(H)))] - \mathcal{P}(U_{\le N}(\varnothing)(H,y,M(H)))\right| \\
&\quad= 2d_{\mathrm{TV}}(\mathbb{E}\mathcal{P}_n, \mathcal{P}),
\end{aligned}$$

where $\mathbb{E}\mathcal{P}_n$ is the distribution given by $\mathbb{E}[\mathcal{P}_n(\cdot)]$, and the last term is the total variation (TV) distance between $\mathcal{P}$ and $\mathbb{E}\mathcal{P}_n$. Since $\tilde{\mathcal{G}}_\star$ is discrete, convergence in distribution implies convergence in TV distance, so that $2d_{\mathrm{TV}}(\mathbb{E}\mathcal{P}_n,\mathcal{P}) = o(1)$. The fact that the term $\mathbb{P}(\widehat{R}^{(N)}(H,y,M(H)) > r)$ is the same for the graph sequence and the limit comes from the fact we are looking at expectations of i.i.d. random variables on a given structure $(H,y,M(H))$.

The bound in (6.5.2) is enough to conclude that the generalized PageRank with $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ independent of the graph, and themselves independent i.i.d. sequences, converges in distribution. Here no further assumptions are made on the distributions C and B. Such result does not apply to the convergence in probability, since (6.5.1) is an expectation with respect to the random graph.

In this setting, for $N\in\mathbb{N}$, the limiting distribution $R^{(N)}_{\varnothing}$ of the Nth approximation of PageRank is again a weighted sum of all paths of length at most N that ends at the root $\varnothing$. In particular, $R^{(N)}_{\varnothing}$ is given by

$$R^{(N)}_{\varnothing} = \sum_{k=0}^{N} \sum_{\ell\in \mathrm{path}_{\varnothing}(k)} B_{\ell_k} \prod_{h=1}^{k} \frac{C_{\ell_h}}{m^{(\mathrm{out})}_{\ell_h}}.$$

where now a path $\ell \in \mathrm{path}_{\varnothing}(k)$ contributes with the weight $B_{\ell_k}\prod_{h=1}^{k} C_{\ell_h}/m^{(\mathrm{out})}_{\ell_h}$, and again, all the appearing $(C_i)_{i\ge1}$ and $(B_i)_{i\ge1}$ are independent i.i.d. sequences.

6.5.3. Extended directed LW convergence

The advantage of (6.5.1) is that, once the structure $(H,y,M(H))$ is fixed, the probability that PageRank exceeds r is given by an expectation in terms of $(C_i)_{i\in[n]}$ and

$(B_i)_{i\in[n]}$. Equation (6.5.1) does not extend to convergence in probability, since we are taking expectations. In fact, when considering convergence in probability, we have to prove that the second term in (6.4.14) converges to zero in probability. With a similar argument as the one that we have used to get (6.5.1), for any $(H, y, M(H))$ finite marked directed rooted graph,

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{R_i^{(N)}(n) > r,\ U_{\leq N}(i) \cong (H, y, M(H))\} \\ - \mathcal{P}\left(R_\varnothing^{(N)} > r,\ U_{\leq N}(\varnothing) \cong (H, y, M(H))\right).$$

Here, the convergence in probability of the graph sequence is not enough to conclude that the sum over all possible finite structures $(H, y, M(H))$ is small.

In order to prove this convergence in probability, we need to include $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ as additional marks in the definition of directed marked rooted graphs. In the exploration process described in Definition 6.3.3, to every explored vertex v we assign a mark $m_v^{(\mathrm{out})}$ that is equal to the mark of v in the starting graph. Assuming that $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ take *discrete values*, we can assign a multi-mark $(m_v^{(\mathrm{out})}, C_v, B_v)$ to vertices found in the exploration process. Here, we then need no independence assumptions on $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ w.r.t. the graph G_n, but beware that the notion of multi-marked LWC has become significantly stronger.

This leads to an extended definition of local weak convergence on directed multi-marked rooted graphs, where now the definition of isomorphism (as in Definition 6.3.2) includes the preservation of the multi-marks. More precisely, an isomorphisms between two directed multi-marked rooted graphs $(G, \varnothing, M(G))$ and $(G', \varnothing', M(G'))$ is a map $\gamma : G \to G'$ such that it satisfies Definition 6.3.2 and , for every $v \in V(G)$, $C_{\gamma(v)} = C_v$ and $B_{\gamma(v)} = B_v$.

It is easy to verify that the construction of the extended directed local weak convergence is the same as the one presented in Section 6.3, where now instead of marks we consider multi-marks. As a consequence, the family of functions $(\mathbb{1}\{R_\varnothing^{(N)} > r\})_{N\in\mathbb{N}}$ is continuous with respect to the topology of the extended directed LW convergence, therefore (6.4.14) follows immediately. In the next section, we formalize these two different approaches to LWC.

6.5.4. Formulation of the result for generalized PageRank

We can summarize the results discussed for the generalized PageRank in the following theorem:

Theorem 6.5.1 (Asymptotic generalized PageRank distribution)**.** *Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of directed random graphs. Consider the generalized PageRank as in* (6.1.4)*, where, for $j \in [n]$, $A_j = C_j/D_j^{(\mathrm{out})}$, where C_j's are random variables bounded by $c < 1$ and the random vector $(B_i)_{i\in[n]}$ satisfies $\mathbb{E}(B_1) = 1 - c$ and is independent of G_n. Then, the following holds:*

(a) Assume that $(C_i)_{i\in[n]}$ are i.i.d., $(C_i)_{i\in[n]}$ is independent of $(B_i)_{i\in[n]}$,$(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ are independent i.i.d. sequences that are independent of G_n. If G_n converges

LW in distribution in the sense of Definition 6.3.9, then there exists a distribution $R_\varnothing$ such that $R_{V_n}(n) \xrightarrow{d} R_\varnothing$;

(b) Assume that $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ take discrete values. Then, Theorem 6.2.1 holds for the extended LWC for multi-marked directed graphs defined in Section 6.5.3.

Theorem 6.5.1(a) is given by the independent setting in Section 6.5.2. This method is simpler, in the sense that it does not require additional constructions than the ones used to prove Theorem 6.2.1. On the other hand, it gives a weaker result, since the convergence holds in distribution. Also, we need to assume that $(C_i)_{i\in[n]}$ are i.i.d. and they are independent of $(B_i)_{i\in[n]}$ and the graph G_n. In this case, it is not clear what the appropriate conditions are under which LWC in probability holds.

Theorem 6.5.1(b) depends on the extended LWC notion of Section 6.5.3. The reformulation of LWC requires less assumptions, in the sense that now we allow the distribution $(D^{(\mathrm{in})}, D^{(\mathrm{out})}, C, B)$ to have dependent components. The disadvantage is that, to incorporate $(C_i)_{i\in[n]}$ and $(B_i)_{i\in[n]}$ in the definition of isomorphism, we require them to take discrete values, and the notion of LWC is stronger. This might not be suitable for applications. We next remark about a possible way to avoid this unnatural discreteness assumption:

Remark 6.5.2 (Weighted rooted graphs)**.** Benjamini, Lyons and Schramm [19] consider undirected LWC in the case of weighted edges. In particular, they define a different metric on the space of weighted rooted graphs, that includes the distance between edge weights. This construction can be extended to vertex weights, and it would lead to a different approach to investigate generalized PageRank. This requires additional work, for example due to the fact that the metric in [19] is not a simple extension of the metrics that we consider in Sections 4.1 and 6.3. We refrain from studying this further.

6.6. Examples of directed local weak convergence

6.6.1. Directed configuration model

The directed configuration model (DCM) is a version of the configuration model defined in Section 3.1.1, where half-edges are labeled as in- and out-half-edges. In this setting, DCM_n is a directed graph of size $n \in \mathbb{N}$ with prescribed in- and out-degree sequences. We denote the in-degree sequence by $\boldsymbol{D}_n^{(\mathrm{in})} = (D_1^{(\mathrm{in})}, \dots, D_n^{(\mathrm{in})})$ and the out-degree sequence by $\boldsymbol{D}_n^{(\mathrm{out})} = (D_1^{(\mathrm{out})}, \dots, D_n^{(\mathrm{out})})$. We call $(\boldsymbol{D}_n^{(\mathrm{out})}, \boldsymbol{D}_n^{(\mathrm{in})})$ the *bi-degree sequence* of the graph.

For a precise description of DCM, we refer to [47, 46]. The graph is defined as follows: let $n \in \mathbb{N}$ be the size of the graph, and fix a bi-degree sequence $(\boldsymbol{D}_n^{(\mathrm{out})}, \boldsymbol{D}_n^{(\mathrm{in})})$. The graph is generated by fixing a free outgoing half edge and we pair it uniformly at random with a free incoming half edge. In this process, self loops and multiple edges are allowed. Until the pairing is made uniformly, it is not relevant in which order we choose the free outgoing half-edge. In this setting, the total in-degree and

the out-degree of the graph have to be equal. In the case of random in- and out-degrees, this is a rare event. The algorithm presented in [47] generates an admissible bi-degree sequence in a finite number of steps, and approximates the initial degree distributions.

Condition 6.6.1 (Bi-degree regularity conditions)**.** *Let* $(\boldsymbol{D}_n^{(\mathrm{out})}, \boldsymbol{D}_n^{(\mathrm{in})})$ *be a bi-degree sequence. Then, the* bi-degree regularity conditions *are as follows:*

(a) There exists a distribution $(p(h,l))_{h,l\in\mathbb{N}}$ *such that, for every* $h, l \in \mathbb{N}$*, as* $n \to \infty$*,*

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{D_i^{(\mathrm{out})}=h, D_i^{(\mathrm{in})}=l\}} \longrightarrow p(h,l); \tag{6.6.1}$$

(b) Denote by $(\mathcal{D}^{(\mathrm{out})}, \mathcal{D}^{(\mathrm{in})})$ *a pair of random variables with distribution* $(p(h,l))_{h,l\in\mathbb{N}}$ *as in* (6.6.1)*. Then, as* $n \to \infty$*,*

$$\frac{1}{n}\sum_{i\in[n]} h\mathbb{1}_{\{D_i^{(\mathrm{out})}=h\}} \longrightarrow \mathbb{E}\left[\mathcal{D}^{(\mathrm{out})}\right], \qquad \frac{1}{n}\sum_{i=1}^{n} l\mathbb{1}_{\{D_i^{(\mathrm{in})}=l\}} \longrightarrow \mathbb{E}\left[\mathcal{D}^{(\mathrm{in})}\right], \tag{6.6.2}$$

and $\mathbb{E}\left[\mathcal{D}^{(\mathrm{in})}\right] = \mathbb{E}\left[\mathcal{D}^{(\mathrm{out})}\right]$*;*

(c) For $L_n = D_1^{(\mathrm{out})} + \cdots + D_n^{(\mathrm{out})}$*, as* $n \to \infty$*,*

$$\frac{1}{n}\sum_{i\in[n]} \frac{h}{L_n}\mathbb{1}_{\{D_i^{(\mathrm{out})}=h, D_i^{(\mathrm{in})}=l\}} \longrightarrow \frac{k}{\mathbb{E}[\mathcal{D}^{(\mathrm{out})}]} p(h,l) =: p^\star(h,l). \tag{6.6.3}$$

Denote by $(\mathcal{D}^{\star(\mathrm{out})}, \mathcal{D}^{(\mathrm{in})})$ *a pair of random variable with distribution* $(p^\star(h,l))_{h,l\in\mathbb{N}}$*.*

Condition 6.6.1(a) implies that the empirical bi-degree distribution converges to a limiting distribution given by $(p(h,l))_{h,l\in\mathbb{N}}$ as in (6.6.1). Condition 6.6.1(b) implies that both the in- and out-degree distributions have finite first moment, equal to the one of $(p(h,l))_{h,l\in\mathbb{N}}$. Condition 6.6.1(c) implies that the out-degree size-biased distribution converges to a limiting distribution $(p^\star(h,l))_{h,l\in\mathbb{N}}$ as in (6.6.3).

With Condition 6.6.1, we are ready to state the convergence result on DCM:

Proposition 6.6.2. *Consider a directed configuration model* DCM_n *such that the bi-degree sequence* $(\boldsymbol{D}_n^{(\mathrm{out})}, \boldsymbol{D}_n^{(\mathrm{in})})$ *satisfies Condition 6.6.1. Then,* DCM_n converges in probability in the directed LW sense *to the law* $\mathcal{P}$ *of a* marked *Galton-Watson tree, where*

(1) edges are directed from children to parents;

(2) the mark and the in-degree of the root are distributed as $(\mathcal{D}^{(\mathrm{out})}, \mathcal{D}^{(\mathrm{in})})$ *as in* (6.6.1)*;*

(3) the mark and the in-degree of any other vertex are independent across the tree vertices, and are distributed according to $(\mathcal{D}^{\star(\mathrm{out})}, \mathcal{D}^{(\mathrm{in})})$ *as in* (6.6.3)*.*

The proof of Proposition 6.6.2 is an adaptation of the proof for the undirected case as presented in [87, Section 2.2.2]. The proof is divided in two parts. First, we use a

coupling argument to prove that DCM_n converges in distribution to the prescribed limit. The second part consists in the application of the second moment method on the number of vertices in DCM_n with a fixed finite neighborhood structure, to prove that the number of such vertices is concentrated around its mean.

We start with the coupling argument:

Lemma 6.6.3 (LW convergence of DCM in distribution)**.** *Fix a finite marked rooted tree $(H, y, M(H))$. Under the assumptions of Proposition 6.6.2 there exists a marked Galton-Watson tree $\mathrm{GW}^{(n)}$ such that*

$$\mathbb{P}\big(U_{\leq k}(V_n) \cong (H, y, M(H))\big) = \mathbb{P}\left(\mathrm{GW}^{(n)}_{\leq k} \cong (H, y, M(H))\right) + o(1), \tag{6.6.4}$$

where $\mathrm{GW}^{(n)}_{\leq k}$ denote the first k generations of $\mathrm{GW}^{(n)}$. Further, $\mathrm{GW}^{(n)} \to \mathcal{P}$ locally weakly in distribution, where $\mathcal{P}$ is the limit in Proposition 6.6.2. As a consequence, $\mathrm{DCM}_n \to \mathcal{P}$ locally weakly in distribution.

Proof. We prove that, for every finite $k \in \mathbb{N}$ and n large enough, the k-neighborhood of a uniform chosen vertex in DCM_n has approximately the same distribution as the first k generations of a marked Galton-Watson tree $\mathrm{GW}^{(n)}$, where marks and offspring distributions in $\mathrm{GW}^{(n)}$ depends on n. Define $(p_n(h,l))_{h,l\in\mathbb{N}}$ and $(p^\star_n(h,l))_{ih,l\in\mathbb{N}}$ by

$$\begin{aligned} p_n(h,l) &= \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{D_i^{(\mathrm{out})}=h, D_i^{(\mathrm{in})}=l\}}, \\ p^*_n(h,l) &= \frac{1}{n}\sum_{i\in[n]} \frac{h}{L_n}\mathbb{1}_{\{D_i^{(\mathrm{out})}=h, D_i^{(\mathrm{in})}=l\}}, \end{aligned} \tag{6.6.5}$$

where $L_n = D_1^{(\mathrm{out})} + \cdots + D_n^{(\mathrm{out})}$.

The coupling is constructed as follows: the mark and the degree of the root both in $U_{\leq k}(V_n)$ and in $\mathrm{GW}^{(n)}$ are chosen according to the distribution p_n as in (6.6.5). Therefore, $U_{\leq 0}(V_n)$ and the 0-generation of $\mathrm{GW}^{(n)}$ (which both consist only of the root and its mark) are the same.

We have to construct $U_{\leq k}(V_n)$ and $\mathrm{GW}^{(n)}_{\leq k}$ at the same time. Conditioning on $U_{\leq k-1}(V_n)$ and $\mathrm{GW}^{(n)}_{\leq k-1}$, the new exploration step from $U_{\leq k-1}(V_n)$ to $U_{\leq k}(V_n)$ is made as follows: assuming that during the exploration up to distance $k-1$ we have created t edges, take the first unpaired incoming half-edge x_{t+1}, that we pair to a uniformly chosen outgoing half-edge that is not paired yet. We choose this outgoing half-edge y_{t+1} uniformly at random among all outgoing half-edges, independently from the previously matched half-edges.

Let W_{t+1} be the vertex in DCM_n to which y_{t+1} is incident. Then, in $\mathrm{GW}^{(n)}_{\leq k}$ we assign to a new vertex mark and in-degree equal to $(D^{(\mathrm{out})}_{W_{t+1}}, D^{(\mathrm{in})}_{W_{t+1}})$. Notice that in this case the pair $(D^{(\mathrm{out})}_{W_{t+1}}, D^{(\mathrm{in})}_{W_{t+1}})$ is distributed as $p^\star_n$ given in (6.6.5).

In $U_{\leq k-1}(V_n)$ we have to be careful since the half-edge y_{t+1} might have already been paired. If y_{t+1} has not been paired yet, then we pair x_{t+1} to y_{t+1} to create an

edge. If y_{t+1} has already been paired, then we draw a new outgoing half-edge y'_{t+1} chosen uniformly from the unpaired ones.

We do this procedure for every ingoing half-edge $x_{t+1}, \dots, x_{t+s}$, where s is the number of unpaired ingoing half-edges in $U_{\leq k-1}(V_n)$. We can have differences between the exploration process in DCM_n and $\mathrm{GW}^{(n)}$. Differences can happen in two ways:

(1) the outgoing half-edge that we select to create a new edge has already been paired;
(2) the outgoing half-edge that we select to create a new edge has not been paired yet, but it is incident to a vertex already found in the exploration process.

These two contributions have small probability. In fact, after creating t edges, the probability that we select an outgoing half-edge that is already used is equal t/L_n, where L_n is the total number of outgoing edges. This means that the probability that in the first s steps we use the same out-going half-edge twice is bounded by

$$\sum_{t=0}^{s} \frac{t}{L_n} = \frac{s(s+1)}{2L_n}. \tag{6.6.6}$$

Thanks to Condition 6.6.1(b), L_n is of order n, so the expression in (6.6.6) is $o(1)$ whenever $s = o(\sqrt{n})$. The probability of selecting a vertex i when choosing an outgoing half-edge is $D_i^{(\mathrm{out})}/L_n$. Then, the probability that a vertex i is selected at least twice when t edges are created is bounded by

$$\frac{t(t+1)}{2} \frac{(D_i^{(\mathrm{out})})^2}{L_n^2} \tag{6.6.7}$$

Using (6.6.7) and the union bound, the probability that a vertex is selected twice when T edges are created is bounded by

$$\frac{t(t+1)}{2} \sum_{i=1}^{n} \frac{(D_i^{(\mathrm{out})})^2}{L_n^2} \leq \frac{t(t+1)}{2L_n} D_{max}^{(\mathrm{out})}, \tag{6.6.8}$$

where $D_{max}^{(\mathrm{out})}$ is the maximum out-degree in the bi-degree sequence. In this case, the expression in (6.6.8) is $o(1)$ when $s = o(\sqrt{n/D_{max}^{(\mathrm{out})}})$. Further, $D_{max}^{(\mathrm{out})}$ under Condition 6.6.1 is $o(n)$.

The two bounds in (6.6.6) and (6.6.8) together holds for $s = s(n)$, with $s(n) \to \infty$ sufficiently slowly. Since any finite tree H is made by a finite number of edges S, we can take n large enough such that $s(n) \geq S$. This implies (6.6.4). Note that from (6.6.4) it directly follows that

$$\mathbb{P}\left(U_{\leq k}(V_n) \cong \mathrm{GW}_{\leq k}^{(n)}\right) = 1 - o(1).$$

Finally, since the distributions p_n and $p_n^\star$ converge respectively to p and $p^\star$ as defined

in Condition 6.6.1, and (6.6.4) holds for any finite marked rooted tree $(H, y, M(H))$, we have proved that DCM_n converges locally weakly in distribution to $\mathcal{P}$. □

Next we prove the convergence in probability, using the second moment method on the number of vertices in DCM_n with a prescribed neighborhood $(H, y, M(H))$.

Lemma 6.6.4 (Second moment method). *Fix $k \in \mathbb{N}$ and a finite structure $(H, y, M(H))$ for the root neighborhood. Let $N_k(H, y, M(H))$ be the number of vertices i in DCM_n such that $U_{\leq k}(i) \cong (H, y, M(H))$. Then, as $n \to \infty$,*

$$\frac{1}{n^2}\mathbb{E}\left[N_k(H, y, M(H))^2\right] \longrightarrow \mathcal{P}\left(U_{\leq k}(\varnothing) \cong (H, y, M(H))\right)^2. \tag{6.6.9}$$

Proof. We can rewrite

$$\frac{\mathbb{E}\left[N_k(H, y, M(H))^2\right]}{n^2} = \mathbb{P}\left(U_{\leq k}(V_n^1) \cong (H, y, M(H)), U_{\leq k}(V_n^2) \cong (H, y, M(H))\right),$$

where V_n^1 and V_n^2 are two vertices chosen uniformly at random in DCM_n. Since we fix $k \in \mathbb{N}$, we can take n large enough such that, with high probability, V_n^2 is not a vertex found in the exploration up to distance $2k$ from V_n^1. Then we can rewrite the probability in the right-hand side of (6.6.9) as

$$\mathbb{P}\left(U_{\leq k}(V_n^1) \cong (H, y, M(H)), U_{\leq k}(V_n^2) \cong (H, y, M(H)), V_n^2 \notin U_{\leq 2k}(V_n^1)\right) + o(1),$$

where the factor $2k$ comes from the fact that we look at the structure $(H, y, M(H))$ for the two neighborhoods when they are disjoint. With a similar argument to the one just used, since k is fixed,

$$\mathbb{P}\Big(U_{\leq k}(V_n^1) \cong (H, y, M(H)), V_n^2 \notin U_{\leq 2k}(V_n^1)\Big) \longrightarrow \mathcal{P}\Big(U_{\leq k}(\varnothing) \cong (H, y, M(H))\Big). \tag{6.6.10}$$

We now use the fact that, conditioning on the existence of a tree in DCM_n, the probability to have a second tree disjoint from the first one is equal to have a tree in a different configuration model with different size and bi-degree distribution. More precisely, conditioning on $\{U_{\leq k}(V_n^1) \cong (H, y, M(H)), V_n^2 \notin U_{\leq 2k}(V_n^1)\}$, we want to evaluate the probability of having a second tree $U_{\leq k}(V_n^2) \cong (H, y, M(H))$, disjoint from $U_{\leq k}(V_n^1) \cong (H, y, M(H))$. We have that

$$\begin{aligned}&\mathbb{P}\Big(\, U_{\leq k}(V_n^2) \cong (H, y, M(H)) \big|\, U_{\leq k}(V_n^1) \cong (H, y, M(H)), V_n^2 \notin U_{\leq 2k}(V_n^1)\Big) \\ &= \mathbb{P}\left(\widehat{U}_{\leq k}(\widehat{V}_n^2) \cong (H, y, M(H)),\, \widehat{j} \notin \widehat{U}_{\leq k}(\widehat{V}_n^2)\right),\end{aligned} \tag{6.6.11}$$

where $\widehat{U}_{\leq k}(\widehat{V}_n^2)$ is the k-neighborhood of a vertex $\widehat{V}_n^2$ chosen uniformly at random in a different configuration model $\widehat{\mathrm{DCM}}_n$, and $\widehat{j}$ is a particular vertex in $\widehat{\mathrm{DCM}}_n$ whose characteristics are specified below.

The vertices set and bi-degree sequence of $\widehat{\mathrm{DCM}}_n$ are defined as follows:

(1) if $i \notin U_{\leq k}(V_n^1)$, then i is a vertex in $\widehat{\mathrm{DCM}}_n$ with the same in- and out-degree $(D_i^{(\mathrm{out})}, D_i^{(\mathrm{in})})$;

(2) if $i \in U_{\leq k}(V_n^1)$, then i is not present in $\widehat{\mathrm{DCM}}_n$;

(3) define an additional vertex $\widehat{j}$ in $\widehat{\mathrm{DCM}}_n$, with in- and out-degree $(\widehat{D^{(\mathrm{out})}}_j, \widehat{D^{(\mathrm{in})}}_j)$, where $\widehat{D^{(\mathrm{out})}}_j$ equals the sum of the unpaired outgoing half-edges in $U_{\leq k}(V_n^1)$, and $\widehat{D^{(\mathrm{in})}}_j$ equals the number of unpaired ingoing half-edges in $U_{\leq k}(V_n^1)$. We point out that $\widehat{U}_{\leq k}(\widehat{V}_n^2)$ needs to avoid $\widehat{j}$.

Notice that the unpaired incoming half-edges in $U_{\leq k}(V_n^1)$ are incident only to vertices at distance k from the root, while the unpaired outgoing half-edges are incident to all vertices in $U_{\leq k}(V_n^1)$. We have that $\widehat{\mathrm{DCM}}_n$ is a graph with $n - |U_{\leq k}(V_n^1)| + 1$ vertices, and a different bi-degree sequence.

The graph $\widehat{\mathrm{DCM}}_n$ is then created by pairing an incoming half-edge to a uniformly chosen outgoing half-edge, as usual as in the regular DCM_n. The probability to observe a structure in $\widehat{\mathrm{DCM}}_n$ that is disjoint from the vertex $\widehat{j}$ is exactly the same as in the regular DCM_n, conditioning on the structure of $U_{\leq k}(V_n^1)$. This explain the equality in (6.6.11).

It is immediate to verify that the bi-degree sequence of $\widehat{\mathrm{DCM}}_n$ satisfies Condition 6.6.1, since we modify a negligible fraction of vertices (recall that k is fixed). As a consequence,

$$\mathbb{P}\left(\widehat{U}_{\leq k}(\widehat{V}_n^2) \cong (H, y, M(H))\right) \longrightarrow \mathcal{P}\Big(U_{\leq k}(\varnothing) \cong (H, y, M(H))\Big). \tag{6.6.12}$$

Using together (6.6.10) and (6.6.12), we complete the proof of (6.6.9). □

DCM with independent in- and out-degrees. In [46] the limiting distribution of PageRank in DCM has been obtained when the size-biased in- and out-degrees are independent:

$$p^\star(h,l) = \frac{h}{\mathbb{E}[\mathcal{D}^{(\mathrm{out})}]}\mathbb{P}\left(\mathcal{D}^{\star(out)} = h\right)\mathbb{P}\left(\mathcal{D}^{(\mathrm{in})} = l\right).$$

Notice that $\mathcal{D}^{(\mathrm{out})}$ and $\mathcal{D}^{(\mathrm{in})}$ can, in general, be dependent, that is, $\mathcal{D}^{(\mathrm{in})}$ may have a different distribution conditioned on the event $\{\mathcal{D}^{(\mathrm{out})} \neq 0\}$, because the vertices with zero out-degrees do not contribute in PageRank of other vertices.

The local weak convergence for this case follows from [46, Lemma 5.4], hence, our Theorem 6.2.1 provides an alternative argument for the existence of the limiting PageRank distribution. It has been proved in [46], under some technical assumptions, that in the limit the PageRank is distributed as

$$\mathcal{R} \stackrel{d}{=} \sum_{i=1}^{\mathcal{N}} \frac{c}{\mathcal{D}_i^{\star(\mathrm{out})}}\mathcal{R}_i^\star + (1-c), \tag{6.6.13}$$

where $\mathcal{R}^\star$ are independent realizations of the endogenous solution of the stochastic fixed-point equation

$$\mathcal{R}^\star \stackrel{d}{=} \sum_{i=1}^{\mathcal{N}^\star} \frac{c}{\mathcal{D}_i^{\star(\mathrm{out})}} \mathcal{R}_i^\star + (1-c). \tag{6.6.14}$$

The recursion (6.6.14) has been studied in a number of papers, see [101, 162], and further references in [46]. The argument in [46] is more general, in fact the authors consider generalized PageRank as solution of a more general equation than (6.6.14), where the $(1-c)$ is replaced by a random variable $\mathcal{B}$. In particular, if $\mathcal{D}^{(\mathrm{in})}$ is regularly varying with a tail heavier than the tail of $\mathcal{B}$, then the limiting PageRank $\mathcal{R}$ follows a power law with the same exponent as the in-degree $\mathcal{D}^{(\mathrm{in})}$.

6.6.2. Inhomogeneous random graphs

In the directed inhomogeneous random graphs, each vertex i receives an in-weight $W_i^{(\mathrm{in})}$ and an out-weight $W_i^{(\mathrm{out})}$. There is a directed edge from vertex i to vertex j with probability $w_{ij}^{(n)}$, which depends on $W_i^{(\mathrm{out})}$ and $W_j^{(\mathrm{in})}$. Lee and Olvera-Cravioto [110] study PageRank in the class of inhomogeneous random graphs that satisfy the assumption

$$w_{ij}^{(n)} = \min\left\{1, \frac{W_i^{(\mathrm{out})} W_j^{(\mathrm{in})}}{\theta\, n}(1+\phi_{ij}(n))\right\},$$

where $\phi_{ij}(n)$ satisfies some technical conditions, and is in fact vanishing as $n \to \infty$ for most natural models. This formulation includes Erdős-Rényi model, the Chung-Lu model, the Poissonian random graph and the generalized random graph. For a detailed analysis of the properties of such directed graphs we refer to [39].

LWC for this class of graphs follows directly from [110, Theorem 3.6] under general conditions, including that the in- and out-weights are allowed to be dependent. Hence, our results imply that PageRank converges in this model as well, to the PageRank of the limiting random graph.

In the case when the in- and out-weights are asymptotically independent, it is proved in [110] that the PageRank converges to the attracting endogenous solution of stochastic recursion (6.6.14). In particular, a power-law distribution of in-weights implies the power-law distribution of PageRank.

6.6.3. Directed CTBP trees

CTBPs, as defined in in Section 2.1, are typically seen as undirected trees. The trees defined by CTBPs can be easily seen as directed trees, by assuming that edges are directed from children to parents. With this remark we can formulate the convergence result:

Proposition 6.6.5 (LWC for CTBPs trees)**.** *Consider a supercritical and Malthusian birth process $(\xi_t)_{t\geq 0}$. Denote the corresponding CTBP by $\boldsymbol{\xi}$. Let $\mathcal{T}(t)$ be the directed random tree defined by $\boldsymbol{\xi}$ at time t, where edges are directed from children to parents. Then, on the event $\{|\mathcal{T}(t)| \to \infty\}$, $\mathcal{T}(t)$* converges $\mathbb{P}$-a.s. in the LW sense *to the law of $\mathcal{T}(T_{\alpha^*})$, where*

(1) all marks are 1;

(2) edges are directed from children to parents;

(3) T_{α^*} is an exponentially distributed random variable with parameter α^* (the Malthusian parameter of the CTBP).

Proof. First of all, at every $t \in \mathbb{R}^+$, $\mathcal{T}(t)$ is a directed finite tree. We can equivalently prove the result on the discrete sequence $(\mathcal{T}_n)_{n\in\mathbb{N}}$, where $\mathcal{T}_n = \mathcal{T}(\tau_n)$, for $(\tau_n)_{n\in\mathbb{N}}$ the sequence of birth times of the CTBP.

Denote the vertices in $\mathcal{T}_n$ by their birth order, which means that the root of $\mathcal{T}_n$ in the sense of CTBP is vertex 1. First of all, notice that, for every $i \in [n]$ and $N \in \mathbb{N}$, the N neighborhood $U_{\leq N}(i)$ in the directed marked rooted graphs $(\mathcal{T}_n, i, 1)$ is just the subtree rooted at i composed by the descendants of i only up to generation N (from i). Notice that every vertex has out-degree 1 except for vertex 1 since it has out-degree 0.

What we need to prove is that, for any finite directed rooted tree (H, y) of depth N and with mark 1 for every vertex, we have, as $n \to \infty$,

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}\left\{U_{\leq N}(i) \cong (H, y)\right\} \overset{\mathbb{P}-a.s.}{\longrightarrow} \mathbb{P}\left(U_{\leq N}(\varnothing) \cong (H, y)\right), \tag{6.6.15}$$

where $U_{\leq N}(\varnothing)$ is the N-neighborhood of the root $\varnothing$ in the random tree $\mathcal{T}(T_{\alpha^*})$. For every $i \in [n]$ the indicator function inside the expectation satisfies the definition of random characteristic, since it is a bounded function that, for every individual i in the branching population, depends only on the birth time τ_i and on the randomness associated to i and its descendants. As a consequence, the result follows by (2.1.4). □

This result resembles the subtree counting result in [152, Theorem 2]. Notice that the limiting rooted graph in Proposition 6.6.5 is finite with probability 1. This is rather different than the undirected settings, where typically the limiting rooted graph is infinite when considering a sequence of graphs with growing size.

Remark 6.6.6 (Non-recursive property of PageRank). the behavior of PageRank is often investigated starting from the recursive distributional equation in (6.1.4). In particular, the solution of (6.1.4) is constructed using a weighted Galton-Watson tree. This construction is based on the fact that the subtree rooted at every vertex is again a Galton-Watson tree with the same distribution.

In some cases, the construction is adapted to allow the root to have different degree and mark, but all other vertices have i.i.d. characteristics. As an example, we refer to [46], where PageRank on directed configuration model is investigated (in the independent case, see Section 6.6.1).

When we consider CTBPs, we have proved that the graph-normalized PageRank converges to the PageRank value of the root in a tree with distribution $\mathcal{T}(T_{\alpha^*})$. In particular, the processes $\{(\xi_t)_{t\geq 0}^x\}_{x\in\mathcal{U}}$ that define $\mathcal{T}(T_{\alpha^*})$ are i.i.d., but they are evaluated at random dependent times $(T_{\alpha^*} - \tau_x)_{x\in\mathcal{U}}$. Thus, the solution based on a weighted

Galton-Watson tree does not apply to the PageRank in CTBPs, as the CTBP is inhomogeneous.

6.6.4. Preferential attachment model

We can split PAMs in two cases, according to the fact that we assume $m = 1$ or $m > 1$. In the case $m = 1$, we can use embedding birth processes as in Definition 2.1.14 (see also Remark 2.1.15) to interpret a CTBP as the continuous-time version of the PAM with out-degree 1, then the directed local weak limit is given by Proposition 6.6.5.

For $m \geq 2$, PAM is no longer a tree, so we have to use the knowledge of the undirected LW limit as in Chapter 4. In fact, we can define a directed version of LW limit of PAMs, the Pólya point tree. The directed version of the Pólya point tree is the following:

Definition 6.6.7 (Directed Pólya point tree)**.** *The* directed Pólya point tree *is an infinite marked rooted random tree constructed as follows: let $m \geq 1$ and $\delta > -m$ be parameters for a preferential attachment model $(\mathrm{PA}_t(m, \delta))_{t\in\mathbb{N}}$. Let*

(a) $\chi = (m + \delta)/(2m + \delta)$, $\phi = (1 - \chi)/\chi$;
(b) Γ_{in} *denote a Gamma distribution with parameters $m + \delta$ and 1;*

Vertices in the graph have three characteristics:

(a) a label i *in the Ulam-Harris set;*
(b) a position $x \in [0, 1]$*;*
(c) a positive number γ called strength*;*

In addition, every vertex has mark m (in the sense of Definition 6.3.3). Assign to $\varnothing$ a position $x_\varnothing = U^\chi$, where U is a uniform random variable on $[0, 1]$, and a strength $\gamma_\varnothing \sim \Gamma_{in}$. Set $\varnothing$ as unexplored. Then, recursively over the elements in the set of unexplored vertices, according to the shortlex order:

(1) let i denote the current unexplored vertex;
(2) assign to i a strength value $\gamma_i \sim \Gamma_{in}$;
(3) let $u_{i1}, \ldots, u_{iD_i^{(\mathrm{in})}}$ be the random $D_i^{(\mathrm{in})}$ points given by an independent Poisson process on $[u_i, 1]$ with density

$$\rho_i(x) = \gamma_i \frac{\phi x^{\phi-1}}{x_i^\phi}.$$

(4) draw an edge from each one of the vertices $i1, \ldots, id_i^{(\mathrm{in})}$ to i;
(5) set $x_{i1}, \ldots, x_{id_i^{(\mathrm{in})}}$ unexplored and i as explored.

Definition 6.6.7 is obtained by the definition of the undirected Pólya point tree in Section 4.3.1, where the exploration of the neighborhood of a vertex is limited to the

exploration of *younger vertices*. In other words, the exploration from a vertex i is made only over vertices with index $j > i$.

With the definition of the directed Pólya point tree, we can state the directed LWC result for PAMs:

Proposition 6.6.8 (LW limit of directed PAM)**.** *Fix $m \geq 1$ and $\delta > -m$. Let $(\mathrm{PA}_t(m,\delta))_{t\in\mathbb{N}}$ be a PAM (of any type (a)-(f) listed in Section 4.3). Denote by $(\mathrm{DPA}_t(m,\delta))_{t\in\mathbb{N}}$ the directed version of $(\mathrm{PA}_t(m,\delta))_{t\in\mathbb{N}}$, where edges are directed from young to old vertices. Then,* $\mathrm{DPA}_t(m,\delta)$ converges in probability in the directed LW sense *to the directed Pólya point tree as in Definition 6.6.7.*

The proof of Proposition 6.6.8 follows immediately from [21, Theorem 2.2] and the fact that the exploration process in DPA_t corresponds to exploring only younger vertices.

Remark 6.6.9 (**Non-recursive property of PageRank).** Similarly to Remark 6.6.6 about CTBPs, we point out that the PageRank value of the root of a directed Pólya point tree does not satisfy the recursive property that is necessary to consider it as a solution of (6.1.4). Notice that the Poisson point process assigned to vertex i in Definition 6.6.7 is defined on the interval $[x_i, 1]$, where the position x_i depends on the ancestors (in the Ulam-Harris sense) of i.

Another way to interpret this is that the family of Poisson point process in Definition 6.6.7 is composed by i.i.d. processes parametrized by the positions of vertices, that are dependent random variables. This suggests that the positions in the Pólya point tree play the same role as the birth times in CTBPs.

7

Fitness and aging in PA trees

Content and structure of the chapter

This chapter is mainly focused on the condition under which a continuous-time branching process (CTBP) can generate a random tree where the in-degree distribution obeys a power law. In particular, moving from observations of citation networks as discussed in Section 1.8, we focus our attention on the setting where the rate at which an individual produces children depends on the number of previous children, the age of the individual itself and an individual characteristic that we call fitness. In particular, we identify the conditions that are necessary to generate power-law distributions.

The chapter is structured as follows: in Section 7.1 we describe our setting and we introduce the main results of the chapter, i.e., the characterization of in-degree limiting distributions of CTBPs. More precisely, in Section 7.1.1 we describe the model, in Section 7.1.2 we describe the result in the presence of age-dependence but no fitness, and Section 7.1.3 the result with both age-dependence and fitness. Section 7.2 contains a discussion about necessary conditions on the fitness distribution. Section 7.3 contains the proof of the existence of the limiting degree distribution for a CTBP. In Section 7.4.1 and Section 7.4.3 a detailed asymptotic analysis of the limting degree distributions are presented. The novel results in this chapter are based on [73].

7.1. Introduction and main results

In this chapter we introduce the effect of aging and fitness in CTBP populations, giving rise to directed trees. Our model is motivated by the study of *citation networks*,

which can be seen as directed graphs. Trees are the simplest case in which we can see the effects of aging and fitness.

Previous works have shown that PAMs can be obtained from PA trees by collapsing, and their general degree structure can be quite well understood from those in trees. For example, PAMs with fixed out-degree $m \geq 2$ and affine PA function $f(k) = k + \delta$ can be defined through a collapsing procedure, where a vertex in the multigraph is formed by $m \in \mathbb{N}$ vertices in the tree (see [85, Section 8.2]). In this case, the limiting degree distribution of the PAM preserve the structure of the tree case ([85, Section 8.4], [22, Section 5.7]). In Chapter 2 we have shown that this relation between the tree case and $m \geq 2$ holds also in continuous-time. This explains the relevance of the tree case results for the study of the effect of aging and fitness in PAMs.

7.1.1. A CTBP model

The starting idea of our model of citation networks is that, given the history of the process up to time t, the rate $\lambda(i, k, t)$ of an individual i of age t and k children is

$$\lambda(i, k, t) \; = \; \eta_i f_k g(t), \tag{7.1.1}$$

where f_k is a non-decreasing PA function of the degree, g is an integrable function of time, and η is a positive random variable called fitness. Therefore, the likelihood to generate children increases by having many children and/or a high fitness, while it is reduced by age.

Recalling Figure 1.22, we assume that the PA function f is affine, so $f_k = ak + b$. In terms of a PA scheme, this implies

$$\mathbb{P}\left(\text{a paper cites another with past } k \text{ citations} \mid \text{past}\right) \approx \frac{n(k)(ak+b)}{A},$$

where $n(k)$ denotes the number of papers with k past citations, and A is the normalization factor. Such behavior has already been observed by Redner [147] and Barabási et al. [102].

In general we assume that the aging function g is integrable. In fact, we start with the fact that the age of cited papers is lognormally distributed (recall Figure 1.23). By normalizing such a distribution by the average increment in the number of citations of papers in the selected time window, we identify a universal function $g(t)$. Such function can be approximated by a lognormal shape with field-dependent parameters. In particular, from the procedure used to define $g(t)$, we observe that

$$g(t) \approx \frac{\text{number of references to year } t}{\text{number of papers of age } t} \, \frac{\text{total number of papers considered}}{\text{total number of references considered}},$$

which means in terms of PA mechanisms that

$$\mathbb{P}\left(\text{a paper cites another of age } t \mid \text{past}\right) \approx \frac{n(t)g(t)}{B},$$

where B is the normalization factor, while this time $n(t)$ is the number of papers of age t. This suggests that the citing probability depends on age through a lognormal aging function $g(t)$, which is integrable. This is one of the main assumptions in our model, as we discuss in Section 1.9.2.

The presence of fitness assures that the behavior of the degree (number of citations) of individuals is different from individual to individual (recall Figure 1.21). In particular, as it is clear in the sequel, we will find that the most suitable case to model citation networks is to consider fitnesses with *unbounded support*. In the citation networks perspective, viewing fitnesses as intrinsic potential of papers, there is no maximum value for fitnesses, i.e., *there is no maximum value to the potential attractiveness of a paper*.

Goal and structure of the chapter. It is known from the literature [9, 22, 151, 152] that CTBPs show power-law limiting degree distributions when the infinitesimal rates of jump depend only on a sequence $(f_k)_{k\in\mathbb{N}}$ that is asymptotically *linear*. Our main aim is to investigate whether power-laws can also arise in branching processes that include aging and fitness.

The results in the present chapter are organized as follows. In Section 7.1.2, we discuss the results for CTBPs with aging in the absence of fitness. In Section 7.1.3, we present the results with aging and fitness. In Section 7.1.4, we specialize to fitness with distributions with exponential tails, where we show that the limiting degree distribution is a power law with a *dynamic* power-law exponent.

7.1.2. Results with aging without fitness

In this section, we focus on aging in PA trees in the absence of fitness. The aging process can then be viewed as a time-changed stationary birth process (see Definition 2.1.14). A stationary birth process is a stochastic process $(\zeta_t)_{t\geq 0}$ such that, for h small enough,

$$\mathbb{P}\left(\zeta_{t+h} = k+1 \mid \zeta_t = k\right) = f_k h + o(h).$$

In general, we assume that $k \mapsto f_k$ is increasing. The *affine case* arises when $f_k = ak + b$ with $a, b > 0$. By our observations in Figure 1.22, as well as related works [102, 147], the affine case is a reasonable approximation for the attachment rates in citation networks.

For a stationary birth process $(\zeta_t)_{t\geq 0}$, under the assumption that it is supercritical and Malthusian, the limiting degree distribution $(p_k)_{k\in\mathbb{N}}$ of the corresponding branching process is given by

$$p_k^{(1)} = \frac{\alpha^*}{\alpha^* + f_k} \prod_{i=0}^{k-1} \frac{f_i}{\alpha^* + f_i}.$$

For a more detailed description, we refer to Section 2.1.2. Branching processes defined by stationary processes (with no aging effect) have a so-called *old-get-richer* effect, i.e, old vertices have high degree that keeps increasing. As this is not what we ob-

serve in citation networks (recall Figure 1.21), we want to introduce *aging* in the reproduction process of individuals. The aging process arises by adding age-dependence in the infinitesimal transition probabilities:

Definition 7.1.1 (Aging birth processes)**.** *Consider a non-decreasing PA sequence $(f_k)_{k\in\mathbb{N}}$ of positive real numbers and an aging function $g\colon \mathbb{R}^+ \to \mathbb{R}^+$. We call a stochastic process $(\bar{\zeta}_t)_{t\geq 0}$ an* aging birth process *(without fitness) when*

(1) $\bar{\zeta}_0 = 0$, and $\bar{\zeta}_t \in \mathbb{N}$ for all $t \in \mathbb{N}$;
(2) $\bar{\zeta}_t \leq N_s$ for every $t \leq s$;
(3) for fixed $k \in \mathbb{N}$ and $t \geq 0$, as $h \to 0$,

$$\mathbb{P}\left(\bar{\zeta}_{t+h} = k+1 \mid \bar{\zeta}_t = k\right) = f_k g(t) h + o(h).$$

Aging processes are time-rescaled versions of the corresponding stationary process defined by the same sequence $(f_k)_{k\in\mathbb{N}}$. In particular, for any $t \geq 0$, $\bar{\zeta}_t$ has the same distribution as $\zeta_{G(t)}$, where $G(t) = \int_0^t g(s)ds$. For this reason, we denote an aging birth process as $(\zeta_{G(t)})_{t\geq 0}$, assuming that $G(t) = \int_0^t g(s)ds$.

In general, we assume that the aging function is *integrable*, which means that $G(\infty) := \int_0^\infty g(s)ds < \infty$. This implies that the number of children of a single individual in its entire lifetime has distribution $\zeta_{G(\infty)}$, which is finite in expectation. In terms of citation networks, this assumption is reasonable since we do not expect papers to receive an infinite number of citations ever (recall Figure 1.20). Instead, for the stationary process $(\zeta_t)_{t\geq 0}$ in Definition 2.1.14, we have that $\mathbb{P}$-a.s. $\zeta_t \to \infty$, so that also the aging process diverges $\mathbb{P}$-a.s. when $G(\infty) = \infty$.

For aging processes, the main result is the following theorem, proven in Section 7.3. In its statement, we rely on the Laplace transform of a function. For a precise definition of this notion, we refer to Section 2.1:

Theorem 7.1.2 (Limiting distribution for aging branching processes)**.** *Consider an integrable aging function g and a PA sequence $(f_k)_{k\in\mathbb{N}}$. Denote the corresponding aging birth process by $(\zeta_{G(t)})_{t\geq 0}$. Then, assuming that $(\zeta_{G(t)})_{t\geq 0}$ is supercritical and Malthusian, the limiting degree distribution of the branching process $\boldsymbol{\zeta}_G$ defined by the birth process $(\zeta_{G(t)})_{t\geq 0}$ is given by*

$$p_k^{(1)} = \frac{\alpha^*}{\alpha^* + f_k \hat{\mathcal{L}}^g(k,\alpha^*)} \prod_{i=0}^{k-1} \frac{f_i \hat{\mathcal{L}}^g(i,\alpha^*)}{\alpha^* + f_i \hat{\mathcal{L}}^g(i,\alpha^*)}, \tag{7.1.2}$$

where α^* is the Malthusian parameter of $\boldsymbol{\zeta}_G$. Here, the sequence of coefficients $(\hat{\mathcal{L}}^g(k,\alpha^*))_{k\in\mathbb{N}}$ appearing in (7.1.2) *is given by*

$$\hat{\mathcal{L}}^g(k,\alpha^*) = \frac{\mathcal{L}(\mathbb{P}\left(\zeta_{G(\cdot)} = k\right) g(\cdot))(\alpha^*)}{\mathcal{L}(\mathbb{P}\left(\zeta_{G(\cdot)} = k\right))(\alpha^*)}, \tag{7.1.3}$$

where, for $h\colon \mathbb{R}^+ \to \mathbb{R}$, $\mathcal{L}(h(\cdot))(\alpha)$ denotes the Laplace transform of h.
Further, considering a fixed individual in the branching population, the total number of children in its entire lifetime is distributed as $\zeta_{G(\infty)}$, where $G(\infty)$ is the L^1-norm of g.

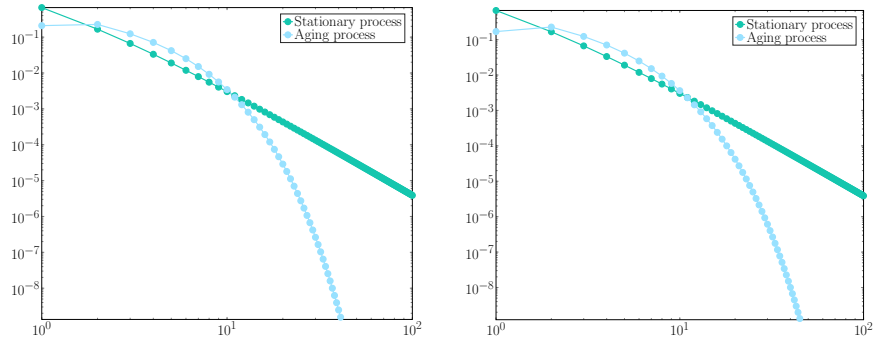


FIGURE 7.1: Examples of limiting degree distribution defined by a stationary process $(\zeta_t)_{t\geq 0}$ with affine PA sequence (red lines), and two aging processes $(\zeta_{G_1(t)})_{t\geq 0}$ and $(\zeta_{G_2(t)})_{t\geq 0}$ defined by the same PA sequence and two different aging functions g_1 and g_2. Aging functions are the same as in Figure 1.24.

The limiting degree distribution maintains a product structure as in the stationary case (see (2.1.10) for comparison). Unfortunately, the analytic expression for the probability distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ in (7.1.2) given by the previous theorem is not explicit. In the stationary case, the form reduces to the simple expression in (2.1.10).

In general, the asymptotics of the coefficients $(\hat{\mathcal{L}}^g(k,\alpha^*))_{k\in\mathbb{N}}$ is unclear, since it depends both on the aging function g as well as the PA sequence $(f_k)_{k\in\mathbb{N}}$ itself in an intricate way. In particular, we have no explicit expression for the ratio in (7.1.3), except in special cases. In this type of birth process, the cumulative advantage given by $(f_k)_{k\in\mathbb{N}}$ and the aging effect given by g cannot be separated from each other.

Numerical examples in Figure 7.1 show how aging destroys the power-law degree distribution. In each of the two plots, the limiting degree distribution of a stationary process with affine PA function gives a power-law degree distribution, while the process with two different integrable aging functions does not.

In the examples we have used $g(t) = \mathrm{e}^{-\lambda t}$ and $g(t) = (1+t)^{-\lambda}$ for some $\lambda > 1$, and we observe the insensitivity of the limiting degree distribution with respect to g. The distribution given by (7.1.2) can be seen as the limiting degree distribution of a CTBP defined by the sequence $(f_k\hat{\mathcal{L}}^g(k,\alpha^*))_{k\in\mathbb{N}}$. This suggests that $f_k\hat{\mathcal{L}}^g(k,\alpha^*)$ is not asymptotically linear in k.

In Section 7.5.2, we investigate the two examples in Figure 7.1, showing that the limiting degree distribution has exponential tails, a fact that we know in general just as an upper bound (see Lemma 7.4.3).

In order to apply the general CTBP result in Theorem 2.1.11 below, we need to prove that an aging process $(\zeta_{G(t)})_{t\geq 0}$ is supercritical and Malthusian. We show in Section 7.3 that, for an integrable aging function g, the corresponding process is su-

percritical if and only if

$$\lim_{t\to\infty} \mathbb{E}\left[\zeta_{G(t)}\right] = \mathbb{E}\left[\zeta_{G(\infty)}\right] > 1. \tag{7.1.4}$$

Condition (7.1.4) heuristically suggests that the process $(\zeta_{G(t)})_{t\geq 0}$ has a Malthusian parameter if and only if the expected number of children in the entire lifetime of a fixed individual is larger than one, which seems quite reasonable. In particular, such a result follows from the fact that if g is integrable, then the Laplace transform is always finite for every $\alpha > 0$. In other words, since $\zeta_{T_{\alpha^*}}$ has the same distribution as $\zeta_{G(T_{\alpha^*})}$, $\mathbb{E}[\zeta_{T_{\alpha^*}}]$ is always bounded by $\mathbb{E}[\zeta_{G(\infty)}]$. This implies that $G(\infty)$ cannot be too small, as otherwise the Malthusian parameter would not exist, and the CTBP would die out $\mathbb{P}$-a.s..

The aging effect obviously slows down the birth process, and makes the limiting degree distribution have exponential tails for affine PA function. One may wonder whether the power-law degree distribution could be restored when $(f_k)_{k\in\mathbb{N}}$ grows super-linearly instead. Here, we say that a sequence of $(f_k)_{k\in\mathbb{N}}$ grows super-linearly when $\sum_{k\geq 1} 1/f_k < \infty$ (see Definition 7.2.1). In the super-linear case, however, the branching process is *explosive*, i.e., for every individual the probability of generating an infinite number of children in finite time is 1. In this situation, the Malthusian parameter does not exist, since the Laplace transform of the process is always infinite. One could ask whether, by using an integrable aging function, this explosive behavior is destroyed. The answer to this question is given by the following theorem:

Theorem 7.1.3 (Explosive aging branching processes for super-linear PA function)**.** *Consider a stationary process $(\zeta_t)_{t\geq 0}$ defined by a super-linear sequence $(f_k)_{k\in\mathbb{N}}$. For any aging function g, the corresponding non-stationary process $(\zeta_{G(t)})_{t\geq 0}$ is explosive.*

The proof of Theorem 7.1.3 is rather simple, and is given in Section 7.3.2. We investigate the case of affine PA function $f_k = ak + b$ in more detail in Section 7.4.1. Under a hypothesis on the regularity of the integrable aging function, in Proposition 7.4.2, we give the asymptotic behavior of the corresponding limiting degree distribution. In particular, as $k \to \infty$,

$$p_k^{(1)} = C_1 \frac{\Gamma(k + b/a)}{\Gamma(k+1)} \mathrm{e}^{-C_2 k} \mathcal{G}(k, g)(1 + o(1)),$$

for some positive constants C_1, C_2. The term $\mathcal{G}(k, g)$ is a function of k, the aging function g and its derivative. The precise behavior of such term depends crucially on the aging function. Apart from this, we notice that aging generates an exponential term in the distribution, which explains the two examples in Figure 7.1. In Section 7.5.2, we prove that the two limiting degree distributions in Figure 7.1 indeed have exponential tails.

In addition to the general result on the degree distribution of CTBPs with affine f and integrable g, we prove that such CTBPs satisfy the hypothesis of Theorem 2.2.2 (and Theorem 2.3.2) we presented in Chapter 2. We formulate the result as follows:

Corollary 7.1.4 (Aging CBPs). *Fix $m \geq 2$, $\delta > -m$, and define the sequence $(k+1+\delta/m)_{k\in\mathbb{N}}$. Denote the corresponding embedding birth process by $(\xi_t)_{t\geq 0}$. Let g be an aging function as in Definition 7.1.1, such that $g(t) \leq \bar{g}$ for some constant $\bar{g} > 0$ and for every $t \geq 0$. Assume that $\lim_{t\to\infty} \mathbb{E}[\xi_{G(t)}] > 1$. Then, the CBP obtained by the CTBP defined by the aging process satisfies Theorem 2.2.2 (and Theorem 2.3.2). As a consequence, the limiting degree distribution $(p_k^{(m)})_{k\in\mathbb{N}}$ satisfies*

$$p_k^{(m)} = \frac{\Gamma(k+m+\delta)}{\Gamma(k+1)} \mathrm{e}^{-Ck}(1+o(1)), \tag{7.1.5}$$

where $C = |\log(1-\mathrm{e}^{-\int_0^\infty g(t)dt})|$.

In particular, it is possible to show that the transition probabilities of the discrete-time version $(\mathrm{CBP}^{(m)}_{\tau_{(n,j)}})_{n\in\mathbb{N},j\in[m]}$ of a CBP defined by an aging process satisfies

$$\mathbb{P}\left(n \overset{j+1}{\to} i \mid \mathrm{CBP}^{(m)}_{\tau_{(n,j)}}, \tau_{(n,j+1)}\right) \approx \frac{(D_i(\tau_{(n,j)})+\delta)g(\tau_{(n,j+1)}-\tau_{(i,1)})}{\sum_{h=1}^n (D_h(\tau_{(n,j)})+\delta)g(\tau_{(n,j+1)}-\tau_{(h,1)})}, \tag{7.1.6}$$

where $D_i(t)$ denotes the total degree of vertex i in $\mathrm{CBP}^{(m)}_t$ and the approximation is due to the fact that we consider $\tau_{(i,1)}$ as the birth time of all the m individuals collapsed to generate vertex i. The expression in (7.1.6) for the attachment rule in the presence of aging resembles the ones given in other works about aging in PAMs [75, 168, 173].

7.1.3. Results with aging and fitness

The analysis of birth processes becomes harder when we also consider fitness. First of all, we define the birth process with aging and fitness as follows:

Definition 7.1.5 (Aging birth process with fitness). *Consider a birth process $(\zeta_t)_{t\geq 0}$. Let $g\colon \mathbb{R}^+ \to \mathbb{R}^+$ be an aging function, and η a positive random variable. The process $\bar{\zeta}_t := \zeta_{\eta G(t)}$ is called a birth process with* aging and fitness.

Definition 7.1.5 implies that the infinitesimal jump rates of the process $(\zeta_{\eta G(t)})_{t\geq 0}$ are as in (7.1.1), so that the birth probabilities of an individual depend on the PA function, the age of the individual and on its fitness. Assuming that the process $(\zeta_{\eta G(t)})_{t\geq 0}$ is supercritical and Malthusian, we can prove the following theorem:

Theorem 7.1.6 (Limiting degree distribution for aging and fitness). *Consider a process $(\zeta_{\eta G(t)})_{t\geq 0}$ with integrable aging function g, fitnesses that are i.i.d. across the population, and assume that it is supercritical and Malthusian with Malthusian parameter α^*. Then, the limiting degree distribution for the corresponding branching process is given by*

$$p_k^{(1)} = \mathbb{E}\left[\frac{\alpha^*}{\alpha^* + f_k\eta\hat{\mathcal{L}}(k,\alpha^*,\eta)} \prod_{i=0}^{k-1} \frac{f_i\eta\hat{\mathcal{L}}(i,\alpha^*,\eta)}{\alpha^* + f_i\eta\hat{\mathcal{L}}(i,\alpha^*,\eta)}\right].$$

For a fixed individual, the distribution $(q_k)_{k\in\mathbb{N}}$ of the number of children it generates over its

entire lifetime is given by

$$q_k = \mathbb{P}\left(\zeta_{\eta G(\infty)} = k\right).$$

Similarly to Theorem 7.1.2, the sequence $(\hat{\mathcal{L}}(k, \alpha^*, \eta))_{k\in\mathbb{N}}$ is given by

$$\hat{\mathcal{L}}(k, \alpha^*, \eta) = \left(\frac{\mathcal{L}(\mathbb{P}\left(\zeta_{uG(\cdot)} = k\right) g(\cdot))(\alpha^*)}{\mathcal{L}(\mathbb{P}\left(\zeta_{uG(\cdot)} = k\right))(\alpha^*)}\right)_{u=\eta},$$

where again $\mathcal{L}(h(\cdot))(\alpha)$ denotes the Laplace transform of a function h. Notice that in this case, the presence of the fitness η, this sequence is no longer deterministic but random instead. We still have the product structure for $(p_k^{(1)})_{k\in\mathbb{N}}$ as in the stationary case, but now we have to average over the fitness distribution.

We point out that Theorem 7.1.2 is a particular case of Theorem 7.1.6, when we consider $\eta \equiv 1$. We state the two results as separate theorems to improve the logic of the presentation. We prove Theorem 7.1.6 in Section 7.3.1. In Section 7.3.2 we show how Theorem 7.1.2 can be obtained from Theorem 7.1.6, and in particular how Condition (7.1.4) is obtained from the analogous Condition (7.1.7) stated below for general fitness distributions.

With affine PA function, in Proposition 7.4.5, we can identify the asymptotics of the limiting degree distribution we obtain. This is proved by similar techniques as in the case of aging only, even though the result cannot be expressed so easily. In particular, we prove that

$$\begin{aligned} p_k^{(1)} =& \frac{\Gamma(k + b/a)}{\Gamma(b/a)\Gamma(k + 1)} \frac{2\pi}{\sqrt{\det(kH_k(t_k, s_k))}} \\ &\times \mathrm{e}^{-k\Psi_k(t_k,s_k)} \mathbb{P}\left(\mathcal{N}_1 \geq -t_k, \mathcal{N}_2 \geq -s_k\right)(1 + o(1)), \end{aligned}$$

where the function $\Psi_k(t, s)$ depends on the aging function, the density μ of the fitness and k. The point (t_k, s_k) is the absolute minimum of $\Psi_k(t, s)$, $H_k(t, s)$ is the Hessian matrix of $\Psi_k(t, s)$, and $(\mathcal{N}_1, \mathcal{N}_2)$ is a bivariate normal vector with covariance matrix related to $H_k(t, s)$. We do not know the necessary and sufficient conditions for the existence of such a minimum (t_k, s_k). However, in Section 7.4.4, we consider two examples where we can apply this result, and we show that it is possible to obtain power-laws for them.

In the case of aging and fitness, the supercriticality condition in (7.1.4) is replaced by the analogous condition that

$$\mathbb{E}\left[\zeta_{\eta G(t)}\right] < \infty \quad \text{for every } t \geq 0 \qquad \text{and} \qquad \lim_{t\to\infty} \mathbb{E}\left[\zeta_{\eta G(t)}\right] > 1. \tag{7.1.7}$$

Borgs et al. [35] and Dereich [53, 58] prove results on stationary CTBPs with fitness. In these works, the authors investigate models with affine dependence on the degree and bounded fitness distributions. This is necessary since unbounded distributions with affine PA function are explosive and thus *do not have Malthusian parameter*. We refer to Section 7.2 for a more precise discussion of the conditions on fitness

distributions.

In the case of integrable aging and fitness, it is possible to consider affine PA function, even with unbounded fitness distributions, as exemplified by (7.1.7). In particular, for $f_k = ak + b$,

$$\mathbb{E}[\zeta_t] = \frac{b}{a}\left(\mathrm{e}^{at} - 1\right).$$

As a consequence, Condition (7.1.7) can be written as

$$\forall t \geq 0 \quad \mathbb{E}\left[\mathrm{e}^{a\eta G(t)}\right] < \infty \qquad \text{and} \qquad \lim_{t\to\infty} \mathbb{E}\left[\mathrm{e}^{a\eta G(t)}\right] > 1 + \frac{a}{b}. \tag{7.1.8}$$

The expected value $\mathbb{E}\left[\mathrm{e}^{a\eta G(t)}\right]$ is the moment generating function of η evaluated in $aG(t)$. In particular, a necessary condition to have a Malthusian parameter is that the moment generating function is finite on the interval $[0, aG(\infty))$. As a consequence, denoting $\mathbb{E}[\mathrm{e}^{s\eta}]$ by $\varphi_\eta(s)$, we have effectively moved from the condition of having bounded distributions to the condition

$$\varphi_\eta(x) < +\infty \quad \text{on} \quad [0, aG(\infty)), \qquad \text{and} \qquad \lim_{x\to aG(\infty)} \varphi_\eta(x) > \frac{a+b}{a}. \tag{7.1.9}$$

Condition (7.1.9) is weaker than assuming a bounded distribution for the fitness η, which means we can consider a larger class of distributions for the aging and fitness birth processes. Particularly for citation networks, it seems reasonable to have unbounded fitnesses, as the relative popularity of papers varies substantially.

7.1.4. Dynamical power-laws for exponential fitness and integrable aging

In Section 7.4.4 we introduce three different classes of fitness distributions, for which we give the asymptotics for the limiting degree distribution of the corresponding CTBP.

The first class is called *heavy-tailed*. Recalling (7.1.9), any distribution η in this class satisfies, for any $t > 0$,

$$\varphi_\eta(t) = \mathbb{E}\left[\mathrm{e}^{t\eta}\right] = +\infty. \tag{7.1.10}$$

These distributions have a tail that is thicker than exponential. For instance, power-law distributions belong to this first class. Similarly to unbounded distributions in the stationary regime, such distributions generate *explosive* birth processes, independent of the choice of the integrable aging functions.

The second class is called *sub-exponential*. The density μ of a distribution η in this class satisfies

$$\forall\, \beta > 0, \qquad \lim_{s\to+\infty} \mu(s)\mathrm{e}^{\beta s} = 0. \tag{7.1.11}$$

An example of this class is the density $\mu(s) = C\mathrm{e}^{-\theta s^{1+\varepsilon}}$, for some $\varepsilon, C, \theta > 0$. For such density, we show in Proposition 7.4.7 that the corresponding limiting degree distribution has a thinner tail than a power-law.

The third class is called *general-exponential*. The density μ of a distribution η in

this class is of the form

$$\mu(s) = Ch(s)\mathrm{e}^{-\theta s}, \tag{7.1.12}$$

where $h(s)$ is a twice differentiable function such that, as $s \to \infty$,

$$h'(s)/h(s) \to 0, \quad \text{and} \quad h''(s)/h(s) \to 0,$$

and C is a normalization constant. For instance, exponential and Gamma distributions belong to this class. From (7.1.9), we know that in order to obtain a non-explosive process, it is necessary to consider the exponential rate $\theta > aG(\infty)$. We will see that the limiting degree distribution obeys a power law as $\theta > aG(\infty)$ with tails becoming thinner when θ increases.

For a distribution in the general exponential class, as proven in Proposition 7.4.6, the limiting degree distribution of the corresponding CTBP has a power-law term, with slowly-varying corrections given by the aging function g and the function h. We do not state Propositions 7.4.6 and 7.4.7 here, as these need notation and results from Section 7.4.1. For this reason, we only state the result for the special case of purely exponential fitness distribution:

Corollary 7.1.7 (Exponential fitness distribution)**.** *Let the fitness distribution η be exponentially distributed with parameter θ, and let g be an integrable aging function. Assume that the corresponding birth process $(\zeta_{\eta G(t)})_{t\geq 0}$ is supercritical and Malthusian. Then, the limiting degree distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ of the corresponding CTBP $\boldsymbol{\zeta}_{\eta G}$ is*

$$p_k^{(1)} = \mathbb{E}\left[\frac{\theta}{\theta + f_k G(T_{\alpha^*})}\prod_{i=0}^{k-1}\frac{f_i G(T_{\alpha^*})}{\theta + f_i G(T_{\alpha^*})}\right].$$

The distribution $(q_k)_{k\in\mathbb{N}}$ of the number of children of a fixed individual in its entire lifetime is given by

$$q_k = \frac{\theta}{\theta + G(\infty)f_k}\prod_{i=0}^{k-1}\frac{G(\infty)f_i}{\theta + G(\infty)f_i}.$$

Using exponential fitness makes the computation of the Laplace transform and the limiting degree distribution easier. We refer to Section 7.4.5 for the precise proof. In particular, the sequence defined in Corollary 7.1.7 is very similar to the limiting degree distribution of a stationary process with a bounded fitness. Let $(\zeta_t^\eta)_{t\geq 0}$ be a birth process with defined by $(f_k)_{k\in\mathbb{N}}$ and fitness η with bounded support. As proved in [58, Corollary 2.8], and as we show in Section 7.2, the limiting degree distribution of the corresponding branching process, assuming that $(\zeta_t^\eta)_{t\geq 0}$ is supercritical and Malthusian, has the form

$$p_k^{(1)} = \mathbb{E}\left[\frac{\alpha^*}{\alpha^* + \eta f_k}\prod_{i=0}^{k-1}\frac{\eta f_i}{\alpha^* + \eta f_i}\right] = \mathbb{P}\left(\zeta_{T_{\alpha^*}}^\eta = k\right).$$

We notice the similarities with the limiting degree sequence given by Corollary 7.1.7.

When g is integrable, the random variable $G(T_{\alpha^*})$ has bounded support. In particular, we can rewrite the sequence of the Corollary 7.1.7 as

$$p_k^{(1)} = \mathbb{P}\left(\zeta_{T_\theta}^{G(T_{\alpha^*})} = k\right).$$

As a consequence, the limiting degree distribution of the process $(\zeta_{\eta G(t)})_{t\geq 0}$ equals that of a stationary process with fitness $G(T_{\alpha^*})$ and Malthusian parameter θ.

In the case where η has exponential distribution and the PA function is affine, we can also investigate the occurrence of *dynamical power laws*. In fact, with $(\zeta_{\eta G(t)})_{t\geq 0}$ such a process, the exponential distribution η leads to

$$\begin{aligned} P_k[\zeta_{\eta G}](t) &= \mathbb{P}\left(\zeta_{\eta G(t)} = k\right) \\ &= \frac{\theta}{\theta + f_k G(t)} \prod_{i=0}^{k-1} \frac{f_i G(t)}{\theta + f_i G(t)} \\ &= \frac{\theta}{aG(t)} \frac{\Gamma((b+\theta)/(aG(t)))}{\Gamma(aG(t))} \frac{\Gamma(k + b/(aG(t)))}{\Gamma(k + b/(aG(t)) + 1 + \theta/(aG(t)))}. \end{aligned} \tag{7.1.13}$$

Here, $\zeta_{\eta G(t)}$ describes the number of children of an individual of age t. In other words, $(\mathbb{P}(\zeta_{\eta G(t)} = k))_{k\in\mathbb{N}}$ is a distribution such that, as $k \to \infty$,

$$P_k[\zeta_{\eta G}](t) = \mathbb{P}\left(\zeta_{\eta G(t)} = k\right) = k^{-(1+\theta/aG(t))}(1 + o(1)).$$

This means that for every time $t \geq 0$, the random variable $\zeta_{\eta G(t)}$ has a power-law distribution with exponent $\tau(t) = 1 + \theta/aG(t) > 2$. In particular, for every $t \geq 0$, $\zeta_{\eta G(t)}$ has finite expectation. We call this behavior where power laws occur that vary with the age of the individuals a *dynamical power law*. This occurs not only in the case of pure exponential fitness, but in general for every distribution as in (7.1.12), as shown in Proposition 7.4.6 below.

Further, we see that when $t \to \infty$, the dynamical power-law exponent coincides with the power-law exponent of the entire population. Indeed, the limiting degree distribution equals

$$\begin{aligned} p_k^{(1)} =& \mathbb{E}\Bigg[\theta/(aG(T_{\alpha^*}))) \frac{\Gamma(\theta/(aG(T_{\alpha^*})) + b/(aG(T_{\alpha^*}))}{\Gamma(b/(aG(T_{\alpha^*})))} \\ &\times \frac{\Gamma(k + b/(aG(T_{\alpha^*})))}{\Gamma(k + b/(aG(T_{\alpha^*})) + 1 + \theta/(aG(T_{\alpha^*})))}\Bigg]. \end{aligned} \tag{7.1.14}$$

In Figure 1.25 we compare a numerical example of the dynamical power-law, for a process with exponential fitness distribution and affine PA function, and the degree distribution in citation networks. When time increases, the power-law exponent monotonically decreases to the limiting exponent $\tau \equiv \tau(\infty) > 2$, which means that the limiting distribution still has a finite first moment.

When $t \to \infty$, the power-law exponent converges, and also $\zeta_{\eta G(t)}$ converges in

distribution to a limiting random distribution $\zeta_{\eta G(\infty)}$ given by

$$\begin{aligned} q_k &= \mathbb{P}\left(\zeta_{\eta G(\infty)} = k\right) \\ &= \frac{\theta}{aG(\infty)} \frac{\Gamma((b+\theta)/(aG(\infty)))}{\Gamma(b/(aG(\infty)))} \frac{\Gamma(k+b/(aG(\infty)))}{\Gamma(k+b/(aG(\infty))+1+\theta/(aG(\infty)))}. \end{aligned} \tag{7.1.15}$$

$\zeta_{\eta G(\infty)}$ has a power-law distribution, where the power-law exponent now is

$$\tau = \lim_{t\to\infty} \tau(t) = 1 + \theta/(aG(\infty)) > 2.$$

In particular, since $\tau > 2$, a fixed individual has finite expected number of children also in its entire lifetime, unlike the stationary case with affine PA function. In terms of citation networks, this type of processes predicts that papers do not receive an infinite number of citations after they are published (recall Figure 1.20), but papers receiving enormously many citations do exist.

Figure 7.1 shows the effect of aging on the stationary process with affine PA function, where the power-law is lost due to the aging effect. Thus, aging *slows down* the stationary process, and it is not possible to create the amount of high-degree vertices that are present in power-law distributions. Fitness can *speed up* the aging process to gain high-degree vertices, so that the power-law distribution is restored. This is shown in Figure 1.24, where aging is combined with exponential fitness for the same aging functions as in Figure 7.1.

In the stationary case, it is not possible to use unbounded distributions for the fitness to obtain a Malthusian process if the PA function f is affine. In fact, using unbounded distributions, the expected number of children at exponential time T_α is not finite *for any* $\alpha > 0$, i.e., the branching process is *explosive*. The aging effect allows us to relax the condition on the fitness, and the restriction to bounded distributions is relaxed to a condition on its moment generating function.

7.2. The Malthusian parameter

The existence of the Malthusian parameter is a necessary condition to have a branching process growing at exponential rate. In particular, the Malthusian parameter does not exist in two cases: when the process is subcritical and grows slower than exponential, or when it is explosive. In the first case, the branching population might either die out or grow indefinitely with positive probability, but slower than at exponential rate. In the second case, the population size explodes in finite time with probability one. In both cases, the behavior of the branching population is different from what we observe in citation networks (Figure 1.17). For this reason, we focus on supercritical processes, i.e., on the case where the Malthusian parameter exists.

Denote by $(\zeta_t)_{t\geq 0}$ a stationary birth process defined by the sequence $(f_k)_{k\in\mathbb{N}}$. In general, we assume that $f_k \to \infty$. Denote the sequence of jump times by $(T_k)_{k\in\mathbb{N}}$. As

we quote in Section 2.1.2, the Laplace transform of a birth process $(\zeta_t)_{t\geq 0}$ is given by

$$\mathcal{L}(\mathbb{E}\zeta(d\cdot))(\alpha) = \mathbb{E}\left[\sum_{k\in\mathbb{N}} \mathrm{e}^{-\alpha T_k}\right] = \mathbb{E}\left[\zeta_{T_\alpha}\right] = \sum_{k\in\mathbb{N}} \prod_{i=0}^{k-1} \frac{f_i}{\alpha + f_i}.$$

This expression comes from the fact that, in the stationary regime, T_k is the sum of k independent exponential random variables. We can write

$$\sum_{k\in\mathbb{N}} \exp\left(-\sum_{i=0}^{k-1} \log\left(1+\frac{\alpha}{f_i}\right)\right) = \sum_{k\in\mathbb{N}} \exp\left(-\alpha \sum_{i=0}^{k-1} \frac{1}{f_i}(1+o(1))\right).$$

The behavior of the Laplace transform depends on the asymptotic behavior of the PA function f. We now define the terminology we use:

Definition 7.2.1 (Superlinear PA function)**.** *Consider a PA sequence $(f_k)_{k\in\mathbb{N}}$. We say that the sequence f is* superlinear *if $\sum_{i=0}^{\infty} 1/f_i < \infty$. Equivalently, we say that the PA function f is superlinear.*

As a general example, consider $f_k = ak^q + b$, where $q > 0$. In this case, the sequence is affine when $q = 1$, superlinear when $q > 1$ and sublinear when $q < 1$.

When f is superlinear, since $C = \sum_{i=0}^{\infty} 1/f_i < \infty$, we have

$$\sum_{k\in\mathbb{N}} \exp\left(-\alpha \sum_{i=0}^{k-1} \frac{1}{f_i}(1+o(1))\right) \geq \sum_{k\in\mathbb{N}} \exp\left(-\alpha C\right) = +\infty. \tag{7.2.1}$$

This holds for every $\alpha > 0$. As a consequence, the Laplace transform $\mathcal{L}(\mathbb{E}\zeta(d\cdot))(\alpha)$ is always infinite, and there exists no Malthusian parameter. In particular, if we denote by $T_\infty = \lim_{k\to\infty} T_k$, then $T_\infty < \infty$ a.s.. This means that the birth process $(\zeta_t)_{t\geq 0}$ explodes in a finite time.

When f is at most linear, the bound in (7.2.1) does not hold anymore. In fact, consider as example affine PA function $f_k = ak+b$. We have that $\sum_{i=0}^{k-1} \frac{1}{f_i} = (1/a)\log k(1+o(1))$. As a consequence, the Laplace transform can be written as

$$\sum_{k\in\mathbb{N}} \exp\left(-\frac{\alpha}{a}\log k(1+o(1))\right) = \sum_{k\in\mathbb{N}} k^{-\frac{\alpha}{a}}(1+o(1)). \tag{7.2.2}$$

In this case, the Laplace transform is finite for $\alpha > a$. For the sublinear case, for which $\sum_{i=0}^{k-1} 1/f_i = Ck^{(1-q)}(1+o(1))$, we obtain

$$\sum_{k\in\mathbb{N}} \exp\left(-C\alpha k^{1-q}\right).$$

This sum is finite for any $\alpha > 0$.

Remark 7.2.2. *Consider the process $(\zeta_t)_{t\geq 0}$ defined by the sequence $(f_k)_{k\in\mathbb{N}}$ as in Section 2.1.2. For $u \in \mathbb{R}^+$ we denote by $(\zeta_t^u)_{t\geq 0}$ the process defined by the sequence $(uf_k)_{k\in\mathbb{N}}$. It is*

easy to show that

$$\mathcal{L}(\mathbb{E}\zeta^u(d\cdot))(\alpha) = \mathcal{L}(\mathbb{E}\zeta(d\cdot))(\alpha/u).$$

The behavior of the degree sequence of $(\zeta_t^u)_{t\geq 0}$ is the same as that of the process ζ_t.

Remark 7.2.2 shows a sort of monotonicity of the Laplace transform with respect to the sequence $(f_k)_{k\in\mathbb{N}}$. This is very useful to describe the Laplace transform of a birth process with fitness, which we define now:

Definition 7.2.3 (Stationary fitness birth processes)**.** *Consider a birth process $(\zeta_t)_{t\geq 0}$ defined by a sequence $(f_k)_{k\in\mathbb{N}}$. Let η be a positive random variable. We call the process $(\zeta_t^\eta)_{t\geq 0}$ a stationary fitness birth processes defined by the random sequence $(\eta f_k)_{k\in\mathbb{N}}$, i.e., conditionally on η,*

$$\mathbb{P}\left(\zeta_{t+h}^\eta = k+1 \mid \zeta_t^\eta = k, \eta\right) = \eta f_k h + o(h).$$

By Definition 7.2.3, it is obvious that the properties of the process $(\zeta_t^\eta)_{t\geq 0}$ are related to the properties of $(\eta_t)_{t\geq 0}$. Since we consider a random fitness η independent of the process $(\zeta_t)_{t\geq 0}$, from Remark 7.2.2, it follows that

$$\mathcal{L}(\mathbb{E}\zeta^\eta(d\cdot))(\alpha) = \mathbb{E}\left[\mathcal{L}(\mathbb{E}\zeta^u(d\cdot))(\alpha)_{u=\eta}\right] = \mathbb{E}\left[\sum_{k\in\mathbb{N}}\prod_{i=0}^{k-1}\frac{\eta f_i}{\alpha+\eta f_i}\right]. \tag{7.2.3}$$

For affine PA functions f the fitness distribution needs to be bounded, as discussed in Section 7.1.4. In this section we give a qualitative explanation of this fact. Consider the sum in the expectation in the right hand term of (7.2.3). We can rewrite the sum as

$$\begin{aligned}\sum_{k\in\mathbb{N}}\prod_{i=0}^{k-1}\frac{\eta f_i}{\alpha+\eta f_i} &= \sum_{k\in\mathbb{N}}\exp\left(-\sum_{i=0}^{k-1}\log\left(1+\frac{\alpha}{\eta f_i}\right)\right)\\ &= \sum_{k\in\mathbb{N}}\exp\left(-\frac{\alpha}{\eta}\sum_{i=0}^{k-1}\frac{1}{f_i}(1+o(1))\right).\end{aligned} \tag{7.2.4}$$

The behavior depends sensitively on the asymptotic behavior of the PA function. In particular, a necessary condition for the existence of the Malthusian parameter is that the sum in (7.2.3) is finite on an interval of the type $(\tilde{\alpha}, +\infty)$. In other words, since the Laplace transform is a decreasing function (when finite), we need to prove the existence of a minimum value $\tilde{\alpha}$ such that it is finite for every $\alpha > \tilde{\alpha}$. Using (7.2.4) in (7.2.3), we just need to find a value α such that the right hand side of (7.2.4) equals 1.

In the case of an affine PA function $f_k = ak + b$, we have $\sum_{i=0}^{k-1}\frac{1}{f_i} = C\log k(1+o(1))$, for a constant C. As a consequence, (7.2.4) is equal to

$$\mathbb{E}\left[\sum_{k\in\mathbb{N}}\exp\left(-C\frac{\alpha}{\eta}\log k\right)\right] = \mathbb{E}\left[\sum_{k\in\mathbb{N}}k^{-C\alpha/\eta}\right]. \tag{7.2.5}$$

The sum inside the last expectation is finite only on the event $\{\eta < C\alpha\}$. If η has an unbounded distribution, then for every value of $\alpha > 0$, we have that $\{\eta \geq C\alpha\}$ is an event of positive probability. As a consequence, for every $\alpha > 0$, the Laplace transform of the birth process $(\zeta_t^\eta)_{t\geq 0}$ is infinite, which means that there exists no Malthusian parameter. This is why a bounded fitness distribution is necessary to have a Malthusian parameter using affine PA function.

The situation is different in the case of sublinear PA functions. For example, consider $f_k = (1+k)^q$, where $q \in (0,1)$. Then, the difference to the affine case is that now $\sum_{i=0}^{k-1} 1/f_i = Ck^{1-q}(1+o(1))$. Using this in (7.2.4), we obtain

$$\mathbb{E}\left[\sum_{k\in\mathbb{N}} \exp\left(-C\frac{\alpha}{\eta}k^{(1-q)}\right)\right].$$

In this case, since both α and η are always positive, the last sum is finite with probability 1, and the expectation might be finite under appropriate moment assumptions on η.

Assume now that the fitness η satisfies the necessary conditions, so that the process $(\zeta_t^\eta)_{t\geq 0}$ is supercritical and Malthusian with parameter α^*. We can evaluate the limiting degree distribution. Conditioning on η, the Laplace transform of $\mathbb{E}\zeta^\eta(dx)$ is

$$\sum_{k\in\mathbb{N}} \prod_{i=0}^{k-1} \frac{\eta f_i}{\alpha + \eta f_i},$$

so, as a consequence, the limiting degree distribution of the branching processes is

$$p_k^{(1)} = \mathbb{E}\left[\frac{\alpha^*}{\alpha^* + \eta f_k}\prod_{i=0}^{k-1}\frac{\eta f_i}{\alpha^* + \eta f_i}\right]. \tag{7.2.6}$$

It is possible to see that the right-hand side of (7.2.6) is similar to the distribution of the simpler case with no fitness given by (2.1.10). We still have a product structure for the limiting distribution, but in the fitness case it has to be averaged over the fitness distribution. This result is similar to [58, Theorem 2.7, Corollary 2.8].

Considering an affine PA function $f_k = ak + b$, we can rewrite (7.2.6) as

$$p_k^{(1)} = \mathbb{E}\left[\frac{\Gamma((\alpha^*+b)/(a\eta))}{\Gamma(b/(a\eta))}\frac{\Gamma(k+b/(a\eta))}{\Gamma(k+b/(a\eta)+1+\alpha^*/(a\eta))}\right].$$

Asymptotically in k, the argument of the expectation in the previous expression is random with a power-law exponent $\tau(\eta) = 1+\alpha^*/(a\eta)$. For example, averaging over the fitness distribution, in this case it is possible to obtain power laws with logarithmic corrections (see eg [22, Corollary 32]).

7.3. Existence of limiting distributions

In this section, we give the proof of Theorems 7.1.2, 7.1.3 and 7.1.6, proving that the branching processes defined in Section 7.1 do have a limiting degree distribution. As mentioned, we start by proving Theorem 7.1.6, and then explain how Theorem 7.1.2 follows as special case.

Before proving the result, we do need some remarks on the processes we consider. Birth process with aging alone and aging with fitness are defined respectively in Definition 7.1.1 and 7.1.5. Consider then a process with aging and fitness $(\zeta_{\eta G(t)})_{t\geq 0}$ as in Definition 7.1.5. Let $(T_k)_{k\in\mathbb{N}}$ denote the sequence of birth times, i.e.,

$$T_k = \inf\left\{t \geq 0 \colon \zeta_{\eta G(t)} \geq k\right\}.$$

It is an immediate consequence of the definition that, for every $k \in \mathbb{N}$,

$$\mathbb{P}\left(T_k \leq t\right) = \mathbb{P}\left(\bar{T}_k \leq \eta G(t)\right), \tag{7.3.1}$$

where $(\bar{T}_k)_{k\in\mathbb{N}}$ is the sequence of birth times of a stationary birth process $(\zeta_t)_{t\geq 0}$ defined by the same PA function f.

Consider then the sequence of functions $(P_k[\zeta](t))_{k\in\mathbb{N}}$ associated with the stationary process $(\zeta_t)_{t\geq 0}$ defined by the same sequence $(f_k)_{k\in\mathbb{N}}$ (see Proposition 2.1.16). As a consequence, for every $k \in \mathbb{N}$, $\mathbb{P}(\zeta_{\eta G(t)} = k) = \mathbb{E}[P_k[\zeta](\eta G(t))]$, and the same holds for an aging process just considering $\eta \equiv 1$. Formula (7.3.1) implies that the aging process is the stationary process with a deterministic time-change given by $G(t)$. A process with aging and fitness is the stationary process with a random time-change given by $\eta G(t)$.

Assume now that g is integrable, i.e. $\lim_{t\to\infty} G(t) = G(\infty) < \infty$. Using (7.3.1) we can describe the limiting degree distribution $(q_k)_{k\in\mathbb{N}}$ of a fixed individual in the branching population, i.e., the distribution $\zeta_{G(\infty)}$ (or $\zeta_{\eta G(\infty)}$) of the total number of children an individual will generate in its entire lifetime. In fact, for every $k \in \mathbb{N}$,

$$\lim_{t\to\infty} \mathbb{P}\left(\zeta_{G(t)} = k\right) = \lim_{t\to\infty} P_k[\zeta](G(t)) = \mathbb{P}\left(\zeta_{G(\infty)} = k\right). \tag{7.3.2}$$

With fitness,

$$\lim_{t\to\infty} \mathbb{P}\left(\zeta_{\eta G(t)} = k\right) = \lim_{t\to\infty} \mathbb{E}[P_k[\zeta](\eta G(t))] = \mathbb{P}\left(\zeta_{\eta G(\infty)} = k\right).$$

For example, in the case of aging only, this is rather different from the stationary case, where the number of children of a fixed individual diverges as the individual gets old (see e.g [9, Theorem 2.6]).

7.3.1. Proof of Theorem 7.1.6

Birth processes with continuous aging effect and fitness are defined in Definition 7.1.5. We now identify conditions on the fitness distribution to have a Malthusian

parameter:

Lemma 7.3.1 (Condition (7.1.7))**.** *Consider a stationary process $(\zeta_t)_{t\geq 0}$, an integrable aging function g and a random fitness η. Assume that $\mathbb{E}[\zeta_t] < \infty$ for every $t \geq 0$. Then the process $(\zeta_{\eta G(t)})_{t\geq 0}$ is supercritical if and only if Condition* (7.1.7) *holds, i.e.,*

$$\mathbb{E}\left[\zeta_{\eta G(t)}\right] < \infty \quad \textit{for every } t \geq 0 \qquad \textit{and} \quad \lim_{t\to\infty} \mathbb{E}\left[\zeta_{\eta G(t)}\right] > 1.$$

Proof. For the if part, we need to prove that

$$\lim_{\alpha\to 0^+} \mathbb{E}\left[\zeta_{\eta G(T_{\alpha^*})}\right] > 1 \qquad \text{and} \qquad \lim_{\alpha\to\infty} \mathbb{E}\left[\zeta_{\eta G(T_{\alpha^*})}\right] = 0.$$

As before, $(\bar{T}_k)_{k\in\mathbb{N}}$ are the jump times of the process $(\zeta_{\eta G(t)})_{t\geq 0}$. Then

$$\mathbb{E}\left[\zeta_{\eta G(T_{\alpha^*})}\right] = \sum_{k\in\mathbb{N}} \mathbb{E}\left[\mathrm{e}^{-\alpha \bar{T}_k/\eta}\right].$$

When $\alpha \to 0$, we have $\mathbb{E}\left[\mathrm{e}^{-\alpha\bar{T}_k}\right] \to \mathbb{P}\left(\bar{T}_k/\eta < \infty\right)$. Now,

$$\sum_{k\in\mathbb{N}} \mathbb{P}\left(\bar{T}_k < \infty\right) = \lim_{t\to\infty} \sum_{k\in\mathbb{N}} \mathbb{P}\left(\bar{T}_k/\eta \leq t\right) = \lim_{t\to\infty} \mathbb{E}\left[\zeta_{\eta G(t)}\right] > 1.$$

For $\alpha \to \infty$,

$$\int_0^\infty \alpha \mathrm{e}^{-\alpha t} \mathbb{E}\left[\zeta_{\eta G(t)}\right] dt = \int_0^\infty \mathrm{e}^{-u} \mathbb{E}\left[\zeta_{\eta G(u/\alpha)}\right] du.$$

When $\alpha \to \infty$ we have $\mathbb{E}\left[\zeta_{\eta G(u/\alpha)}\right] \to 0$. Then, fix $\alpha_0 > 0$ such that $\mathbb{E}\left[\zeta_{\eta G(u/\alpha)}\right] < 1$ for every $\alpha > \alpha_0$. As a consequence, $\mathrm{e}^{-u}\mathbb{E}\left[\zeta_{\eta G(u/\alpha)}\right] du \leq \mathrm{e}^{-u}$ for any $\alpha > \alpha_0$. By dominated convergence,

$$\lim_{\alpha\to\infty} \int_0^\infty \alpha \mathrm{e}^{-\alpha t} \mathbb{E}\left[\zeta_{\eta G(t)}\right] dt = 0.$$

Now suppose Condition (7.1.7) does not hold. This means that $\mathbb{E}[\zeta_{\eta G(t_0)}] = +\infty$ for some $t_0 \in [0, G(\infty))$ or $\lim_{t\to\infty} \mathbb{E}[\zeta_{\eta G(t_0)}] \leq 1$.

If the first condition holds, then there exists $t_0 \in (0, aG(\infty))$ such that $\mathbb{E}\left[\zeta_{\eta G(t)}\right] = +\infty$ for every $t \geq t_0$ (recall that $\mathbb{E}\left[\zeta_{\eta G(t)}\right]$ in an increasing function in t). As a consequence, for every $\alpha > 0$, we have $\mathbb{E}\left[\zeta_{\eta G(T_\alpha)}\right] = +\infty$, which means that the process is explosive.

If the second condition holds, then for every $\alpha > 0$ the Laplace transform of the process is strictly less than 1, which means there exists no Malthusian parameter. □

Lemma 7.3.1 gives a weaker condition on the distribution η than requiring it to be bounded. Now, we want to investigate the degree distribution of the branching process, assuming that the process $(\zeta_{\eta G(t)})_{t\geq 0}$ is supercritical and Malthusian. Denote the Malthusian parameter by α^*. The above allows us to complete the proof of

Theorem 7.1.6:

Proof of Theorem 7.1.6. We start from

$$p_k^{(1)} = \mathbb{E}\left[P_k[\zeta](\eta G(T_{\alpha^*}))\right]. \tag{7.3.3}$$

Conditioning on η and integrating by parts in the integral given by the expectation in (7.3.3), gives

$$-f_k\eta\int_0^\infty \mathrm{e}^{-\alpha^* t}P_k[\zeta](\eta G(t))g(t)dt + f_{k-1}\eta\int_0^\infty \mathrm{e}^{-\alpha^* t}P_{k-1}[\zeta](\eta G(t))g(t)dt.$$

Now, we define

$$\hat{\mathcal{L}}(k,\alpha^*,\eta) = \left(\frac{\mathcal{L}(\mathbb{P}\left(\zeta_{uG(\cdot)}=k\right)g(\cdot))(\alpha^*)}{\mathcal{L}(\mathbb{P}\left(\zeta_{uG(\cdot)}=k\right))(\alpha^*)}\right)_{u=\eta} \tag{7.3.4}$$

Notice that the sequence $(\hat{\mathcal{L}}(k,\alpha^*,\eta))_{k\in\mathbb{N}}$ is a sequence of random variables. Multiplying both sides of the equation by α^*, on the right hand side we have

$$-f_k\eta\hat{\mathcal{L}}(k,\alpha^*,\eta)\mathbb{E}\left[P_k[\zeta](uG(T_{\alpha^*}))\right]_{u=\eta} + f_{k-1}\eta\hat{\mathcal{L}}(k-1,\alpha^*,\eta)\mathbb{E}\left[P_{k-1}[\zeta](uG(T_{\alpha^*}))\right]_{u=\eta},$$

while on the left hand side we have

$$\alpha^*\mathbb{E}\left[P_k[\zeta](uG(T_{\alpha^*}))\right]_{u=\eta}.$$

As a consequence,

$$\mathbb{E}\left[P_k[\zeta](uG(T_{\alpha^*}))\right]_{u=\eta} = \frac{f_{k-1}\eta\hat{\mathcal{L}}(k-1,\alpha^*,\eta)}{\alpha^* + f_k\eta\hat{\mathcal{L}}(k,\alpha^*,\eta)}\mathbb{E}\left[P_{k-1}[\zeta](uG(T_{\alpha^*}))\right]_{u=\eta}. \tag{7.3.5}$$

We start from p_0, that is given by

$$\mathbb{E}\left[P_0[\zeta](uG(T_{\alpha^*}))\right]_{u=\eta} = \frac{\alpha^*}{\alpha^* + f_0\eta\hat{\mathcal{L}}(0,\alpha^*,\eta)}.$$

Recursively using (7.3.5), gives

$$\mathbb{E}\left[P_k[\zeta](uG(T_{\alpha^*}))\right]_{u=\eta} = \frac{\alpha^*}{\alpha^* + f_k\eta\hat{\mathcal{L}}(k,\alpha^*,\eta)}\prod_{i=0}^{k-1}\frac{f_i\eta\hat{\mathcal{L}}(i,\alpha^*,\eta)}{\alpha^* + f_i\eta\hat{\mathcal{L}}(i,\alpha^*,\eta)}.$$

Taking expectation on both sides gives

$$p_k^{(1)} = \mathbb{E}\left[\frac{\alpha^*}{\alpha^* + f_k\eta\hat{\mathcal{L}}(k,\alpha^*,\eta)}\prod_{i=0}^{k-1}\frac{f_i\eta\hat{\mathcal{L}}(i,\alpha^*,\eta)}{\alpha^* + f_i\eta\hat{\mathcal{L}}(i,\alpha^*,\eta)}\right].$$

□

Now the sequence $(\hat{\mathcal{L}}(k, \alpha^*, \eta))_{k\in\mathbb{N}}$ creates a relation among the PA function, the aging function and the fitness distribution, so that these three ingredients are deeply related.

7.3.2. Proof of Theorems 7.1.2 and 7.1.3

As mentioned, Theorem 7.1.2 follows immediately by considering $\eta \equiv 1$. The proof in fact is the same, since we can express the probabilities $\mathbb{P}(\zeta_{G(t)} = k)$ as function of the stationary process $(\zeta_t)_{t\geq 0}$ defined by the same PA function f.

Condition (7.1.4) immediately follows from Condition (7.1.7). In fact, considering $\eta \equiv 1$, Condition (7.1.7) becomes

$$\mathbb{E}\left[\zeta_{G(t)}\right] < \infty \quad \text{for every } t \geq 0 \qquad \text{and} \qquad \lim_{t\to\infty} \mathbb{E}\left[\zeta_{G(t)}\right] > 1. \tag{7.3.6}$$

The first inequality in general true for the type of stationary process we consider (for instance with f affine). The second inequality is exactly Condition (7.1.4).

The expression of the sequence $(\hat{\mathcal{L}}^g(k, \alpha^*))_{k\in\mathbb{N}}$ is simpler than the general case given in (7.3.4). In fact, in (7.3.4), the sequence $(\hat{\mathcal{L}}(k, \alpha^*, \eta))_{k\in\mathbb{N}}$ is actually a sequence of random variables. In the case of aging alone,

$$\hat{\mathcal{L}}^g(k, \alpha^*) = \frac{\mathcal{L}(\mathbb{P}\left(\zeta_{G(\cdot)} = k\right) g(\cdot))(\alpha^*)}{\mathcal{L}(\mathbb{P}\left(\zeta_{G(\cdot)} = k\right))(\alpha^*)},$$

which is a deterministic sequence.

Remark 7.3.2. *Notice that $\hat{\mathcal{L}}^g(k, \alpha^*) = 1$ when $g(t) \equiv 1$, so that $G(t) = t$ for every $t \in \mathbb{R}^+$ and there is no aging, and we retrieve the stationary process $(\zeta_t)_{t\geq 0}$.*

Unfortunately, the explicit expression of the coefficients $(\hat{\mathcal{L}}^g(k, \alpha^*))_{k\in\mathbb{N}}$ is not easy to find, even though they are deterministic.

Theorem 7.1.3, which states that even if g is integrable, the aging does not affect the explosive behavior of a birth process with superlinear PA function, is a direct consequence of (7.3.2):

Proof of Theorem 7.1.3. Consider a birth process $(\zeta_t)_{t\geq 0}$, defined by a superlinear sequence P $(f_k)_{k\in\mathbb{N}}$ (in the sense of Definition 7.2.1), and an integrable aging function g. Then, for every $t > 0$, $\mathbb{P}\left(\zeta_{G(t)} = \infty\right) > 0$. Since this holds for every $t > 0$, the process $(\zeta_{G(t)})_{t\geq 0}$ is explosive. As a consequence, for any $\alpha > 0$, $\mathbb{E}\left[\zeta_{G(T_\alpha)}\right] = \infty$, which means that there exists no Malthusian parameter. □

7.4. Affine PA sequence and adapted Laplace method

7.4.1. Aging and no fitness

In this section, we consider affine PA functions, i.e., we consider $f_k = ak + b$. The main aim is to identify the asymptotic behavior of the limiting degree distribution of the branching process with aging. Consider a stationary process $(\zeta_t)_{t\geq 0}$, where $f_k = ak+b$. Then, for any $t \geq 0$, it is possible to show by induction and the recursions in (2.1.8) and (2.1.9) that

$$P_k[\zeta](t) = \mathbb{P}\left(\zeta_t = k\right) = \frac{1}{\Gamma(b/a)} \frac{\Gamma(k+b/a)}{\Gamma(k+1)} \mathrm{e}^{-bt} \left(1 - \mathrm{e}^{-at}\right)^k. \tag{7.4.1}$$

We omit the proof of (7.4.1). As a consequence, since the corresponding aging process is $(\zeta_{G(t)})_{t\geq 0}$, the limiting degree distribution is given by

$$p_k^{(1)} = \mathbb{P}\left(\zeta_{G(t)} = k\right) = \frac{\Gamma(k+b/a)}{\Gamma(b/a)\Gamma(k+1)} \int_0^\infty \alpha^* \mathrm{e}^{-\alpha^* t} \mathrm{e}^{-bG(t)} \left(1 - \mathrm{e}^{-aG(t)}\right)^k dt. \tag{7.4.2}$$

We can obtain an immediate upper bound for $p_k^{(1)}$, in fact

$$\begin{aligned} p_k^{(1)} &= \frac{\Gamma(k+b/a)}{\Gamma(b/a)\Gamma(k+1)} \int_0^\infty \alpha^* \mathrm{e}^{-\alpha^* t} \mathrm{e}^{-bG(t)} \left(1 - \mathrm{e}^{-aG(t)}\right)^k dt \\ &\leq \frac{\Gamma(k+b/a)}{\Gamma(b/a)\Gamma(k+1)} (1 - \mathrm{e}^{-aG(\infty)})^k, \end{aligned}$$

which implies that the distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ has at most an exponential tail. A more precise analysis is hard. Instead we will give an asymptotic approximation, by adapting the Laplace method for integrals to our case.

The Laplace method states that, for a function f that is twice differentiable and with a unique absolute minimum $x_0 \in (a, b)$, as $k \to \infty$,

$$\int_a^b \mathrm{e}^{-k\Psi(x)} dx = \sqrt{\frac{2\pi}{k\Psi''(x_0)}} \mathrm{e}^{-k\Psi(x_0)} (1 + o(1)). \tag{7.4.3}$$

In this situation, the interval $[a, b]$ can be infinite. The idea behind this result is that, when $k \gg 1$, the major contribution to the integral comes from a neighborhood of x_0 where $\mathrm{e}^{-k\Psi(x)}$ is maximized. In the integral in (7.4.2), we do not have this situation, since we do not have an integral of the type (7.4.3). Defining

$$\Psi_k(t) := \frac{\alpha^*}{k} t + \frac{b}{k} G(t) - \log\left(1 - \mathrm{e}^{-aG(t)}\right), \tag{7.4.4}$$

we can rewrite the integral in (7.4.2) as

$$I(k) := \int_0^\infty \alpha^* \mathrm{e}^{-k\Psi_k(t)} dt. \tag{7.4.5}$$

The derivative of the function $\Psi_k(t)$ is

$$\Psi_k'(t) = \frac{\alpha^*}{k} + \frac{b}{k} g(t) - \frac{a g(t) \mathrm{e}^{-aG(t)}}{1 - \mathrm{e}^{-aG(t)}}. \tag{7.4.6}$$

In particular, if there exists a minimum t_k, then it depends on k. In this framework, we cannot directly apply the Laplace method. We now show that we can apply a result similar to (7.4.3) even to our case:

Lemma 7.4.1 (Adapted Laplace method 1)**.** *Consider $\alpha, a, b > 0$. Let the integrable aging function g be such that*

(1) for every $t \geq 0$, $0 < g(t) \leq A < \infty$;

(2) g is differentiable on $\mathbb{R}^+$, and g' is finite almost everywhere;

(3) there exists a positive constant $B < \infty$ such that $g(t)$ is decreasing for $t \geq B$;

(4) assume that the solution t_k of $\Psi_k'(t) = 0$, for $\Psi_k'(t)$ as in (7.4.6)*, is unique, and also $g'(t_k) < 0$.*

Then, for $\sigma_k^2 = (k\Psi_k''(t_k))^{-1}$, there exists a constant C such that, as $k \to \infty$,

$$I(k) = C\sqrt{2\pi\sigma_k^2} \mathrm{e}^{-k\Psi_k(t_k)} \left(\frac{1}{2} + \mathbb{P}\left(\mathcal{N}(0, \sigma_k^2) \geq t_k \right) \right) (1 + o(1)),$$

where $\mathcal{N}(0, \sigma_k^2)$ denotes a normal distribution with zero mean and variance σ_k^2.

Since Lemma 7.4.1 is an adapted version of the classical Laplace method, we move the proof to Section 7.6. We can use the result of Lemma 7.4.1 to prove:

Proposition 7.4.2 (Asymptotics - affine sequence, aging, no fitness)**.** *Consider the affine PA function $f_k = ak + b$, an integrable aging function g, and denote the limiting degree distribution of the corresponding branching process by $(p_k^{(1)})_{k \in \mathbb{N}}$. Then, under the hypotheses of Lemma 7.4.1, there exists a constant $C > 0$ such that, as $k \to \infty$,*

$$p_k^{(1)} = \frac{\Gamma(k + b/a)}{\Gamma(k+1)} \left(C g(t_k) - \frac{g'(t_k)}{g(t_k)} \right)^{1/2} \mathrm{e}^{-\alpha^* t_k} (1 - \mathrm{e}^{-aG(\infty)})^k D_k(g)(1 + o(1)), \tag{7.4.7}$$

where

$$D_k(g) = \frac{1}{2} + \frac{1}{2\sqrt{\pi}} \int_{-C_k(g)}^{C_k(g)} \mathrm{e}^{-\frac{u^2}{2}} du,$$

and $C_k(g) = t_k \left(C g(t_k) - \frac{g'(t_k)}{g(t_k)} \right)^{1/2}$.

7.4.2. Proof of Corollary 7.1.4

Here we prove the result on aging processes stated in Corollary 7.1.4. The result follows immediately from the proof of Corollary 2.2.3 and the definition of the aging process. In fact, an aging process is defined as $(\xi_{G(t)})_{t\geq 0}$, where $(\xi_t)_{t\geq 0}$ is an embedding process defined by the sequence $(k+1+\delta/m)_{k\in\mathbb{N}}$. As a simple consequence of the chain rule, from Proposition 2.1.16 it follows that

$$\frac{d}{dt}P_k[\xi](G(t)) = \Big[-(k+1+\delta/m)P_k[\xi](G(t)) + (k+\delta/m)P_{k-1}[\xi](G(t))\Big]g(t).$$

Assuming that the aging function g is bounded almost everywhere, Condition 2.3.1 is satisfied for $\ell = k\sup_{t\geq 0}|g(t)|$. The condition $\lim_{t\to\infty}\mathbb{E}[\xi_{G(t)}] > 1$ is necessary and sufficient for the existence of the Malthusian parameter α^* (see [73, Lemma 4.1]).

Since the sum of m processes $\xi_t^1 + \cdots + \xi_t^m$ is distributed as a single embedding process defined by the sequence $(k+m+\delta)_{k\in\mathbb{N}}$, it follows that $\xi_{G(t)}^1 + \cdots + \xi_{G(t)}^m$ is distributed as a single aging process with the same aging function g and sequence $(k+m+\delta)_{k\in\mathbb{N}}$. (7.1.5) is then a consequence of Proposition 7.4.2.

7.4.3. Aging and fitness case

In this section, we investigate the asymptotic behavior of the limiting degree distribution of a CTBP, in the case of affine PA function. The method we use is analogous to that in Section 7.4.1.

We assume that the fitness η is absolutely continuous with respect to the Lebesgue measure, and we denote its density function by μ. The limiting degree distribution of this type of branching process is given by

$$\begin{aligned} p_k^{(1)} &= \mathbb{P}\left(\zeta_{\eta G(T_{\alpha^*})} = k\right) \\ &= \frac{\Gamma(k+b/a)}{\Gamma(b/a)\Gamma(k+1)} \int_{\mathbb{R}^+\times\mathbb{R}^+} \alpha^* \mathrm{e}^{-\alpha^* t}\mu(s)\mathrm{e}^{-bsG(t)}\left(1-\mathrm{e}^{-asG(t)}\right)^k dsdt. \end{aligned} \tag{7.4.8}$$

We immediately see that the degree distribution has exponential tails when the fitness distribution is bounded:

Lemma 7.4.3 (Exponential tails for integrable aging and bounded fitnesses)**.** *When there exists γ such that $\mu([0,\gamma]) = 1$, i.e., the fitness has a bounded support, then*

$$p_k^{(1)} \leq \frac{\Gamma(k+b/a)}{\Gamma(b/a)\Gamma(k+1)}\left(1-\mathrm{e}^{-a\gamma G(\infty)}\right)^k.$$

In particular, $p_k^{(1)}$ has exponential tails.

Proof. Obvious. □

Like in the situation with only aging, the explicit solution of the integral in (7.4.8) may be hard to find. We again have to adapt the Laplace method to estimate the

asymptotic behavior of the integral. We write

$$I(k) := \int_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{e}^{-k\Psi_k(t,s)} ds dt, \tag{7.4.9}$$

where

$$\Psi_k(t,s) := \frac{\alpha^*}{k} t + \frac{b}{k} s G(t) - \frac{1}{k} \log \mu(s) - \log(1 - \mathrm{e}^{-saG(t)}). \tag{7.4.10}$$

As before, we want to minimize the function Ψ_k. We state here the lemma:

Lemma 7.4.4 (Adapted Laplace method 2)**.** *Let $\Psi_k(t,s)$ as in* (7.4.10)*. Assume that*

(1) g satisfies the assumptions of Lemma 7.4.1;
(2) μ is twice differentiable on $\mathbb{R}^+$;
(3) there exists a constant $B' > 0$ such that, for every $s \geq B'$, μ is monotonically decreasing;
(4) (t_k, s_k) is the unique point where both partial derivatives are zero;
(5) (t_k, s_k) is the absolute minimum for $\Psi_k(t,s)$;
(6) the hessian matrix $H_k(t_k, s_k)$ of $\Psi_k(t,s)$ evaluated in (t_k, s_k) is positive definite.

Then,

$$I(k) = \mathrm{e}^{-k\Psi_k(t_k,s_k)} \frac{2\pi}{\sqrt{\det(kH_k(t_k,s_k))}} \mathbb{P}\left(\mathcal{N}_1(k) \geq -t_k, \mathcal{N}_2(k) \geq -s_k\right)(1 + o(1)),$$

where $(\mathcal{N}_1(k), \mathcal{N}_2(k)) := \mathcal{N}(\mathbf{0}, (kH_k(t_k,s_k))^{-1})$ is a bivariate normal distributed vector and $\mathbf{0} = (0,0)$.

The proof of Lemma 7.4.4 can be found in Section 7.6.1. Using Lemma 7.4.4 we can describe the limiting degree distribution $(p_k^{(1)})_{k\in\mathbb{N}}$:

Proposition 7.4.5 (Asymptotics - affine sequence, aging, fitness)**.** *Consider affine PA function $f_k = ak + b$, an integrable aging function g and a fitness distribution density μ. Assume that the corresponding branching process is supercritical and Malthusian. Under the hypotheses of Lemma 7.4.4, the limiting degree distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ of the corresponding* CTBP *satisfies*

$$p_k^{(1)} = \frac{k^{b/a-1}}{\Gamma(b/a)} \frac{2\pi}{\sqrt{\det(kH_k(t_k,s_k))}} \mathrm{e}^{-k\Psi_k(t_k,s_k)} \mathbb{P}\left(\mathcal{N}_1 \geq -t_k, \mathcal{N}_2 \geq -s_k\right)(1 + o(1)).$$

7.4.4. Three classes of fitness distributions

Proposition 7.4.5 in Section 7.4.3 gives the asymptotic behavior of the limiting degree distribution of a CTBP with integrable aging and fitness. Lemma 7.4.4 requires conditions under which the function $\Psi_k(t,s)$ as in (7.4.10) has a unique minimum

point denoted by (t_k, s_k). In this section we consider the three different classes of fitness distributions that we have introduced in Section 7.1.4.

For the heavy-tailed class, i.e., for distributions with tail thicker than exponential, there is nothing to prove. In fact, (7.1.8) immediately implies that such distributions are explosive.

For the other two cases, we apply Proposition 7.4.5, giving the precise asymptotic behavior of the limiting degree distributions of the corresponding CTBPs. Propositions 7.4.6 and 7.4.7 contain the results respectively on the general-exponential and sub-exponential classes. The proof of these propositions are moved to Section 7.7.

Proposition 7.4.6. *Consider a general exponential fitness distribution as in* (7.1.12)*. Let* $(\zeta_{\eta G(t)})_{t\geq 0}$ *be the corresponding birth process. Denote the unique minimum point of* $\Psi_k(t,s)$ *as in* (7.4.10) *by* (t_k, s_k)*. Then*

(1) for every $t \geq 0$, $\zeta_{\eta G(t)}$ *has a dynamical power law with exponent* $\tau(t) = 1 + \frac{\theta}{aG(t)}$;

(2) the asymptotic behavior of the limiting degree distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ *is given by*

$$p_k^{(1)} = \mathrm{e}^{-\alpha^* t_k} h(s_k) \left(\tilde{C} - \alpha^* \frac{g'(t_k)}{g(t_k)} \right)^{-1/2} k^{-(1+\theta/(aG(\infty)))} (1 + o(1)),$$

where the power law term has exponent $\tau = 1 + \theta/aG(\infty)$;

(3) the distribution $(q_k)_{k\in\mathbb{N}}$ *of the total number of children of a fixed individual has a power law behavior with exponent* $\tau = 1 + \theta/aG(\infty)$.

By (7.1.9) it is necessary to consider the exponential rate $\theta > aG(\infty)$ to obtain a non-explosive process. In particular, this implies that, for every $t \geq 0$, $\tau(t)$, as well as τ, are strictly larger than 2. As a consequence, the three distributions $(\mathbb{E}[P_k[\zeta](\eta G(t))])_{k\in\mathbb{N}}$, $(p_k^{(1)})_{k\in\mathbb{N}}$ and $(q_k)_{k\in\mathbb{N}}$ have finite first moment. Increasing the value of θ leads to power-law distributions with exponent larger than 3, so with finite variance.

A second observation is that, independently of the aging function g, the point s_k is of order $\log k$. In particular, this has two consequences. First the correction to the power law given by $h(s_k)$ is a power of $\log k$. Since $h'(s)/h(s) \to 0$ as $s \to \infty$. Second the power-law term $k^{-(1+\theta/(aG(\infty)))}$ arises from $\mu(s_k)$. This means that the exponential term in the fitness distribution μ not only is necessary to obtain a non-explosive process, but also generates the power law.

The third observation is that the behavior of $(\mathbb{E}[P_k[\zeta](\eta G(t))])_{k\in\mathbb{N}}$, $(p_k^{(1)})_{k\in\mathbb{N}}$ and $(q_k)_{k\in\mathbb{N}}$ depends on the integrability of the aging function, *but does only marginally depends on its precise shape*. The contribution of the aging function g to the exponent of the power law in fact is given only by the value $G(\infty)$. The other terms that depend directly on the shape of g are $\mathrm{e}^{-\alpha^* t_k}$ and the ratio $g'(t_k)/g(t_k)$. The ratio g'/g does not contribute for any function g whose decay is in between power law and exponential. The term $\mathrm{e}^{-\alpha^* t_k}$ depends on the behavior of t_k, that can be seen as roughly $g^{-1}(1/\log k)$. For any function between power law and exponential, $\mathrm{e}^{-\alpha^* t_k}$ is asymptotic to a power of $\log k$.

The last observation is that every distribution in the general exponential class shows a *dynamical power law* as for the pure exponential distribution, as shown in Section 7.4.5. The pure exponential distribution is a special case where we consider $h(s) \equiv 1$. Interesting is the fact that τ actually does not depend on the choice of $h(s)$, but only on the exponential rate $\theta > aG(\infty)$. In particular, Proposition 7.4.6 proves that the limiting degree distribution of the two examples in Figure 1.24 have power-law decay.

We move to the class of sub-exponential fitness. We show that the power law is lost due to the absence of a pure exponential term. We prove the result using densities of the form

$$\mu(s) = C\mathrm{e}^{-s^{1+\varepsilon}}, \tag{7.4.11}$$

for $\varepsilon > 0$ and C the normalization constant. The result is the following:

Proposition 7.4.7 (Sub-exp fitness)**.** *Consider a sub-exponential fitness distribution as in* (7.4.11)*. Let $(\zeta_{\eta G(t)})_{t\geq 0}$ be the corresponding birth process. Denote the minimum point of $\Psi_k(t,s)$ as in* (7.4.10) *by (t_k, s_k). Then*

(1) for every $t \geq 0$, M_t satisfies

$$\mathbb{P}\left(\zeta_{\eta G(t)} = k\right) = k^{-1}(\log k)^{-\varepsilon/2}\mathrm{e}^{-\frac{\theta}{(aG(t))^{1+\varepsilon}}(\log k)^{1+\varepsilon}}(1+o(1));$$

(2) the limiting degree distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ of the CTBP *has asymptotic behavior given by*

$$p_k^{(1)} = \mathrm{e}^{-\alpha^* t_k}k^{-1}\left(C_1 - s_k^{\varepsilon}\frac{g'(t_k)}{g(t_k)}\right)\mathrm{e}^{-\frac{\theta}{(aG(\infty))^{1+\varepsilon}}(\log k)^{1+\varepsilon}}(1+o(1));$$

(3) the distribution $(q_k)_{k\in\mathbb{N}}$ of the total number of children of a fixed individual satisfies

$$q_k = k^{-1}(\log k)^{-\varepsilon/2}\mathrm{e}^{-\frac{\theta}{(aG(\infty))^{1+\varepsilon}}(\log k)^{1+\varepsilon}}(1+o(1)).$$

In Proposition 7.4.7 the distributions $(\mathbb{E}[P_k[\zeta](\eta G(t))])_{k\in\mathbb{N}}$, $(p_k^{(1)})_{k\in\mathbb{N}}$ and $(q_k)_{k\in\mathbb{N}}$ decay faster than a power law. This is due to the fact that a sub-exponential tail for the fitness distribution does not allow the presence of sufficiently many individuals in the branching population whose fitness value is sufficiently high to restore the power law.

In this case, we have that s_k is roughly $c_1 \log k - c_2 \log\log k$. Hence, as first approximation, s_k is still of logarithmic order. The power-law term is lost because there is no pure exponential term in the distribution μ. In fact, in this case $\mu(s_k)$ generates the dominant term $\mathrm{e}^{-\theta(\log k)^{1+\varepsilon}}$.

7.4.5. The case of exponentially distributed fitness: Proof of Corollary 7.1.7

The case when the fitness η is exponentially distributed turns out to be simpler. In this section, denote the fitness by T_θ, where θ is the parameter of the exponential

distribution. First of all, we investigate the Laplace transform of the process. In fact, we can write

$$\mathbb{E}\left[\zeta_{\eta G(T_\alpha)}\right] = \int_0^\infty \theta \mathrm{e}^{-\theta s} \mathbb{E}\left[\zeta_{sG(T_\alpha)}\right] ds,$$

which is the Laplace transform of the stationary process $(\zeta_{sG(T_\alpha)})_{s\geq 0}$ with bounded fitness $G(T_\alpha)$ in θ. As a consequence,

$$\mathbb{E}\left[\zeta_{\eta G(T_\alpha)}\right] = \sum_{k\in\mathbb{N}} \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{f_i G(T_\alpha)}{\theta + f_i G(T_\alpha)}\right].$$

Suppose that there exists a Malthusian parameter α^*. This means that, for fixed $(f_k)_{k\in\mathbb{N}}$, g and θ, α^* is the unique value such that $\mathbb{E}\left[\zeta_{\eta G(T_{\alpha^*})}\right] = 1$. As a consequence, if we fix $(f_k)_{k\in\mathbb{N}}$, g and α^*, θ is the unique value such that

$$\sum_{k\in\mathbb{N}} \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{f_i G(T_\alpha)}{\theta + f_i G(T_\alpha)}\right] = 1.$$

Therefore θ is the Malthusian parameter of the process $(V_{sG(T_\alpha)})_{s\geq 0}$. We are now ready to prove Corollary 7.1.7:

Proof of Corollary 7.1.7. We have to evaluate the Laplace transform of $\mathbb{P}\left(\zeta_{sG(t)} = k\right)$ in θ. Using (7.4.1) the first part follows immediately by simple calculations. For the second part, we just need to take the limit as $t \to \infty$. For the sequence $(p_k^{(1)})_{k\in\mathbb{N}}$, the result is immediate since $p_k^{(1)} = \mathbb{E}[P_k[\zeta](\eta G(T_{\alpha^*}))]$. □

The case of affine PA function $f_k = ak + b$ is particularly nice. As already mentioned in Section 7.1, the process $(\zeta_{\eta G(t)})_{t\geq 0}$ has a power-law distribution at every $t \in \mathbb{R}^+$ and (7.1.13) follows immediately. Further, (7.1.14) and (7.1.15) follow directly. □

7.5. Limiting distribution with aging effect, no fitness

In this section, we analyze the limiting degree distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ of CTBPs with aging but no fitness. In Section 7.5.1 we prove the adapted Laplace method for the general asymptotic behavior of $p_k^{(1)}$. In Section 7.5.2 we consider some examples of aging function g, giving the asymptotics for the corresponding distributions.

7.5.1. Proofs of Lemma 7.4.1 and Proposition 7.4.2

Proof of Lemma 7.4.1. First of all, we show that t_k is actually a minimum. In fact,

$$\lim_{t\to 0} \frac{d}{dt}\Psi_k(t) = -\infty, \qquad \text{and} \qquad \lim_{t\to\infty} \frac{d}{dt}\Psi_k(t) = \frac{\alpha}{k} > 0.$$

As a consequence, t_k is a minimum. Then,

$$\lim_{k\to\infty} g(t_k)\left(\frac{\alpha^*}{k}\frac{1-\mathrm{e}^{-aG(\infty)}}{a\mathrm{e}^{-aG(\infty)}}\right)^{-1} = \lim_{k\to\infty} g(t_k)\frac{ak}{\alpha^*(\mathrm{e}^{aG(\infty)}-1)} = 1. \tag{7.5.1}$$

In particular, $g(t_k)$ is of order $1/k$. Then, since t_k is the actual minimum, and g is monotonically decreasing for $t \geq B$,

$$\Psi_k''(t_k) = \frac{b}{k}g'(t_k) + g(t_k)^2\frac{a^2\mathrm{e}^{-aG(t_k)}(2-\mathrm{e}^{-aG(t_k)})}{(1-\mathrm{e}^{-aG(t_k)})^2} - g'(t_k)\frac{a\mathrm{e}^{-aG(t_k)}}{1-\mathrm{e}^{-aG(t_k)}} > 0. \tag{7.5.2}$$

We use the fact that we are evaluating the second derivative in the point t_k where the first derivative is zero. This means

$$g(t_k)\frac{a\mathrm{e}^{-aG(t_k)})}{1-\mathrm{e}^{-aG(t_k)}} = \frac{\alpha}{k} + \frac{b}{k}g(t_k).$$

We use this in (7.5.2) to obtain

$$\begin{aligned} k\Psi_k''(t_k) &= bg'(t_k) + g(t_k)\frac{a(2-\mathrm{e}^{-aG(t_k)})}{1-\mathrm{e}^{-aG(t_k)}}(\alpha+bg(t_k)) - \frac{g'(t_k)}{g(t_k)}(\alpha+bg(t_k)) \\ &= g(t_k)\frac{a(2-\mathrm{e}^{-aG(t_k)})}{1-\mathrm{e}^{-aG(t_k)}}(\alpha+bg(t_k)) - \alpha\frac{g'(t_k)}{g(t_k)}. \end{aligned} \tag{7.5.3}$$

Now, we use Taylor expansion around t_k of $\Psi_k(t)$ in the integral in (7.4.5). Since we use the expansion around t_k, which is the minimum of $\Psi_k(t)$, the first derivative of Ψ_k is zero. As a consequence, we have

$$I(k) = \int_0^\infty \mathrm{e}^{-k\left(\Psi_k(t_k)+\frac{1}{2}\Psi_k''(t_k)(t-t_k)^2+o((t-t_k)^2)\right)}dt.$$

First of all, notice that the contribution of the terms with $|t-t_k| \gg 1$ is negligible. In fact, we have

$$\mathrm{e}^{-k\Psi_k(t)} \leq \mathrm{e}^{-\alpha^* t}(1-\mathrm{e}^{-aG(\infty)})^k,$$

which means that such terms are exponentially small, so we can ignore them. Now we make a change of variable $u = t - t_k$. Then

$$I(k) = \int_{-t_k}^\infty \mathrm{e}^{-k\left(\Psi_k(t_k)+\frac{1}{2}\Psi_k''(t_k)u^2+o(u^2)\right)}du.$$

In particular, since the term $\mathrm{e}^{-k\Psi_k(t_k)}$ does not depend on u, we can write

$$I(k) = \mathrm{e}^{-k\Psi_k(t_k)}\int_{-t_k}^\infty \mathrm{e}^{-k\left(\frac{1}{2}\Psi_k''(t_k)u^2+o(u^2)\right)}du.$$

We use the notation $k\Psi_k(t_k) = \frac{1}{\sigma_k^2}$, which means we can rewrite the integral as

$$\mathrm{e}^{-k\Psi_k(t_k)}\sqrt{2\pi\sigma_k^2}\int_{-\infty}^{t_k}\frac{1}{\sqrt{2\pi\sigma_k^2}}\mathrm{e}^{-\frac{u^2}{2\sigma_k^2}}dt = \mathrm{e}^{-k\Psi_k(t_k)}\sqrt{2\pi\sigma_k^2}\mathbb{P}\left(\mathcal{N}(0,\sigma_k^2)\leq t_k\right).$$

Since the distribution $\mathcal{N}(0,\sigma_k^2)$ is symmetric with respect to 0, for every $k\in\mathbb{N}$,

$$\mathbb{P}\left(\mathcal{N}(0,\sigma_k^2)\leq t_k\right) = \frac{1}{2}\left[1+\frac{1}{\sqrt{\pi}}\int_{-t_k/\sigma_k}^{t_k/\sigma_k}\mathrm{e}^{-\frac{u^2}{2}}du\right]. \tag{7.5.4}$$

The behavior of the above integral depends on the ratio t_k/σ_k, which is bounded between 0 and 1. As a consequence, the term $\mathbb{P}\left(\mathcal{N}(0,\sigma_k^2)\leq t_k\right)$ is bounded between $1/2$ and 1. □

Using Lemma 7.4.1, we can prove Proposition 7.4.2:

Proof of Proposition 7.4.2. Recall that $\sigma_k^2 = (k\Psi(t_k)'')^{-1}$. Using (7.5.3), the fact that g is bounded almost everywhere, and $g'(t_k)<0$, we can write

$$k\Psi(t_k)'' = \alpha\left(\frac{a(2-\mathrm{e}^{-aG(\infty)})}{1-\mathrm{e}^{-aG(\infty)}}g(t_k) - \frac{g'(t_k)}{g(t_k)}\right)(1+o(1)). \tag{7.5.5}$$

Notice that in (7.5.5) the terms $g(t_k) - \frac{g'(t_k)}{g(t_k)}$ are always strictly positive, since $g(t)$ is decreasing and $t_k\to\infty$ as $k\to\infty$. As a consequence, we can replace the term $\sqrt{2\pi/\sigma_k^2}$ by $\left(Cg(t_k) - \frac{g'(t_k)}{g(t_k)}\right)^{1/2}$, for $C = \frac{a(2-\mathrm{e}^{-aG(\infty)})}{1-\mathrm{e}^{-aG(\infty)}}$. We also have that

$$\begin{aligned}\mathrm{e}^{-k\Psi_k(t_k)} &= \exp\left[-\alpha^* t_k - bG(t_k) + k\log\left(1-\mathrm{e}^{-aG(t_k)}\right)\right]\\ &= \mathrm{e}^{-\alpha^* t_k}(1-\mathrm{e}^{-aG(\infty)})^k(1+o(1)),\end{aligned}$$

since $G(t_k)$ converges to $G(\infty)$. For the term in (7.5.4), it is easy to show that it is asymptotic to $D_k(g)$. This completes the proof. □

7.5.2. Examples of aging functions

In this section, we analyze two examples of aging functions, in order to give examples of the limiting degree distribution of the branching process. We consider affine PA function $f_k = ak+b$, and three different aging functions:

$$g(t) = \mathrm{e}^{-\lambda t}, \qquad g(t) = (1+t)^{-\lambda}, \qquad \text{and} \qquad g(t) = \lambda_1\mathrm{e}^{-\lambda_2(\log(t+1)-\lambda_3)^2}.$$

We assume that in every case the aging function g is integrable, so we consider $\lambda>0$ for the exponential case, $\lambda>1$ for the power-law case and $\lambda_1,\lambda_2,\lambda_3>0$ for the lognormal case. We assume that g satisfies Condition (7.1.4) in order to have a supercritical process.

We now apply (7.4.7) to these three examples, giving their asymptotics. In general, we approximate t_k with the solution of, for $c_1 = \frac{a\mathrm{e}^{-aG(\infty)}}{1-\mathrm{e}^{-aG(\infty)}}$,

$$\frac{\alpha^*}{k} + \frac{b}{k} g(t) - c_1 g(t) = 0. \tag{7.5.6}$$

We start considering the exponential case $g(t) = \mathrm{e}^{-\lambda t}$. In this case, from (7.5.6) we obtain that, ignoring constants,

$$t_k = \log k(1 + o(1)). \tag{7.5.7}$$

As we expected, $t_k \to \infty$. We now use (7.5.5), which gives a bound on σ_k^2 in (7.5.1) in terms of g and its derivatives. As a consequence,

$$\left(g(t_k) - \frac{g'(t_k)}{g(t_k)}\right)^{-1/2} = \left(\mathrm{e}^{-\lambda t_k} + \lambda\right)^{1/2} \sim \lambda^{1/2}(1 + o(1)).$$

Looking at $\mathrm{e}^{-k\Psi_k(t_k)}$, it is easy to compute that, with t_k as in (7.5.7),

$$\exp\left[-\alpha^* \log k + bG(t_k) + k \log(1 - \mathrm{e}^{-aG(t_k)})\right] = k^{-\alpha^*}(1 + o(1)).$$

Since $t_k/\sigma_k \to \infty$, then $\mathbb{P}\left(\mathcal{N}(0, \sigma_k^2) \le t_k\right) \to 1$, so that

$$p_k^{(1)} = \frac{\Gamma(k + b/a)}{\Gamma(b/a)} \frac{1}{\Gamma(k+1)} C_1 k^{-\alpha^*} \mathrm{e}^{-C_2 k}(1 + o(1)),$$

which means that $p_k^{(1)}$ has an exponential tail with power-law corrections.

We now apply the same result to the power-law aging function, so $g(t) = (1+t)^{-\lambda}$, and $G(t) = \frac{1}{\lambda - 1}(1+t)^{1-\lambda}$. In this case

$$(1 + t_k) = \left(\frac{\alpha^*}{c_1 k}\right)^{-1/\lambda} (1 + o(1)).$$

We use again (7.5.1), so

$$\left(g(t_k) - \frac{g'(t_k)}{g(t_k)}\right) = \left(\frac{\alpha^*}{c_1 k} + \lambda \left(\frac{c_1 k}{\alpha^*}\right)^{1/\lambda}\right)^{1/2} \sim k^{\alpha^*/2\lambda}(1 + o(1)).$$

In conclusion,

$$p_k^{(1)} = \frac{\Gamma(k + b/a)}{\Gamma(b/a)} \frac{1}{\Gamma(k+1)} k^{\alpha^*/2\lambda} \mathrm{e}^{-\alpha^* \left(\frac{\alpha^*}{c_1 k}\right)^{-1/\lambda} - C_2 k}(1 + o(1)),$$

which means that also in this case we have a power-law with exponential truncation.

In the case of the lognormal aging function, (7.5.6) implies that

$$[\log(t_k+1)-\lambda_3]^2 \approx +\frac{1}{\lambda_2}\log\left(\frac{c_1}{\alpha^*}k\right).$$

By (7.5.1) we can say that

$$\begin{aligned}\left(g(t_k)-\frac{g'(t_k)}{g(t_k)}\right) &= \left(\lambda_1\log\left(\frac{c_1}{\alpha^*}k\right)+2\lambda_2\frac{\log(t_k+1)}{t_k+1}\right)(1+o(1))\\ &= \lambda_1\log\left(\frac{c_1}{\alpha^*}k\right)(1+o(1)).\end{aligned}$$

We conclude then, for some constant $C_3>0$,

$$p_k^{(1)} = \frac{\Gamma(k+b/a)}{\Gamma(b/a)}\frac{1}{\Gamma(k+1)}\left(\lambda_1\log\left(\frac{c_1}{\alpha^*}k\right)\right)^{1/2}\mathrm{e}^{-\alpha^*\mathrm{e}^{(\log(\frac{c_1}{\alpha^*}k))^{1/2}}}\mathrm{e}^{-C_3k}(1+o(1)).$$

7.6. Limiting distribution with aging and fitness

In this section, we consider birth processes with aging and fitness. We prove Lemma 7.4.4, used in the proof of Proposition 7.4.5. Then we give examples of limiting degree distributions for different aging functions and exponentially distributed fitness.

7.6.1. Proofs of Lemma 7.4.4 and Proposition 7.4.5

Proof of Lemma 7.4.4. We use again second order Taylor expansion of the function $\Psi_k(t,s)$ centered in (t_k,s_k), where the first order partial derivatives are zero. As a consequence we write

$$\exp\left[-k\Psi_k(t,s)\right] = \exp\left[-k\Psi_k(t_k,s_k)+\frac{1}{2}\boldsymbol{x}^T\left(kH_k(t_k,s_k)\right)\boldsymbol{x}+o(||\boldsymbol{x}||^2)\right],$$

where

$$\boldsymbol{x}=\begin{bmatrix}t-t_k\\ s-s_k\end{bmatrix},\quad\text{and}\quad H_k(t_k,s_k)=\begin{bmatrix}\frac{\partial^2\Psi_k}{\partial t^2}(t_k,s_k) & \frac{\partial^2\Psi_k}{\partial s\partial t}(t_k,s_k)\\ \frac{\partial^2\Psi_k}{\partial s\partial t}(t_k,s_k) & \frac{\partial^2\Psi_k}{\partial s^2}(t_k,s_k)\end{bmatrix}.$$

As for the proof of Lemma 7.4.1, we start by showing that we can ignore the terms where $||\boldsymbol{x}||^2\gg 1$. In fact,

$$\mathrm{e}^{-k\Psi_k(t,s)}\le\exp\left(-\alpha^*t-bsG(t)+k\log(\mu(s))\right).$$

Since μ is a probability density, $\mu(s)<1$ for $s\gg 1$. As a consequence, $\log(\mu(s))<0$, which means that the above bound is exponentially decreasing whenever t and s are

very large. As a consequence, we can ignore the contribution given by the terms where $|t - t_k| \gg 1$ and $|s - s_k| \gg 1$.

The term $\mathrm{e}^{-k\Psi_k(t_k,s_k)}$ is independent of t and s, so we do not consider it in the integral. Writing $u = t - t_k$ and $v = s - s_k$, we can write

$$\int_{\mathbb{R}^+\times\mathbb{R}^+} \mathrm{e}^{-\frac{1}{2}\boldsymbol{x}^T(kH_k(t_k,s_k))\boldsymbol{x}} ds dt = \int_{-t_k}^{\infty}\int_{-s_k}^{\infty} \mathrm{e}^{-\frac{1}{2}\boldsymbol{y}^T(kH_k(t_k,s_k))\boldsymbol{y}} du dv,$$

where this time $\boldsymbol{y}^T = [u\ v]$. As a consequence,

$$\begin{aligned}\int_{-t_k}^{\infty}\int_{-s_k}^{\infty} &\mathrm{e}^{-\frac{1}{2}\boldsymbol{y}^T(kH_k(t_k,s_k))\boldsymbol{y}} du dv \\ &\times = \frac{2\pi}{\sqrt{\det(kH_k(t_k,s_k))}}\mathbb{P}\left(\mathcal{N}_1(k) \geq -t_k, \mathcal{N}_2(k) \geq -s_k\right),\end{aligned} \tag{7.6.1}$$

provided that the covariance matrix $(kH_k(t_k,s_k))^{-1}$ is positive definite. □

As a consequence, we can use (7.6.1) to obtain that, for the corresponding limiting degree distribution of the branching process $(p_k^{(1)})_{k\in\mathbb{N}}$, as $k \to \infty$,

$$\begin{aligned}p_k^{(1)} = &\frac{\Gamma(k+b/a)}{\Gamma(b/a)}\frac{1}{\Gamma(k+1)}\frac{2\pi}{\sqrt{\det(kH_k(t_k,s_k))}} \\ &\times \mathrm{e}^{-k\Psi_k(t_k,s_k)}\mathbb{P}\left(\mathcal{N}_1(k) \geq -t_k, \mathcal{N}_2(k) \geq -s_k\right)(1+o(1)).\end{aligned}$$

This results holds if the point (t_k, s_k) is the absolute minimum of Ψ_k, and the Hessian matrix is positive definite at (t_k, s_k).

7.6.2. The Hessian matrix of $\Psi_k(t,s)$

First of all, we need to find a point (t_k, s_k) which is the solution of the system

$$\frac{\partial\Psi_k}{\partial t} = \frac{\alpha^*}{k} + \frac{b}{k}sg(t) - \frac{sag(t)\mathrm{e}^{-saG(t)}}{1-\mathrm{e}^{-saG(t)}} = 0, \tag{7.6.2}$$

$$\frac{\partial\Psi_k}{\partial s} = \frac{b}{k}G(t) - \frac{1}{k}\frac{\mu'(s)}{\mu(s)} - \frac{aG(t)\mathrm{e}^{-saG(t)}}{1-\mathrm{e}^{-saG(t)}} = 0. \tag{7.6.3}$$

Denote the solution by (t_k, s_k). Then

$$\begin{aligned}\frac{\partial^2\Psi_k}{\partial t^2} &= \frac{b}{k}sg'(t_k) + g(t_k)^2\frac{s^2a^2\mathrm{e}^{-asG(t_k)}}{(1-\mathrm{e}^{-asG(t_k)})^2} - g'(t_k)\frac{as\mathrm{e}^{-asG(t_k)}}{1-\mathrm{e}^{-asG(t_k)}}, \\ \frac{\partial^2\Psi_k}{\partial s^2} &= -\frac{1}{k}\frac{\mu''(s)\mu(s)-\mu'(s)^2}{\mu(s)^2} + \frac{a^2G(t)^2\mathrm{e}^{-saG(t)}}{(1-\mathrm{e}^{-saG(t)})^2}, \\ \frac{\partial^2\Psi_k}{\partial s\partial t} &= \frac{b}{k}g(t_k) + \left(1-\frac{1}{as_k}\right)\left(\frac{b}{k}G(t_k) - \frac{1}{k}\frac{\mu'(s_k)}{\mu(s_k)}\right)\left(\frac{\alpha^*}{k} + \frac{b}{k}s_kg(t_k)\right).\end{aligned}$$

From (7.6.2) and (7.6.3) we know

$$\begin{aligned}\frac{\alpha^*}{k}+\frac{b}{k}s_k g(t_k) &= \frac{s_k a g(t_k)\mathrm{e}^{-s_k aG(t_k)}}{1-\mathrm{e}^{-s_k aG(t_k)}},\\ \frac{b}{k}G(t_k)-\frac{1}{k}\frac{\mu'(s_k)}{\mu(s_k)} &= \frac{aG(t_k)\mathrm{e}^{-s_k aG(t_k)}}{1-\mathrm{e}^{-s_k aG(t_k)}}.\end{aligned} \tag{7.6.4}$$

Using (7.6.4) in the expressions for the second derivatives,

$$\begin{aligned}\frac{\partial^2\Psi_k}{\partial t^2} &= \frac{b}{k}s_k g'(t_k)+\frac{as_k g(t_k)}{(1-\mathrm{e}^{-as_k G(t_k)})}\left(\frac{\alpha}{k}+\frac{b}{k}s_k g(t_k)\right)-\frac{g'(t_k)}{g(t_k)}\left(\frac{\alpha}{k}+\frac{b}{k}s_k g(t_k)\right)\\ &= \frac{as_k g(t_k)}{(1-\mathrm{e}^{-as_k G(t_k)})}\left(\frac{\alpha}{k}+\frac{b}{k}s_k g(t_k)\right)-\frac{\alpha}{k}\frac{g'(t_k)}{g(t_k)},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\Psi_k}{\partial s^2} &= -\frac{1}{k}\frac{\mu''(s_k)\mu(s_k)-\mu'(s_k)^2}{\mu(s_k)^2}+\frac{aG(t_k)}{1-\mathrm{e}^{-as_k G(t_k)}}\left(\frac{b}{k}G(t_k)-\frac{1}{k}\frac{\mu'(s_k)}{\mu(s_k)}\right)\\ &= -\frac{1}{k}\frac{\mu''(s_k)}{\mu(s_k)}+\frac{1}{k}\left(\frac{\mu'(s_k)}{\mu(s_k)}\right)^2-\frac{1}{k}\frac{\mu'(s_k)}{\mu(s_k)}\frac{aG(t_k)}{1-\mathrm{e}^{-as_k G(t_k)}}+\frac{1}{k}\frac{abG(t_k)^2}{1-\mathrm{e}^{-as_k G(t_k)}}.\end{aligned}$$

In conclusion, the matrix $kH_k(t_k,s_k)$ is given by

$$(kH_k(t_k,s_k))_{1,1} = \frac{as_k g(t_k)}{(1-\mathrm{e}^{-as_k G(t_k)})}\left(\alpha+bs_k g(t_k)\right)-\alpha\frac{g'(t_k)}{g(t_k)}; \tag{7.6.5}$$

$$(kH_k(t_k,s_k))_{2,2} = -\frac{\mu''(s_k)}{\mu(s_k)}+\left(\frac{\mu'(s_k)}{\mu(s_k)}\right)^2-\frac{\mu'(s_k)}{\mu(s_k)}\frac{aG(t_k)}{1-\mathrm{e}^{-as_k G(t_k)}}+\frac{abG(t_k)^2}{1-\mathrm{e}^{-as_k G(t_k)}};$$

$$(kH_k(t_k,s_k))_{2,1} = bg(t_k)+\left(1-\frac{1}{as_k}\right)\left(\frac{b}{k}G(t_k)-\frac{1}{k}\frac{\mu'(s_k)}{\mu(s_k)}\right)\left(\alpha^*+bs_k g(t_k)\right).$$

We point out that, solving (7.6.3) in terms of s, it follows that

$$s = \frac{1}{aG(t)}\log\left(1+k\frac{aG(t)}{bG(t)-\frac{\mu'(s)}{\mu(s)}}\right). \tag{7.6.6}$$

As a consequence,

$$s_k g(t_k) = \alpha^* G(t_k)\frac{\mu(s_k)}{\mu'(s_k)}. \tag{7.6.7}$$

We use (7.6.6), (7.6.7) and the expressions for the elements of the Hessian matrix given in (7.6.5) for the examples in Section 7.6.3. We also use the formulas of this section in the proof of Propositions 7.4.6 and 7.4.7 given in Section 7.7.

7.6.3. Examples of aging functions

Here we give examples of limiting degree distributions. We consider the same three examples of aging functions we considered in Section 7.5.2, so

$$g(t) = \mathrm{e}^{-\lambda t}, \qquad g(t) = (1+t)^{-\lambda}, \qquad \text{and} \qquad g(t) = \lambda_1 \mathrm{e}^{-\lambda_2(\log(t+1)-\lambda_3)^2}.$$

We consider exponentially distributed fitness, so $\mu(s) = \theta \mathrm{e}^{-\theta s}$. In order to have a supercritical and Malthusian process, we can rewrite Condition (7.1.9) for exponentially distributed fitness as $aG(\infty) < \theta < (a+b)G(\infty)$.

In general, we identify the minimum point (t_k, s_k), then use (7.7.6). For all three examples, replacing $G(t)$ by $G(\infty)$ and using (7.6.6), it holds that

$$s_k \approx \frac{1}{aG(\infty)} \log\left(k \frac{aG(\infty)}{bG(\infty)+\theta}\right),$$

and $s_k g(t_k) \approx \alpha^* G(\infty)\theta$. For the exponential aging function, using (7.6.7), it follows that $\mathrm{e}^{-\lambda t_k} \approx \log k$. In this case, since $g'(t)/g(t) = -\lambda$, the conclusion is that, ignoring the constants,

$$p_k^{(1)} = k^{-(1+\lambda\theta/a)} (\log k)^{\alpha^*/\lambda} (1+o(1)).$$

For the inverse-power aging function $t_k \approx (\log k)^{1/\lambda}$, which implies (ignoring again the constants) that

$$p_k^{(1)} = k^{-(1+(\lambda-1)\theta/a)} \mathrm{e}^{-\alpha^* (\log k)^{1/\lambda}} (1+o(1)),$$

where we recall that, for g being integrable, $\lambda > 1$. For the lognormal case,

$$t_k \approx \mathrm{e}^{(\log k)^{1/2}},$$

which means that

$$p_k^{(1)} = k^{-(1+\theta/aG(\infty))} \mathrm{e}^{-\alpha^* \mathrm{e}^{(\log k)^{1/2}}} (1+o(1)).$$

7.7. Dynamical power law: proof of propositions 7.4.6 and 7.4.7

In the present section, we prove Propositions 7.4.6 and 7.4.7. These proofs are applications of Proposition 7.4.5, and mainly consist of computations. In the proof of the two propositions, we often refer to Appendix 7.6.2 for expressions regarding the Hessian matrix of $\Psi_k(t,s)$ as in (7.4.10).

7.7.1. Proof of Proposition 7.4.6

We start by proving the existence of the dynamical power-law. We already know that

$$\mathbb{P}\left(M_t = k\right) = \frac{\Gamma(k+b/a)}{\Gamma(b/a)\Gamma(k+1)} \int_0^\infty \mu(s)\mathrm{e}^{-bsG(t)} \left(1-\mathrm{e}^{-asG(t)}\right)^k ds. \tag{7.7.1}$$

We write

$$J(k) = \int_0^\infty \mathrm{e}^{-k\psi_k(s)} ds, \tag{7.7.2}$$

where

$$\psi_k(s) = \frac{bG(t)}{k} s - \frac{1}{k}\log(\mu(s)) - \log\left(1-\mathrm{e}^{-asG(t)}\right). \tag{7.7.3}$$

In order to give asymptotics on $J(k)$ as in (7.7.2), we can use a Laplace method similar to the one used in the proof of Lemma 7.4.1, but the analysis is simpler since in this case $\psi_k(s)$ is a function of only one variable. The idea is again to find a minimum point s_k for $\psi_k(s)$, and to use Taylor expansion inside the integral, so

$$\psi_k(s) = \psi_k(s_k) + \frac{1}{2}{\psi''}_k(s_k)(s-s_k)^2 + o((s-s_k)^2).$$

We can ignore the contribution of the terms where $(s-s_k)^2 \gg 1$, since $\mathrm{e}^{-k\psi_k(s)} \leq \mathrm{e}^{-bsG(t)}$, so that the error is at most exponentially small. As a consequence,

$$J(k) = \sqrt{\frac{\pi}{{\psi''}_k(s_k)}} \mathrm{e}^{-k\psi_k(s_k)}(1+o(1)).$$

The minimum s_k is a solution of

$$\frac{d\psi_k(s)}{ds} = \frac{bG(t)}{k} - \frac{1}{k}\frac{\mu'(s)}{\mu(s)} - \frac{aG(t)\mathrm{e}^{-saG(t)}}{1-\mathrm{e}^{-asG(t)}} = 0.$$

In particular, s_k satisfies the following equality, which is similar to (7.6.6):

$$s_k = \frac{1}{aG(t)} \log\left(1 + k\frac{aG(t)}{bG(t) - \mu'(s_k)/\mu(s_k)}\right).$$

When $\mu(s) = Ch(s)\mathrm{e}^{-\theta s}$,

$$\frac{\mu'(s)}{\mu(s)} = \frac{h'(s)\mathrm{e}^{-\theta s} - \theta h(s)\mathrm{e}^{-\theta s}}{h(s)\mathrm{e}^{-\theta s}} = -\theta\left(1 - \frac{h'(s)}{\theta h(s)}\right) \approx -\theta.$$

In particular, this implies

$$s_k = \frac{1}{aG(t)} \log\left(1 + k\frac{aG(t)}{bG(t)+\theta}\right)(1+o(1)).$$

Similarly to the element $(kH_k(t_k, s_k))_{2,2}$ in (7.6.5),

$$k\frac{d^2\psi_k(s_k)}{ds^2} = -\frac{\mu''(s_k)}{\mu(s_k)} + \left(\frac{\mu'(s_k)}{\mu(s_k)}\right)^2 - \frac{\mu'(s_k)}{\mu(s_k)}\frac{aG(t)}{1-\mathrm{e}^{-as_kG(t)}} + \frac{abG(t)^2}{1-\mathrm{e}^{-as_kG(t)}}.$$

For the general exponential class, the ratio

$$\frac{\mu''(s_k)}{\mu(s_k)} = \frac{h''(s)}{h(s)} - 2\theta + \theta^2.$$

As a consequence, $k\frac{d^2\psi_k(s)}{ds^2}$ converges to a positive constant, which means that s_k is an actual minimum. Then $J(k) = c_1\mathrm{e}^{-k\psi_k(s_k)}(1+o(1))$. Using this in (7.7.1) and ignoring the constants,

$$\begin{aligned}\mathbb{P}(M_t = k) &= \frac{\Gamma(k+b/a)}{\Gamma(b/a)\Gamma(k+1)}\mathrm{e}^{-s_kbG(t)}\mu(s_k)(1+o(1))\\ &= k^{-1}k^{-b/a}k^{b/a}h(s_k)k^{-\theta/aG(t)}(1+o(1))\\ &= h(s_k)k^{-(1+\theta/aG(t))}(1+o(1)),\end{aligned}$$

which is a power-law distribution with exponent $\tau(t) = 1+\theta/aG(t)$, and minor corrections given by $h(s_k)$. This holds for every $t \geq 0$. In particular, considering $G(\infty)$ instead of $G(t)$, with the same argument we can also prove that the distribution of the total number of children obeys a power-law tail with exponent $\tau(\infty) = 1+\theta/aG(\infty)$.

We now prove the result on the limiting distribution $(p_k^{(1)})_{k\in\mathbb{N}}$ of the CTBP, for which we apply directly Proposition 7.4.5, using the analysis on the Hessian matrix given in Section 7.6.2. First of all, from (7.6.6) it follows that

$$s_k = \frac{1}{aG(t_k)}\log\left(1+k\frac{aG(t_k)}{bG(t_k)+\theta}\right)(1+o(1)), \tag{7.7.4}$$

and by (7.6.7)

$$s_kg(t_k) \overset{k\to\infty}{\longrightarrow} \alpha\frac{G(\infty)}{\theta}. \tag{7.7.5}$$

For the Hessian matrix, using (7.7.4) and (7.7.4) in (7.6.5), for any integrable aging function g we have

$$(kH_k(t_k, s_k))_{2,2} = C_2 + o(1) > 0, \qquad \text{and} \qquad (kH_k(t_k, s_k))_{2,1} = o(1),$$

but $(kH_k(t_k, s_k))_{1,1}$ behaves according to $g'(t_k)/g(t_k)$. If this ratio is bounded, then $(kH_k(t_k, s_k))_{1,1} = C_1 + o(1) > 0$, while $(kH_k(t_k, s_k))_{1,1} \to \infty$ whenever $g'(t_k)/g(t_k)$ diverges. In both cases, (t_k, s_k) is a minimum. In particular, again ignoring the multiplicative constants and using (7.7.4) and (7.7.5) in the definition of $\Psi_k(t, s)$, the lim-

iting degree distribution of the CTBP is asymptotic to

$$k^{-(1+\theta/(aG(t_k)))}h(s_k)\mathrm{e}^{-\alpha^* t_k}\left(\tilde{C}-\alpha^*\frac{g'(t_k)}{g(t_k)}\right)^{-1/2}, \tag{7.7.6}$$

where the term $\left(\tilde{C}-\alpha^*\frac{g'(t_k)}{g(t_k)}\right)^{-1/2}$, which comes from the determinant of the Hessian matrix, behaves differently according to the aging function. With this, the proof of Proposition 7.4.6 is complete. □

7.7.2. Proof of Proposition 7.4.7

This proof is identical to the proof of Proposition 7.4.6, but this time we consider a sub-exponential distribution. First, we start looking at the distribution of the birth process at a fixed time $t \geq 0$. We define $\psi_k(s)$ and $J(k)$ as in (7.7.3) and (7.7.2). We use again (7.7.1), so

$$s_k = \frac{1}{aG(t)}\log\left(1+k\frac{aG(t)}{bG(t)-\mu'(s_k)/\mu(s_k)}\right).$$

In this case, we have

$$\frac{\mu'(s)}{\mu(s)} = -\theta(1+\varepsilon)s^{\varepsilon}. \tag{7.7.7}$$

Then s_k satisfies

$$s_k = \frac{1}{aG(t)}\log\left(1+k\frac{aG(t)}{bG(t)+\theta(1+\varepsilon)s_k^{\varepsilon}}\right).$$

By substitution, it is easy to check that s_k is approximately $c_1\log k - c_2\log\log k = \log k(1-\frac{\log\log k}{\log k})$, for some positive constants c_1 and c_2. This means that as first order approximation, s_k is still of logarithmic order. Then,

$$\frac{\mu''(s)}{\mu(s)} = \theta^2(1+\varepsilon)^2 s^{2\varepsilon} - \theta(1+\varepsilon)\varepsilon s^{\varepsilon-1}. \tag{7.7.8}$$

Using (7.7.7) and (7.7.8), we can write

$$\begin{aligned} k\frac{d^2\psi_k(s)}{ds^2} =& \theta(1+\varepsilon)\varepsilon s_k^{\varepsilon-1} + \theta(1+\varepsilon)s_k^{\varepsilon}\frac{aG(t)}{1-\mathrm{e}^{-as_kG(t)}} + \frac{abG(t)^2}{1-\mathrm{e}^{-as_kG(t)}} \\ =& \theta(1+\varepsilon)\varepsilon s_k^{\varepsilon-1} + \theta(1+\varepsilon)\frac{s_k^{\varepsilon}}{k}(bG(t)+\theta(1+\varepsilon)s_k^{\varepsilon}) \\ &+ \frac{bG(t)}{k}(bG(t)+\theta(1+\varepsilon)s_k^{\varepsilon}). \end{aligned}$$

The dominant term is $c_1 s_k^{\varepsilon}$, for some constant c_1. This means $k\frac{d^2\psi_k(s)}{ds^2}$ is of order $(\log k)^{\varepsilon}$. Now,

$$\begin{aligned} J(k) &= \left(k\frac{d^2\psi_k(s)}{ds^2}\right)^{-1/2} C\mathrm{e}^{-bG(t)s_k-\theta s_k^{1+\varepsilon}+k\log(1-\mathrm{e}^{-aG(t)s_k})}(1+o(1)) \\ &= (\log k)^{-\varepsilon/2}k^{-b/a}\mathrm{e}^{-\theta(\log k)^{1+\varepsilon}}(1+o(1)). \end{aligned}$$

As a consequence,

$$\mathbb{P}\left(M_t = k\right) = k^{-1}(\log k)^{-\varepsilon/2}\mathrm{e}^{-\theta(\log k)^{1+\varepsilon}}(1+o(1)),$$

which is not a power-law distribution. Again using similar arguments, we show that the limiting degree distribution of the CTBP does not show a power-law tail. In this case

$$s_k = \frac{1}{aG(t_k)}\log\left(1+k\frac{aG(t_k)}{bG(t_k)+\theta(1+\varepsilon)s_k^{\varepsilon}}\right),$$

and

$$s_k g(t_k) = \frac{\alpha G(t_k)}{\theta(1+\varepsilon)s_k^{\varepsilon}} = \frac{\alpha G(t_k)}{\log^{\varepsilon} k}(1+o(1)) \to 0.$$

The Hessian matrix elements are

$$\begin{aligned} (kH_k(t_k,s_k))_{1,1} &= \frac{a\alpha^2 G(t_k)}{s_k^{\varepsilon}} - a\alpha\frac{g'(t_k)}{g(t_k)} + o(1), \\ (kH_k(t_k,s_k))_{2,2} &= \theta(1+\varepsilon)\varepsilon s_k^{\varepsilon-1} + \theta s_k^{\varepsilon} aG(\infty) + abG(\infty)^2 + o(1), \\ (kH_k(t_k,s_k))_{1,2} &= o(1). \end{aligned}$$

This implies that

$$\det\left(kH_k(t_k,s_k)\right) = C_1 - s_k^{\varepsilon}\frac{g'(t_k)}{g(t_k)} + o(1) > 0.$$

As a consequence, (t_k, s_k) is an actual minimum. Then using the definition of $\Psi_k(t,s)$,

$$\begin{aligned} p_k^{(1)} &= \mathrm{e}^{-\alpha^* t_k}k^{-1+b/a}k^{-b/a}\mu(s_k) \\ &= \mathrm{e}^{-\alpha^* t_k}k^{-1}\left(C_1 - s_k^{\varepsilon}\frac{g'(t_k)}{g(t_k)}\right)\mathrm{e}^{-\frac{\theta}{(aG(\infty))^{1+\varepsilon}}(\log k)^{1+\varepsilon}}(1+o(1)). \end{aligned}$$

This completes the proof. □

8

Conclusions and open problems

In this thesis, we have discussed mathematical models for complex networks. We have focused our analysis on PAMs, since they are dynamic models that can represent the evolution *over time* of networks. With a simple mechanism, we are able to generate graphs with asymptotic properties that resembles most of the ones that we observe in real-world networks.

In this last chapter, we briefly summarize the results of this thesis, and we discuss some questions, open problems and possible future research topics related to our results.

8.1. Collapsed branching processes

Real-world networks often show power-law degree distributions. PAMs are known to be scale free whenever the PA function is affine, i.e., of the form $f(k) = k + \delta$, for some constant δ that allows to tune the power-law exponent τ. In the case of fixed $m \geq 1$, and $\delta > -m$, we have that $\tau = 3 + \delta/m$.

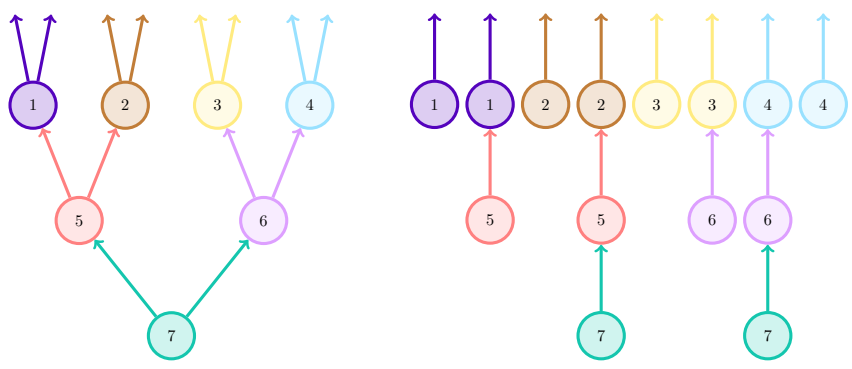
In Chapter 2 we have defined *collapsed branching processes* (CBPs), that are multigraph models in continuous-time. The main property of CBPs is that they extend the continuous-time embedding of PAMs to the case $m \geq 2$ (for $m = 1$, this was a known result on continuous-time branching processes (CTBPs) [9, 10, 22, 152]). In particular, we prove that we can describe the limiting degree distribution of a CBP using the properties of the birth process that defines the underlying CTBP. This general result is stated in Theorem 2.2.2.

We show that we can describe PA model (b') (see Section 4.3) as a CBP, that is the result is given in Theorem 2.2.4. We obtain an alternative expression for the degree distribution. In fact, (2.2.2) in Theorem 2.2.2, when applied to the CBP in Corollary 2.2.3, coincides with the known degree distribution of many PAMs as introduced in (1.3.1).

Theorem 2.2.2 is rather general, and it can be applied to other PA functions, and not only the affine case. In Corollary 2.2.5 we investigate the degree distribution of the *random recursive graph* (the extension of the random recursive tree to the case $m \geq 2$). A relevant extension of Theorem 2.2.2 is given in Corollary 7.1.4, where we investigate the degree distribution of a PAM where the PA function depends on the degree *and the age of a vertex*, partially extending to $m \geq 2$ the analysis we make in Chapter 7 on generalized PA functions.

We want to underline that the properties of a CBP we have described resemble the properties of the CTBP that defines the CBP itself (recall Theorem 2.1.11). More precisely, the proof of Theorem 2.2.2 is based on the more technical Theorem 2.3.2. The statement of Theorem 2.3.2 resembles Theorem 2.1.11. Notice in fact the similarities between, for example, (2.1.3) and (2.3.1) (the exponential growth of the size of a CTBP and a CBP respectively), or between (2.1.4) and (2.3.2) (a CTBP evaluated with a random characteristic and the number of vertices with a fixed degree in CBP).

Neighborhoods in CBP. We might ask what topological properties of a CBP we can describe using properties of the underlying CTBP. For example, we might want to investigate the local structure of neighborhoods as investigated in Chapter 4, where we discuss the local weak convergence (LWC) of PAMs. Since PAMs are locally treelike graphs, as shown in Chapter 4, we can ask the question whether this is true because



(a) CBP (b) CTBP

FIGURE 8.1: An example of minimally-2-connected neighborhood in the sense of Definition 3.4.2 in a CBP with $m = 2$, and a realization of the corresponding structure in CTBP that generates it. Notice that different realizations in the CTBP can generate the same graph in CBP. The difficulty here lies in the fact that, similarly to the analysis of the degree of a vertex in CBP, this construction is not a random characteristics in the sense of CTBP theory.

PAMs can be defined as CBPs.

For example, a minimally-k-connected vertex v as in Definition 3.4.2 in CBP is generated by *chains* of individuals in the corresponding CTBP. Proposition 3.4.6 proves that for a suitable sequence $k_t^* \to \infty$, the number of minimally-k_t^*-connected vertices diverges, thus implying that this also holds for any fixed $k \in \mathbb{N}$. In terms of the CTBP, it is necessary to look for structures similar to the one in Figure 8.1. It would be interesting to investigate the topological properties of the neighborhoods of vertices in CBP from the CTBP perspective, to see whether and, if so, how they depend on the corresponding CTBP.

CBPs and Pólya point tree. In the same line as the observations above, it would also be interesting to compare the local structure of CBPs with the Pólya point tree introduced in Section 4.3.1. In particular, it would be interesting to check if it is possible to prove the LWC of CBPs that embed PAMs to the Pólya point tree using CTBPs techniques. We already know that PAMs converge locally weakly in probability to the Pólya point tree, and so does the corresponding CBP. We might wonder if this convergence can be checked without using discrete-time PAMs, but directly in continuous-time.

The situation is different when we look at the *directed* LWC that we define in Chapter 6. In fact, in Theorem 6.2.4 we prove that CTBPs and PAMs converge in the directed LW sense. In Proposition 6.6.5 we prove that a CTBP converges *almost surely* in the *directed* local weak sense to the law of itself at the random time T_{α^*}, where α^* is the Malthusian parameter of the CTBP, and T_{α^*} is an exponentially distributed random variable with mean $1/\alpha^*$. For any $m \geq 1$, instead, Proposition 6.6.8 shows that a PAM converges *in probability* in the directed LW sense to the directed version of the Pólya point tree.

For $m = 1$ an embedded CTBP and a directed PAM have the same distribution, thus they must converge LW to the same limit (recall that the directed LWC, as well as the undirected one, is defined on a Polish space). For $m \geq 2$, though, an embedded CBP and the corresponding PAM also have the same distribution. The almost-sure convergence though is stronger than the convergence in probability, thus raising the question whether the almost-sure convergence is limited to the $m = 1$ case, or it can be extended to any $m \geq 2$.

Random out-degree. An interesting extension of the present work is the case of *random out-degree* graphs. Instead, our CBP have *fixed out-degree* $m \geq 2$. However, the collapsing procedure is well defined for any sequence of out-degrees $(m_n)_{n\in\mathbb{N}}$, both deterministic or random. Results are known for PAM with random out-degree (see [52]), suggesting that CBPs with random out-degree are the continuous-time versions of PAMs with random out-degrees.

More general PA functions and fitnesses. When collapsing, the degree $D_n^{(\text{in})}(t)$ of a vertex n in CBP is distributed as the sum of m independent birth process $(\xi_t)_{t\geq 0}$. When we consider an affine PA function of the type $f(k) = ak + b$, with $a \geq 0$ and

$b > 0$, the sum of the m weights corresponding to the m individuals becomes $a(D_1 + \ldots + D_m) + mb$, i.e., the collapsed individuals become *indistinguishable*. This is still true when we consider an affine PA function f and aging g, because of the linearity of f and the fact that the error given by the difference of birth times is negligible.

This is no longer true when the PA function is not affine and/or in the presence of multiplicative fitness. In multiplicative fitness models every individual x is assigned an independent realization η_x from a fitness distribution, and it produces children according to the sequence of PA weights $(\eta_x f(k))_{k\in\mathbb{N}}$ (see [22, 24, 35, 73]). In this case, individuals with different fitness values are not indistinguishable anymore. Assigning the same fitness value to m different individuals would define a process that is not a CTBP in the sense of Definition 2.1.4.

In the case of *additive fitness*, i.e., the weight of an individual x is $D_x^{(in)}(t) + \eta_x$, the problem has an immediate solution. In fact, for m different individuals $x_1, \ldots, x_m$, we have that the corresponding collapsed vertex v has weight

$$\sum_{j=1}^{m} D_{x_j}^{(in)} + \eta_{x_j} = D_v^{(in)}(t) + \hat{\eta}_v, \qquad \text{where } \hat{\eta}_v = \sum_{j=1}^{m} \eta_{x_j}.$$

In other words, the additive fitness of vertex v is the sum of the m fitnesses of the m individuals that generated it. Notice that in this case the weight of vertex v is still affine with respect to the degree $D_v^{(in)}(t)$.

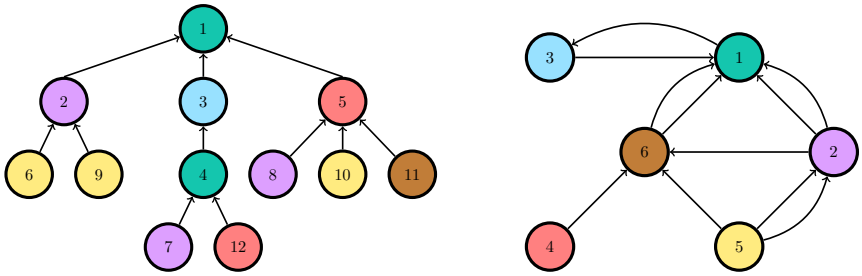
To overcome the problem in the multiplicative fitness case, a possible solution can be a modification of the collapsing procedure. In the case of discrete-valued fitnesses, we might collapse individuals according to their fitness values and not according to their birth order. This might be applied also to CTBPs with fitness and aging as introduced in Chapter 7, where, as discussed, problems are depending on the presence of fitness and not aging.

8.2. Ultra-small diameter

Real-world networks are typically small worlds, so distances between vertices in the networks are small compared to the size of the network itself. Mathematically, this has been formalized by saying that a graph of size n is a small world when its typical distances and diameter are of order $\log n$, and ultra-small world when they are of order $\log\log n$.

PAMs can be small or ultra-small worlds, according the value of the power-law exponent τ. We focus our analysis on the diameter of PAM model (a) of Section 4.3 when $m \geq 2$ and $\delta \in (-m, 0)$, thus obtaining $\tau \in (2, 3)$ (finite-mean and infinite-variance degree).

More precisely, in Chapter 3 we prove that the diameter of PAM model (a) and CM scales as $\log\log t$ (recall (1.4.1)), identifying the constant defined in (1.4.2). Remarkably, the proofs of the two convergence results, namely Theorem 3.1.3 for CM and Theorem 3.1.5 for PAM, consist of the same high-level steps, as shown in Section 3.2. This shows some universality property of our argument.



(a) CBP (b) CTBP

FIGURE 8.2: An example of randomized collapsing, according to fitness values. In this figure, colors represent the value of the fitness assigned to individuals. Individuals are divided in groups by fitness value, then within each group individuals are collapsed m by m according to birth order. For example, individuals 6,9 and 10 all have the same fitness value, so they are collapsed together (in this case, $m \geq 3$). If $m = 2$, in this case individuals 6 and 9 would be collapsed together, while 10 would be collapsed with the next individual with the same fitness value.

The constant in (1.4.2), that is model-dependent, consists of two terms: one coming from the exploration of the periphery of the graph, and one depending on the core of the graph.

The periphery constant is a function of the minimum degree of the graph. More precisely, we obtain this constant by performing an exploration process starting at vertices with low degree. We use the fact that this exploration process is a tree up to a $\log\log t$ distance (both for the upper and lower bound).

The core constant depends on the power-law exponent of the degree distribution. It is obtained by performing a lazy exploration of high-degree vertices, starting at the boundary of the exploration process just mentioned above. At every step we increase the degree of the vertex along a path, thus reaching the hubs of the graph. We show that connecting the two boundaries takes $\log\log t$ steps again.

Truncated degrees. In [92], van der Hofstad and Komjáthy investigate typical distances for configuration models and $\tau \in (2,3)$ in great generality, extending the results in [91] beyond the setting of i.i.d. degrees. Interestingly, they also investigate the effect of truncating the degrees at n^{β_n} for values of $\beta_n \to 0$. It would be of interesting to also extend our diameter results to this setting.

Fluctuations. An interesting open problem is the study of fluctuations of the diameters in CM_n and PA_t around the asymptotic mean, i.e., the study of the difference between the diameter of the graph and the asymptotic behavior (for these two mod-

els, the difference between the diameter and the right multiple of $\log\log n$).

In the literature results on fluctuations for the diameter of random graph models are rare. Bollobás in [28], and, later, Riordan and Wormald in [148] give precise estimates on the diameter of the Erdős-Renyi random graph.

In [90], the authors prove that in graphs with i.i.d. power-law degrees with $\tau \in (2,3)$, the difference Δ_n between the typical distance and the asymptotic behavior $2\log\log n/|\log(\tau-2)|$ does not converge in distribution, even though it is *tight* (i.e., for every $\epsilon > 0$ there is $M < \infty$ such that $\mathbb{P}(|\Delta_n| \leq M) > 1-\epsilon$ for all $n \in \mathbb{N}$). These results have been significantly improved in [92].

Conditional diameter. In the Pólya urn interpretation of PAMs presented in Chapter 4 we assign an intensity ψ_v to every vertex $v \in [t]$ related to the urn experiment. As already mentioned, we can see the random variables $(\psi_v)_{v\in[t]}$ as hidden weights associated to the vertices. This procedure somehow resembles the construction of rank-1 inhomogeneous random graphs (see Section 1.2).

Similarly to Chapter 5, we can see the graph as a stochastic process with two levels of randomness: one given by the weights, and one given by the assignment of edges. In Section 5.1 we have introduced the idea of conditional convergence of the number of subgraphs in PAMs, and we can do something similar for the diameter.

In other words, we can look at the law of PAM model (a) conditioning on the sequence $(\psi_v)_{v\in[t]}$. In this case the edges are inserted in the graph according to the value of (conditionally) independent Bernoulli random variables, similarly to the rank-1 inhomogeneous random graphs, or the Erdős-Rényi random graph.

The diameter of the inhomogeneous random graphs has been investigated. Bollobás, Janson and Riordan [30] prove that the diameter of a broad class of inhomogeneous random graphs is of logarithmic order. We might investigate the reasons why this is different and whether there exists any setting where the conditional diameter of PAMs and inhomogeneous random graphs can be similar.

8.3. Urns and local weak convergence

Real-world networks are highly clustered. In general, small subgraphs occur frequently. This is not true for PAMs, since they are locally treelike. This is formalized in terms of local weak convergence, and PAMs converge locally-weakly (LW) to an inhomogeneous random rooted tree called the Pólya point tree.

PAMs can be interpreted as Pólya urn experiments, where balls represent the degree of vertices in the graph. More precisely, the graph of size t is represented as a Pólya urn experiment with t urns, where the balls in urn $i \in [t]$ represent the degree of vertex i.

We extend the LWC result of Berger, Borgs, Chayes and Saberi [21] to many different versions of PAMs (listed in Section 4.3). We give a sketch of the proof structure in Section 4.7, identifying the necessary conditions that a PAM has to satisfy to converge LW to the Pólya point tree. We list these conditions in Proposition 4.9.1.

In particular, to adapt the argument in [21] to PAMs in Section 4.3, we generalize

the construction of Pólya urns (so the Pólya point graph in Definition 4.4.2) to unit graphs, defined in Definition 4.6.1.

We show that, for a version of PAMs defined through collapsing, if for $m = 1$ it converges LW to the Pólya point tree with parameters 1 and δ/m, then the same PAM for $m \geq 2$ converges to the Pólya point tree with parameters m and δ. To resembles the results of Chapter 2 about CBPs, since we are able to express properties of the case $m \geq 1$ in terms of the case $m = 1$.

Random initial degrees. Models (a)-(g) listed in Section 4.3 differ for minor modifications, such as the presence of self-loops and the initial graph. Other modifications are possible. For example, we can consider PAMs where the number of initial connections can follow a prescribed distribution M [52]. This would lead to a change in the urn definitions, but we believe that the exchangeability property still holds in this case, if we condition on the sequence of initial degrees $(m_t)_{t\in\mathbb{N}}$.

If the distribution M of the initial degrees is regular enough, then we can use the law of large numbers and/or central limit theorem to have regularity of the parameters of the distribution of the Beta random variables as in Definition 4.4.2. In fact, the urn version of PAMs is defined in terms of Beta random variables $(\psi_k)_{k\in\mathbb{N}}$, where ψ_k is a Beta random variable with parameters $a_k = m + \delta$ and $b_k \approx (2m + \delta)k$. In case of random initial degrees $(m_t)_{t\in\mathbb{N}}$, we would have that $a_k = m_k + \delta$ and $b_k \approx (2\sum_{j=1}^{k-1} m_j) + delta(k-1)$. If M is regular enough, then b_k would approximately be $(2\mathbb{E}[M] + \delta)k$, as in the Pólya urn graph in Definition 4.4.2.

The corresponding LW limit in this case would not be precisely the Pólya point tree defined in Section 4.3.1. In fact, every vertex $n \in \mathcal{U}$ has m or $m-1$ children of type O according to its own type. In the case of random initial degree, every vertex $n \in \mathcal{U}$ would have a random number of children of type O, distributed as M or $M-1$ according to the type of n. It is still not clear to us what would be the precise definition of the LW limit, but we believe that the definition would be similar to the one given in Section 4.3.1.

Additive fitness. We can consider a PA function where the parameter δ is a random variable. More precisely, assume that we fix a distribution δ, and then instead of considering the same value of δ for every vertex, we consider an i.i.d. sequence $(\delta_t)_{t\in\mathbb{N}}$ sampled from the distribution δ.

Similarly to the random initial degree setting just discussed, we believe that the PAM still satisfies the exchangeability property required to apply De Finetti's Theorem. In this case, for $k \in \mathbb{N}$, the random variable ψ_k would have a Beta distribution with parameters $a_k = m + \delta_k$ and $b_k \approx m(2k-3) + (k-1)\mathbb{E}[\delta]$. Also in this case, we believe that the LW limit would be a modification of the Pólya point tree, but we do not precisely know its definition.

The conditionally independent model. Model (h) in Section 4.3 is called PAM with conditional independent edges. Here, conditional independence has a different mean-

ing than, for instance, the conditional convergence of subgraph occurrences of Chapter 5.

Model (h), investigated by Dereich and Mörters in a series of papers [54, 55, 56], is called the conditionally independent model because every vertex $t \in \mathbb{N}$, conditionally on $\mathrm{PA}_{t-1}^{(h)}$, is connected to possibly all vertices $[t-1]$ independently of each other. In this case, the initial degree of vertex t is a sum of (conditionally) independent Bernoulli random variables with different probabilities of success.

Results in [54, 55, 56] establish that, for any PA function f that satisfies some regularity conditions, $\mathrm{PA}_t^{(h)}(f)$ converges LW to an inhomogeneous random tree, that is not the Pólya point tree. Remarkably, in contrast with the results in Chapter 4, based on affine urn schemes as in Section 4.4, these results hold for any (admissible) function f. We might wonder whether we can establish a relation between the LW limit of model (h) and the Pólya point tree.

8.4. Subgraph occurrences

Even though PAMs are locally treelike, in contrast of real-world networks, we might still want to compute the number of occurrences of small subgraphs in PAMs. Since PAMs are constructed recursively by adding vertices one at a time, we have to considered ordered subgraphs, where we specify the order according to which the vertices appear in the graph.

In Chapter 5 we investigate the expected number of times a graph H appears as a subgraph of a PAM for any degree exponent τ. We find the scaling of the expected number of such subgraphs in terms of the graph size t and the degree exponent τ by defining an optimization problem that finds the optimal structure of the subgraph in terms of the ages of the vertices that form subgraph H and by using the interpretation of the PAM as a Pólya urn graph given in Chapter 4.

We derive the asymptotic scaling of the number of subgraphs in Theorem 5.1.2. This result, as (5.1.2) suggests, is rather weak, in the sense that in principle (5.1.2) does not imply the existence of such a limit. This is a consequence of the fact that the optimization problem in 5.1.1 is defined to identify the order of magnitude (as function of t) of the indices of vertices.

We do not have precise constants, but we are able to make a step forward, defining the conditional convergence for the number of occurrences of subgraphs, seeing the intensities of the Pólya urn graph as hidden weights associated to the vertices of the graph. Proposition 5.1.4 gives a criterion to establish whether a subgraph H is conditionally concentrated, based on the number of occurrences of the overlapping but distinct copies of H.

Since the triangle is the most studied subgraph, we use it as an example of the technical difficulties that the identification of the precise scaling constant in (5.1.2) presents. We obtain more precise asymptotics in Theorem 5.1.5, that extend the known results for $\tau \geq 3$ to $\tau > 2$.

Precise asymptotics. As we did for triangles, it would be interesting to obtain precise asymptotics of the expected number of other types of subgraphs as well. In particular, this is necessary to compute the variance of the number of subgraphs, which may allow us to derive laws of large numbers for the number of subgraphs. We show that different subgraphs may have significantly different concentration properties. Therefore, identifying the distribution of the number of rescaled subgraphs for any type of subgraph remains a challenging open problem.

Extension to other PAMs. Another interesting extension would be to investigate whether our result still holds for other types of PAMs, for example models that allow for self-loops, or models that include extra triangles. We believe that the results of Chapter 5 should hold also for the PAMs (a)-(g) listed in Section 4.3. More precisely, a particular version of PAM has to satisfy a bound of the type (5.2.2) (or equivalently (5.2.4)) given by Lemma 5.2.1, that links a subgraph H to the optimization problem (5.1.1).

Different class of subgraphs. We prove results for the number of subgraphs of fixed size k, while the graph size tends to infinity. It would also be interesting to let the subgraph size grow with the graph size, for example by counting the number of cycles of a certain length that grows in the graph size. Another example of such subgraphs would be the minimally-k_t^*-connected neighborhoods defined in Chapter 3. Proposition 3.4.6 shows that the number of such subgraphs diverges with t, but we do not know the scaling constant.

Also, we investigate the number of times that H appears as a subgraph of a PAM. It is also possible to count the number of times H appears as an *induced* subgraph instead, forbidding edges that are not present in H to be present in the larger graph. It would be interesting to see whether the optimal subgraph structure is different from the optimal induced subgraph structure.

Hierarchical PAMs. As shown, PAMs are locally treelike and subgraphs such as triangles or small cliques do not occur frequently. As mentioned in Section 1.5 , Prokhorenkova et al. [139, 140, 141] obtained a linear number of triangles in PAMs by modifying the attachment probabilities and the number of edges in the graph.

A different way of increasing the number of triangles and cliques in PAMs would be to introduce these small subgraphs artificially. In a series of papers [93, 94, 156, 157] van der Hofstad, van Leeuwaarden and Stegehuis work on a modification of the configuration model, where they introduced *communities*. More precisely, they fix a shape for a community, like a small complete graph, and they fix the number of edges that go outside from each community. Then, edges across communities are generated by pairing half-edges incident to communities, as in the standard configuration model. This creates a graph with scale-free degree distribution but with a higher number of small subgraphs.

We can think of a similar mechanism for PAMs. More precisely, we can fix the community shape, for instance a complete graph on k vertices. Then we can define a

sequence of graphs $(\mathrm{HPA}_t(m,k,\delta))_{t\in\mathbb{N}}$ called *hierarchical PA model* as follows: HPA_1 consists of a single community (so HPA_1 has size k). Recursively, for $t \geq 2$, a new community C_t appears with m outgoing edges. Then, the jth edge of the C_t is attached to an existing community C_i that is chosen according to the following probability:

$$\mathbb{P}(C_t \xrightarrow{j} C_i \mid \mathrm{HPA}_{t-1}) = \frac{D_{C_i}(t-1,j-1)+k\delta}{Z(t-1,j-1)}, \tag{8.4.1}$$

where $D_{C_i}(t-1,j-1)$ is the total degree of C_i after the $(j-1)$th edge of C_t has been attached, and $Z(t-1,j-1)$ is a normalization constant. Once an edge is attached to a community, then it is assigned to a uniformly chosen vertex in that community, so that we obtain a graph. Notice that the chosen vertex inside the community does not modify the probabilities in (8.4.1).

The hierarchical PAM has some flexibility in terms of the type of communities, and the parameters m and δ, and by definition, it shows more subgraphs than a standard PAM. Apart from this, we do not know any other property of hierarchical PAM. This would be an interesting object to investigate.

8.5. Limiting PageRank distribution

The PageRank hypothesis claims that if a complex network shows a power-law (in-)degree distribution, then the graph-normalized PageRank distribution obeys a power-law distribution with the same exponent as the (in-)degree. In order to avoid finite-dimension problems, in Chapter 6 we extend the notion of LWC to directed graphs, in order to investigate the limiting PageRank distribution of PAMs.

Our argument, stated in Theorem 6.2.1, is rather general, and shows that for any sequence of directed random graphs that converge in the directed LW sense, the limiting PageRank distribution can be retrieved by looking at the directed LW limit. This helps because it allows us to avoid the difficulties arising by the finite size of a graph. To show that our construction is applicable to many models, we prove that the directed CM (and consequently, directed inhomogeneous random graphs) and directed CTBPs converge LW also in the directed case. This convergence results are stated in Theorem 6.2.4.

In particular, the limiting PageRank distribution $R_\varnothing$ of PAMs is given by a function of the neighborhood of the root $\varnothing$ in the directed LW limit. The random variable $R_\varnothing$ is defined in Proposition 6.4.3. A direct consequence of Theorem 6.2.1 and the definition $R_\varnothing$ is that, as stated in Remark 6.2.2, for PAMs the tail of the distribution $R_\varnothing$ is bounded from below by a multiple of the tail of the in-degree, which indeed obeys a power-law. This gives a lower bound that partially proves the power-law PageRank hypothesis for PAMs.

Extension to exploration of outgoing edges. Since we are motivated by the interest in PageRank algorithms on random graphs, we build our definition on the *exploration of incoming edges* in their opposite direction, i.e., an edge (i,j) is explored from i to j. The outgoing edges are considered as marks and we do not explore them. In the

same way, it is possible to define the exploration process according to the natural direction of the edges. In this case, we consider *outgoing neighborhoods* instead. The definition of LW convergence would just be a consequence of symmetry. This second interpretation might be useful, for instance, in the study of diffusion processes on graphs, such as epidemic spreads. An interesting and more complex extension would be to explore the incoming and outgoing neighborhoods *at the same time*.

PageRank on limiting graphs. We are able to prove that, under relatively general assumptions, a sequence of random directed graphs admits a limiting distribution for the PageRank of a uniformly chosen vertex. In this way, we have moved the analysis of a graph's PageRank distribution from a whole sequence of graphs to a single (possibly infinite) rooted directed marked graph. Note that we prove the existence of this distribution, but we do not always have a convenient description of it. It will be interesting to investigate the behavior of this limiting distribution. In particular, it is interesting to investigate the conditions under which the rank of the root in the limiting graph shows a *power-law tail*, and thus confirm the power-law hypothesis.

Stochastic fixed point equations. In the case of the directed CM, much more detailed asymptotics are known on the tail of $R_\varnothing$. In our setting, the directed CM converges to a marked Galton-Watson tree (see Proposition 6.6.2). On the limiting tree, $R_\varnothing$ satisfies a recursive distributional equation, also called a stochastic fixed point

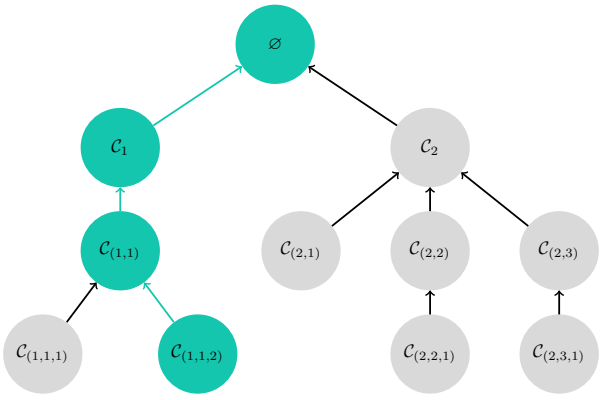


FIGURE 8.3: Example of solution to the SFPE constructed on an infinite branching tree. To every vertex $n \in \mathcal{U}$ (the Ulam-Harris set) of the tree we assign a weight $\mathcal{C}_n$, and the solution $\mathcal{R}$ is the sum over all the self-avoiding paths ending at the root $\varnothing$ of the product of the weights along the path. Notice that this construction is very similar to the formulation of PageRank in terms of weighted sum of paths. For example, the weight corresponding to the green path is $\mathcal{C}_{(1,1,2)}\mathcal{C}_{(1,1)}\mathcal{C}_1$. In the case of the Pólya point tree, the construction is the same but the tree is *finite*.

equation (SFPE).

An equation of the type (1.7.2) has a solution $\mathcal{R}$ that can be constructed using a Galton-Watson tree with weights (we refer to [162] for the precise construction). The idea is briefly explained in Figure 8.3. In this setting, under suitable hypothesis, if $\mathcal{D}^{(\text{in})}$ obeys a power law with exponent τ, so does $\mathcal{R}$.

In Remark 6.6.9 we explain that this no longer applies to the directed Pólya point tree. This is due to the fact that the corresponding equation, in this case, we should consider i.i.d. copies of the time-dependent process $\mathcal{R}_\cdot$, but they are evaluated at different times. In simpler words, any subtree in the directed Pólya point tree has a distribution that depends on the birth time of its sub-root, thus they are not identically distributed.

It would certainly be interesting to study whether the construction of the solution of the SFPE in (1.7.2) using tree can be extended to the case of the Pólya point tree, i.e., to the case of finite branching trees.

PageRank in undirected graphs. As mentioned in Section 6.4.5, our convergence result based on the directed LWC can be immediately extended to PageRank defined on undirected graphs. In this case, due to the fact that we now look at a weighted sum of undirected paths, the analysis of PageRank is even harder than the directed setting. Our methodology might simplify the study of PageRank in undirected graphs if, for example, the corresponding LW limit is simple (for example, a Galton-Watson tree).

8.6. Modeling citation networks

Different complex networks can be represented by different models. This might imply that a PA mechanism depending only on the degree of vertices might not be feasible to describe different networks.

Because of this, we consider citation networks in more detail, in order to investigate what *factors* should have influence on the PA function in order to obtain a graph model with similar characteristics as the ones that we observe in the data. We refer to Section 1.8 for a more detailed qualitative analysis of citation networks data.

We generalize the construction of PA tree to general PA functions. In particular, we define PA functions depending on three main factors: the degree (number of citations), the age of a vertex (how much time has passed since a paper was published), and a fitness (representing intrinsic potential of a paper).

We start by investigating PA trees with aging but without fitness. We prove in Proposition 7.4.2 (particular case of Theorem 7.1.2) that any integrable aging function g destroys the power-law degree distribution of the PA tree, independently of the particular shape of g. The assumption about the integrability of g is crucial if we want to represent the fact that we expect papers to receive a finite number of citations.

In order to restore the power law, we have to assign fitnesses to the vertices in the tree. We make a detailed analysis on the necessary assumptions on the fitness distribution that are necessary to obtain a power-law in-degree distribution. Of the three general classes of distributions we define in Section 7.1.4, we show that the right

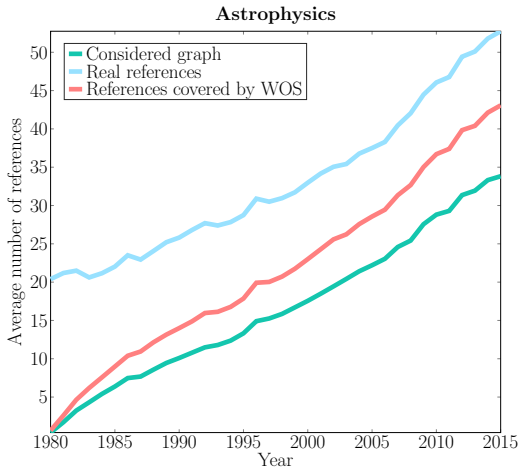


FIGURE 8.4: Plot of the average out-degree in the Astrophysics dataset. Notice that *Web of Science* does not cover all the references contained in the papers that we have selected to be part of the dataset. In addition, the way that we have constructed the dataset considering only references between Astrophysics papers ignores some of the references, even though they are covered by WoS.

one is the class of distributions with exponential tail. This class also has the additional property of the dynamical power law, as we observe in the data.

Beyond the tree setting. Chapter 7 contains results only on the *tree setting*, which is clearly unrealistic for citation networks. However, the analysis of PAMs made, for example, in Chapters 2 and 4 has shown that the qualitative features of the degree distribution for PAMs are identical to those in the tree setting.

We already have discussed the randomized collapsing procedure, i.e., the collapsing of vertices with same fitness values. We believe this can be a way to overcome the fact that the simple collapsing does not work in this setting, due to the presence of fitness. Of course, in order to do this, we should assign discrete-valued fitnesses with exponential tail, for example sampling from a geometric distribution.

Fitting the data. We started our analysis from citation network data. It would be natural to check how accurate our model is in describing the data. In other words, we might want to produce simulations of our model to match the data not only qualitatively, but also quantitatively.

To do this, we should find a way to estimate the right parameters from the data. This is actually a hard problem, due to the complicated nature of our model, the fact

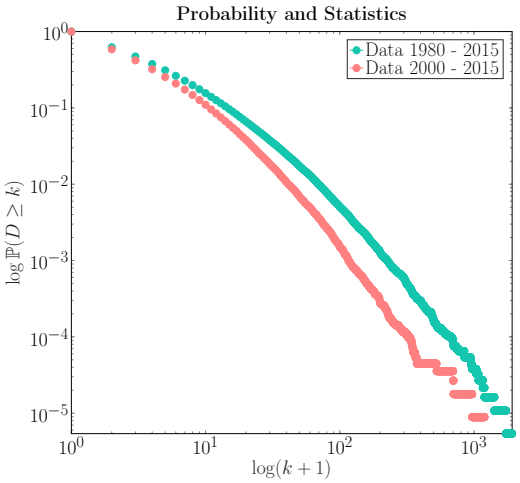


FIGURE 8.5: Example of incompleteness of data. In this figure, we have the loglog plot of the in-degree distribution tail of the number of citations in PS dataset. The light green line is the distribution obtained using the complete PS dataset we have. The dark green line is the distribution obtained by using the data corresponding to papers published after 2000. We observe that the two distributions clearly have a different power-law exponent. Data from *Web of Science*.

that we do not have a theoretical description of the case where $m \geq 1$, and the fact that the data is not complete.

Unfortunately, there exists no database that covers all the publications ever made in all scientific fields. When we start looking at the data, we have to pick a database and restrict ourselves to a part of it. This creates different problems. For instance, what is a good definition of a scientific field? Our selection of fields within Web of Science made us ignore part of the data. An example of this is given in Figure 8.5.

The fact that data is incomplete also has another consequence. Assume that we have a way to estimate the power-law exponent from the data (this is already highly non-trivial). We would like to produce a graph with that same exponent. Figure 8.5 shows that the power-law exponent changes according to the starting point in time that we consider for our data. Since our data starts in 1980 anyway (thus we ignore everything before that), we might be trying to produce a citation distribution with the wrong power-law exponent. How to approach this is still an open and difficult problem.

Bibliography

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling or, power-law degree distributions in regular graphs. *STOC'05: Proceedings of the 37th annual ACM symposium on theory of Computing*, 56(4):694–703, 2002.

[2] R. Albert, and A .L. Barabási. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Reviews of modern Physics*, 74(1):47–97, 2002.

[4] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electronic Journal of Probability*, 12:1454–1508, 2007.

[5] D. Aldous and J. M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. *Probability on Discrete Structures*, Springer Berlin Heidelberg, 1–72, 2004.

[6] N. Alon, and J. Spencer . The probabilistic method, Second Edition. *Wiley-Interscience series in discrete Mathematics and Optimization*, John Wiley & Sons, 2000.

[7] A. D. Anastasiadis, M. P. de Albuquerque, M. P. de Albuquerque, D. B. Mussi. Tsallis q-exponential describes the distribution of scientific citations - a new characterization of the impact. *Scientometrics*, 83:205–218, 2010.

[8] R. Andersen, F. Chung, and K. Lang Local graph partitioning using PageRank vectors. *Proceedings of the 47th annual IEEE symposium on foundations of Computer Science*, pages 475–486, 2006.

[9] K. B. Athreya. Preferential attachment random graphs with general weight function. *Internet Mathemathics*, 4(4):401–418, 2007.

[10] K. B. Athreya, A. P. Ghosh, and S. Sethuraman. Growth of preferential attachment random graphs via continuous-time branching processes. *Proceedings - Mathematical sciences*, 118(3):473–494, 2008.

[11] K. B. Athreya and P. E. Ney. Branching processes. *Dover Publications, NY*, 2004. Reprint of the 1972 original [Springer, New York].

[12] K. Avrachenkov, A. Kadavankandy, L. O. Prokhorenkova, and A. Raigorodskii. Pagerank in undirected random graphs. *International workshop on algorithms and models for the Web-graph (WAW)*, 151–163, 2015.

[13] K. Avrachenkov and D. Lebedev. Pagerank of scale-free growing networks. *Internet Mathematics*, 3(2):207–231, 2006.

[14] K. Avrachenkov and N. Litvak. The effect of new links on google pagerank. *Stochastic Models*, 22(2):319–331, 2006.

[15] B. Bahmani, A. Chowdhury, and A. Goel. Fast incremental and personalized pagerank. *Proceedings of the VLDB endowment*, 4(3):173–184, 2010.

[16] A. L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272:173–187, 1999.

[17] A. L. Barabási, and M. Pósfai. Network science. *Cambridge University Press*, 2016.

[18] A. Barrat, and M. Weigt. On the properties of small-world networks. *European Physical Journal B* 13:547–560, (2000).

[19] I. Benjamini, R. Lyons, and O. Schramm. Unimodular random trees. *Ergodic Theory and Dynamical Systems* , 35(2):359–373, 2015.

[20] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electronic Journal of Probability*, 6:1–13, 2001.

[21] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. Asymptotic behavior and distributional limits of preferential attachment graphs. *Annals of Probability*, 42(1):1–40, 2014.

[22] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and Local analysis. *Preprint*, `http://www.unc.edu/~bhamidi/preferent.pdf`, 2007.

[23] M. Bianchini, M. Gori, and F. Scarselli. Inside pagerank. *ACM Transactions on Internet Technology*, 5(1):92–128, 2005.

[24] G. Bianconi, and A. L. Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(4):436, 2001.

[25] M. Bloznelis. A note on loglog distances in a power-law random intersection graph. *Preprint*, arXiv.org:0911.5127 [math.PR], 2009.

[26] M. Boguñá, and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):036112, 2003.

[27] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.

[28] B. Bollobás. The diameter of random graphs. *Transactions of the American Mathematical Society*, 267(1):41–52, 1981.

[29] B. Bollobás. Random graphs (second edition). *Volume 73 of Cambridge Studies in Advanced Mathematics*, Cambridge University Press, 2001.

[30] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[31] B. Bollobás and O. Riordan. Mathematical results on scale-free random graphs. *Handbook of Graphs and Networks*, Wiley–WCH, 1–37, 2002.

[32] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[33] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[34] M. J. Bommarito, and D. M. Katz. A mathematical approach to the study of the united states code. *Physica A: Statistical Mechanics and its Applications*, 389:4195–4200, 2010.

[35] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch. First to market is not everything: an analysis of preferential attachment with fitness. *STOC'07: Proceedings of the 39th annual ACM symposium on theory of Computing*, 135–144, 2007.

[36] C. Borgs, J. T. Chayes, H. Kesten, and J. Spencer. The birth of the infinite cluster: finite-size scaling in percolation. *Communications in Mathematical Physics*, 224(1):153–204, 2001.

[37] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics*, 124(6):1377–1397, 2006.

[38] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. *Computer Networks*, 33(1-6):309–320, 2000.

[39] J. Cao and M. Olvera-Cravioto. On a general class of inhomogeneous random digraphs. *Preprint*, arXiv:1712.03319 [math.PR], 2017.

[40] F. Caravenna, A. Garavaglia, and R. van der Hofstad. Diameter in ultra-small scale-free random graphs. *Random Structures & Algorithms*, 1–55, 2018.

[41] A. Clauset, and C. Moore. Traceroute sampling makes random graphs appear to have power law degree distributions. *Preprint*, arXiv:cond-mat/0312674 [cond-mat.dis-nn], 2003.

[42] A. Clauset, C. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[43] G. Csárdi. Dynamics of citation networks. *International conference on artificial neural networks*, 4131:698–709, 2006.

[44] D. Donato, L. Laura, S. Leonardi, and S. Millozi. Large scale properties of the Webgraph. European Physical Journal B, 38(2):239–243, 2004.

[45] N. Chen, N. Litvak, and M. Olvera-Cravioto. Pagerank in scale-free random graphs. *International workshop on algorithms and models for the Web-graph (WAW)*, 120–131, 2014.

[46] N. Chen, N. Litvak, and M. Olvera-Cravioto. Generalized pagerank on directed configuration networks. *Random Structures & Algorithms*, 51(2):237–274, 2017.

[47] N. Chen and M. Olvera-Cravioto. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013.

[48] P. Chen, H. Xie, S. Maslov, and S. Redner. Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, 1(1):8–15, 2007.

[49] F. Chung and L. Lu. The diameter of sparse random graphs. *Advances in Applied Mathematics*, 26(4):257–279, 2001.

[50] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Internet Mathematics*, 1:15879–15882, 2002.

[51] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145, 2002.

[52] M. Deijfen, H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. A preferential attachment model with random initial degrees. *Arkiv för Matematik*, 47(1):41–72, 2008.

[53] S. Dereich. Preferential attachment with fitness: unfolding the condensate. *Electronic Journal of Probability*, 21(3),1–38, 2016.

[54] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electronic Journal of Probability*, 14:1222–1267, 2009.

[55] S. Dereich and P. Mörters. Random networks with concave preferential attachment rule. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 113(1):21–40, 2011.

[56] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: the giant component. *Annals of Probabability*, 41(1):329–384, 01 2013.

[57] S. Dereich, C. Mönch, and P. Mörters. Typical distances in ultrasmall random networks. *Advances in Applied Probability*, 44(2):583–601, 2012.

[58] S. Dereich, and M. Ortgiese. Robust analysis of preferential attachment models with fitness. *Combinatorics, Probability and Computing*, 23(3):386–411, 2014.

[59] S. Dommers, R. van der Hofstad, and G. Hooghiemstra. Diameters in preferential attachment models. *Journal of Statistical Physics*, 139(1):72–107, 2010.

[60] R. Durrett Random graph dynamics. Cambridge series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2007.

[61] N. Eggemann, and S. D. Noble. The clustering coefficient of a scale-free random graph. *Discrete Applied Mathematics*, 159(10):953–965, 2011.

[62] L. Egghe. Power laws in the information production process: Lotkaian informetrics. *Elsevier*, Oxford, 2005.

[63] P. Erdős, and A. Rényi. On Random Graphs I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[64] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs. *Journal of Statistical Physics*, 133(1):169–202, 2008.

[65] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review*, 29(4):251–262, 1999.

[66] D. Fernholz and V. Ramachandran. The diameter of sparse random graphs. *Random Structures & Algorithms*, 31(4):482–516, 2007.

[67] S. Fortunato, M Boguñá, and A. F. F. Menczer. How to make the top ten: Approximating PageRank from in-degree. *International workshop on algorithms and models for the Web-graph (WAW)*, 59–71, 2006.

[68] K. Gabrysch. Convergence of directed random graphs to the Poisson-weighted infinite tree. *Journal of Applied Probability*, 53:463–474, 2016.

[69] C. Gao, and J. Lafferty. Testing network structure using relations between small subgraph probabilities. *Preprint*, arXiv:1704.06742 [stat.ME], 2017.

[70] A. Garavaglia and R. van der Hofstad. From trees to graphs: collapsing continuous-time branching processes. *Journal of Applied Probability*, 55(3):900–919, 2018.

[71] A. Garavaglia, and R. van der Hofstad. Universality of the local weak limit of preferential attachment models. *In preparation*, 2019+.

[72] A. Garavaglia, R. van der Hofstad, and N. Litvak. Local weak convergence for PageRank. *Preprint*, arXiv:1803.06146 [math.PR], 2018.

[73] A. Garavaglia, R. van der Hofstad, and G. Woeginger. The dynamics of power laws: fitness and aging in preferential attachment trees. *Journal of Statistical Physics*, 168(6):1137–1179, 2017.

[74] A. Garavaglia, and C. Stegehuis. Subgraphs in preferential attachment models. *Preprint*, arXiv:1806.10406 [math.PR], 2018.

[75] X. Geng, and Y. Wang. Degree correlations in citation networks model with aging. *Europhysics Letters*, 88(3):38002, 2009.

[76] E. N. Gilbert. Random graphs *Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[77] G. R. Grimmett. Random labelled trees and their branching networks. *Journal of the Australian Mathematical Society*, 30(2):229–237, 1980.

[78] B. M. Gupta. Analysis of distribution of the age of citations in theoretical population genetics. *Scientometrics*, 40(1):139–162, 1997.

[79] Z. Gyongyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with trustrank *30th international conference on very large data bases*, 576–587, 2004.

[80] O. Hagberg, and C. Wiuf. Convergence properties of the degree distribution of some growing network models. *Bulletin of Mathematical Biology*, 68(6):1275, 2006.

[81] K. B. Hajra, and P. Sen. Aging in citation networks. *Physica A: Statistical Mechanics and its Applications*, 346(1-2):44–48, 2005.

[82] K. B. Hajra, and P. Sen. Modelling aging characteristics in citation networks. *Physica A: Statistical Mechanics and its Applications*, 368(2):575–582, 2006.

[83] T. H. Haveliwala. Topic-sensitive pagerank. *Proceedings of the 11th international conference on World-Wide Web*, 517–526, 2002.

[84] M. J. Hazoglou, V. Kulkarni, S. S. Skiena, and K. K. Dill. Citation histories of papers: sometimes the rich get richer, sometimes they don't. *Preprint*, arXiv:1703.04746 [cs.DL], 2017.

[85] R. van der Hofstad. Random graphs and complex networks, Volume 1. *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, 2017.

[86] R. van der Hofstad. Random graphs and complex networks, Volume 2. *Book in preparation*, `http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf`, 2018+.

[87] R. van der Hofstad. Stochastic processes on random graphs. *Lecture notes for the Saint Flour Summer School in Probability 2017*, 2017+.

[88] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. On the covariance of the level sizes in random recursive trees. *Random Structures & Algorithms*, 20(4):519–539, 2002.

[89] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures & Algorithms*, 27(1):76–123, 2005.

[90] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electronic Journal of Probability*, 12(25):703–766, 2007.

[91] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. A phase transition for the diameter of the configuration model. *Internet Mathematics*, 4(1):113–128, 2007.

[92] R. van der Hofstad and J. Komjáthy. When is a scale-free graph ultra-small? *Journal of Statistical Physics*, 169(2):223–264, 2017.

[93] R. van der Hofstad , J. S. H. van Leeuwaarden, and C. Stegehuis. Optimal subgraph structures in scale-free networks. *Preprint*, arXiv:1709.03466 [math.PR], 2017.

[94] R. van der Hofstad , J. S. H. van Leeuwaarden, and C. Stegehuis. Mesoscopic scales in hierarchical configuration models. *Stochastic processes and their applications*, 2018.

[95] P. Jagers. Branching processes with biological application. Wiley & Sons, 1975.

[96] P. Jagers and O. Nerman. The growth and composition of branching populations. *Advances in Applied Probability*, 16(2):221–259, 1984.

[97] S. Janson. Asymptotic degree distribution in random recursive trees. *Random Structures & Algorithms*, 26(1-2):69–83, 2005.

[98] S. Janson. The probability that a random multigraph is simple. *Combinatorics, Probability and Computing*, 18(1-2):205–225, 2009.

[99] S. Janson. The probability that a random multigraph is simple II. *Journal of Applied Probability*, 51A:123–137 , 2014.

[100] S. Janson, T. Luczak, and A. Rucinski. Random graphs. *Wiley-Interscience series in discrete Mathematics and Optimization*, John Wiley & Sons, 2000.

[101] P. R. Jelenković and M. Olvera-Cravioto. Information ranking and power laws on trees. *Advances in Applied Probability*, 42(4):1057–1093, 2010.

[102] H. Jeong, Z. Néda, and A. L. Barabási. Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61(4): 567–572, 2003.

[103] J. Jordan. The degree sequences and spectra of scale-free random graphs. *Random Structures & Algorithms*, 29(2):226–242, 2006.

[104] H. Kesten, and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *Journal of Mathematical Analysis and Applications*, 17(2):309–338, 1967.

[105] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.

[106] V. Lariviére, É. Archambault, and Y. Gingras. Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology*, 59(2):288–296, 2008.

[107] V. Lariviére, Y. Gingras, and É. Archambault. The decline in the concentration of citations, 1900–2007. *Journal of the American Society for Information Science and Technology*, 60(4):858–862, 2009.

[108] C. P. C. Lee, G. H. Golub, and S. A. Zenios. A fast two-stage algorithm for computing pagerank and its extensions. *Scientific Computation and Computational Mathematics*, 1(1):1–9, 2003.

[109] R. Oliveira, and J. Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Mathematics*, 2(2):121–163, 2005.

[110] J. Lee and M. Olvera-Cravioto. Pagerank on inhomogeneous random digraphs. *Preprint*, arXiv:1707.02492 [math.PR], 2017.

[111] J. Leskovec, and A. Krevl. *SNAP Datasets*: Stanford Large Network Dataset Collection. `http://snap.stanford.edu/data`, June 2014.

[112] N. Litvak, W. R. W. Scheinhardt, and Y. Volkovich. In-degree and pagerank: why do they follow similar power laws? *Internet Mathematics*, 4(2-3):175–198, 2007.

[113] Q. Liu. Fixed Points of a Generalized Smoothing Transformation and Applications to the Branching Random Walk. *Advances in Applied Probability*, 30(1):85–112, 1998.

[114] Q. Liu. Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stochastic Processes and their Applications*, 95(1):83–107, 2001.

[115] A. Martin-Löf. Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *Journal of Applied Probability*, 23(2):265–282, 1986.

[116] A. J. Lotka. The frequency distribution of scientific productivity. *Journal of Washington Academy Sciences*, 16(12):317–323, 1926.

[117] N. Ma, J. Guan, and Y. Zhao. Bringing pagerank to the citation analysis. *Information Processing & Management*, 44(2):800–810, 2008.

[118] D. Marcus, and Y. Shavitt. Rage - A rapid graphlet enumerator for large networks. *Computer Networks*, 56(2):810–819, 2012.

[119] E. Matricciani. The probability distribution of the age of references in engineering papers. *IEEE transactions on professional communication*, 34(1):7–12, 1991.

[120] P-A. G. Maugis, C. E. Priebe, S. C. Olhede, and P. J. Wolfe. Statistical inference for network samples using subgraph counts. *Preprint*, arXiv:1701.00505 [stat.ME], 2017.

[121] A. De Meyer, and J. Teugels. On the asymptotic behaviour of the distributions of the busy period and service time in M/G/1. *Journal of Applied Probability*, 17(3):802–813, 1980.

[122] S. Milgram. The small world problem. *Psychology Today*, Ziff-Davis Publishing Company, 1967.

[123] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594): 824–827, 2002.

[124] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004.

[125] M. Molloy, and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–179, 1995.

[126] M. Molloy, and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability and Computing*, 7(3):295–305, 1998.

[127] T. F. Móri. On random trees. *Studia Scientiarum Mathematicarum Hungarica*, 39(1-2):143–155, 2002.

[128] T. F. Móri. The maximum degree of the Barabási-Albert random tree. *Combinatorics, Probability and Computing*, 14(3):339–348, 2005.

[129] O. Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Probability Theory and Related Fields*, 57(3):365–395, 1981.

[130] M. E. J. Newman. Networks: An introduction. *Oxford University Press*, 2010.

[131] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167-256, 2003.

[132] R. Oliveira, and J. Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Mathematics*, 2(2):121–163, 2005.

[133] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski. Intensity and coherence of motifs in weighted complex networks. *Phyical Review E*, 71:065103, 2005.

[134] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: bringing order to the web. *Proceedings of the 7th international conference on World-Wide Web*, 161–172, 1998.

[135] G. Pandurangan, P. Raghavan, and E. Upfal. Using PageRank to characterize web structure. *Computing and Combinatorics: 8th annual international conference*, 330–339, 2002.

[136] D. J. de Solla Price. Quantitative measures of the development of science. *Archives Internationales d'Histoire des Sciences*, 14:85–93, 1951.

[137] D. J. de Solla Price. Networks of scientific papers. *Science*, 149(3683):510–515, 1965.

[138] D. J. de Solla Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5):292–306, 1976.

[139] L. O. Prokhorenkova, A. Ryabchenko, and E. Samosvat. Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient. *International workshop on algorithms and models for the Web-graph (WAW)*, 185–202, 2013.

[140] L. O. Prokhorenkova. General results on preferential attachment and clustering coefficient. *Optimization Letters*, 11(2):279–298, 2016.

[141] L. O. Prokhorenkova, and A. V. Krot. Local clustering coefficients in preferential attachment models. *Doklady Mathematics*, 94(3):623–626, 2016.

[142] F. F. J. van Raan. Two-step competition process leads to quasi powerlaw income distributions - Application to scientific publication and citation distributions. *Physica A: Statistical Mechanics and its Applications*, 298(3-4):530–536, 2001.

[143] F. F. J. van Raan. Competition amongst scientists for publication status: toward a model of scientific publication and citation distributions. *Scientometrics*, 51(1):347–357, 2001.

[144] F. Radicchi, S. Fortunato, and C. Castellano. Universality of citation distributions: towards an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the USA*, 105(45):17268–17272, 2008.

[145] E. Ravasz, and A. L. Barabási. Hierarchical organization in complex networks. Physical Review E, 67(2):026112, 2003.

[146] E. Ravasz, A. L. Somera, D. A. Mongru, Z. Oltvai, and A. L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science* 297(5586):1551–1555, 2002.

[147] S. Redner. Citation statistics from 110 years of Physical Review. *Physics Today*, 58(6):49–54, 2005.

[148] O. Riordan and N. Wormald. The diameter of sparse random graphs. *Combinatorics, Probability and Computing*, 19(5-6):835–926, 2010.

[149] S. Rosengren. A multi-type preferential attachment model. *Preprint*, arXiv:1704.03256 [math.PR], 2017.

[150] N. Ross Power laws in preferential attachment graphs and Stein's method for the negative binomial distribution. *Advances in Applied Probability*, 45(3):876–893, 2013.

[151] A. Rudas. Random tree growth with general weight function. *Preprint*, arXiv:math/0410532 [math.PR], 2004.

[152] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures & Algorithms*, 31(2):186–202, 2007.

[153] P. O. Seglen. The skewness of science. *Journal of the American Society for Information Science*, 43(9):628–638, 1992.

[154] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, 31(1):64–68, 2002.

[155] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos. Power laws and the AS-level internet topology. *IEEE/ACM transactions on networking*, 11(4):514–524, 2003.

[156] C. Stegheuis, R. van der Hofstad, and J. S. H. van Leeuwaarden. Power-law relations in random networks with communities. *Physical Review E*, 94:012302, 2016.

[157] C. Stegehuis, R. van der Hofstad, and J. S. H. van Leeuwaarden. Variational principle for scale-free network motifs. *In preparation*, 2018+.

[158] J. Szymanski. Concentration of vertex degrees in a scale-free random graph process. Random Structures & Algorithms, 26(3):224–236, 2005.

[159] J. Travers, and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[160] A. Vázquez. Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations. *Physical Review E*, 67(5):056104, 2003.

[161] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the Internet. *Physical Review E*, 65(6):12, 2001.

[162] Y. Volkovich and N. Litvak. Asymptotic analysis for personalized web search. *Advances in Applied Probability*, 42(2):577–604, 2010.

[163] Y. Volkovich, N. Litvak, and D. Donato. Determining factors behind the PageRank loglog plot. *nternational workshop on algorithms and models for the Web-graph (WAW)*, 108–123, 2007.

[164] M. L. Wallace, V. Larivire, and Y. Gingras. Modeling a century of citation distributions. *Journal of Informetrics*, 3(4):296–303, 2009.

[165] L. Waltman and N. J. van Eck. The relation between eigenfactor, audience factor, and influence weight. *Journal of the American Society for Information Science and Technology*, 61(7):1476–1486, 2010.

[166] D. Wang, C. Song, and A. L. Barabási. Quantifying long-term scientific impact. *Science*, 342(6154):127–132, 2013.

[167] M. Wang, G. Yu, and D. Yu. Measuring the preferential attachment mechanism in citation networks. *Physica A: Statistical Mechanics and its Applications*, 387(18):4692–4698, 2008.

[168] M. Wang, G. Yu, and D. Yu. Effect of the age of papers on the preferential attachment in citation networks. *Physica A: Statistical Mechanics and its Applications*, 388(19):4273–4276, 2009.

[169] R. Wang, W. Zhang, H. Deng, N. Wang, Q. Miao, and X. Zhao. Discover community leader in social network with pagerank. *Advances in Swarm Intelligence: 4th International Conference*, 154–162, 2013.

[170] D. J. Watts. Small worlds: the dynamics of networks between order and randomness. *Princeton University Press*, 1999.

[171] D. J. Watts, and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[172] S. Wernicke, and F. Rasche. FANMOD: a tool for fast network motif detection. *Bioinformatics*, 22(9): 1152–1153, 2006.

[173] Y. Wu, T. J. Z. Fu,and D. M. Chiu. Generalized preferential attachment considering aging. *Journal of Informetrics*, 8(3):650–658, 2014.

[174] S. Wuchty, Z. Oltvai, and A. L. Barabási. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genetics*, 35(2):176–179, 2003.

Summary

Preferential attachment models for dynamic networks

The theory of complex networks and graphs gained a lot of attention in the last two decades. It has been observed that many real-world networks show common features, even though the networks represent completely different systems. Since real–world networks can be rather large, mathematical models are useful to describe and explain the features observed in complex networks.

Random graphs turn out to be a very powerful tool in network science, and they can be divided in two large classes: static and dynamical graph. The first class is composed by random graph models where the number of vertices is fixed, and connections are created between existing vertices. Dynamical random graphs are mathematical models where the number of vertices increases over time.

In this dissertation, we focus on preferential attachment mechanisms in dynamical random graphs. Preferential attachment models are dynamical networks that grow obeying a relatively simple rule: at every time-step, a new vertex appears in the graph, and it connects to one (or more) existing vertices, chosen with probability proportional to some function of the degree of the vertices. We can then have random trees (when new vertices connect to only one old vertex) or graphs (when new vertices have more than one connection). This dynamic has been generalized in the literature. We focus on the so-called affine preferential attachment model, where the attachment probabilities are proportional to the degrees plus a constant. This additive constant turns out to affect the graph tremendously, and thus allows us to tune properties of the network.

The aim of the dissertation is to compare features observed in real-world networks with properties of preferential attachment models. In particular we focus on the following features: degree distributions, distances, subgraph frequency and PageRank distributions.

Affine preferential attachment models are known to show power-law degrees (the fraction of vertices with degree k decreases as an inverse power of k), similarly to

real-world networks. We develop new techniques that allows us to prove results in affine preferential attachment models. In particular, we extend the embedding in continuous time of preferential attachment models, that was known only for the tree setting. We are able to extend this to graphs, with the advantage that our technique applies to more general preferential attachment functions.

We prove that preferential attachment models are ultra-small worlds when the degrees have infinite variance. We define the distance between two vertices as the minimum number of edges that creates a connecting path, and the diameter of the graph is then the maximum of all distances. We prove that, with high probability, the diameter of a preferential attachment model with infinite-variance degrees is a constant times the loglog the size of the graph, and we identify the precise constant.

We further describe the subgraph frequencies in preferential attachment models. We prove that the expected subgraph occurrence of any subgraph scales as a function of the size of the graph. The precise function is a solution of an optimization problem that depends on the internal structure of the subgraph. In particular we prove that the number of triangles in preferential attachment models is sublinear, thus these models are not clustered, in contrast with real-world networks.

The relatively low occurrence of subgraphs with cycles is called the treelike-property. Using the notion of local weak convergence, we formalize this concept and we prove that different versions of preferential attachment models with affine preferential attachment functions converge locally weakly to the same inhomogeneous multi-type branching process called the Pólya Point Tree.

PageRank is a well-known algorithm to rank vertices in a network. It is observed that in real-world networks the degree distribution and the PageRank distribution show a remarkably similar power-law behavior. We develop a technique based on local weak convergence that allows us to investigate the asymptotic PageRank distribution in directed random graphs. In particular, we apply this to preferential attachment models, showing that in this case the PageRank distribution tail is at least as large as that of the degrees.

After investigating the affine preferential attachment model, we consider applications of such models to citation networks. Citation networks are networks where scientific papers are vertices and references are directed edges. We focus our attention to observable properties that can give information about the preferential attachment mechanism in these networks, based on data obtained from Web of Science. We observe that a simple affine preferential attachment mechanism is not suitable to model citation networks.

Starting from the data analysis, we define a generalized preferential attachment model, where the probability of choosing a paper not only depends on the degree (highly cited papers are more likely to be cited again), but also on the age of the paper (old papers are less cited than new papers) and an intrinsic potential, that reflects the high inhomogeneity of papers (some papers are forgotten quickly, while other are cited even after decades). In particular, we identify the conditions that are necessary to generate a power-law degree distribution in this generalized setting, as observed in citation networks data.

Curriculum vitae

Alessandro Garavaglia was born on March 2nd, 1990 in Rho, Italy. After finishing his secondary education at Liceo Scientifico Statale E. Majorana in Rho, in 2009 he began his studies in Mathematics at the University of Milan-Bicocca. He received his Bachelor degree in 2012 and in the same year he started his Master. In 2013 he won one of the EXTRA scholarships by the University of Milan-Bicocca, that he used to spend three months in 2014 at Eurandom working on his final project. He got then his Master degree (Cum Laude) in July 2014. During these years he volunteered as an educator for teenagers at St. Paul's Church in Rho.

In January 2015 he started a PhD research project within the Stochastic (STO) Section at Eindhoven University of Technology (TU/e), under the supervision of Prof. Remco van der Hofstad and Prof. Nelly Litvak. The topic was the study of preferential attachment models and in particular their application to citation networks. His research project was part of the NETWORKS research program.

During his PhD Alessandro was involved in teaching activities, he supervised a bachelor student in his final project, and he participated to several workshops in Eindhoven, Austin, Munich, Como, Edinburgh and Oberwolfach. He was one of the organisers of the "NETWORKS PhD days" for the NETWORKS program. He joined the PhD/PdEngs Council of the Mathematics and Computer Science Department at TU/e, chairing it between June 2017 and June 2018.