Ranking Algorithms on Directed Configuration Networks

Ningyuan Chen

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Ranking Algorithms on Directed Configuration Networks

Ningyuan Chen

In recent decades, complex real-world networks, such as social networks, the World Wide Web, financial networks, etc., have become a popular subject for both researchers and practitioners. This is largely due to the advances in computing power and big-data analytics. A key issue of analyzing these networks is the centrality of nodes. Ranking algorithms are designed to achieve the goal, e.g., Google's PageRank. We analyze the asymptotic distribution of the rank of a randomly chosen node, computed by a family of ranking algorithms on a random graph, including PageRank, when the size of the network grows to infinity.

The thesis is based on [Chen and Olvera-Cravioto, 2013; Chen *et al.*, 2014; Chen and Olvera-Cravioto, 2014; Chen and Olvera-Cravioto, 2015]. In [Chen and Olvera-Cravioto, 2013], we propose a configuration model generating the topological structure of a directed graph given in- and out-degree distributions of the nodes. The algorithm guarantees the generated graph to be simple (without self-loops and multiple edges in the same direction) for a broad spectrum of degree distributions, including power-law distributions. Power-law degree distribution is referred to as scale-free property and observed in many real-world networks. On the random graph $\mathcal{G}_n = (V_n, E_n)$ generated by the configuration model, we study the distribution of the ranks, which solves

$$R_i = \sum_{j:(j,i)\in E_n} C_j R_j + Q_i$$

for all node i, some weight C_i and personalization value Q_i. In [Chen *et al.*, 2014], we show that as the size of the graph $n \to \infty$, the rank of a randomly chosen node converges weakly

to the endogenous solution of the stochastic fixed-point equation:

$$\mathcal{R} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{\mathcal{N}} \mathcal{C}_i \mathcal{R}_i + \mathcal{Q},$$

where $(\mathcal{Q}, \mathcal{N}, \{\mathcal{C}_i\})$ is a random vector and $\{\mathcal{R}_i\}$ are i.i.d. copies of $\mathcal{R}$, independent of $(\mathcal{Q}, \mathcal{N}, \{\mathcal{C}_i\})$. This main result is divided into three steps. First, we show that the rank of a randomly chosen node can be approximated by applying the ranking algorithm on the graph for finite iterations. Second, by coupling the graph to a branching tree that is governed by the empirical size-biased distribution, we approximate the finite iteration of the ranking algorithm by the root node of the branching tree. Finally, we prove that the rank of the root of the branching tree converges to that of a limiting weighted branching process, which is independent of n and solves the stochastic fixed-point equation. The technical detail of the third step and its generalization of coupling a sequence of branching processes are presented in [Chen and Olvera-Cravioto, 2014]. Our result formalizes the well-known heuristics, that a network often locally possesses a tree-like structure. We conduct a numerical example showing that the approximation is very accurate for English Wikipedia pages (over 5 million).

To draw a sample from the endogenous solution of the stochastic fixed-point equation, one can run linear branching recursions on a weighted branching process. We provide an iterative simulation algorithm based on bootstrap. Compared to the naive Monte Carlo, our algorithm reduces the complexity from exponential to linear in the number of recursions. We show that as the boostrap sample size tends to infinity, the sample drawn according to our algorithm converges to the target distribution in the Kantorovich-Rubinstein distance and the estimator is consistent.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Prof. Mariana Olvera-Cravioto for the continuous support of my Ph.D study and research. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Jose Blanchet, Prof. Antonius Dieker, Prof. Predrag Jelenkovic, and Prof. Karl Sigman, for their time to participate in my defense and insightful comments.

My sincere thanks also go to Prof. Mark Broadie, Prof. Guillermo Gallego, and Prof. Steven Kou, for offering guidance in research projects. Their diligence, dedication shape my way of doing research.

Last but not the least, I would like to thank my family: my parents Dongling Pan and Zhenbao Shen, for giving birth to me at the first place and supporting me spiritually throughout my life.

I dedicate my thesis to my wife Fang Zhu, and my daughter Ariel Yuzhu Chen.

Chapter 1

Introduction

The information age has seen a booming development of large real-world networks, such as the World Wide Web (WWW), social networks, interbank financial networks. Thanks to unprecedentedly intensive social and economic interactions, these networks usually possess several common features: they are huge, having thousands or even millions of nodes and edges; they are very dynamic, growing and adapting every day. Fortunately, advanced computing and data-storage power makes it possible to analyze these networks.

Many methods that are used to deal with complex networks are deterministic, i.e. given a particular graph structure (e.g., the adjacency matrix), those methods analyze the properties of the graph. However, deterministic methodologies have its limitations. First, the study doesn't tell how and why the network evolves to what it looks now. For example, how does a superstar emerge on Twitter. Second, the similarity and distinction between networks make it difficult to generalize. For example, what is the difference between Facebook and a collaboration network, can an algorithm designed for the former be applied to the latter. Third, the sheer complexity due to the size of the networks makes many algorithms infeasible. These difficulties popularize the study of random graph. Unlike deterministic methodologies, random graph tries to focus on the construction of networks. Imagine numerous parallel universes. In each universe we generate a Twitter network according to a random mechanism. Would there be a universe in which the Twitter looks exactly (or very similar to) the one we have? If the designed random mechanism can indeed describe how Twitter is generated in our universe, then we can study the properties of the random graph

instead, which may yield more general and deeper results. Random graph can usually give insight about the evolution of the graph by specifying the random process that generates. It also provides connections between complex networks: If two networks, even differing tremendously in size, structure and application, are generated by the same mechanism, then they may still be sharing some common features.

In the first part of the thesis, we introduce a directed random graph model. The model describes how a graph can be constructed (randomly) according to our proposed mechanism and what properties it has. More precisely, given the distributions of the in- and out-degrees of the nodes (either p.m.f.s or realized observations), our model generates a directed graph with no self-loops and multiple edges in the same direction between two nodes. The resulted graph is uniformly distributed in the set of all such graphs with the same degrees. See Section 1.1 for detailed introduction and literature review of this topic.

I then focus on the analysis of ranking algorithms on random networks. Ranking algorithms output a value, the so-called rank, of a node in a graph, according to its topological strucutre, i.e. how nodes are connected to each other. To demonstrate, in a star graph, the central node will have a high rank compared to its leaves; in a complete graph, on the other hand, all nodes will have the same rank. One of the most famous applications of ranking algorithms is Google's PageRank, which is used to rank webpages in the WWW. Some webpages are more "important", because a lot of other pages link to this page, e.g., the frontpage of the Wall Street Journal. These important pages are ranked high when you use Google to search relevant topics. Another example is social networks, in which celebraties and sportstars have millions of followers and thus a huge impact on the community.

To rank the nodes according to their centrality, note that a node would be ranked high if it is connected to a lot of other nodes, or its connected neighbors are important. Ranking algorithms solve a linear equation system of the size of the network, and the solution (a vector of the same size) represents the ranks of all nodes in the graph. Even though the computation is not hard, it leaves open some interesting questions: How is the distribution of the ranks related to the graph, what common attributes do those nodes of high ranks have, how can we design customized ranking algorithms if we want the resulted rank to follow a particular distribution.

Random graph allows us to analyze the ranking algorithm from a new perspective: Rather than concerning about the structure of the graph, it studies the underlying mechanism that generates the graph, and hence greatly simplifies the computation of the ranks. It is based on the heuristics that if we zoom in and analyze the local property of a network, the structure resembles a branching process [Newman *et al.*, 2001]. We can thus characterize the distribution of the rank of a randomly chosen node when the graph size grows to infinity. See Section 1.2 for detailed introduction on this topic.

My last topic is efficient simulation of linear branching recursions. As appearing in the asymptotic analysis of ranking algorithms, weighted branching process turns out to be a useful tool for approximating graphs. The simulation of branching recursions allows us to numerically compute the endogenous solution of a stochastic fixed-point equation on a weighted branching process. However, naive Monte Carlo simulation is extremely time- and memory-consuming, because of the exponential complexity due to branching. Our algorithm can be used to approximately simulate the target distribution. See Section 1.3 for detailed introduction to the algorithm.

1.1 Directed random graphs

One of the earliest random graph models, the Erdős-Rényi model, can be dated back to 1960s [Erdős and Rényi, 1959]. In this model, given n nodes, any pair of nodes are connected with probability p. The Erdős-Rényi graph yields many analytical properties, such as the existence of large components, the connectedness of the graph, etc. However, a significant drawback makes the model unfit for real-world networks: It doesn't admit heavy tails of the degrees, as is the case in many real networks. More precisely, it only gives Poisson degree distribution when $n \to \infty$ and np remains fixed.

Starting from 1990s and 2000s, complex networks have attracted the attention of scholars from biology, physics, computer science, and mathematics, e.g., [Albert and Barabási, 1999; Watts and Strogatz, 1998; Albert and Barabási, 2002; Boccaletti *et al.*, 2006]. Among them, the Watts-Strogatz model [Watts and Strogatz, 1998] and Barabasi-Albert model [Albert and Barabási, 2002] provide two popular alternatives of modeling random graphs. In the

Watts-Strogatz model, an undirected graph is constructed given the number of nodes n and mean degree k. The model first constructs a ring lattice, with each node connected to k neighbors, $k/2$ on each side. Then for each node, we take each of its edge and rewire it with probability β, where the new node is chosen uniformly from all possible values that avoid self-loops. However, it is still unable to produce heavy-tail degree distributions. The Barabasi-Albert model uses a preferential attachment mechanism and thus results in *scale-free* networks, i.e., the degree distribution follows a power law. The network begins with an initial connected network of n_0 nodes. When a new node is added, it is connected to m of the existing nodes, with a probability proportional to the number of edges the nodes already have. The intuition is that a node with high degree is more likely to be connected to new nodes, and hence the degree distribution has a heavy tail. However, the index of the power law is always 3 and the inflexibility imposes a significant limitation of the model. Other classical examples are the Chung-Lu model [Chung and Lu, 2002b], and the Configuration Model [van der Hofstad, 2014, Chapter 7]. New models continue to appear, tuned to the properties of specific networks. For example, an interesting "super-star" model was recently developed to describe retweet graphs [Bhamidi *et al.*, 2012]. We refer to [van der Hofstad, 2014; Durrett, 2007; Newman, 2010] for a more detailed discussion of random graph models for complex networks.

To study the behavior of undirected random graphs with arbitrary degree distributions, [Newman *et al.*, 2001] introduce an intuitive and helpful heuristic that is consistent with the motivation of our work. They avoid analyzing the "big picture" of the graph but rather explore the graph from any single node. The depth-first exploration sequentially adds the neighbors of the node, the neighbors of neighbors, etc. For huge networks, this exploration may proceed for many layers until a cycle is encountered. In other words, a graph can be regarded as a branching process, at least locally. This heuristic allows the authors to derive moment generating functions of various quantities.

As mentioned, some of the models don't have enough flexibility to model real-world networks. The main reason is that their resulted graphs fail to match desired degree distributions. In real-world networks, it is often found that the fraction of nodes with (in- or out-) degree k is $\approx c_0 k^{-\alpha-1}$, usually $\alpha \in (1,3)$ (e.g., [Kleinberg *et al.*, 1999;

Broder *et al.*, 2000; Brin and Page, 1998; Newman, 2010]). The reason is simple: There are usually "superstar" nodes in the graph that is far more influential than other nodes, causing a heavy-tail degree distribution. Hence the ability to match degree distributions to real graphs is perhaps the first one would desire from a random graph model. There are several models that accomplish this for undirected graphs being proposed in the recent literature [McKay and Wormald, 1990b; Chung and Lu, 2002b; Chung and Lu, 2002a; Britton *et al.*, 2006]. However not much has been done for the directed case, except in [Amini *et al.*, 2013] the authors briefly describe a directed configuration model and its properties. Our work [Chen and Olvera-Cravioto, 2013] fills in the gap.

We first briefly introduce the undirected case in [Britton *et al.*, 2006], which is closely related to our directed model. Given a probability distribution F, the goal is to provide an algorithm to generate a simple random graph (doesn't have self-loops or multiple edges between two nodes) whose degree distribution is approximately F. Two of the models presented in [Britton *et al.*, 2006], are in turn related to the well-known configuration model [Wormald, 1978; Bollobás, 1980], where nodes are given stubs or half-edges according to a degree sequence $\{d_i\}$ and these stubs are then randomly paired to form edges. To obtain a prescribed degree distribution, the degree sequence $\{d_i\}$ is chosen as i.i.d. random variables having distribution F. This method allows great flexibility in terms of the generality of F, which is very important in the applications we have in mind. The most general of the results presented here require only that the degree distributions have finite $(1+\epsilon)$th moment, and are therefore applicable to a great variety of examples, including the WWW.

For a directed random graph there are two distributions that need to be chosen, the in-degree and out-degree distributions, denoted respectively $F = \{f_k : k \geq 0\}$ and $G = \{g_k : k \geq 0\}$. The in-degree of a node corresponds to the number of edges pointing to it, while the out-degree is the number of edges pointing out. To follow the ideas from [van der Hofstad *et al.*, 2005; Britton *et al.*, 2006], we propose to draw the in-degree and out-degree sequences as i.i.d. observations from distributions F and G. Unlike the undirected case where the only main problem with this approach is that the sum of the degrees might not be even, which is necessary to draw an undirected graph, in the directed case the corresponding condition is that the sum of the in-degrees and the sum of the out-degrees be the same.

Since the probability that two i.i.d. sequences will have the same sum, even if their means are equal, converges to zero as the number of nodes grows to infinity. [Chen and Olvera-Cravioto, 2013] first focuses on how to construct valid degree sequences without significantly destroying their i.i.d. properties. Once we have valid degree sequences the problem is how to obtain a simple graph, since the random pairing may produce self-loops and multiple edges in the same direction. This problem is addressed in two ways, the first of which consists in showing sufficient conditions under which the probability of generating a simple graph through random pairing is strictly positive, which in turn suggests repeating the pairing process until a simple graph is obtained. The theoretical foundation of this method is laid out in [Blanchet and Stauffer, 2013]. The second approach is to simply erase the self-loops and multiple edges of the resulting graph. In both cases, one must show that the degree distributions in the final simple graph remain essentially unchanged. In particular, if we let $f_k^{(n)}$ be the probability that a randomly chosen node from a graph of size n has in-degree k, and let $g_k^{(n)}$ be the corresponding probability for the out-degree, then we will show that,

$$f_k^{(n)} \to f_k \qquad \text{and} \qquad g_k^{(n)} \to g_k,$$

as $n \to \infty$. We also prove a similar result for the empirical distributions.

The question of whether a given pair of in- and out-degree sequences $(\{m_i\}, \{d_i\})$ is graphical, i.e., from which it is possible to draw a simple directed graph, has been recently studied in [Erdös *et al.*, 2010], where algorithms to realize such graphs have also been analyzed. Random directed graphs with arbitrary degree distributions have been studied in [Newman *et al.*, 2001] via generating functions, which can be used to formalize concepts such as "in-components" and "out-components" as well as to estimate their average size. Models of growing networks that can be calibrated to mimic the power-law behavior of the WWW have been analyzed using statistical physics techniques in [Krapivsky *et al.*, 2001; Krapivsky and Redner, 2002]. The approach followed in [Chen and Olvera-Cravioto, 2013] focuses on one hand on the generation of in- and out-degree sequences that are close to being i.i.d. and that are graphical with high probability, and on the other hand on providing conditions under which a simple graph can be obtained through random pairing. The directed configuration model (DCM) with (close to) i.i.d. degree sequences, although not a growing network model, has the advantage of being analytically tractable and easy to

simulate. This topic is covered in Chapter 2.

1.2 Ranking algorithms

Based on the DCM proposed in Chapter 2, we are able to study the ranking of nodes according to their centrality in a complex network such as the Internet, the World Wide Web, and other social and biological networks. For a comprehensive overview of the vast literature on rankings in networks we refer the reader to [Langville and Meyer, 2011], and more recently to [Boldi and Vigna, 2014] for a thorough up-to-date mathematical classification of centrality measures.

In this thesis we analyze a family of ranking algorithms which includes Google's PageRank, the algorithm proposed by Brin and Page [Brin and Page, 1998], and which is arguably the most influential technique for computing rankings of nodes in large directed networks. The original definition of PageRank is the following. Let $\mathcal{G}_n = (V_n, E_n)$ be a directed graph, with a set of (numbered) vertices $V_n = \{1, \dots, n\}$, and a set of directed edges E_n. Choose a constant $c \in (0, 1)$, which is called a *damping factor*, and let $\mathbf{q} = (q_1, q_2, \dots, q_n)$ be a *personalization* probability vector, i.e., $q_i \geq 0$ and $\sum_{i=1}^{n} q_i = 1$. Denote by $d_i = |\{j : (i, j) \in E_n\}|$ the out-degree of node $i \in V_n$. Then the PageRank vector $\mathbf{r} = (r_1, \dots, r_n)$ is the unique solution to the following system of linear equations:

$$r_i = \sum_{j:(j,i)\in E_n} \frac{c}{d_j} r_j + (1 - c)q_i, \quad i = 1, \dots, n. \tag{1.1}$$

Google's PageRank was designed to rank Web pages based on the network's structure, rather than their content. The idea behind (1.1) is that a page is important if many important pages have a hyperlink to it. Furthermore, by tuning the personalization values, q_i's, one can, for instance, give preference to specific topics [Haveliwala, 2002] or penalize spam pages [Gyöngyi *et al.*, 2004].

In the original definition, $\mathbf{r}$ is normalized so that $||\mathbf{r}||_1 = 1$, where the norm $||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$ denotes the l_1 norm in $\mathbb{R}^n$. The vector $\mathbf{r}$ can be thought of as the stationary distribution of a random walk on the graph. At any node, the random walk travels to all the nodes that are connected from the current node with equal probability (summing up to c), or restarts at a random nodes with probability $(1 - c)\mathbf{q}$. Therefore, the rank of nodes

has an intuitive interpretation: If in steady state, the random walk spends more time on a node, then the node is deemed to have a higher rank.

Since the average PageRank in $\mathbf{r}$ scales as $O(1/n)$, it is more convenient for our purposes to work with a scaled version of PageRank:

$$n\mathbf{r} =: \mathbf{R} = (R_1, R_2, \ldots, R_n).$$

Then, also using the notation C_j for c/d_j, and notation Q_i for $n(1-c)q_i$, we rewrite (1.1) to obtain

$$R_i = \sum_{j:(j,i)\in E_n} C_j\, R_j + Q_i, \quad i = 1, \ldots, n. \tag{1.2}$$

Throughout the thesis, we will refer to $\mathbf{R}$ as the PageRank vector and to $\mathbf{Q} = (Q_1, Q_2, \ldots, Q_n)$ as the personalization vector.

The basic definition (1.1) has many modifications and generalizations. Our analysis will cover a wide range of them by allowing a general form of the coefficients in (1.2). For example, our model admits a random damping factor as studied in [Constantine and Gleich, 2009]. Numerous applications of PageRank and its modifications include graph clustering [Andersen *et al.*, 2006], spam detection [Gyöngyi *et al.*, 2004], and citation analysis [Chen *et al.*, 2007; Waltman and van Eck, 2010].

Note that to solve $\mathbf{R}$ from (1.2) for c, $\mathbf{Q}$ and a given graph, the PageRank is a linear system of size n. Hence the solution of PageRank is not difficult, at least computationally. However, an interesting phenomenon requires more effective analytical characterization of the ranking vector: From the work of Pandurangan et al. [Pandurangan *et al.*, 2002], and many papers that followed, the following hypothesis has always been confirmed by the data.

The power law hypothesis: *If the in-degree distribution in a network follows a power law then the PageRank scores in this network will also follow a power law with the same exponent.*

The power law hypothesis is plausible because in (1.1) the number of terms in the summation on the right-hand side is just the in-degree of i, so the in-degree provides a 'mean-field' approximation for PageRank [Fortunato *et al.*, 2008]. However, this argument is not exact nor accurate enough, which is confirmed by the fact that the top-ranked nodes in PageRank are not exactly those with the largest in-degrees [Chen *et al.*, 2007; Volkovich *et*

al., 2009; Vigna, 2014]. Exact mathematical evidence supporting the power law hypothesis is surprisingly scarce. As one of the few examples, [K. Avrachenkov, 2006] obtains the power law behavior of average PageRank scores in a preferential attachment graph by using Polya's urn scheme and advanced numerical methods.

In a series of papers, Volkovich et al. [Litvak *et al.*, 2007; Volkovich *et al.*, 2007; Volkovich and Litvak, 2010] suggested an analytical explanation for the power law behavior of PageRank by comparing the PageRank of a randomly chosen node to the endogenous solution of a stochastic fixed point equation (SFPE) that mimics (1.2):

$$R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q. \tag{1.3}$$

Here N (in-degree) is a nonnegative integer random variable having a power law distribution with exponent α, Q (personalization) is an arbitrary positive random variable, and the C_i's are random coefficients that in [Volkovich and Litvak, 2010] equal c/D_i, with D_i being the out-degree of a node provided $D_i \geq 1$. The symbol $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. Assuming that N is regularly varying and using Laplace transforms, it was proved in [Volkovich and Litvak, 2010] that R has a power law with the same exponent as N if N has a heavier tail than Q, whereas the tail of R is determined by Q if it is heavier than N. The same result was also proved independently in [Jelenković and Olvera-Cravioto, 2010] using a sample-path approach.

However, the SFPE does not fully explain the behavior of PageRank in networks since it implicitly assumes that the underlying graph is an infinite tree, a condition that is never true in real-world networks. We complete the argument when the underlying network is a DCM by showing that the distribution of the PageRank in the graph converges to the endogenous solution of a SFPE. Our techniques are likely to be useful in the analysis of PageRank in other locally tree-like graphs. Methodologically, our results provide insights in the following respects. First, branching processes have been heuristically used to approximate graphs; but very little is known about what types of graphs will suit this method. We show that the DCM is a proper model that the heuristic can be applied to. Second, we provide a coupling argument and show that the tree and the graph are likely to decouple at the size of $O(\sqrt{n})$. This argument can be used to analytically justify the heuristic for other graph models.

Third, we derive a bound for linear recursions on two weighted branching processes, given that the branching vectors are close. For large networks, the previous coupling technique leads to the analysis of weighted branching processes. The bound we derive ensures a random variable that serves as the limit of a certain quantity on the graph, as its size grows to infinity.

The proof of the main result consists of the following three steps:

1. *Finite approximation* (Section 3.2.2). Show that the class of rankings that we study can be approximated in the DCM with any given accuracy by a finite (independent of the graph size n) number of matrix iterations. The DCM plays a crucial role in this step since it implies that the ranks of all the nodes in the graph have the same distribution.

2. *Coupling with a tree* (Section 3.3). Construct a coupling of the DCM graph and a "thorny branching tree" (TBT). In a TBT each node with the exception of the root has one outbound link to its parent and possibly several other unpaired outbound links. During the construction, all nodes in both the graph and the tree are also assigned a weight. The main result in this section is the Coupling Lemma 12, which states that the coupling between the graph and the tree will hold for a number of generations in the tree that is logarithmic in n. The locally tree-like property of the DCM is important for this step.

3. *Convergence to a weighted branching process* (Section 3.4). Show that the rank of the root node of the TBT converges weakly to the SFPE. This last step requires the weak convergence of the random distributions that define the TBT in the previous step. The convergence requires explicit bounds for the Kantorovich-Rubinstein distance between two random variables constructed according to the representation for the endogenous solution to (1.3); these bounds are given in terms of the Kantorovich-Rubinstein distance between their generic branching vectors. We then use these bounds to obtain the convergence of a sequence of such random variables in the same distance.

This main result is derived in Chapter 3.

1.3 Efficient simulation for branching linear recursions

Consider a general SFPE

$$R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q. \tag{1.4}$$

where $(Q, N, C_1, C_2, \dots)$ is a real-valued random vector with $N \in \mathbb{N} = \{0, 1, 2, \dots\} \cup \{\infty\}$, $\{R_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. copies of R, independent of $(Q, N, C_1, C_2, \dots)$, and $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. In deriving the main result in Section 3.3, we generalize the idea of coupling, which turns out to be of independent interest. Besides ranking algorithms, the stochastic fixed-point equation (1.4) is closely related to the complexity analysis of divide and conquer algorithms such as Quicksort [Rösler, 1991; Fill and Janson, 2001; Rösler and Rüschendorf, 2001], in addition to the information ranking algorithm. More precisely, the number of comparisons required in Quicksort for sorting an array of length n, properly normalized, satisfies in the limit as the array's length grows to infinity a distributional equation of the form in (1.4).

As further motivation for the study of branching fixed-point equations, we mention the maximum equation

$$R \stackrel{\mathcal{D}}{=} Q \vee \bigvee_{r=1}^{N} C_r R_r, \tag{1.5}$$

with $(Q, N, C_1, C_2, \dots)$ nonnegative, which has been shown to appear in the analysis of the waiting time distribution in large queueing networks with parallel servers and synchronization requirements [Karpelevich *et al.*, 1994; Olvera-Cravioto and Ruiz-Lacedelli, 2014]. In this setting, $W = \log R$ represents the waiting time in stationarity of a job, that upon arrival to the network, is split into a number of subtasks requiring simultaneous service from a random subset of servers. Computing the distribution and the moments of W is hence important for evaluating the performance of such systems (e.g., implementations of MapReduce and similar algorithms in today's cloud computing). We focus in this thesis only on (1.4), but we mention that the algorithm we provide can easily be adapted to approximately simulate the solutions to (1.5).

Although the study of (1.4) has received considerable attention in the recent years [Rösler, 1991; Biggins, 1998; Fill and Janson, 2001; Rösler and Rüschendorf, 2001; Aldous and Bandyopadhyay, 2005; Alsmeyer *et al.*, 2012; Alsmeyer and Meiners, 2012; Alsmeyer

and Meiners, 2013; Jelenković and Olvera-Cravioto, 2012b; Jelenković and Olvera-Cravioto, 2012a; Jelenković and Olvera-Cravioto, 2015], the current literature only provides results on the characterization of the solutions to (1.4), the tail asymptotics, and in some instances, their integer moments, which is not always enough for the applications mentioned above. It is therefore of practical importance to have a numerical approach to estimate both the distribution and the general moments of R.

As a mathematical observation, we mention that (1.4) is known to have multiple solutions (see e.g. [Biggins, 1998; Alsmeyer *et al.*, 2012; Alsmeyer and Meiners, 2012; Alsmeyer and Meiners, 2013] and the references therein for the characterization of the solutions). However, in applications we are often interested in the so-called endogenous solution. This endogenous solution is the unique limit under iterations of the distributional recursion

$$R^{(k+1)} \stackrel{\mathcal{D}}{=} \sum_{r=1}^{N} C_r R_r^{(k)} + Q, \tag{1.6}$$

where $(Q, N, C_1, C_2, \dots)$ is a real-valued random vector with $N \in \mathbb{N}$, and $\{R_i^{(k)}\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. copies of $R^{(k)}$, independent of $(Q, N, C_1, C_2, \dots)$, provided one starts with an initial distribution for $R^{(0)}$ with sufficient finite moments (see, e.g., Lemma 4.5 in [Jelenković and Olvera-Cravioto, 2012a]). Moreover, asymptotics for the tail distribution of the endogenous solution R are available under several different sets of assumptions for $(Q, N, C_1, C_2, \dots)$ [Jelenković and Olvera-Cravioto, 2010; Jelenković and Olvera-Cravioto, 2012b; Jelenković and Olvera-Cravioto, 2012a; Olvera-Cravioto, 2012b].

As will be discussed later, the endogenous solution to (1.4) can be explicitly constructed on a weighted branching process. Thus, drawing some similarities with the analysis of branching processes, and the Galton-Watson process in particular, one could think of using the Laplace transform of R to obtain its distribution. Unfortunately, the presence of the weights $\{C_i\}$ in the Laplace transform

$$\varphi(s) = E\left[\exp\left(-sR\right)\right] = E\left[\exp\left(-sQ\right)\prod_{i=1}^{N}\varphi(sC_i)\right]$$

makes its inversion problematic, making a simulation approach even more necessary.

The first observation we make regarding the simulation of R, is that when $P(Q = 0) < 1$ it is enough to be able to approximate $R^{(k)}$ for fixed values of k, since both $R^{(k)}$ and R can

be constructed in the same probability space in such a way that the difference $|R^{(k)} - R|$ is geometrically small. More precisely, under very general conditions, there exist positive constants $K < \infty$ and $c < 1$ such that

$$E\left[\left|R^{(k)} - R\right|^{\beta}\right] \leq Kc^{k+1}. \tag{1.7}$$

Our goal is then to simulate $R^{(k)}$ for a suitably large value of k.

The simulation of $R^{(k)}$ is not that straightforward either, since the naive approach of simulating i.i.d. copies of $(Q, N, C_1, C_2, \dots)$ to construct a single realization of a weighted branching process, up to say k generations, is of order $(E[N])^k$. Considering that in the examples mentioned earlier we typically have $E[N] > 1$ ($N \equiv 2$ for Quicksort, $E[N] \approx 30$ in many information ranking applications, and $E[N]$ in the hundreds for MapReduce implementations), this approach is prohibitive. Instead, we propose in this thesis an iterative bootstrap algorithm that outputs a sample pool of observations $\{\hat{R}_i^{(k,m)}\}_{i=1}^m$ whose empirical distribution converges, in the Kantorovich-Rubinstein distance, to that of $R^{(k)}$ as the size of the pool $m \to \infty$. This mode of convergence is equivalent to weak convergence and convergence of the first absolute moments (see, e.g., [Villani, 2009]). Moreover, the complexity of our proposed algorithm is linear in k. A variant of the algorithm has been used in statistical physics [Abou-Chacra *et al.*, 1973; Mézard and Parisi, 2001; Mezard and Montanari, 2009], but to my knowledge, no previous results derive the explicit bound for the error of the algorithm and point out the exact mode of convergence. The detail of the algorithm is presented in Chapter 4.

Part I

Directed random graphs

Chapter 2

Directed random graphs with given degree distributions

As mentioned in the introduction, the goal of this chapter is to provide an algorithm for generating a random directed graph with n nodes with the property that its in-degrees and out-degrees have some prespecified distributions F and G, respectively. Moreover, we would like the resulting graph to be *simple*, that is, it should not contain self-loops or multiple edges in the same direction. The two models that we propose are based on the so-called configuration or pairing model, which produces a random undirected graph from a degree sequence $\{d_1, d_2, \ldots, d_n\}$.

This chapter is organized as follows. In Section 2.1 we introduce a model to construct in- and out-degree sequences that are very close to being two independent sequences of i.i.d. random variables having distributions F and G, respectively, but whose sums are the same; in the same spirit as the results in [Arratia and Liggett, 2005] we also show that the suggested method produces with high probability a graphical pair of degree sequences. In Section 2.3.1 we prove sufficient conditions under which the probability that the directed configuration model will produce a simple graph will be bounded away from zero, and show that conditional on the resulting graph being simple, the degree sequences have asymptotically the correct distributions. In Section 2.3.2 we show that under very mild conditions, the process of simply erasing self-loops and multiple edges results in a graph whose degree

distributions are still asymptotically F and G.

2.1 Graphs and degree sequences

Following the same idea of using a sequence of i.i.d. random variables to generate the degree sequence of an undirected graph, the natural extension to the directed case would be to draw two i.i.d. sequences from given distributions F and G. We note that in the undirected setting the two main problems with this approach are: 1) that the sum of the degrees may be odd, in which case it is impossible to draw a graph, and 2) that there may not exist a simple graph having the prescribed degree sequence. The first problem is easily fixed by either sampling the i.i.d. sequence until its sum is even (which will happen with probability 1/2 asymptotically), or simply adding one to the last random number in the sequence. The second problem, although related to the verification of graphicality criteria (e.g., the Erdös-Gallai criterion [Erdös and Gallai, 1960]), turns out to be negligible as the number of nodes goes to infinity, as the work in [Arratia and Liggett, 2005] shows. For directed graphs a graphicality criterion also exists, and the second problem turns out to be negligible for large graphs just as in the undirected case. Nonetheless, the equivalent of the first problem is now that the potential in-degree and out-degree sequences must have the same sum, which is considerably harder to fix. Before proceeding with the formulation of our proposed algorithm we give some basic definitions which will be used throughout the chapter.

Definition 1. *We denote by $\vec{G}(V, \vec{E})$ a directed graph on n nodes or vertices, $V = \{v_1, v_2, \ldots, v_n\}$, connected via the set of directed edges $\vec{E}$.*

Definition 2. *We say that $\vec{G}(V, \vec{E})$ is* simple *if any pair of nodes are connected by at most one edge in each direction, and if there are no edges in between a node and itself.*

Definition 3. *The in-degree m_i, respectively, out-degree d_i, of node $v_i \in V$ is the total number of edges from other nodes to v_i, respectively, from v_i to other nodes. The pair of sequences $(\mathbf{m}, \mathbf{d}) = (\{m_1, m_2, \ldots, m_n\}, \{d_1, d_2, \ldots, d_n\})$ of nonnegative integers is called a* bi-degree-sequence *if m_i and d_i correspond to the in-degree and out-degree, respectively, of node v_i.*

Definition 4. *A bi-degree-sequence* $(\mathbf{m}, \mathbf{d})$ *is said to be* graphical *if there exists a simple directed graph* $\vec{G}(V, \vec{E})$ *on the set of nodes* V *such that the in-degree and out-degree sequences together form* $(\mathbf{m}, \mathbf{d})$*. In this case we say that* $\vec{G}$ *realizes the bi-degree-sequence.*

In view of these definitions our goal is to generate the sequences $\{m_i\}$ and $\{d_i\}$ from i.i.d. samples of given distributions $F = \{f_k : k \geq 0\}$ and $G = \{g_k : k \geq 0\}$, respectively. Both F and G are assumed to be probability distributions with support on the nonnegative integers with a finite common mean μ. Note that although the Strong Law of Large Numbers (SLLN) guarantees that if we simply sample i.i.d. random variables $\{\gamma_1, \ldots, \gamma_n\}$ from F and, independently, i.i.d. random variables $\{\xi_1, \ldots, \xi_n\}$ from G, then

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \gamma_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \xi_i\right) = 1,$$

it is also true that in general

$$\lim_{n\to\infty} P\left(\sum_{i=1}^{n} \gamma_i - \sum_{i=1}^{n} \xi_i = 0\right) = 0.$$

One potential idea to fix the problem is to sample one of the two sequences, say the in-degrees, as i.i.d. observations $\{\gamma_1, \ldots, \gamma_n\}$ from F and then sample the second sequence from the conditional distribution G given that its sum is $\Gamma_n = \sum_{i=1}^{n} \gamma_i$. This approach has the major drawback that this conditional distribution may be ill-behaved, in the sense that the probability of the conditioning event, the sum being equal to Γ_n, converges to zero in most cases. It follows that we need a different mechanism to sample the degree sequences. The precise algorithm we propose is described below; we focus on first sampling two independent i.i.d. sequences and then add in- or out-degrees as needed to match their sums.

The following definition will be needed throughout the rest of the chapter.

Definition 5. *We say that a function* $L(\cdot)$ *is slowly varying at infinity if* $\lim_{x\to\infty} L(tx)/L(x) = 1$ *for all fixed* $t > 0$*. A distribution function* F *is said to be regularly varying with index* $\alpha > 0$*,* $F \in \mathcal{R}_{-\alpha}$*, if* $\overline{F}(x) = 1 - F(x) = x^{-\alpha}L(x)$ *with* $L(\cdot)$ *slowly varying.*

We will also use the notation $\Rightarrow$ to denote convergence in distribution, $\xrightarrow{P}$ to denote convergence in probability, and $\mathbb{N} = \{1, 2, 3, \ldots\}$ to refer to the positive integers.

2.1.1 Algorithm to generate degree sequences

We assume that the target degree distributions F and G have support on the nonnegative integers and have common mean $\mu > 0$. Moreover, suppose that there exist slowly varying functions $L_F(\cdot)$ and $L_G(\cdot)$ such that

$$\overline{F}(x) = \sum_{k>x} f_k \leq x^{-\alpha} L_F(x) \qquad \text{and} \qquad \overline{G}(x) = \sum_{k>x} g_k \leq x^{-\beta} L_G(x), \tag{2.1}$$

for all $x \geq 0$, where $\alpha, \beta > 1$.

We refer the reader to [Bingham *et al.*, 1987] for all the properties of slowly varying functions that will be used in the proofs. However, we do point out here that the tail conditions in (2.1) ensure that F has finite moments of order s for all $0 < s < \alpha$, and G has finite moments of order s for all $0 < s < \beta$. The constant

$$\kappa = \min\{1 - \alpha^{-1}, 1 - \beta^{-1}, 1/2\},$$

will play an important role throughout the chapter. The algorithm is given below.

1. Fix $0 < \delta_0 < \kappa$.

2. Sample an i.i.d. sequence $\{\gamma_1, \ldots, \gamma_n\}$ from distribution F; let $\Gamma_n = \sum_{i=1}^n \gamma_i$.

3. Sample an i.i.d. sequence $\{\xi_1, \ldots, \xi_n\}$ from distribution G; let $\Xi_n = \sum_{i=1}^n \xi_i$.

4. Define $\Delta_n = \Gamma_n - \Xi_n$. If $|\Delta_n| \leq n^{1-\kappa+\delta_0}$ proceed to step 5; otherwise repeat from step 2.

5. Choose randomly $|\Delta_n|$ nodes $\{i_1, i_2, \ldots, i_{|\Delta_n|}\}$ without remplacement and let

$$M_i = \gamma_i + \tau_i, \qquad D_i = \xi_i + \chi_i, \qquad i = 1, 2, \ldots, n,$$

where

$$\chi_i = \begin{cases} 1 & \text{if } \Delta_n \geq 0 \text{ and } i \in \{i_1, i_2, \ldots, i_{\Delta_n}\}, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and}$$

$$\tau_i = \begin{cases} 1 & \text{if } \Delta_n < 0 \text{ and } i \in \{i_1, i_2, \ldots, i_{|\Delta_n|}\}, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 1. *(i) This algorithm constructs a bi-degree-sequence* $(\mathbf{M}, \mathbf{D})$ *having the property that* $L_n = \sum_{i=1}^n M_i = \sum_{i=1}^n D_i$. *(ii) Note that we have used the capital letters* M_i *and* D_i *to denote the in-degree and out-degree, respectively, of node* i*, as opposed to using the notation* m_i *and* d_i *from Definition 4; we do this to emphasize the randomness of the bi-degree-sequence itself. (iii) Clearly, neither* $\{M_1, \ldots, M_n\}$ *nor* $\{D_1, \ldots, D_n\}$ *are i.i.d. sequences, nor are they independent of each other, but we will show in the next section that asymptotically as* n *grows to infinity they have the same joint distribution as* $(\{\gamma_i\}, \{\xi_i\})$. *(iv) Regarding the condition* $|\Delta_n| \leq n^{1-\kappa+\delta_0}$ *in step 4, we note that it provides a way to ensure that the number of in-degrees or out-degrees that we add in step 5 is negligible with respect to* n*; the polynomial rate at which we are requiring* $|\Delta_n|/n$ *to converge to zero is nevertheless not essential, but it has the advantage of allowing us to keep the calculations throughout the chapter simple. We will show that the probability of satisfying* $|\Delta_n| \leq n^{1-\kappa+\delta_0}$ *converges to one as* n *grows in the following section. (v) Note that we always choose to add degrees, rather than fixing one sequence and always adjust the other one, to avoid having problems with nodes with in- or out-degree zero.*

2.1.2 Asymptotic behavior of the degree sequence

We now provide some results about the asymptotic behavior of the bi-degree-sequence obtained from the algorithm we propose. The first thing we need to prove is that the algorithm will always end in finite time, and the only step where we need to be careful is in step 4, since it may not be obvious that we can always draw two independent i.i.d. sequences satisfying $|\Delta_n| \leq n^{1-\kappa+\delta_0}$ in a reasonable amount of time. The first lemma we give establishes that this is indeed the case by showing that the probability of satisfying condition $|\Delta_n| \leq n^{1-\kappa+\delta_0}$ converges to one as the size of the graph grows. All the proofs in this section can be found in Appendix A.

Lemma 1. *Define* $\mathcal{D}_n = \{|\Delta_n| \leq n^{1-\kappa+\delta_0}\}$, *then*

$$\lim_{n\to\infty} P(\mathcal{D}_n) = 1.$$

We point out that it is possible to construct a bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ such that $|\Delta_n|/n$ converges to zero in probability under the weaker assumption that $\alpha, \beta \geq 1$ and

F and G have finite mean. This weaker condition would also be necessary, since one can construct examples where $\alpha = \beta = 1$ and either F or G have infinite mean, such that Δ_n/n converges in distribution to a non-degenerate random variable. Our condition (2.1) with $\alpha, \beta > 1$ is therefore close to the best possible in terms of ensuring that $|\Delta_n|/n$ converges to zero, and it is necessary to obtain the polynomial rate $n^{-\kappa+\delta_0}$, which greatly simplifies the calculations throughout the chapter.

Since with our proposed construction the sums of the in-degrees and out-degrees are the same, we can always draw a graph, but this is not enough to guarantee that we can draw a simple graph. In other words, we need to determine with what probability will the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ be graphical, and to do this we first need an appropriate criterion, e.g., a directed version of the Erdös-Gallai criterion for undirected graphs. The following result (Corollary 1 on p. 110 in [Berge, 1976]) gives necessary and sufficient conditions for a bi-degree-sequence to be graphical; the original statement is for more general p-graphs, where up to p parallel edges in the same direction are allowed. The notation $|A|$ denotes the cardinality of set A.

Theorem 2. *Given a set of n vertices $V = \{v_1, \ldots, v_n\}$, having bi-degree-sequence $(\mathbf{m}, \mathbf{d}) = (\{m_1, \ldots, m_n\}, \{d_1, \ldots, d_n\})$, a necessary and sufficient condition for $(\mathbf{m}, \mathbf{d})$ to be graphical is*

1. $\sum_{i=1}^{n} m_i = \sum_{i=1}^{n} d_i$, *and*
2. $\sum_{i=1}^{n} \min\{d_i, |A - \{v_i\}|\} \geq \sum_{v_i \in A} m_i$ *for any* $A \subseteq V$.

We now state a result that shows that for large n, the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ constructed in Section 2.1.1 is with high probability graphical. Related results for undirected graphs can be found in [Arratia and Liggett, 2005], which includes the case when the degree distribution has infinite mean.

Theorem 3. *For the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ constructed in Section 2.1.1 we have*

$$\lim_{n\to\infty} P\left((\mathbf{M}, \mathbf{D}) \text{ is graphical}\right) = 1.$$

The second property of $(\mathbf{M}, \mathbf{D})$ that we want to show is that despite the fact that the sequences $\{M_i\}$ and $\{D_i\}$ are no longer independent nor individually i.i.d., they are still asymptotically so as the number of vertices n goes to infinity. The intuition behind this result is that the number of degrees that need to be added to one of the i.i.d. sequences $\{\gamma_i\}$ or $\{\xi_i\}$ to match their sum is small compared to n, and therefore the sequences $\{M_i\}$ and $\{D_i\}$ are *almost* i.i.d. and independent of each other. This feature makes the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ we propose an approximate equivalent of the i.i.d. degree sequence considered in [Arratia and Liggett, 2005; van der Hofstad *et al.*, 2005; Britton *et al.*, 2006] for undirected graphs.

Theorem 4. *The bi-degree-sequence* $(\mathbf{M}, \mathbf{D})$ *constructed in Section 2.1.1 satisfies that for any fixed* $r, s \in \mathbb{N}$*,*

$$(M_{i_1}, \ldots, M_{i_r}, D_{j_1}, \ldots, D_{j_s}) \Rightarrow (\gamma_1, \ldots, \gamma_r, \xi_1, \ldots, \xi_s)$$

as $n \to \infty$*, where* $\{\gamma_i\}$ *and* $\{\xi_i\}$ *are independent sequences of i.i.d. random variables having distributions* F *and* G*, respectively.*

To end this section, we give a result that establishes regularity conditions of the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ which will be important in the sequel.

Proposition 5. *The bi-degree-sequence* $(\mathbf{M}, \mathbf{D})$ *constructed in Section 2.1.1 satisfies*

$$\frac{1}{n}\sum_{k=1}^{n} 1(M_k = i, D_k = j) \xrightarrow{P} f_i g_j, \quad \textit{for all } i, j \in \mathbb{N} \cup \{0\},$$

$$\frac{1}{n}\sum_{i=1}^{n} M_i \xrightarrow{P} E[\gamma_1], \qquad \frac{1}{n}\sum_{i=1}^{n} D_i \xrightarrow{P} E[\xi_1], \qquad \textit{and} \qquad \frac{1}{n}\sum_{i=1}^{n} M_i D_i \xrightarrow{P} E[\gamma_1 \xi_1],$$

as $n \to \infty$*, and provided* $E[\gamma_1^2 + \xi_1^2] < \infty$*,*

$$\frac{1}{n}\sum_{i=1}^{n} M_i^2 \xrightarrow{P} E[\gamma_1^2], \qquad \textit{and} \qquad \frac{1}{n}\sum_{i=1}^{n} D_i^2 \xrightarrow{P} E[\xi_1^2],$$

as $n \to \infty$*.*

2.2 The undirected configuration model

In the previous section we introduced a model for the generation of a bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ that is close to being a pair of independent sequences of i.i.d. random variables, but

yet has the property of being graphical with probability close to one as the size of the graph goes to infinity. We now turn our attention to the problem of obtaining a realization of such sequence, in particular, of drawing a simple graph having $(\mathbf{M}, \mathbf{D})$ as its bi-degree-sequence.

The approach that we follow is a directed version of the *configuration model*. The configuration, or *pairing model*, was introduced in [Bollobás, 1980; Wormald, 1978], although earlier related ideas based on symmetric matrices with $\{0, 1\}$ entries go back to the early 70's; see [Wormald, 1999; Bollobás *et al.*, 2001] for a survey of the history as well as additional references. The configuration model is based on the following idea: given a degree sequence $\mathbf{d} = \{d_1, \ldots, d_n\}$, to each node v_i, $1 \leq i \leq n$, assign d_i stubs or half-edges, and then pair half-edges to form an edge in the graph by randomly selecting with equal probability from the remaining set of unpaired half-edges. This procedure results in a multigraph on n nodes having $\mathbf{d}$ as its degree sequence, where the term multigraph refers to the possibility of self-loops and multiple edges. Although this algorithm does not produce a multigraph uniformly chosen at random from the set of all multigraphs having degree sequence $\mathbf{d}$, a simple graph uniformly chosen at random can be obtained by choosing a pairing uniformly at random and discarding the outcome if it has self-loops or multiple edges [Wormald, 1999]. The question that becomes important then is to estimate the probability with which the pairing model will produce a simple graph. For the undirected graph setting we have described, such results were given in [Bender and Canfield, 1978; Wormald, 1978; Bollobás, 1980; McKay and Wormald, 1991] for regular d-graphs (graphs where each node has exactly degree d), and in [McKay and Wormald, 1990a; McKay and Wormald, 1991; van der Hofstad, 2014] for general graphical degree sequences. From the previous discussion, it should be clear that it is important to determine conditions under which the probability of obtaining a simple graph in the pairing model is bounded away from zero as $n \to \infty$. Such conditions are essentially bounds on the rate of growth of the maximum (minimum) degree and/or the existence of certain limits (see, e.g., [McKay and Wormald, 1990a; McKay and Wormald, 1991; van der Hofstad, 2014]). The set of conditions given below is taken from [van der Hofstad, 2014], and we include it here as a reference for the directed version discussed in this chapter.

Condition 1. *Given a degree sequence* $\mathbf{d} = \{d_1, \ldots, d_n\}$*, let* $D^{[n]}$ *be the degree of a randomly*

chosen node in the corresponding undirected graph, i.e.,

$$P(D^{[n]} = k) = \frac{1}{n}\sum_{i=1}^{n} 1(d_i = k).$$

1. Weak convergence. *There exists a finite random variable D taking values on the positive integers such that*

$$D^{[n]} \Rightarrow D, \qquad n \to \infty.$$

2. Convergence of the first moment.

$$\lim_{n\to\infty} E[D^{[n]}] = E[D].$$

3. Convergence of the second moment.

$$\lim_{n\to\infty} E[(D^{[n]})^2] = E[D^2].$$

Remark 2. *It is straightforward to verify that if the degree sequence is chosen as an i.i.d. sample $\{D_1, \ldots, D_n\}$ from some distribution F on the positive integers having finite first moment, then parts (a) and (b) of Condition 1 are satisfied, and if F has finite second moment then also part (c) is satisfied; the adjustment made to ensure that the sum of the degrees is even, if needed, can be shown to be negligible.*

Condition 1 guarantees that the probability of obtaining a simple graph in the pairing model is bounded away from zero (see, e.g., [van der Hofstad, 2014]), in which case we can obtain a uniformly simple realization of the (graphical) degree sequence $\{d_i\}$ by repeating the random pairing until a simple graph is obtained. When part (c) of Condition 1 fails, then an alternative is to simply erase the self-loops and multiple edges. These two approaches give rise to the *repeated* an *erased* configuration models, respectively.

2.3 The directed configuration model

Having given a brief description of the configuration model for undirected graphs, we will now discuss how to adapt it to draw directed graphs. The idea is basically the same, given a bi-degree-sequence $(\mathbf{m}, \mathbf{d})$, to each node v_i assign m_i inbound half-edges and d_i

outbound half-edges; then, proceed to match inbound half-edges to outbound half-edges to form directed edges. To be more precise, for each unpaired inbound half-edge of node v_i choose randomly from all the available unpaired outbound half-edges, and if the selected outbound half-edge belongs to node, say, v_j, then add a directed edge from v_j to v_i to the graph; proceed in this way until all unpaired inbound half-edges are matched. The following result shows that conditional on the graph being simple, it is uniformly chosen among all simple directed graphs having bi-degree-sequence $(\mathbf{m}, \mathbf{d})$. All the proofs of Section 2.3 can be found in Appendix A.

Proposition 6. *Given a graphical bi-degree-sequence* $(\mathbf{m}, \mathbf{d})$*, generate a directed graph according to the directed configuration model. Then, conditional on the obtained graph being simple, it is uniformly distributed among all simple directed graphs having bi-degree-sequence* $(\mathbf{m}, \mathbf{d})$.

The question is now under what conditions will the probability of obtaining a simple graph be bounded away from zero as the number of nodes, n, goes to infinity. When this probability is bounded away from zero we can repeat the random pairing until we draw a simple graph: the repeated model; otherwise, we can always erase the self-loops and multiple edges in the same direction to obtain a simple graph: the erased model. These two models are discussed in more detail in the following two subsections, where we also provide sufficient conditions under which the probability of obtaining a simple graph will be bounded away from zero.

We end this section by mentioning that another important line of problems related to the drawing of simple graphs (directed or undirected) is the development of efficient simulation algorithms, see for example the recent work in [Blitzstein and Diaconis, 2011] using importance sampling techniques for drawing a simple graph with prescribed degree sequence $\{d_i\}$; similar ideas should also be applicable to the directed model.

2.3.1 Repeated directed configuration model

In this section we analyze the directed configuration model using the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ constructed in Section 2.1.1. In order to do so we will first need to establish sufficient

conditions under which the probability that the directed configuration model produces a simple graph is bounded away from zero as the number of nodes goes to infinity. Since this property does not directly depend on the specific bi-degree-sequence $(\mathbf{M}, \mathbf{D})$, we will prove the result for general bi-degree-sequences $(\mathbf{m}, \mathbf{d})$ satisfying an analogue of Condition 1. As one may expect, we will require the existence of certain limits related to the (joint) distribution of the in-degree and out-degree of a randomly chosen node. Also, since the sequences $\{m_i\}$ and $\{d_i\}$ need to have the same sum, we prefer to consider a sequence of bi-degree-sequences, i.e., $\{(\mathbf{m}_n, \mathbf{d}_n)\}_{n\in\mathbb{N}}$ where $(\mathbf{m}_n, \mathbf{d}_n) = (\{m_{n1}, \ldots, m_{nn}\}, \{d_{n1}, \ldots, d_{nn}\})$, since otherwise the equal sum constraint would greatly restrict the type of sequences we can use. For example, suppose that the bi-degree sequence $(\{m_1, \ldots, m_i\}, \{d_1, \ldots, d_i\})$ satisfies the equal sums condition, then the only possible choice for the $(i+1)$th node would be $m_{i+1} = d_{i+1}$, so a bi-degree-sequence satisfying the equal sums condition would need to have $m_i = d_i$ for all $i \in \mathbb{N}$. Note that for the undirected case the equivalent condition would be to require that the sum of the degrees is always even, a problem that can be avoided by simply ignoring those values of n for which the sum of $\{d_1, \ldots, d_n\}$ is odd (e.g., in the case of i.i.d. degrees, roughly half of the values of n). The use of a sequence of degree sequences rather than a single degree sequence is nevertheless not new, even for undirected graphs (see, e.g., [Molloy and Reed, 1995]).

The corresponding version of Condition 1 for the directed case is given below. We point out that in [Blanchet and Stauffer, 2013], the sufficient and necessary condition of the degree sequence is given for a similar result. We only give a sufficient condition because it already allows us to find a large class of f and g that the repeated model can be used.

Condition 2. *Given a sequence of bi-degree-sequences* $\{(\mathbf{m}_n, \mathbf{d}_n)\}_{n\in\mathbb{N}}$ *satisfying*

$$\sum_{i=1}^{n} m_{ni} = \sum_{i=1}^{n} d_{ni} \qquad \text{for all } n,$$

let $(M^{[n]}, D^{[n]})$ *denote the in-degree and out-degree of a randomly chosen node, i.e.,*

$$P((M^{[n]}, D^{[n]}) = (i, j)) = \frac{1}{n}\sum_{k=1}^{n} 1(m_{nk} = i, d_{nk} = j).$$

1. Weak convergence. *There exist finite random variables* γ *and* ξ *taking values on the*

nonnegative integers and satisfying $E[\gamma] = E[\xi] > 0$ *such that*

$$(M^{[n]}, D^{[n]}) \Rightarrow (\gamma, \xi), \qquad n \to \infty.$$

2. Convergence of the first moments.

$$\lim_{n\to\infty} E[M^{[n]}] = E[\gamma] \qquad \textit{and} \qquad \lim_{n\to\infty} E[D^{[n]}] = E[\xi].$$

3. Convergence of the covariance.

$$\lim_{n\to\infty} E[M^{[n]} D^{[n]}] = E[\gamma\xi].$$

4. Convergence of the second moments.

$$\lim_{n\to\infty} E[(M^{[n]})^2] = E[\gamma^2] \qquad \textit{and} \qquad \lim_{n\to\infty} E[(D^{[n]})^2] = E[\xi^2].$$

We now state a result that says that the number of self-loops and the number of multiple edges produced by the random pairing converge jointly, as $n \to \infty$, to a pair of independent Poisson random variables. As a corollary we obtain that the probability of the resulting graph being simple converges to a positive number, and is therefore bounded away from zero. The proof is an adaptation of the proof of Proposition 7.9 in [van der Hofstad, 2014].

Consider the multigraph obtained through the directed configuration model from the bi-degree-sequence $(\mathbf{m}_n, \mathbf{d}_n)$, and let S_n be the number of self-loops and T_n be the number of multiple edges in the same direction, that is, if there are $k \geq 2$ (directed) edges from node v_i to node v_j, they contribute $(k-1)$ to T_n.

Proposition 7. *(Poisson limit of self-loops and multiple edges) If* $\{(\mathbf{m}_n, \mathbf{d}_n)\}_{n\in\mathbb{N}}$ *satisfies Condition 2 with* $E[\gamma] = E[\xi] = \mu > 0$*, then*

$$(S_n, T_n) \Rightarrow (S, T)$$

as $n \to \infty$*, where* S *and* T *are two independent Poisson random variables with means*

$$\lambda_1 = \frac{E[\gamma\xi]}{\mu} \qquad \textit{and} \qquad \lambda_2 = \frac{E[\gamma(\gamma-1)]E[\xi(\xi-1)]}{2\mu^2},$$

respectively.

Since the probability of the graph being simple is $P(S_n = 0, T_n = 0)$, we obtain as a consequence the following theorem.

Theorem 8. *Under the assumptions of Proposition 7,*

$$\lim_{n\to\infty} P(\text{graph obtained from } (\mathbf{m}_n, \mathbf{d}_n) \text{ is simple}) = e^{-\lambda_1-\lambda_2} > 0.$$

It is clear from Proposition 5 that Condition 2 is satisfied by the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ proposed in Section 2.1.1 whenever F and G have finite variance. This implies that one way of obtaining a simple directed graph on n nodes is by first sampling the bi-degree-sequence $(\mathbf{M}, \mathbf{D})$ according to Section 2.1.1, then checking if it is graphical, and if it is, use the directed pairing model to draw a graph, discarding any realizations that are not simple. Alternatively, since the probability of $(\mathbf{M}, \mathbf{D})$ being graphical converges to one, then one could skip the verification of graphicality and re-sample $(\mathbf{M}, \mathbf{D})$ each time the pairing needs to be repeated. The algorithm is summarized below:

1. Generate bi-degree-sequence according to Section 2.1.1, with F and G having finite variance.
2. (Optional) Verify graphicality using Theorem 2.
3. Randomly pair the in-degrees and out-degrees.
4. If the resulting graph is not simple, repeat from step 3 (or from step 1 if skipping step 2).

The last thing we show in this section is that the degree distributions of the resulting simple graph will have with high probability the prescribed degree distributions F and G, as required. More specifically, if we let $(\mathbf{M}^{(r)}, \mathbf{D}^{(r)})$ be the bi-degree-sequence of the final simple graph obtained through the repeated directed configuration model with bi-degree-sequence $(\mathbf{M}, \mathbf{D})$, then we will show that the joint distribution

$$h^{(n)}(i,j) = \frac{1}{n}\sum_{k=1}^{n} P(M_k^{(r)} = i, D_k^{(r)} = j) \qquad i,j = 0,1,2,\ldots,$$

converges to $f_i g_j$, and the empirical distributions,

$$\widehat{f_k}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} 1(M_i^{(r)} = k) \qquad \text{and} \qquad \widehat{g_k}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} 1(D_i^{(r)} = k) \qquad k = 0,1,2,\ldots,$$

converge in probability to f_k and g_k, respectively. The same result was shown in [Britton *et al.*, 2006] for the undirected case with i.i.d. degree sequence $\{D_i\}$.

Proposition 9. *For the repeated directed configuration model with bi-degree-sequence* $(\mathbf{M}, \mathbf{D})$*, as constructed in Section 2.1.1 with* F *and* G *having finite variance, we have:*

1. $h^{(n)}(i,j) \to f_i g_j$ *as* $n \to \infty$, $i, j = 0, 1, 2, \ldots$, *and*

2. *for all* $k = 0, 1, 2, \ldots$,

$$\widehat{f_k}^{(n)} \xrightarrow{P} f_k \qquad \text{and} \qquad \widehat{g_k}^{(n)} \xrightarrow{P} g_k, \qquad n \to \infty.$$

Remark 3. *Note that by the continuous mapping theorem, (a) implies that the marginal distributions of the in-degrees and out-degrees,*

$$f^{(n)}(i) = \frac{1}{n}\sum_{k=1}^{n} P(M_k^{(r)} = i) \qquad \text{and} \qquad g^{(n)}(j) = \frac{1}{n}\sum_{k=1}^{n} P(D_k^{(r)} = j),$$

converge to f_i *and* g_j*, respectively. The same arguments used in the proof also give that the joint empirical distribution converges to* $f_i g_j$ *in probability.*

2.3.2 Erased directed configuration model

In this section we consider the erased directed configuration model, which is particularly useful when the probability of drawing a simple graph converges to zero as the number of nodes increases, which could happen, for example, when F or G doesn't have finite variance and Condition 2 (d) fails. Given a bi-degree-sequence $(\mathbf{m}, \mathbf{d})$, the erased model consists in first obtaining a multigraph according to the directed configuration model and then erase all self-loops and merge multiple edges in the same direction into a single edge, with the result being a simple graph. Note that the graph obtained through this process no longer has $(\mathbf{m}, \mathbf{d})$ as its bi-degree-sequence. The algorithm is summarized below:

1. Generate bi-degree-sequence according to Section 2.1.1.

2. Randomly pair the in-degrees and out-degrees.

3. Erase self-loops and merge multiple edges in the same direction.

As for the repeated model, let $(\mathbf{M}^{(e)}, \mathbf{D}^{(e)})$ be the bi-degree-sequence of the simple graph obtained through the erased directed configuration model with bi-degree-sequence $(\mathbf{M}, \mathbf{D})$. Define the joint distribution

$$h^{(n)}(i,j) = \frac{1}{n}\sum_{k=1}^{n} P(M_k^{(e)} = i, D_k^{(e)} = j) \qquad i,j = 0,1,2,\ldots,$$

and the empirical distributions,

$$\widehat{f_k}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} 1(M_i^{(e)} = k) \qquad \text{and} \qquad \widehat{g_k}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} 1(D_i^{(e)} = k) \qquad k = 0,1,2,\ldots.$$

The following result is the analogue of Proposition 9 for the erased model; note that in this case we do not require F and G to have finite variance.

Proposition 10. *For the erased directed configuration model with bi-degree-sequence* $(\mathbf{M}, \mathbf{D})$*, as constructed in Section 2.1.1, we have:*

1. $h^{(n)}(i,j) \to f_i g_j$ *as* $n \to \infty$, $i,j = 0,1,2,\ldots,$ *and*

2. *for all* $k = 0,1,2,\ldots,$

$$\widehat{f_k}^{(n)} \xrightarrow{P} f_k \qquad \textit{and} \qquad \widehat{g_k}^{(n)} \xrightarrow{P} g_k, \qquad n \to \infty.$$

Part II

Ranking algorithms

Chapter 3

Ranking algorithms on directed configuration networks

We first give an overview of this chapter. Let $\mathcal{G}_n = (V_n, E_n)$ be a directed graph. We number the nodes $V_n = \{1, 2, \ldots, n\}$ in an arbitrary fashion and let $R_1 =: R_1^{(n)}$ denote the PageRank of node 1, as defined by (1.2). The in-degree of node 1 is then a random variable N_1 picked uniformly at random from the in-degrees of all n nodes in the graph (i.e., from the empirical distribution). Next, we use the notation N_{i+1} to denote the in-degree of the ith inbound neighbor of node 1 (i.e., $(i+1, 1) \in E_n$), and note that although the $\{N_i\}_{i\geq 2}$ have the same distribution, it is not necessarily the same of N_1 since their corresponding nodes implicitly have one or more out-degrees. More precisely, the distribution of the $\{N_i\}_{i\geq 2}$ is an empirical *size-biased* distribution where nodes with high out-degrees are more likely to be chosen. The two distributions can be significantly different when the number of dangling nodes (nodes with zero out-degrees) is a positive fraction of n and their in-degree distribution is different than that of nodes with one or more out-degrees. Similarly, let Q_1 and $\{Q_i\}_{i\geq 2}$ denote the personalization values of node 1 and of its neighbors, respectively, and let $\{C_i\}_{i\geq 2}$ denote the coefficients, or weights, of the neighbors.

As mentioned in the introduction, we will assume throughout the chapter that $\mathcal{G}_n$ is constructed according to the directed configuration model (DCM). To briefly explain the construction of the DCM (the details can be found in Chapter 2 and Section 3.1) consider

a bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n) = \{(N_i, D_i) : 1 \leq i \leq n\}$ of nonnegative integers satisfying $\sum_{i=1}^n N_i = \sum_{i=1}^n D_i$. To draw the graph think of each node, say node i, as having N_i inbound and D_i outbound half-edges or stubs, then pair each of its inbound stubs with a randomly chosen outbound stub from the set of unpaired outbound stubs. The resulting graph is in general what is called a multigraph, i.e., it can have self-loops and multiple edges in the same direction.

Our main result requires us to make some assumptions on the bi-degree sequence used to construct the DCM, as well as on the coefficients $\{C_i\}$ and the personalization values $\{Q_i\}$, which we will refer to as the extended bi-degree sequence. The first set of assumptions (see Assumption 1) requires the existence of certain limits in the spirit of the weak law of large numbers, including $\frac{1}{n}\sum_{i=1}^n D_i^2$ to be bounded in probability (which essentially imposes a finite variance on the out-degrees). This first assumption will ensure the local tree-like structure of the graph. The second set of assumptions (see Assumption 2 in Section **??**) requires the convergence of certain empirical distributions, derived from the extended bi-degree sequence, to proper limits as the graph size goes to infinity. This type of weak convergence assumption is typical in the analysis of random graphs [van der Hofstad, 2014]. We point out that the two sets of assumptions mentioned above are rather weak, and therefore our result is very general. Moreover, as an example, we provide in Section 3.5 an algorithm to generate an extended bi-degree sequence from a set of prescribed distributions that satisfies both assumptions.

To state our main result let $(\mathcal{N}_0, \mathcal{Q}_0)$ and $(\mathcal{N}, \mathcal{Q}, \mathcal{C})$ denote the weak limits of the joint random distributions of (N_1, Q_1) and (N_2, Q_2, C_2), respectively, as defined in Assumption 2. Let $\mathcal{R}$ denote the endogenous solution to the following SFPE:

$$\mathcal{R} \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\mathcal{N}} \mathcal{C}_j \mathcal{R}_j + \mathcal{Q}, \tag{3.1}$$

where $\{\mathcal{R}_i\}$ are i.i.d. copies of $\mathcal{R}$, independent of $(\mathcal{N}, \mathcal{Q}, \{\mathcal{C}_i\})$, and with $\{\mathcal{C}_i\}$ i.i.d. and independent of $(\mathcal{N}, \mathcal{Q})$. Our main result establishes that under the assumptions mentioned above, we have that

$$R_1^{(n)} \Rightarrow \mathcal{R}^*, \qquad n \to \infty,$$

where $\Rightarrow$ denotes weak convergence and $\mathcal{R}^*$ is given by

$$\mathcal{R}^* := \sum_{j=1}^{\mathcal{N}_0} \mathcal{C}_j \mathcal{R}_j + \mathcal{Q}_0, \tag{3.2}$$

where the $\{\mathcal{R}_i\}$ are again i.i.d. copies of $\mathcal{R}$, independent of $(\mathcal{N}_0, \mathcal{Q}_0, \{\mathcal{C}_i\})$, and with $\{\mathcal{C}_i\}$ independent of $(\mathcal{N}_0, \mathcal{Q}_0)$. Thus, $R_1^{(n)}$ is well approximated by a linear combination of endogenous solutions of a SFPE. Here $\mathcal{R}^*$ represents the PageRank of node 1, and the $\mathcal{R}_i$'s represent the PageRank of its inbound neighbors. We give more details on the explicit construction of $\mathcal{R}$ and comment on why it is called the "endogenous" solution in Section 3.4. Furthermore, since $\mathcal{R}$ has been thoroughly studied in the weighted branching processes literature, we can establish the power law behavior of PageRank in a wide class of DCM graphs.

The proof of our main result is given in several steps, each of them requiring a very different type of analysis. For the convenience of the reader, we include in this section a map of these steps.

We start in Section 3.1 by describing the DCM, which on its own does not require any assumptions on the bi-degree sequence. Then, in Section 3.2 we define a class of ranking algorithms, of which PageRank and its various modifications are special cases. These algorithms produce a vector $\mathbf{R}^{(n)}$ that is a solution to a linear system of equations, where the coefficients are the *weights* $\{C_i\}$ assigned to the nodes. For example, in the classical PageRank scenario, we have $C_i = c/D_i$, if $D_i \neq 0$.

The proof of the main result consists of the following three steps:

1. *Finite approximation* (Section 3.2.2). Show that the class of rankings that we study can be approximated in the DCM with any given accuracy by a finite (independent of the graph size n) number of matrix iterations. The DCM plays a crucial role in this step since it implies that the ranks of all the nodes in the graph have the same distribution. A uniform bound on the sequence $\{C_i D_i\}$ is required to provide a suitable rate of convergence.

2. *Coupling with a tree* (Section 3.3). Construct a coupling of the DCM graph and a "thorny branching tree" (TBT). In a TBT each node with the exception of the root

has one outbound link to its parent and possibly several other unpaired outbound links. During the construction, all nodes in both the graph and the tree are also assigned a weight C_i. The main result in this section is the Coupling Lemma 12, which states that the coupling between the graph and the tree will hold for a number of generations in the tree that is logarithmic in n. The locally tree-like property of the DCM and our first set of assumptions (Assumption 1) on the bi-degree sequence are important for this step.

3. *Convergence to a weighted branching process* (Section 3.4). Show that the rank of the root node of the TBT converges weakly to (3.2). This last step requires the weak convergence of the random distributions that define the TBT in the previous step (Assumption 2).

Finally, Section 3.5 gives an algorithm to construct an extended bi-degree sequence satisfying the two main assumptions.

3.1 The directed configuration model

A formal analysis of the directed configuration model (DCM) with given in- and out-degree distributions is recently presented by Chen and Olvera-Cravioto [Chen and Olvera-Cravioto, 2013] (see Chapter 2). In order to analyze the distribution of ranking scores on the DCM we also need other node attributes besides the in- and out-degrees, such as the coefficients and the personalization values. With this in mind we give the following definition.

Definition 6. *We say that the sequence* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n) = \{(N_i, D_i, C_i, Q_i) : 1 \leq i \leq n\}$ *is an* extended bi-degree sequence *if for all* $1 \leq i \leq n$ *it satisfies* $N_i, D_i \in \mathbb{N} = \{0, 1, 2, 3, \dots\}$, $Q_i, C_i \in \mathbb{R}$, *and is such that*

$$L_n := \sum_{i=1}^{n} N_i = \sum_{i=1}^{n} D_i.$$

In this case, we call $(\mathbf{N}_n, \mathbf{D}_n)$ *a* bi-degree sequence.

Formally, the DCM can be defined as follows.

Definition 7. *Let* $(\mathbf{N}_n, \mathbf{D}_n)$ *be a bi-degree sequence and let* $V_n = \{1, 2, \ldots, n\}$ *denote the nodes in the graph. To each node* i *assign* N_i *inbound half-edges and* D_i *outbound half-edges. Enumerate all* L_n *inbound half-edges, respectively outbound half-edges, with the numbers* $\{1, 2, \ldots, L_n\}$*, and let* $\mathbf{x}_n = (x_1, x_2, \ldots, x_{L_n})$ *be a random permutation of these* L_n *numbers, chosen uniformly at random from the possible* $L_n!$ *permutations. The DCM with bi-degree sequence* $(\mathbf{N}_n, \mathbf{D}_n)$ *is the directed graph* $\mathcal{G}_n = (V_n, E_n)$ *obtained by pairing the* x_i*th outbound half-edge with the* i*th inbound half-edge.*

We point out that instead of generating the permutation $\mathbf{x}_n$ of the outbound half-edges up front, one could alternatively construct the graph in a breadth-first fashion, by pairing each of the inbound half-edges, one at a time, with an outbound half-edge, randomly chosen with equal probability from the set of unpaired outbound half-edges. In Section 3.3 we will follow this approach while simultaneously constructing a coupled TBT.

We emphasize that the DCM is, in general, a multi-graph. It was shown in [Chen and Olvera-Cravioto, 2013] that the random pairing of inbound and outbound half-edges results in a simple graph with positive probability provided both the in-degree and out-degree distributions possess a finite variance. In this case, one can obtain a simple realization after finitely many attempts, a method we refer to as the *repeated* DCM, and this realization will be chosen uniformly at random from all simple directed graphs with the given bi-degree sequence. Furthermore, if the self-loops and multiple edges in the same direction are simply removed, a model we refer to as the *erased* DCM, the degree distributions will remain asymptotically unchanged.

For the purposes of this chapter, self-loops and multiple edges in the same direction do not affect the main convergence result for the ranking scores, and therefore we do not require the DCM to result in a simple graph. A similar observation was made in the paper by van der Hofstad et al. [van der Hofstad *et al.*, 2005] when analyzing distances in the undirected CM.

Throughout the chapter, we will use $\mathscr{F}_n = \sigma((\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n))$ to denote the sigma algebra generated by the extended bi-degree sequence, which does not include information about the random pairing. To simplify the notation, we will use $\mathbb{P}_n(\cdot) = P(\cdot|\mathscr{F}_n)$ and $\mathbb{E}_n[\cdot] = E[\cdot|\mathscr{F}_n]$ to denote the conditional probability and conditional expectation, respectively,

given $\mathscr{F}_n$.

3.2 Spectral ranking algorithms

In this section we introduce the class of ranking algorithms that we analyze in this chapter. Following the terminology from [Boldi and Vigna, 2014], these algorithms belong to the class of *spectral centrality measures*, which 'compute the left dominant eigenvector of some matrix derived from the graph'. We point out that the construction of the matrix of weights and the definition of the rank vector that we give in Section 3.2.1 is not particular to the DCM.

3.2.1 Definition of the rank vector

The general class of spectral ranking algorithms we consider are determined by a matrix of weights $M = M(n) \in \mathbb{R}^{n\times n}$ and a personalization vector $\mathbf{Q} \in \mathbb{R}^n$. More precisely, given a directed graph with $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ as its extended bi-degree sequence, we define the (i,j)th component of matrix M as follows:

$$M_{i,j} = \begin{cases} s_{ij}C_i, & \text{if there are } s_{ij} \text{ edges from } i \text{ to } j, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

The rank vector $\mathbf{R} = (R_1, \ldots, R_n)$ is then defined to be the solution to the system of equations

$$\mathbf{R} = \mathbf{R}M + \mathbf{Q}. \tag{3.4}$$

Remark 4. *In the case of the PageRank algorithm, $C_i = c/D_i$, $Q_i = 1 - c$ for all i, and the constant $0 < c < 1$ is the so-called damping factor.*

3.2.2 Finitely many iterations

To solve the system of equations given in (3.4) we proceed via matrix iterations [Langville and Meyer, 2011]. To initialize the process let $\mathbf{1}$ be the (row) vector of ones in $\mathbb{R}^n$ and let $\mathbf{r}_0 = r_0\mathbf{1}$, with $r_0 \in \mathbb{R}$. Define

$$\mathbf{R}^{(n,0)} = \mathbf{r}_0,$$

and for $k \geq 1$,

$$\mathbf{R}^{(n,k)} = \mathbf{r}_0 M^k + \sum_{i=0}^{k-1} \mathbf{Q} M^i.$$

With this notation, we have that the solution $\mathbf{R}$ to (3.4), provided it exists, can be written as

$$\mathbf{R} = \mathbf{R}^{(n,\infty)} = \sum_{i=0}^{\infty} \mathbf{Q} M^i.$$

We are interested in analyzing a randomly chosen coordinate of the vector $\mathbf{R}^{(n,\infty)}$. The first step is to show that we can do so by using only finitely many matrix iterations. To this end note that

$$\mathbf{R}^{(n,k)} - \mathbf{R}^{(n,\infty)} = \mathbf{r}_0 M^k - \sum_{i=k}^{\infty} \mathbf{Q} M^i = \left(\mathbf{r}_0 - \sum_{i=0}^{\infty} \mathbf{Q} M^i \right) M^k.$$

Moreover,

$$\left|\left| \mathbf{R}^{(n,k)} - \mathbf{R}^{(n,\infty)} \right|\right|_1 \leq \left|\left| \mathbf{r}_0 M^k \right|\right|_1 + \sum_{i=0}^{\infty} \left|\left| \mathbf{Q} M^{k+i} \right|\right|_1 .$$

Next, note that for any row vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)$,

$$\begin{aligned} ||\mathbf{y} M^r||_1 &\leq \sum_{j=1}^{n} |\mathbf{y} (M^r)_{\bullet j}| \leq \sum_{j=1}^{n} \sum_{i=1}^{n} |y_i (M^r)_{ij}| \\ &= \sum_{i=1}^{n} |y_i| \sum_{j=1}^{n} |(M^r)_{ij}| = \sum_{i=1}^{n} |y_i| \cdot ||M^r_{i\bullet}||_1 \\ &\leq ||\mathbf{y}||_1 \, ||M^r||_\infty , \end{aligned}$$

where $A_{i\bullet}$ and $A_{\bullet j}$ are the ith row and jth column, respectively, of matrix A, and $||A||_\infty = \max_{1 \leq i \leq n} ||A_{i\bullet}||_1$ is the operator infinity norm. It follows that if we assume that $\max_{1 \leq i \leq n} |C_i| D_i \leq c$ for some $c \in (0, 1)$, then we have

$$||M^r||_\infty \leq ||M||_\infty^r = \left(\max_{1 \leq i \leq n} |C_i| D_i \right)^r \leq c^r.$$

In this case we conclude that

$$\begin{aligned} \left|\left| \mathbf{R}^{(n,k)} - \mathbf{R}^{(n,\infty)} \right|\right|_1 &\leq ||\mathbf{r}_0||_1 c^k + \sum_{i=0}^{\infty} ||\mathbf{Q}||_1 c^{k+i} \\ &= |r_0| n c^k + ||\mathbf{Q}||_1 \frac{c^k}{1-c}. \end{aligned}$$

Now note that all the coordinates of the vector $\mathbf{R}^{(n,k)} - \mathbf{R}^{(n,\infty)}$ have the same distribution, since by construction, the configuration model makes all permutations of the nodes' labels equally likely. Hence, the randomly chosen node may as well be the first node, and the error that we make by considering only finitely many iterations in its approximation is bounded in expectation by

$$\begin{aligned}\mathbb{E}_n\left[\left|R_1^{(n,k)} - R_1^{(n,\infty)}\right|\right] &= \frac{1}{n}\mathbb{E}_n\left[\left|\left|\mathbf{R}^{(n,k)} - \mathbf{R}^{(n,\infty)}\right|\right|_1\right] \\ &\leq |r_0|c^k + \mathbb{E}_n\left[||\mathbf{Q}||_1\right]\frac{c^k}{n(1-c)} \\ &= \left(|r_0| + \frac{1}{n(1-c)}\sum_{i=1}^{n}|Q_i|\right)c^k.\end{aligned}$$

It follows that if we let

$$B_n = \left\{\max_{1\leq i\leq n}|C_i|D_i \leq c,\ \frac{1}{n}\sum_{i=1}^{n}|Q_i| \leq H\right\} \tag{3.5}$$

for some constants $c \in (0,1)$ and $H < \infty$, then Markov's inequality yields

$$\begin{aligned}&P\left(\left|R_1^{(n,k)} - R_1^{(n,\infty)}\right| > n^{-\epsilon}\middle| B_n\right) \\ &= \frac{1}{P(B_n)}E\left[1(B_n)\mathbb{E}_n\left[1\left(\left|R_1^{(n,k)} - R_1^{(n,\infty)}\right| > n^{-\epsilon}\right)\right]\right] \\ &\leq \frac{1}{P(B_n)}E\left[1(B_n)n^{\epsilon}\mathbb{E}_n\left[\left|R_1^{(n,k)} - R_1^{(n,\infty)}\right|\right]\right] \\ &\leq \left(|r_0| + \frac{1}{1-c}E\left[\frac{1}{n}\sum_{i=1}^{n}|Q_i|\middle| B_n\right]\right)n^{\epsilon}c^k \\ &\leq \left(|r_0| + \frac{H}{1-c}\right)n^{\epsilon}c^k.\end{aligned} \tag{3.6}$$

We have thus derived the following result.

Proposition 11. *Consider the directed configuration graph generated by the extended bi-degree sequence* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ *and let* B_n *be defined according to* (3.5)*. Then, for any* $x_n \to \infty$ *and any* $k \geq 1$*, we have*

$$P\left(\left|R_1^{(n,\infty)} - R_1^{(n,k)}\right| > x_n^{-1}\middle| B_n\right) = O\left(x_n c^k\right)$$

as $n \to \infty$*.*

This completes the first step of our approach. In the next section we will explain how to couple the graph, as seen from a randomly chosen node, with an appropriate branching tree.

3.3 Construction of the graph and coupling with a branching tree

The next step in our approach is to approximate the distribution of $R_1^{(n,k)}$ with the rank of the root node of a suitably constructed branching tree. To ensure that we can construct such a tree we require the extended bi-degree sequence to satisfy some further properties with high probability. These properties are summarized in the following assumption.

Assumption 1. *Let* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ *be an extended bi-degree sequence for which there exists constants* $H, \nu_i > 0$, $i = 1, \ldots, 5$, *with*

$$\mu := \nu_2/\nu_1, \quad \lambda := \nu_3/\nu_1 \qquad \text{and} \qquad \rho := \nu_5\mu/\nu_1 < 1,$$

$0 < \kappa \leq 1$, *and* $0 < c, \gamma, \epsilon < 1$ *such that the events*

$$\begin{aligned}
\Omega_{n,1} &= \left\{ \left| \sum_{r=1}^{n} D_r - n\nu_1 \right| \leq n^{1-\gamma} \right\}, \\
\Omega_{n,2} &= \left\{ \left| \sum_{r=1}^{n} D_r N_r - n\nu_2 \right| \leq n^{1-\gamma} \right\}, \\
\Omega_{n,3} &= \left\{ \left| \sum_{r=1}^{n} D_r^2 - n\nu_3 \right| \leq n^{1-\gamma} \right\}, \\
\Omega_{n,4} &= \left\{ \left| \sum_{r=1}^{n} D_r^{2+\kappa} - n\nu_4 \right| \leq n^{1-\gamma} \right\}, \\
\Omega_{n,5} &= \left\{ \left| \sum_{r=1}^{n} |C_r| D_r - n\nu_5 \right| \leq n^{1-\gamma}, \max_{1 \leq r \leq n} |C_r| D_r \leq c \right\}, \\
\Omega_{n,6} &= \left\{ \sum_{r=1}^{n} |Q_r| \leq Hn \right\},
\end{aligned}$$

satisfy as $n \to \infty$,

$$P\left(\Omega_n^c\right) = P\left(\left(\bigcap_{i=1}^{6} \Omega_{n,i}\right)^c\right) = O\left(n^{-\epsilon}\right).$$

It is clear from (3.5) that $\Omega_n \subseteq B_n$, hence Proposition 11 holds under Assumption 1. We also point out that all six conditions in the assumption are in the spirit of the Weak Law of Large Numbers, and are therefore general enough to be satisfied by many different

constructions of the extended bi-degree sequence. As an example, we give in Section 3.5 an algorithm based on sequences of i.i.d. random variables that satisfies Assumption 1.

In Sections 3.3.1–3.3.4 we describe in detail how to construct a coupling of the directed graph $\mathcal{G}_n$ and its approximating branching tree. We start by explaining the terminology and notation in Section 3.3.1, followed by the construction itself in Section 3.3.2. Then, in Section 3.3.3 we present the Coupling Lemma 12, which is the main result of Section 3.3. Finally, Section 3.3.4 explains how to compute the rank of the root node in the coupled tree.

3.3.1 Terminology and notation

Throughout the remainder of the chapter we will interchangeably refer to the $\{N_i\}$ as the in-degrees/number of offspring/number of inbound stubs, to the $\{D_i\}$ as the out-degrees/number of outbound links/number of outbound stubs, to the $\{C_i\}$ as the weights, and to the $\{Q_i\}$ as the personalization values. We will refer to these four characteristics of a node as the *node attributes*.

The fact that we are working with a directed graph combined with the presence of weights, means that we need to use a more general kind of tree in our coupling than the standard branching process typically used in the random graph literature. To this end, we will define a process we call a Thorny Branching Tree (TBT), where each individual (node) in the tree has a directed edge pointing towards its parent, and also a certain number of unpaired outbound links (pointing, say, to an artificial node outside of the tree). The name 'thorny' is due to these unpaired outbound links, see Figure 3.1. We point out that the structure of the tree (i.e., parent-offspring relations) is solely determined by the number of offspring.

The simpler structure of a tree compared to a general graph allows for a more natural enumeration of its nodes. As usually in the context of branching processes, we let each node in the TBT have a label of the form $\mathbf{i} = (i_1, i_2, \ldots, i_k) \in \mathcal{U}$, where $\mathcal{U} = \bigcup_{k=0}^{\infty} (\mathbb{N}_+)^k$ is the set of all finite sequences of positive integers. Here, the convention is that $\mathbb{N}_+^0 = \{\emptyset\}$ contains the null sequence $\emptyset$. Also, for $\mathbf{i} = (i_1)$ we simply write $\mathbf{i} = i_1$, that is, without the parenthesis. Note that this form of enumeration gives the complete lineage of each

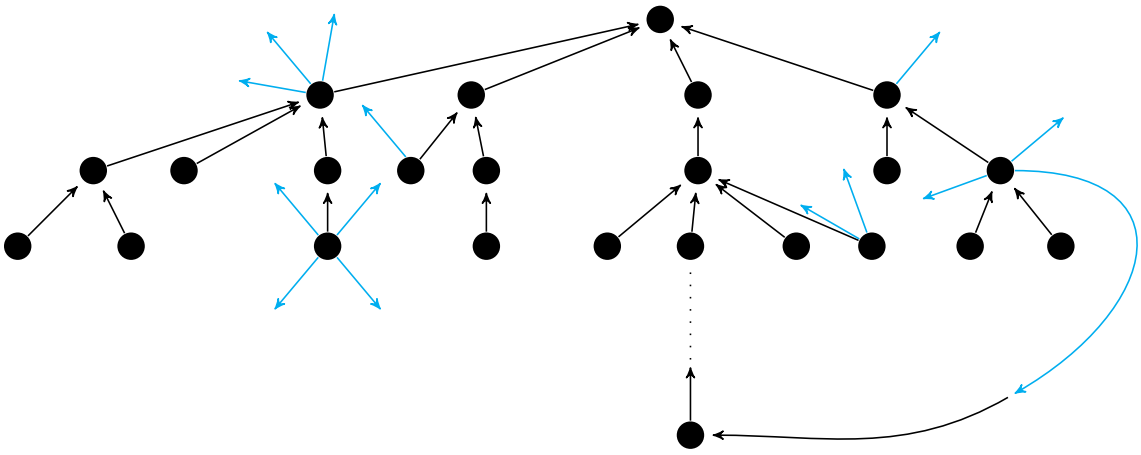

Figure 3.1: Graph construction process. Unpaired outbound links are in blue.

individual in the tree.

We will use the following terminology and notation throughout the chapter.

Definition 8. *We say that a node i in the graph (resp. TBT) is at distance k of the first (resp. root) node if it can reach the first (resp. root) node in k steps, but not in any less than k steps.*

In addition, for $r \geq 0$, we define on the graph/tree the following processes:

- A_r: set of nodes in the graph at distance r of the first node.
- $\hat{A}_r$: set of nodes in the tree at distance r of the root node ($\hat{A}_r$ is also the set of nodes in the rth generation of TBT, with the root node being generation zero).
- Z_r: number of inbound stubs of all the nodes in the graph at distance r of the first node ($Z_r \geq |A_{r+1}|$).
- $\hat{Z}_r$: number of inbound stubs of all the nodes in generation r of the TBT ($\hat{Z}_r = |\hat{A}_{r+1}|$).
- V_r: number of outbound stubs of all the nodes in the graph at distance r of the first node.
- $\hat{V}_r$: number of outbound stubs of all the nodes in generation r of the TBT.

Finally, given the extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$, we introduce two empirical distributions that will be used in the construction of the coupling. The first one

describes the attributes of a randomly chosen node:

$$f_n^*(i,j,s,t) = \sum_{k=1}^{n} 1(N_k = i, D_k = j, C_k = s, Q_k = t)\mathbb{P}_n(\text{node } k \text{ is sampled})$$
$$= \frac{1}{n}\sum_{k=1}^{n} 1(N_k = i, D_k = j, C_k = s, Q_k = t). \tag{3.7}$$

The second one, corresponds to the attributes of a node that is chosen by sampling uniformly at random from all the L_n outbound stubs:

$$f_n(i,j,s,t) = \sum_{k=1}^{n} 1(N_k = i, D_k = j, C_k = s, Q_k = t)\mathbb{P}_n(\text{an outbound stub from node } k \text{ is sampled})$$
$$= \sum_{k=1}^{n} 1(N_k = i, D_k = j, C_k = s, Q_k = t)\frac{D_k}{L_n}. \tag{3.8}$$

Note that this is a size-biased distribution, since nodes with more outbound stubs are more likely to be chosen, whereas nodes with no outbound stubs (dangling nodes) cannot be chosen.

3.3.2 Construction of the coupling

Given an extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ we now explain how to construct the graph $\mathcal{G}_n$ and its coupled TBT through a breadth-first exploration process. From this point onwards we will ignore the implicit numbering of the nodes in the definition of the extended bi-degree sequence and rename them according to the order in which they appear in the graph exploration process.

To keep track of which outbound stubs have already been matched we borrow the approach used in [van der Hofstad *et al.*, 2005] and label them 1, 2, or 3 according to the following rules:

1. Outbound stubs with label 1 are stubs belonging to a node that is not yet attached to the graph.
2. Outbound stubs with label 2 belong to nodes that are already part of the graph but that have not yet been paired with an inbound stub.
3. Outbound stubs with label 3 are those which have already been paired with an inbound stub and now form an edge in the graph.

The graph $\mathcal{G}_n$ is constructed as follows. Right before the first node is sampled, all outbound stubs are labeled 1. To start the construction of the graph, we choose randomly a node (all nodes with the same probability) and call it node 1. The attributes of this first node, denoted by (N_1, D_1, C_1, Q_1), are sampled from distribution (3.7).

After the first node is chosen, its D_1 outbound stubs are labeled 2. We then proceed to pair the first of the $Z_0 = N_1$ inbound stubs of the first node with a randomly chosen outbound stub. The corresponding node is attached to the graph by forming an edge pointing to node 1 using the chosen outbound stub, which receives a label 3, and all the remaining outbound stubs from the new node are labeled 2. Note that it is possible that the chosen node is node 1 itself, in which case the pairing forms a self-loop and no new nodes are added to the graph. We continue in this way until all Z_0 inbound stubs of node 1 have been paired with randomly chosen outbound stubs. Since these outbound stubs are sampled independently and with replacement from all the possible L_n outbound stubs, this corresponds to drawing the node attributes independently from the random distribution (3.8). Note that in the construction of the graph any unfeasible matches will be discarded, and therefore the attributes of nodes in $\mathcal{G}_n$ do not necessarily have distribution (3.8), but rather have the conditional distribution given the pairing was feasible. We will use the vector (N_i, D_i, C_i, Q_i) to denote the attributes of the ith node to be added to the graph.

In general, the kth iteration of this process is completed when all Z_{k-1} inbound stubs have been matched with an outbound stub, and the corresponding node attributes have been assigned. The process ends when all L_n inbound stubs have been paired. Note that whenever an outbound stub with label 2 is chosen a cycle or a double edge is formed in the graph.

Next, we explain how the TBT is constructed. To distinguish the attribute vectors of nodes in the TBT from those of nodes in the graph, we denote them by $(\hat{N}_{\mathbf{i}}, \hat{D}_{\mathbf{i}}, \hat{C}_{\mathbf{i}}, \hat{Q}_{\mathbf{i}})$, $\mathbf{i} \in \mathcal{U}$. We start with the root node (node $\emptyset$) that has the same attributes as node 1 in the graph: $(\hat{N}_\emptyset, \hat{D}_\emptyset, \hat{C}_\emptyset, \hat{Q}_\emptyset) \equiv (N_1, D_1, C_1, Q_1)$, sampled from distribution (3.7). Next, for $k \geq 1$, each of the $\hat{Z}_{k-1}$ individuals in the kth generation will independently have offspring, outbound stubs, weight and personalization value according to the joint distribution $f_n(i, j, s, t)$ given by (3.8).

Now, we explain how the coupling with the graph, i.e., the simultaneous construction of the graph and the TBT, is done.

1) Whenever an outbound stub is sampled randomly in an attempt to add an edge to $\mathcal{G}_n$, then, independently of the stub's label, a new offspring is added to the TBT. This is done to maintain the branching property (i.i.d. node attributes). In particular, if the chosen outbound stub belongs to node j, then the new offspring in the TBT will have $D_j - 1$ outbound stubs (which will remain unpaired), N_j inbound stubs (number of offspring), weight C_j, and personalization value Q_j.

2) If an outbound stub with label 1 is chosen, then both the graph and the TBT will connect the chosen outbound stub to the inbound stub being matched, resulting in a node being added to the graph and an offspring being born to its parent. We then update the labels by giving a 2 label to all the 'sibling' outbound stubs of the chosen outbound stub, and a 3 label to the chosen outbound stub itself.

3) If an outbound stub with label 2 is chosen it means that its corresponding node already belongs to the graph, and a cycle, self-loop, or multiple edge is created. We then relabel the chosen outbound stub with a 3. An offspring is born in the TBT according to 1).

4) If an outbound stub with label 3 is chosen it means that the chosen outbound stub has already been matched. In terms of the construction of the graph, this case represents a failed attempt to match the current inbound stub, and we have to keep sampling until we draw an outbound stub with label 1 or 2. Once we do so, we update the labels according to the rules given above. An offspring is born in the TBT according to 1).

Note that as long as we do not sample any outbound stub with label 2 or 3, the graph $\mathcal{G}_n$ and the TBT are identical. Once we draw the first outbound stub with label 2 or 3 the processes Z_k and $\hat{Z}_k$ may start to disagree. The moment this occurs we say that the coupling has been broken. Nonetheless, we will continue with the pairing process following the rules given above until all L_n inbound stubs have been paired. The construction of the

TBT also continues in parallel by keeping the synchronization of the pairing whenever the inbound stub being matched belongs to a node that is both in the graph and the tree. If the pairing of all L_n inbound stubs is completed after k iterations of the process, then we will have completed k generations in the TBT. Moreover, up to the time the coupling breaks, a node $\mathbf{i} \in \hat{A}_k$ is also the jth node to be added to the graph, where:

$$j = 1 + \sum_{r=0}^{k-2} \hat{Z}_r + \sum_{s=1}^{i_{k-1}-1} \hat{N}_{(i_1,\ldots,i_{k-2},s)} + i_k,$$

with the convention that $\sum_{r=a}^{b} x_r = 0$ if $b < a$.

Definition 9. *Let τ be the number of generations in the TBT that can be completed before the first outbound stub with label 2 or 3 is drawn, i.e., $\tau = k$ if and only if the first inbound stub to draw an outbound stub with label 2 or 3 belonged to a node $\mathbf{i} \in \hat{A}_k$.*

The main result in this section consists in showing that provided the extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ satisfies Assumption 1, the coupling breaks only after a number of generations that is of order $\log n$, which combined with Proposition 11 will allow us to approximate the rank of a randomly chosen node in the graph with the rank of the root node of the coupled TBT.

3.3.3 The coupling lemma

It follows from the construction in Section 3.3.2 that, before the coupling breaks, the neighborhood of node 1 in $\mathcal{G}_n$ and of the root node in the TBT are identical. Recall also from Proposition 11 that we only need a finite number k of matrix iterations to approximate the elements of the rank vector to any desired precision. Furthermore, the weight matrix M is such that the elements $(M^r)_{i,1}$, $1 \leq i \leq n$, $1 \leq r \leq k$, depend only on the k-neighborhood of node 1. Hence, if the coupling holds for $\tau > k$ generations, then the rank score of node 1 in $\mathcal{G}_n$ is exactly the same as that of the root node of the TBT restricted to those same k generations. The following coupling lemma will allow us to complete the appropriate number of generations in the tree to obtain the desired level of precision in Proposition 11. Its proof is rather technical and is therefore postponed to Appendix B.1.

Lemma 12. *Suppose $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ satisfies Assumption 1. Then,*

- *for any* $1 \leq k \leq h \log n$ *with* $0 < h < 1/(2 \log \mu)$, *if* $\mu > 1$,
- *for any* $1 \leq k \leq n^b$ *with* $0 < b < \min\{1/2, \gamma\}$, *if* $\mu \leq 1$,

we have

$$P\left(\tau \leq k | \, \Omega_n\right) = \begin{cases} O\left((n/\mu^{2k})^{-1/2}\right), & \mu > 1, \\ O\left((n/k^2)^{-1/2}\right), & \mu = 1, \\ O\left(n^{-1/2}\right), & \mu < 1, \end{cases}$$

as $n \to \infty$.

Remark 5. *The constant* μ *was defined in Assumption 1, and it corresponds to the limiting expected number of offspring that each node in the TBT (with the exception of the root node) will have. The coupling between the graph and the TBT will hold for any* $\mu > 0$.

We conclude from Lemma 12 that if $\hat{R}^{(n,k)} := \hat{R}_{\emptyset}^{(n,k)}$ denotes the rank of the root node of the TBT restricted to the first k generations, then, for any $\delta > 0$,

$$P\left(\left|R_1^{(n,k)} - \hat{R}^{(n,k)}\right| > n^{-\delta} \Big| \, \Omega_n\right) \leq P(\tau < k | \Omega_n) := \varphi(k, n).$$

Note that the super index n does not refer to the number of nodes in the tree, and is being used only in the definition of the distributions f_n^* and f_n (given in (3.7) and (3.8), respectively).

This observation, combined with Proposition 11, implies that if we let $k_n = \lceil h \log n \rceil$, when $\mu > 1$, and $k_n = n^{\varepsilon}$, when $\mu \leq 1$, where $h = (1-\varepsilon)/(2 \log \mu)$ and $0 < \varepsilon < \min\{1/3, \gamma\}$, then

$$\begin{aligned} P\left(\left|R_1^{(n,\infty)} - \hat{R}^{(n,k_n)}\right| > n^{-\delta} \Big| \, \Omega_n\right) &\leq P\left(\left|R_1^{(n,\infty)} - R_1^{(n,k_n)}\right| > n^{-\delta}/2 \Big| \, \Omega_n\right) \\ &\quad + P\left(\left|R_1^{(n,k_n)} - \hat{R}^{(n,k_n)}\right| > n^{-\delta}/2 \Big| \, \Omega_n\right) \\ &= O\left(n^{\delta} c^{k_n} + \varphi(k_n, n)\right) \\ &= O\left(n^{\delta - h|\log c|} + n^{-\varepsilon/2}\right). \end{aligned} \tag{3.9}$$

In view of (3.9), analyzing the distribution of $R_1^{(n,k)}$ in the graph reduces to analyzing the rank of the root node of the coupled TBT, $\hat{R}^{(n,k)}$. In the next section, we compute $\hat{R}^{(n,k)}$ by relating it to a linear process constructed on the TBT.

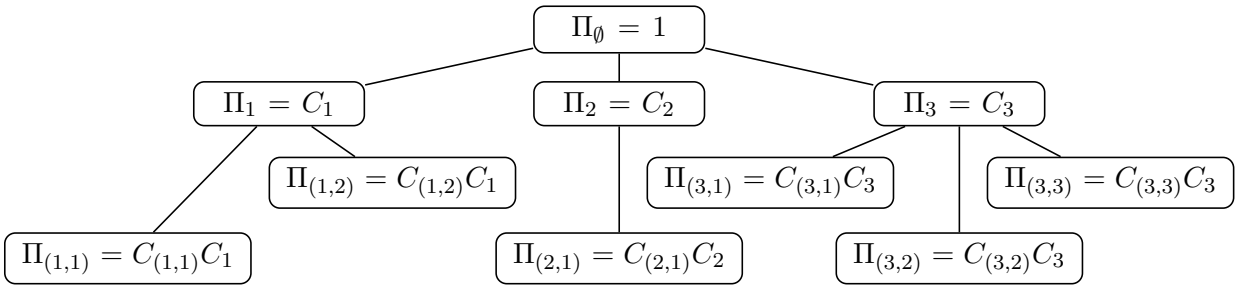

Figure 3.2: Weighted branching tree.

3.3.4 Computing the rank of nodes in the TBT

In order to compute $\hat{R}^{(n,k)}$ we need to introduce a new type of weights. To simplify the notation, for $\mathbf{i} = (i_1, \ldots, i_k)$ we will use $(\mathbf{i}, j) = (i_1, \ldots, i_k, j)$ to denote the index concatenation operation; if $\mathbf{i} = \emptyset$, then $(\mathbf{i}, j) = j$. Each node $\mathbf{i}$ is then assigned a weight $\hat{\Pi}_\mathbf{i}$ according to the recursion

$$\hat{\Pi}_\emptyset \equiv 1 \qquad \text{and} \qquad \hat{\Pi}_{(\mathbf{i},j)} = \hat{\Pi}_\mathbf{i} \hat{C}_{(\mathbf{i},j)}, \quad \mathbf{i} \in \mathcal{U}.$$

Note that the $\hat{\Pi}_\mathbf{i}$'s are the products of all the weights $\hat{C}_\mathbf{j}$ along the path leading to node $\mathbf{i}$, as depicted in Figure 3.2.

Next, for each fixed $k \in \mathbb{N}$ and each node $\mathbf{i}$ in the TBT define $\hat{R}^{(n,k)}_\mathbf{i}$ to be the rank of node $\mathbf{i}$ computed on the subtree that has $\mathbf{i}$ as its root and that is restricted to having only k generations, with each of the $|\hat{A}_k|$ nodes having rank r_0. In mathematical notation,

$$\hat{R}^{(n,k)}_\mathbf{i} = \sum_{j=1}^{\hat{N}_\mathbf{i}} \hat{C}_{(\mathbf{i},j)} \hat{R}^{(n,k-1)}_{(\mathbf{i},j)} + \hat{Q}_\mathbf{i}, \qquad k \geq 1, \qquad \hat{R}^{(n,0)}_\mathbf{j} = r_0. \tag{3.10}$$

Iterating (3.10) gives

$$\begin{aligned} \hat{R}^{(n,k)} &= \sum_{\mathbf{i} \in \hat{A}_1} \hat{\Pi}_\mathbf{i} \hat{R}^{(n,k-1)}_\mathbf{i} + \hat{Q}_\emptyset = \sum_{\mathbf{i} \in \hat{A}_1} \hat{\Pi}_\mathbf{i} \left(\sum_{j=1}^{\hat{N}_\mathbf{i}} \hat{C}_{(\mathbf{i},j)} \hat{R}^{(n,k-2)}_{(\mathbf{i},j)} + \hat{Q}_\mathbf{i} \right) + \hat{Q}_\emptyset \\ &= \sum_{\mathbf{i} \in \hat{A}_2} \hat{\Pi}_\mathbf{i} \hat{R}^{(n,k-2)}_\mathbf{i} + \sum_{\mathbf{i} \in \hat{A}_1} \hat{\Pi}_\mathbf{i} \hat{Q}_\mathbf{i} + \hat{Q}_\emptyset = \cdots = \sum_{\mathbf{i} \in \hat{A}_k} \hat{\Pi}_\mathbf{i} r_0 + \sum_{s=0}^{k-1} \sum_{\mathbf{i} \in \hat{A}_s} \hat{\Pi}_\mathbf{i} \hat{Q}_\mathbf{i}. \end{aligned} \tag{3.11}$$

The last step in our proof of the main result is to identify the limit of $\hat{R}^{(n,k_n)}$ as $n \to \infty$, for a suitable chosen $k_n \to \infty$. This is done in the next section.

3.4 Coupling with a weighted branching process

The last step in the derivation of our approximation for the rank of a randomly chosen node in the graph $\mathcal{G}_n$ is to substitute the rank of the root node in the TBT, which is defined with respect to empirical distributions based on the extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$, with a limiting random variable independent of the size of the graph, n.

The appropriate limit will be given in terms of a solution to a certain stochastic fixed-point equation (SFPE). The appeal of having such a representation is that these solutions have been thoroughly studied in the WBPs literature, and in many cases exact asymptotics describing their tail behavior are available [Jelenković and Olvera-Cravioto, 2010; Jelenković and Olvera-Cravioto, 2012a; Olvera-Cravioto, 2012b]. We will elaborate more on this point after we state our main result.

Our main result shows that

$$R_1^{(n,\infty)} \Rightarrow \mathcal{R}^*$$

as $n \to \infty$, where $\mathcal{R}^*$ can be written in terms of the so-called endogenous solution to a linear SFPE. Before we write the expression for $\mathcal{R}^*$ we will need to introduce a few additional concepts.

3.4.1 Weighted branching processes

We first briefly describe what we will refer to as a weighted branching process. Then we will also explain how to construct a variation of this process that appears in the analysis of random graphs, and that will require a somewhat different treatment.

We start by letting $\mathbb{N}_+ = \{1, 2, 3, \dots\}$ be the set of positive integers and setting $U = \bigcup_{k=0}^{\infty} (\mathbb{N}_+)^k$ to be the set of all finite sequences $\mathbf{i} = (i_1, i_2, \dots, i_n)$, $n \geq 0$, where by convention $\mathbb{N}_+^0 = \{\emptyset\}$ contains the null sequence $\emptyset$. To ease the exposition, for a sequence $\mathbf{i} = (i_1, i_2, \dots, i_k) \in U$ we write $\mathbf{i}|n = (i_1, i_2, \dots, i_n)$, provided $k \geq n$, and $\mathbf{i}|0 = \emptyset$ to denote the index truncation at level n, $n \geq 0$. Also, for $\mathbf{i} \in A_1$ we simply use the notation $\mathbf{i} = i_1$, that is, without the parenthesis. Similarly, for $\mathbf{i} = (i_1, \dots, i_n)$ we will use $(\mathbf{i}, j) = (i_1, \dots, i_n, j)$ to denote the index concatenation operation, if $\mathbf{i} = \emptyset$, then $(\mathbf{i}, j) = j$.

Next, let $(Q, N, C_1, C_2, \dots)$ be a real-valued vector with $N \in \mathbb{N} \cup \{\infty\}$. We will refer

to this vector as the generic branching vector. Now let $\{(Q_{\mathbf{i}}, N_{\mathbf{i}}, C_{(\mathbf{i},1)}, C_{(\mathbf{i},2)}, \dots)\}_{\mathbf{i}\in U}$ be a sequence of i.i.d. copies of the generic branching vector. To construct a weighted branching process we start by defining a tree as follows: let $A_0 = \{\emptyset\}$ denote the root of the tree, and define the nth generation according to the recursion

$$A_n = \{(\mathbf{i}, i_n) \in U : \mathbf{i} \in A_{n-1}, 1 \le i_n \le N_{\mathbf{i}}\}, \quad n \ge 1.$$

Now, assign to each node $\mathbf{i}$ in the tree a weight $\Pi_{\mathbf{i}}$ according to the recursion

$$\Pi_{\emptyset} \equiv 1, \qquad \Pi_{(\mathbf{i},i_n)} = C_{(\mathbf{i},i_n)}\Pi_{\mathbf{i}}, \qquad n \ge 1,$$

see Figure 3.2. Note that the tree's structure, disregarding the weights, is a Galton-Watson process with offspring distribution $f(k) = P(N = k)$, provided $P(N < \infty) = 1$.

Using the same notation described above, consider now constructing this process using a generic branching vector of the form (Q, N, C), with $N \in \mathbb{N}$, and a sequence of i.i.d. copies $\{(Q_{\mathbf{i}}, N_{\mathbf{i}}, C_{\mathbf{i}})\}_{\mathbf{i}\in U}$. As mentioned earlier, we will refer to this construction as a weighted branching tree. The difference lies in the dependence structure that now governs the nodes in the tree, since whereas in a usual weighted branching process the weight $C_{\mathbf{i}}$ of node $\mathbf{i}$ is independent of $(Q_{\mathbf{i}}, N_{\mathbf{i}})$, in a weighted branching tree it may not be. Another important observation is that in a weighted branching tree the weights $\{C_{\mathbf{i}}\}_{\mathbf{i}\in U}$ are i.i.d. random variables, unlike in a weighted branching process where the weights of "sibling" nodes are arbitrarily dependent and not necessarily identically distributed. It follows from these observations that when C is independent of (Q, N), the corresponding weighted branching tree is a special case of a weighted branching process.

We will now explain how to construct the endogenous solution to the linear SFPE

$$R \overset{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q, \tag{3.12}$$

using a weighted branching process.

3.4.1.1 The endogenous solution to the linear SFPE

For a weighted branching process with generic branching vector $(Q, N, C_1, C_2, \dots)$, define the processes $\{W^{(j)} : j \geq 0\}$ and $\{R^{(k)} : k \geq 0\}$ as follows:

$$W^{(0)} = Q_0, \qquad W^{(j)} = \sum_{\mathbf{i} \in A_j} Q_{\mathbf{i}} \Pi_{\mathbf{i}}, \quad j \geq 1, \tag{3.13}$$

$$R^{(k)} = \sum_{j=0}^{k} W^{(j)} = \sum_{j=0}^{k} \sum_{\mathbf{i} \in A_j} Q_{\mathbf{i}} \Pi_{\mathbf{i}}, \quad k \geq 0. \tag{3.14}$$

By focusing on the branching vector belonging to the root node, i.e., $(Q_\emptyset, N_\emptyset, C_1, C_2, \dots)$ we can see that the processes $\{W^{(j)}\}$ and $\{R^{(k)}\}$ satisfy the distributional equations

$$W^{(j)} = \sum_{r=1}^{N_\emptyset} C_r \left(\sum_{(r,\mathbf{i}) \in A_j} Q_{(r,\mathbf{i})} \Pi_{(r,\mathbf{i})} / C_r \right) \stackrel{\mathcal{D}}{=} \sum_{r=1}^{N} C_r W_r^{(j-1)}, \qquad j \geq 1, \tag{3.15}$$

and

$$R^{(k)} = \sum_{r=1}^{N_\emptyset} C_r \left(\sum_{j=1}^{k} \sum_{(r,\mathbf{i}) \in A_j} Q_{(r,\mathbf{i})} \Pi_{(r,\mathbf{i})} / C_r \right) + Q_\emptyset \stackrel{\mathcal{D}}{=} \sum_{r=1}^{N} C_r R_r^{(k-1)} + Q, \qquad k \geq 1, \tag{3.16}$$

where $W_r^{(j-1)}$ are i.i.d. copies of $W^{(j-1)}$ and $R_r^{(k-1)}$ are i.i.d. copies of $R^{(k-1)}$, all independent of $(Q, N, C_1, C_2, \dots)$. Here and throughout the paper the convention is that $XY/Y \equiv 1$ if $Y = 0$.

For the homogeneous case ($Q \equiv 0$ in (3.12)), assume the weights $\{C_i\}$ are nonnegative and redefine the $\{W^{(j)}\}$ process as

$$W^{(0)} = 1, \qquad W^{(j)} = \sum_{\mathbf{i} \in A_j} \Pi_{\mathbf{i}}, \quad j \geq 1.$$

In this case the process $M^{(j)} = W^{(j)}/\rho^j$, $j \geq 0$, where $\rho = E\left[\sum_{i=1}^{N} C_i\right]$ defines a nonnegative martingale. It follows that $M^{(j)}$ converges almost surely, as $j \to \infty$, to a finite limit W with $E[W] \leq 1$. Taking the limit as $j \to \infty$ in (3.15) then gives that W satisfies

$$W \stackrel{\mathcal{D}}{=} \sum_{r=1}^{N} \frac{C_r}{\rho} W_r \stackrel{\triangle}{=} \sum_{r=1}^{N} C_r' W_r,$$

where the $\{W_r\}$ are i.i.d. copies of W, independent of $(N, C_1, C_2, \dots)$. Hence, W is a solution to the homogeneous version of (3.12) and the generic branching vector is $(N, C_1', C_2', \dots)$.

For the non homogeneous case ($P(Q \neq 0) > 0$), one can argue, as was done in [Jelenković and Olvera-Cravioto, 2012a], that provided $E\left[\sum_{i=1}^{N} |C_i|^\beta\right] < 1$ and $E[|Q|^\beta] < \infty$ for some $0 < \beta \leq 1$, then the random variable $R^{(k)}$ converges almost surely, as $k \to \infty$, to a finite limit R. Taking the limit as $k \to \infty$ in (3.16) gives that R is a solution to (3.12). We refer to the random variables W and R described above as the endogenous solutions to (3.12) in the homogeneous and non homogeneous cases, respectively.

3.4.2 The Kantorovich-Rubinstein distance

Before proceeding to the main results in the paper we give a brief description of the Kantorovich-Rubinstein. This distance on the space of probability measures is also known as the minimal l_1 metric or the Wasserstein distance of order one. For the purposes of this paper, we consider the vector space of infinite real sequences $\mathbb{R}^\infty$ having finite l_1 norm, i.e., $\mathbf{x} \in \mathbb{R}^\infty$ such that

$$||\mathbf{x}||_1 = \sum_{i=1}^{\infty} |x_i| < \infty.$$

Since our estimates will be given in terms of the distance between generic branching vectors, it follows that when analyzing a weighted branching process with a generic branching vector satisfying $P(N \leq m) = 1$ for some $m \in \mathbb{N}$, or when analyzing a weighted branching tree, we do not need to consider the space of infinite sequences, rather only the spaces $\mathbb{R}^{m+2}$ or $\mathbb{R}^3$, respectively. In any case, it will become clear from the context whether we are working on $\mathbb{R}^d$ or $\mathbb{R}^\infty$, and $||\mathbf{x}||_1$ will always refer to the corresponding l_1 norm.

Now recall the definition of the Kantorovich-Rubinstein distance: Let $M(\mu, \nu)$ denote the set of joint probability measures on $\mathcal{S} \times \mathcal{S}$ ($\mathcal{S} = \mathbb{R}^d$ or $\mathbb{R}^\infty$) with marginals μ and ν. Then, the Kantorovich-Rubinstein distance between μ and ν is given by

$$d_1(\mu, \nu) = \inf_{\pi \in M(\mu,\nu)} \int_{\mathcal{S}\times\mathcal{S}} ||\mathbf{x} - \mathbf{y}||_1 \, d\pi(\mathbf{x}, \mathbf{y}).$$

We point out that d_1 is only strictly speaking a distance when restricted to the subset of probability measures

$$\mathscr{P}_1(\mathcal{S}) \triangleq \left\{ \mu \in \mathscr{P}(\mathcal{S}) : \int_{\mathcal{S}} ||\mathbf{x}||_1 \, d\mu(\mathbf{x}) < \infty \right\},$$

where $\mathscr{P}(\mathcal{S})$ is the set of Borel probability measures on $\mathcal{S}$. We refer the interested reader to [Villani, 2009] for a thorough treatment of this distance, since the definition is only a special case.

Any construction on the same probability space of the joint vector $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ has marginal distribution μ and $\mathbf{Y}$ has marginal distribution ν, is called a *coupling* of μ and ν. In this notation we can rewrite d_1 as

$$d_1(\mu, \nu) = \inf_{\mathbf{X},\mathbf{Y}} E\left[||\mathbf{X} - \mathbf{Y}||_1\right],$$

where the infimum is taken over all couplings of μ and ν.

It is well known that d_1 is a metric on $\mathscr{P}_1$ and that the infimum is attained, or equivalently, that an optimal coupling $(\mathbf{X}, \mathbf{Y})$ such that

$$d_1(\mu, \nu) = E\left[||\mathbf{X} - \mathbf{Y}||_1\right]$$

always exists (see, e.g., [Villani, 2009], Theorem 4.1). This optimal coupling, nonetheless, is not in general explicitly available. One noteworthy exception is when μ and ν are probability measures on the real line, in which case we have that

$$d_1(\mu, \nu) = \int_0^1 |F^{-1}(u) - G^{-1}(u)|du = \int_{-\infty}^{\infty} |F(x) - G(x)|dx,$$

where F and G are the cumulative distribution functions of μ and ν, respectively, and $f^{-1}(t) = \inf\{x \in \mathbb{R} : f(x) \geq t\}$ denotes the pseudo-inverse of f. It follows that the optimal coupling is given by $(X, Y) = (F^{-1}(U), G^{-1}(U))$ for U uniformly distributed in $[0, 1]$.

Another important property of the Kantorovich-Rubinstein distance is that if $\{\mu_k\}_{k\in\mathbb{N}}$ is a sequence of probability measures in $\mathscr{P}_1$, then convergence in d_1 to a limit $\mu \in \mathscr{P}_1$ is equivalent to weak convergence. Furthermore, it satisfies the useful **duality formula**:

$$d_1(\mu, \nu) = \sup_{||\psi||_{\text{Lip}} \leq 1} \left\{ \int_{\mathcal{S}} \psi(\mathbf{x}) d\mu(\mathbf{x}) - \int_{\mathcal{S}} \psi(\mathbf{x}) d\nu(\mathbf{x}) \right\}$$

for all $\mu, \nu \in \mathscr{P}_1(\mathcal{S})$, where the supremum is taken over all Lipschitz continuous functions $\psi : \mathcal{S} \to \mathbb{R}$ with Lipschitz constant one (see Remark 6.5 in [Villani, 2009]).

3.4.3 Bounds for the Kantorovich-Rubinstein distance

We first present two sets of results; the first one provides explicit bounds for the Kantorovich-Rubinstein distance between two versions of the processes $\{W^{(j)} : j \geq 0\}$ (as defined

by (3.13)) constructed on weighted branching processes, respectively weighted branching trees, using different generic branching vectors. These bounds are given in terms of the Kantorovich-Rubinstein distance between the two generic branching vectors. The second set of results apply the explicit bounds to a sequence of processes $\{W^{(n,j)} : j \geq 0\}$ and $\{R^{(n,k)} : k \geq 0\}$ for $n \geq 1$, to obtain the convergence in the Kantorovich-Rubinstein distance to the endogenous solution to (3.12) in a limiting weighted branching process. The results for weighted branching trees are then used to show that $R^{(n,\infty)_1} \Rightarrow R^*$ in the analysis of the ranking algorithm.

Let $\{W^{(j)} : j \geq 0\}$ and $\{\hat{W}^{(j)} : j \geq 0\}$ be defined according to (3.13) on two different weighted branching processes using the generic branching vectors $(Q, N, C_1, C_2, \dots)$ and $(\hat{Q}, \hat{N}, \hat{C}_1, \hat{C}_2, \dots)$, respectively. As our result will show, it is enough to consider generic branching vectors of the form $(Q, B_1, B_2, \dots)$ and $(\hat{Q}, \hat{B}_1, \hat{B}_2, \dots)$ where $B_i = C_i 1(N \geq i)$ and $\hat{B}_i = \hat{C}_i 1(\hat{N} \geq i)$ for all $i \in \mathbb{N}_+$. Let μ denote the probability measure of $(Q, B_1, B_2, \dots)$ and let $\hat{\mu}$ denote the probability measure of $(\hat{Q}, \hat{B}_1, \hat{B}_2, \dots)$. We assume throughout the paper that

$$\int_{\mathcal{S}} ||\mathbf{x}||_1 d\mu(\mathbf{x}) < \infty \qquad \text{and} \qquad \int_{\mathcal{S}} ||\mathbf{x}||_1 d\hat{\mu}(\mathbf{x}) < \infty. \tag{3.17}$$

To construct the two processes on the same probability space, let π denote any coupling of μ and $\hat{\mu}$ and let $\{(Q_{\mathbf{i}}, B_{(\mathbf{i},1)}, B_{(\mathbf{i},2)}, \dots, \hat{Q}_{\mathbf{i}}, \hat{B}_{(\mathbf{i},1)}, \hat{B}_{(\mathbf{i},2)}, \dots)\}_{\mathbf{i} \in U}$ be a sequence of i.i.d. random vectors distributed according to π. Then, use the vectors $\{(Q_{\mathbf{i}}, B_{(\mathbf{i},1)}, B_{(\mathbf{i},2)}, \dots)\}_{\mathbf{i} \in U}$ to construct $\{W^{(j)} : j \geq 0\}$, as described in Section 3.4.1, and the vectors $\{(\hat{Q}_{\mathbf{i}}, \hat{B}_{(\mathbf{i},1)}, \hat{B}_{(\mathbf{i},2)}, \dots)\}_{\mathbf{i} \in U}$ to construct $\{\hat{W}^{(j)} : j \geq 0\}$. Our first result is stated below.

Throughout the paper, we use $x \wedge y$ and $x \vee y$ to denote the minimum and the maximum, respectively, of x and y, $x^+ = \max\{0, x\}$, and we use the convention that $\sum_{i=a}^{b} x_i \equiv 0$ if $b < a$. The notation $E_\pi[\cdot]$ means that the expectation is taken with respect to the coupling π.

Proposition 13. *For any coupling π of μ and $\hat{\mu}$, and any $j \geq 0$,*

$$E\left[\left|\hat{W}^{(j)} - W^{(j)}\right|\right] \leq \left(\hat{\rho}^j + E[|Q|] \sum_{t=0}^{j-1} \rho^t \hat{\rho}^{j-1-t}\right) \mathcal{E},$$

where $\rho = E\left[\sum_{i=1}^{N} |C_i|\right]$, $\hat{\rho} = E\left[\sum_{i=1}^{\hat{N}} |\hat{C}_i|\right]$ and $\mathcal{E} = E_\pi\left[|\hat{Q} - Q| + \sum_{i=1}^{\infty} |\hat{B}_i - B_i|\right]$.

We point out that the bound provided by Proposition 13 is also a bound for the Kantorovich-Rubinstein distance between $\hat{W}^{(j)}$ and $W^{(j)}$, and if we take π to be the optimal coupling of μ and $\hat{\mu}$ then we have $\mathcal{E} = d_1(\hat{\mu}, \mu)$. It is also worth mentioning that if we let ν and $\hat{\nu}$ be the probability measures of $(Q, N, C_1, C_2, \dots)$ and $(\hat{Q}, \hat{N}, \hat{C}_1, \hat{C}_2, \dots)$, respectively, and assume that $E[N + \hat{N}] < \infty$, then $d_1(\mu, \hat{\mu})$ can be small even if $d_1(\nu, \hat{\nu})$ is not. This is due to the observation that, in general, large disagreements between C_r and $\hat{C}_r$ for values of r for which $P(N > r)$ and $P(\hat{N} > r)$ are negligible do not affect $d_1(\mu, \hat{\mu})$, whereas they do adversely affect $d_1(\nu, \hat{\nu})$.

Our next result provides a similar bound for the case when $\hat{W}^{(j)}$ and $W^{(j)}$ are constructed on weighted branching trees using the generic branching vectors $(\hat{Q}, \hat{N}, \hat{C})$ and (Q, N, C), respectively. As before, let $\hat{\nu}$ and ν denote the probability measures of $(\hat{Q}, \hat{N}, \hat{C})$ and (Q, N, C). Because of the different dependence structure they generate on the tree, we allow the coupling used for the root nodes to be different than all other nodes, i.e., the two trees are constructed using the sequence of i.i.d. vectors $\{(Q_{\mathbf{i}}, C_{\mathbf{i}}, N_{\mathbf{i}}, \hat{Q}_{\mathbf{i}}, \hat{C}_{\mathbf{i}}, \hat{N}_{\mathbf{i}})\}_{\mathbf{i} \in U, \mathbf{i} \neq \emptyset}$ distributed according to a coupling π of ν and $\hat{\nu}$, while $(Q_\emptyset, N_\emptyset, \hat{Q}_\emptyset, \hat{N}_\emptyset)$ is independent of the previous sequences and is distributed according to a coupling π^* of ν^* and $\hat{\nu}^*$, where ν^* is the probability measure of (Q, N) and $\hat{\nu}^*$ is that of $(\hat{Q}, \hat{N})$. We have ignored $C_\emptyset$ and $\hat{C}_\emptyset$ since they do not appear in the definitions of $W^{(j)}$ and $\hat{W}^{(j)}$.

Proposition 14. *For any coupling π of ν and $\hat{\nu}$ and any coupling π^* of ν^* and $\hat{\nu}^*$,*

$$E\left[\left|\hat{W}^{(0)} - W^{(0)}\right|\right] \leq \mathcal{E}^*$$

and for $j \geq 1$,

$$E\left[\left|\hat{W}^{(j)} - W^{(j)}\right|\right] \leq \left(E[\hat{N}] \vee \frac{E[N]E[|CQ|]}{\rho}\right)\left(\sum_{t=0}^{j-1} \hat{\rho}^t \rho^{j-1-t}\right)\mathcal{E} + E[|Q|]\hat{\rho}^{j-1}\mathcal{E}^*,$$

where $\rho = E[N|C|]$, $\hat{\rho} = E[\hat{N}|\hat{C}|]$,

$$\mathcal{E}^* = E_{\pi^*}\left[|\hat{Q} - Q| + |\hat{N} - N|\right] \quad \text{and} \quad \mathcal{E} = E_\pi\left[|\hat{C}\hat{Q} - CQ| + \sum_{i=1}^{\infty} |\hat{C}1(\hat{N} \geq i) - C1(N \geq i)|\right].$$

3.4.4 Convergence to the endogenous solution

Our second set of results considers a sequence of weighted branching processes (respectively, weighted branching trees), each constructed using a generic branching vector having

probability measure ν_n, $n \geq 1$. In other words, for weighted branching processes, ν_n is the probability measure of a vector of the form $(Q^{(n)}, N^{(n)}, C_1^{(n)}, C_2^{(n)}, \dots)$, while for weighted branching trees it corresponds to a vector of the form $(Q^{(n)}, N^{(n)}, C^{(n)})$. On each of them we define the processes $\{W^{(n,j)} : j \geq 0\}$ and $\{R^{(n,k)} : k \geq 0\}$ according to (3.13) and (3.14), and we are interested in providing conditions under which $W^{(n,j)}$ (suitably scaled) and $R^{(n,k)}$ will converge, as n, j, k go to infinity, to the endogenous solution of a linear SFPE of the form in (3.12).

The main conditions for the convergence we seek will be in terms of the sequence of probability measures $\{\mu_n\}_{n\geq 1}$, where μ_n is the probability measure of the vector

$$(Q^{(n)}, C_1^{(n)} 1(N^{(n)} \geq 1), C_2^{(n)} 1(N^{(n)} \geq 2), \dots)$$

for weighted branching processes, and of

$$(C^{(n)} Q^{(n)}, C^{(n)} 1(N^{(n)} \geq 1), C^{(n)} 1(N^{(n)} \geq 2), \dots)$$

for weighted branching trees.

In both cases, we assume that there exists a probability measure μ such that $d_1(\mu_n, \mu) \to 0$. We point out that for a weighted branching process, μ is always the probability measure of a generic branching vector, since each of the μ_n is. However, this is not the case for a weighted branching tree. In order for μ to define a weighted branching process we need C to be independent of (Q, N), in which case the limiting weighted branching process has a generic branching vector of the form $(Q, N, C_1, C_2, \dots)$ with the $\{C_i\}_{i\geq 1}$ i.i.d. and independent of (Q, N); condition (3.17) implies that $E[N] < \infty$.

We refer to the case where we analyze a sequence of weighted branching processes as *Case 1*, and to the case where we analyze a sequence of weighted branching trees as *Case 2*. For *Case 2*, in addition to the measure μ_n defined above, we define ν_n^* to be the probability measure of the vector $(Q^{(n)}, N^{(n)})$ and ν^* to be the probability measure of (Q, N). The symbol $\Rightarrow$ denotes convergence in distribution and $\xrightarrow{d_1}$ denotes convergence in the Kantorovich-Rubinstein distance.

Theorem 15. *Define the processes* $\{W^{(n,j)} : j \geq 0\}$, $n \geq 1$, *and* $\{W^{(j)} : j \geq 0\}$ *according to* (3.13). *Suppose that as* $n \to \infty$,

$$d_1(\mu_n, \mu) \to 0 \quad \textit{(Case 1)} \quad \textit{or} \quad d_1(\nu_n^*, \nu^*) + d_1(\mu_n, \mu) \to 0 \quad \textit{(Case 2)}.$$

Then, for any fixed $j \in \mathbb{N}$

$$\hat{W}^{(j)} \xrightarrow{d_1} W^{(j)}, \qquad n \to \infty.$$

Moreover, if $Q^{(n)} = Q \equiv 1$, *and* $C_j^{(n)}, C_j$ *are nonnegative for all* n *and* j, *then for any* $j_n \in \mathbb{N}$ *such that* $j_n \to \infty$ *and*

$$j_n\, d_1(\mu_n, \mu) \to 0 \quad \textit{(Case 1)} \quad \textit{or} \quad d_1(\nu_n^*, \nu^*) + j_n\, d_1(\mu_n, \mu) \to 0 \quad \textit{(Case 2)},$$

as $n \to \infty$, *we have*

$$\frac{W^{(n,j_n)}}{\rho_n^{j_n}} \Rightarrow \mathcal{W} \qquad \textit{and} \qquad \frac{W^{(n,j_n)}}{\rho^{j_n}} \Rightarrow \mathcal{W},$$

where $\mathcal{W}$ *is the a.s. limit of* $W^{(j)}/\rho^j$ *as* $j \to \infty$.

As pointed out in Section 3.4.1.1, $\mathcal{W}$ is the endogenous solution to the SFPE

$$\mathcal{W} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} \frac{C_i}{\rho}\, \mathcal{W}_i,$$

where the $\{\mathcal{W}_i\}$ are i.i.d. copies of $\mathcal{W}$, independent of $(N, C_1, C_2, \dots)$. See [Liu, 1998] for conditions on when the random variable $\mathcal{W}$, which satisfies $E[\mathcal{W}] \leq 1$, is non-trivial, as well as characterizations of its tail behavior. Furthermore, when $E[\mathcal{W}] = 1$ we can replace the convergence in distribution with convergence in the Kantorovich-Rubinstein distance, i.e.,

$$\frac{W^{(n,j_n)}}{\rho_n^{j_n}} \xrightarrow{d_1} \mathcal{W} \qquad \text{and} \qquad \frac{W^{(n,j_n)}}{\rho^{j_n}} \xrightarrow{d_1} \mathcal{W}, \qquad n \to \infty.$$

We now give a similar result for the non homogeneous equation.

Theorem 16. *Define the processes* $\{R^{(n,k)} : k \geq 0\}$, $n \geq 1$, *and* $\{R^{(k)} : k \geq 0\}$ *according to* (3.14). *Suppose that as* $n \to \infty$,

$$d_1(\mu_n, \mu) \to 0 \quad \textit{(Case 1)} \quad \textit{or} \quad d_1(\nu_n^*, \nu^*) + d_1(\mu_n, \mu) \to 0 \quad \textit{(Case 2)}.$$

Then, for any fixed $k \in \mathbb{N}$,

$$R^{(n,k)} \xrightarrow{d_1} R^{(k)}, \qquad n \to \infty.$$

Moreover, if $\rho < 1$, *then for any* $k_n \in \mathbb{N}$ *such that* $k_n \to \infty$ *as* $n \to \infty$, *we have*

$$R^{(n,k_n)} \xrightarrow{d_1} R, \qquad n \to \infty,$$

where $R = \sum_{k=0}^{\infty} \sum_{\mathbf{i} \in A_k} \Pi_{\mathbf{i}} Q_{\mathbf{i}}$ *is the a.s. limit of* $R^{(k)}$ *as* $k \to \infty$.

In this case, R is the endogenous solution to the SFPE

$$R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q, \tag{3.18}$$

where the $\{R_i\}$ are i.i.d. copies of R, independent of $(Q, N, C_1, C_2, \dots)$. Moreover, the asymptotic behavior of $P(R > x)$ as $x \to \infty$ can be described for several different assumptions on the generic vector $(Q, N, C_1, C_2, \dots)$. We refer the reader to [Jelenković and Olvera-Cravioto, 2012a] and [Olvera-Cravioto, 2012a] for the precise set of theorems.

Note that in *Case 1*, the convergence of $R^{(n,k)}$ as $k \to \infty$ for a fixed n is guaranteed whenever $E\left[\sum_{i=1}^{N^{(n)}} |C_i^{(n)}|^\beta\right] < 1$ for some $0 < \beta \leq 1$ (see Lemma 4.1 in [Jelenković and Olvera-Cravioto, 2012a]), and its limit, $R^{(n)}$ would be the endogenous solution to

$$R^{(n)} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N^{(n)}} C_i^{(n)} R_i^{(n)} + Q^{(n)}. \tag{3.19}$$

For *Case 2*, on the other hand, an adaptation of the proof of Lemma 4.1 in [Jelenković and Olvera-Cravioto, 2012a] would give that $R^{(n,k)}$ converges a.s. to

$$R^{(n)} = \sum_{j=0}^{\infty} W^{(n,j)},$$

as $k \to \infty$, with $R^{(n)}$ finite a.s., provided $E\left[N^{(n)}|C^{(n)}|^\beta\right] < 1$ for some $0 < \beta \leq 1$. However, this random variable $R^{(n)}$ would not necessarily have the interpretation of being a solution to (3.19).

We end this section with a result for the weighted branching tree setting that states that $d_1(\mu_n, \mu)$ converges to zero whenever $d_1(\nu_n, \nu)$ and the moments of $Q^{(n)}C^{(n)}$ and $N^{(n)}C^{(n)}$ do. However, the rates at which $d_1(\nu_n, \nu)$ and $d_1(\mu_n, \mu)$ converge could be different.

Lemma 17. *For Case 2, suppose that as $n \to \infty$, $d_1(\nu_n, \nu) \to 0$, $E[|C^{(n)}Q^{(n)}|] \to E[|CQ|]$ and $E[|C^{(n)}|N^{(n)}] \to E[|C|N]$. Then,*

$$d_1(\mu_n, \mu) \to 0, \qquad n \to \infty.$$

3.4.5 Main Result

We are now ready to state the main result of this chapter, which establishes the convergence of the rank of a randomly chosen node in the DCM to a non-degenerate random variable $\mathcal{R}^*$.

We first give the required assumption. With some abuse of notation, for joint distribution functions $F_n, F \in \mathbb{R}^d$ we write $d_1(F_n, F)$ to denote the Kantorovich-Rubinstein distance between their probability measures μ_n and μ. The symbol $\xrightarrow{P}$ denotes convergence in probability.

Assumption 2. *Given the extended bi-degree sequence* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ *define*

$$F_n^*(m,q) := \frac{1}{n}\sum_{k=1}^{n} 1(N_k \le m, Q_k \le q) \quad \textit{and} \quad F_n(m,q,x) := \sum_{k=1}^{n} 1(N_k \le m, Q_k \le q, C_k \le x)\frac{D_k}{L_n}.$$

Suppose there exist random vectors $(\mathcal{N}_0, \mathcal{Q}_0)$ *and* $(\mathcal{N}, \mathcal{Q})$*, and a random variable* $\mathcal{C}$*, such that*

$$d_1(F_n^*, F^*) \xrightarrow{P} 0 \qquad \textit{and} \qquad d_1(F_n, F) \xrightarrow{P} 0,$$

as $n \to \infty$*, where*

$$F^*(m,q) := P(\mathcal{N}_0 \le m, \mathcal{Q}_0 \le q) \qquad \textit{and} \qquad F(m,q,x) := P(\mathcal{N} \le m, \mathcal{Q} \le q)P(\mathcal{C} \le x).$$

Remark 6. *Note that Assumption 2 and the duality formula imply that*

$$\sup\left\{\mathbb{E}_n\left[\psi(\hat{N}_1, \hat{Q}_1, \hat{C}_1)\right] - E[\psi(\mathcal{N}, \mathcal{Q}, \mathcal{C})] : \psi \text{ is bounded and continuous}\right\}$$

converges to zero in probability, and therefore, by the bounded convergence theorem,

$$E\left[\psi(\hat{N}_1, \hat{Q}_1, \hat{C}_1)\right] \to E[\psi(\mathcal{N}, \mathcal{Q}, \mathcal{C})], \quad n \to \infty,$$

for any bounded and continuous function ψ*, or equivalently,* $(\hat{N}_1, \hat{Q}_1, \hat{C}_1) \Rightarrow (\mathcal{N}, \mathcal{Q}, \mathcal{C})$*; similarly,* $(\hat{N}_\emptyset, \hat{Q}_\emptyset) \Rightarrow (\mathcal{N}_0, \mathcal{Q}_0)$*. The duality formula, combined with Assumption 1, also implies that* $E[\mathcal{N}_0] = \nu_1$*,* $E[\mathcal{N}] = \mu$ *and* $E[\mathcal{C}] = \nu_5/\nu_1$*.*

Theorem 18. *Suppose the extended bi-degree sequence* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ *satisfies Assumptions 1 and 2. Then,*

$$R_1^{(n,\infty)} \Rightarrow \mathcal{R}^*$$

as $n \to \infty$*, where* $\mathcal{R}^*$ *is defined as in* (**??**) *with the weights* $\{\mathcal{C}_i\}$ *i.i.d. and independent of* $(\mathcal{N}_0, \mathcal{Q}_0)$*, respectively of* $(\mathcal{N}, \mathcal{Q})$ *in* (3.1)*.*

Proof. Define Ω_n according to Assumption 1 and note that $P(\Omega_n^c) = O(n^{-\epsilon})$, so it suffices to show that $R_1^{(n,\infty)}$, conditional on Ω_n, converges weakly to $\mathcal{R}^*$. Note that by Assumption 1, $\rho = E[\mathcal{N}]E[|\mathcal{C}|] = \nu_5\mu/\nu_1 < 1$, which is a sufficient condition for $\mathcal{R}$ to be well defined (see Lemma 4.1 in [Jelenković and Olvera-Cravioto, 2012a]). First, when $\mu > 1$, fix $0 < \delta < |\log c|/(2\log\mu)$ and let $k_n = s\log n$, where $\delta/|\log c| < s < 1/(2\log\mu)$. Next, note that by the arguments leading to (3.9),

$$\begin{aligned} P\left(\left|R_1^{(n,\infty)} - \hat{R}^{(n,k_n)}\right| > n^{-\delta}\middle|\,\Omega_n\right) &= O\left(n^{\delta}c^{k_n} + (\mu^{2k_n}/n)^{1/2}\right) \\ &= O\left(n^{\delta - s|\log c|} + n^{(2s\log\mu - 1)/2}\right) = o(1) \end{aligned}$$

as $n \to \infty$. When $\mu \le 1$ we can take $k_n = n^{\varepsilon}$, with $\varepsilon < \min\{1/2, \gamma\}$, to obtain that the probability converges to zero. We then obtain that conditionally on Ω_n,

$$\left|R_1^{(n,\infty)} - \hat{R}^{(n,k_n)}\right| \Rightarrow 0.$$

That $\hat{R}^{(n,k_n)} \Rightarrow \mathcal{R}^*$ conditionally on Ω_n will follow from Theorem 16 and Lemma 17 in Chapter **??** and Assumption 2 once we verify that, as $n \to \infty$,

$$\mathbb{E}_n\left[\hat{N}_1|\hat{C}_1|\right] \xrightarrow{P} E[\mathcal{N}]E[|\mathcal{C}|] \qquad \text{and} \qquad \mathbb{E}_n\left[|\hat{Q}_1\hat{C}_1|\right] \xrightarrow{P} E[|\mathcal{Q}|]E[|\mathcal{C}|]. \tag{3.20}$$

To show that (3.20) holds define $\phi_K(q,x) = (|q| \wedge K)(|x| \wedge 1)$ for $K > 0$, and note that since ϕ_K is bounded and continuous, Assumption 2 and Remark 6 imply that

$$\mathbb{E}_n\left[\phi_K(\hat{Q}_1, \hat{C}_1)\right] \xrightarrow{P} E[\phi_K(\mathcal{Q},\mathcal{C})] = E[|\mathcal{Q}| \wedge K]E[|\mathcal{C}|], \qquad n \to \infty.$$

Next, fix $\epsilon > 0$ and choose K such that $E[|\mathcal{Q}|1(|\mathcal{Q}| > K)] < \epsilon/4$. Then,

$$\begin{aligned} \left|\mathbb{E}_n\left[|\hat{Q}_1\hat{C}_1|\right] - E[|\mathcal{Q}\mathcal{C}|]\right| &\le \left|\mathbb{E}_n\left[\phi_K(\hat{Q}_1,\hat{C}_1)\right] - E[\phi_K(\mathcal{Q},\mathcal{C})]\right| \\ &\quad + \mathbb{E}_n\left[(|\hat{Q}_1| - K)^+|\hat{C}_1|\right] + E[(|\mathcal{Q}| - K)^+|\mathcal{C}|] \\ &\le \left|\mathbb{E}_n\left[\phi_K(\hat{Q}_1,\hat{C}_1)\right] - E[\phi_K(\mathcal{Q},\mathcal{C})]\right| + c\mathbb{E}_n\left[(|\hat{Q}_1| - K)^+\right] + c\epsilon/4, \end{aligned}$$

where we used that both $|\hat{C}_1|$ and $|\mathcal{C}|$ are bounded by $c < 1$. It follows that

$$\lim_{n\to\infty} P\left(\left|\mathbb{E}_n\left[|\hat{Q}_1\hat{C}_1|\right] - E[|\mathcal{Q}\mathcal{C}|]\right| > \epsilon\right) \le \lim_{n\to\infty} P\left(\mathbb{E}_n\left[(|\hat{Q}_1| - K)^+\right] > \epsilon/2\right).$$

To show that this last limit is zero note that $(|x|-K)^+$ is Lipschitz continuous with Lipschitz constant one, so by the duality formula we obtain

$$\mathbb{E}_n\left[(|\hat{Q}_1|-K)^+\right] \xrightarrow{P} E[(|\mathcal{Q}|-K)^+] < \epsilon/4$$

as $n \to \infty$, which gives the desired limit.

The proof for $\mathbb{E}_n\left[|\hat{N}_1\hat{C}_1|\right]$ follows the same steps and is therefore omitted. □

3.4.6 Asymptotic behavior of the limit

We end this section by giving a limit theorem describing the tail asymptotics of $\mathcal{R}^*$; its proof is given in Appendix B.3. This result covers the case where the weights $\{\mathcal{C}_i\}$ are nonnegative and either the limiting in-degree $\mathcal{N}$ or the limiting personalization value $\mathcal{Q}$ have a regularly varying distribution, which in turn implies the regular variation of $\mathcal{R}$. Then, we deduce the asymptotics of $\mathcal{R}^*$ using some results for weighted random sums with heavy-tailed summands. The corresponding theorems can be found in [Olvera-Cravioto, 2012b; Volkovich and Litvak, 2010].

Definition 10. *We say that a function f is regularly varying at infinity with index $-\alpha$, denoted $f \in \mathscr{R}_{-\alpha}$, if $f(x) = x^{-\alpha}L(x)$ for some slowly varying function L; and $L : [0,\infty) \to (0,\infty)$ is slowly varying if $\lim_{x\to\infty} L(\lambda x)/L(x) = 1$ for any $\lambda > 0$.*

We use the notation $f(x) \sim g(x)$ as $x \to \infty$ for $\lim_{x\to\infty} f(x)/g(x) = 1$.

Theorem 19. *Suppose the generic branching vector $(\mathcal{N}, \mathcal{Q}, \mathcal{C}_1, \mathcal{C}_2, \dots)$ is such that the weights $\{\mathcal{C}_i\}$ are nonnegative, bounded i.i.d. copies of $\mathcal{C}$, independent of $(\mathcal{N}, \mathcal{Q})$, $\mathcal{N} \in \mathbb{N}$ and $\mathcal{Q} \in \mathbb{R}$. Define $\rho = E[\mathcal{N}]E[\mathcal{C}]$ and $\rho_\alpha = E[\mathcal{N}]E[\mathcal{C}^\alpha]$ and let $\mathcal{R}$ be defined as in (??).*

- *If $P(\mathcal{N} > x) \in \mathscr{R}_{-\alpha}$, $\alpha > 1$, $\rho \vee \rho_\alpha < 1$, $P(\mathcal{N}_0 > x) \sim \kappa P(\mathcal{N} > x)$ as $x \to \infty$ for some $\kappa > 0$, $E[\mathcal{Q}], E[\mathcal{Q}_0] > 0$, and $E\left[|\mathcal{Q}|^{\alpha+\epsilon} + |\mathcal{Q}_0|^{\alpha+\epsilon}\right] < \infty$ for some $\epsilon > 0$, then*

$$P(\mathcal{R}^* > x) \sim \left(E[\mathcal{N}_0]E[\mathcal{C}^\alpha] + \kappa(1-\rho_\alpha)\right)\frac{(E[\mathcal{Q}]E[\mathcal{C}])^\alpha}{(1-\rho)^\alpha(1-\rho_\alpha)}P(\mathcal{N} > x), \qquad x \to \infty.$$

- *If $P(\mathcal{Q} > x) \in \mathscr{R}_{-\alpha}$, $\alpha > 1$, $\rho \vee \rho_\alpha < 1$, $P(\mathcal{Q}_0 > x) \sim \kappa P(\mathcal{Q} > x)$ as $x \to \infty$ for some $\kappa > 0$, $E[|\mathcal{Q}|^\beta + |\mathcal{Q}_0|^\beta] < \infty$ for all $0 < \beta < \alpha$, and $E\left[|\mathcal{N}|^{\alpha+\epsilon} + |\mathcal{N}_0|^{\alpha+\epsilon}\right] < \infty$ for some $\epsilon > 0$, then*

$$P(\mathcal{R}^* > x) \sim \left(E[\mathcal{N}_0]E[\mathcal{C}^\alpha] + \kappa(1-\rho_\alpha)\right)(1-\rho_\alpha)^{-1}P(\mathcal{Q} > x), \qquad x \to \infty.$$

Remark 7. *(i) For PageRank we have $C_i = c/D_i$ and $Q_i = 1 - c$, where $c \in (0,1)$ is the damping factor. This leads to a limiting weight distribution of the form*

$$P(\mathcal{C} \leq x) = \lim_{n\to\infty} \frac{1}{L_n} \sum_{i=1}^{n} 1(c/D_i \leq x) D_i,$$

which is not the limiting distribution of the reciprocal of the out-degrees, $\{c/D_i\}$, but rather a size-biased version of it.

(ii) Applying Theorem 19 to PageRank when $P(\mathcal{N} > x) \in \mathscr{R}_{-\alpha}$ and $P(\mathcal{N}_0 > x) \sim \kappa P(\mathcal{N} > x)$ for some constant $\kappa > 0$ gives that

$$P(\mathcal{R}^* > x) \sim \kappa' P(\mathcal{N} > x) \qquad \text{as } x \to \infty,$$

where $\kappa' > 0$ is determined by the theorem.

(iii) The theorem above only includes two possible cases of the relations between $(\mathcal{N}_0, \mathcal{Q}_0)$ and $(\mathcal{N}, \mathcal{Q})$. The exact asymptotics of $\mathcal{R}^*$ can be obtained from those of $\mathcal{R}$ in more cases than these using the same techniques; we leave the details to the reader.

(iv) Theorem 19 requires the weights $\{\mathcal{C}_i\}$ to be nonnegative, which is not a condition in Theorem 18. The tail asymptotics of $\mathcal{R}$, and therefore of $\mathcal{R}^*$, in the real-valued case are unknown.

3.5 Algorithm to generate bi-degree sequences

As an example of an extended bi-degree sequence satisfying Assumptions 1 and 2, we give in this section an algorithm based on sequences of i.i.d. random variables. The method for generating the bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n)$ is taken from [Chen and Olvera-Cravioto, 2013], where the goal was to generate a directed random graph with prescribed in- and out-degree distributions.

To define the algorithm we need to first specify target distributions for the in- and out-degrees, which we will denote by $f_k^{\text{in}} = P(\mathscr{N} = k)$, and $f_k^{\text{out}} = P(\mathscr{D} = k)$, $k \geq 0$, respectively. Furthermore, we will assume that these target distributions satisfy $E[\mathscr{N}] = E[\mathscr{D}]$,

$$\overline{F^{\text{in}}}(x) = \sum_{k>x} f_k^{\text{in}} \leq x^{-\alpha} L_{\text{in}}(x) \qquad \text{and} \qquad \overline{F^{\text{out}}}(x) = \sum_{k>x} f_k^{\text{out}} \leq x^{-\beta} L_{\text{out}}(x),$$

for some slowly varying functions L_{in} and L_{out}, and $\alpha > 1, \beta > 2$. To the original construction given in [Chen and Olvera-Cravioto, 2013] we will need to add two additional steps to generate the weight and personalization sequences $\mathbf{C}_n$ and $\mathbf{Q}_n$, for which we need two more distributions $F^{\zeta}(x) = P(\zeta \leq x)$ and $F^{Q}(x) = P(Q \leq x)$ with support on the real line and satisfying

$$P(|\zeta| \leq c) = 1 \text{ for some } 0 < c < 1, \quad \text{and} \quad E[|Q|^{1+\epsilon_Q}] < \infty \text{ for some } 0 < \epsilon_Q \leq 1.$$

Let

$$\kappa_0 = \min\{1 - \alpha^{-1}, 1/2\}.$$

The IID Algorithm:

1. Fix $0 < \delta_0 < \kappa_0$.
2. Sample an i.i.d. sequence $\{\mathscr{N}_1, \ldots, \mathscr{N}_n\}$ from distribution F^{in}; let $\overline{\mathscr{N}}_n = \sum_{i=1}^n \mathscr{N}_i$.
3. Sample an i.i.d. sequence $\{\mathscr{D}_1, \ldots, \mathscr{D}_n\}$ from distribution F^{out}, independent of $\{\mathscr{N}_i\}$; let $\overline{\mathscr{D}}_n = \sum_{i=1}^n \mathscr{D}_i$.
4. Define $\Delta_n = \overline{\mathscr{N}}_n - \overline{\mathscr{D}}_n$. If $|\Delta_n| \leq n^{1-\kappa_0+\delta_0}$ proceed to step 5; otherwise repeat from step 2.
5. Choose randomly $|\Delta_n|$ nodes $\{i_1, i_2, \ldots, i_{|\Delta_n|}\}$ without replacement and let

$$N_i = \begin{cases} \mathscr{N}_i + 1 & \text{if } \Delta_n < 0 \text{ and } i \in \{i_1, i_2, \ldots, i_{|\Delta_n|}\}, \\ \mathscr{N}_i & \text{otherwise,} \end{cases}$$

$$D_i = \begin{cases} \mathscr{D}_i + 1 & \text{if } \Delta_n \geq 0 \text{ and } i \in \{i_1, i_2, \ldots, i_{|\Delta_n|}\}, \\ \mathscr{D}_i & \text{otherwise.} \end{cases}$$

6. Sample an i.i.d. sequence $\{Q_1, \ldots, Q_n\}$ from distribution F^{Q}, independent of $\{\mathscr{N}_i\}$ and $\{\mathscr{D}_i\}$.
7. Sample an i.i.d. sequence $\{\zeta_1, \ldots, \zeta_n\}$ from distribution F^{ζ}, independent of $\{\mathscr{N}_i\}$, $\{\mathscr{D}_i\}$ and $\{Q_i\}$, and set $C_i = \zeta_i / D_i$ if $D_i \geq 1$ or $C_i = c\,\text{sgn}(\zeta_i)$ otherwise.

Remark 8. *Note that since* $E[|\mathscr{N} - \mathscr{D}|^{1+a}] < \infty$ *for any* $0 < a < \min\{\alpha - 1, \beta - 1\}$*, then* $E[|\mathscr{N} - \mathscr{D}|^{1+(\kappa_0-\delta_0)/(1-\kappa_0)}] < \infty$*, and Corollary 32 in Appendix B gives*

$$P\left(|\Delta_n| > n^{1-\kappa_0+\delta_0}\right) = O\left(n^{-\delta_0(\kappa_0-\delta_0)/(1-\kappa_0)}\right) \tag{3.21}$$

as $n \to \infty$*.*

The two propositions below give the desired properties. Their proofs are given in Appendix B.4.

Proposition 20. *The extended bi-degree sequence* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ *generated by the IID Algorithm satisfies Assumption 1 for any* $0 < \kappa < \beta - 2$*, any* $0 < \gamma < \min\{(\kappa_0 - \delta_0)^2/(1 - \delta_0), (\beta - 2 - \kappa)/\beta\}$*,* $\mu = \nu_1 = E[\mathscr{N}] = E[\mathscr{D}]$*,* $\nu_2 = (E[\mathscr{D}])^2$*,* $\nu_3 = E[\mathscr{D}^2]$*,* $\nu_4 = E[\mathscr{D}^{2+\kappa}]$*,* $\nu_5 = E[|\zeta|]P(\mathscr{D} \geq 1)$*,* $H = E[|Q|] + 1$*, and some* $\varepsilon > 0$*.*

Proposition 21. *The extended bi-degree sequence* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ *generated by the IID Algorithm satisfies Assumption 2 with*

$$F^*(m, q) = P(\mathscr{N} \leq m)P(Q \leq q) \qquad \textit{and}$$

$$F(m, q, x) = P(\mathscr{N} \leq m)P(Q \leq q)E[1(\zeta/\mathscr{D} \leq x)\mathscr{D}]/\mu.$$

3.6 Numerical examples

To complement the theoretical contribution of the chapter, we use the IID Algorithm described in the previous section to provide some numerical results showing the accuracy of the WBP approximation to PageRank. To generate the in- and out-degrees we use the zeta distribution. More precisely, we set

$$\mathscr{N}_i = X_{1,i} + Y_{1,i}, \quad \mathscr{D}_i = X_{2,i} + Y_{2,i},$$

where $\{X_{1,i}\}$ and $\{X_{2,i}\}$ are independent sequences of i.i.d. Zeta random variables with parameters $\alpha + 1$ and $\beta + 1$, respectively; $\{Y_{1,i}\}$ and $\{Y_{2,i}\}$ are independent sequences of i.i.d. Poisson random variables with different parameters chosen so that $\mathscr{N}$ and $\mathscr{D}$ have

equal mean. Note that the Poisson distribution has a light tail so that the power law tail behavior of $\mathscr{N}$ and $\mathscr{D}$ is preserved and determined by α and β, respectively.

Once the sequences $\{\mathscr{N}_i\}$ and $\{\mathscr{D}_i\}$ are generated, we use the IID Algorithm to obtain a valid bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n)$. Note that in PageRank, we have $\zeta_i = c$ and $Q_i = 1 - c$. Given this bi-degree sequence we next proceed to construct the graph and the TBT simultaneously, according to the rules described in Section 3.3. To compute $\mathbf{R}^{(n,\infty)}$ we perform matrix iterations with $r_0 = 1$ until $\|\mathbf{R}^{(n,k)} - \mathbf{R}^{(n,k-1)}\|_2 < \varepsilon_0$ for some tolerance ε_0. We only generate the TBT for as many generations as it takes to construct the graph, with each generation corresponding to a step in the breadth first graph exploration process. The computation of the root node of the TBT, $\hat{R}^{(n,k)}$ is done recursively starting from the leaves using

$$\hat{R}_{\mathbf{i}}^{(n,0)} = 1 \text{ for } \mathbf{i} \in \hat{A}_k, \qquad \hat{R}_{\mathbf{i}}^{(n,r)} = \sum_{j=1}^{\hat{N}_{\mathbf{i}}} \frac{c}{\hat{D}_{(\mathbf{i},j)}} \hat{R}_{(\mathbf{i},j)}^{(n,r-1)} + 1 - c, \text{ for } \mathbf{i} \in \hat{A}_r,\ 0 \le r < k.$$

To draw a sample from $\mathcal{R}^*$, note that by Proposition 21, $\mathcal{R}^*$ in the IID Algorithm has the same distribution as $\mathcal{R}$, i.e., the endogenous solution to the SFPE

$$\mathcal{R} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{\mathscr{N}} \mathcal{C}_i \mathcal{R}_i + 1 - c,$$

where $P(\mathcal{C} \le x) = E[1(c/\mathscr{D} \le x)\mathscr{D}]/\mu$. To sample $\mathcal{R}$ we construct a WBP with generic branching vector $(\mathscr{N}, 1-c, \{\mathcal{C}_i\})$, with the $\{\mathcal{C}_i\}$ i.i.d. and independent of $\mathscr{N}$ and proceed as in the computation of $\hat{R}^{(n,k)}$. To simulate samples of $\mathcal{C}$ we use the acceptance-rejection method.

To show the convergence of $R_1^{(n,\infty)}$ to $\mathcal{R}^*$, we let $n = 10$, 100 and 10000. The values of the other parameters are $\alpha = 1.5$, $\beta = 2.5$, $\mathbb{E}[\mathscr{N}] = \mathbb{E}[\mathscr{D}] = 2$, $c = 0.3$. For the TBT, we simulate up to $k_n = \lfloor \log n \rfloor$ generations. For the WBP, we simulate 10 generations. For each n, we draw 1000 samples of $R_1^{(n,\infty)}$, $R_1^{(n,k_n)}$, $\hat{R}^{(n,k_n)}$ and $\mathcal{R}^*$, respectively, to approximate the distribution of these quantities.

Figure 3.3 shows the empirical CDFs of 1000 i.i.d. samples of the true PageRank, $R_1^{(n,\infty)}$; finitely many iterations of PageRank, $R_1^{(n,k_n)}$; and the TBT approximation $\hat{R}^{(n,k_n)}$; it also plots the distribution of the limit $\mathcal{R}^*$ using 1000 simulations. The approximations are so accurate that the CDFs are almost indistinguishable. Figure 3.4 illustrates the weak

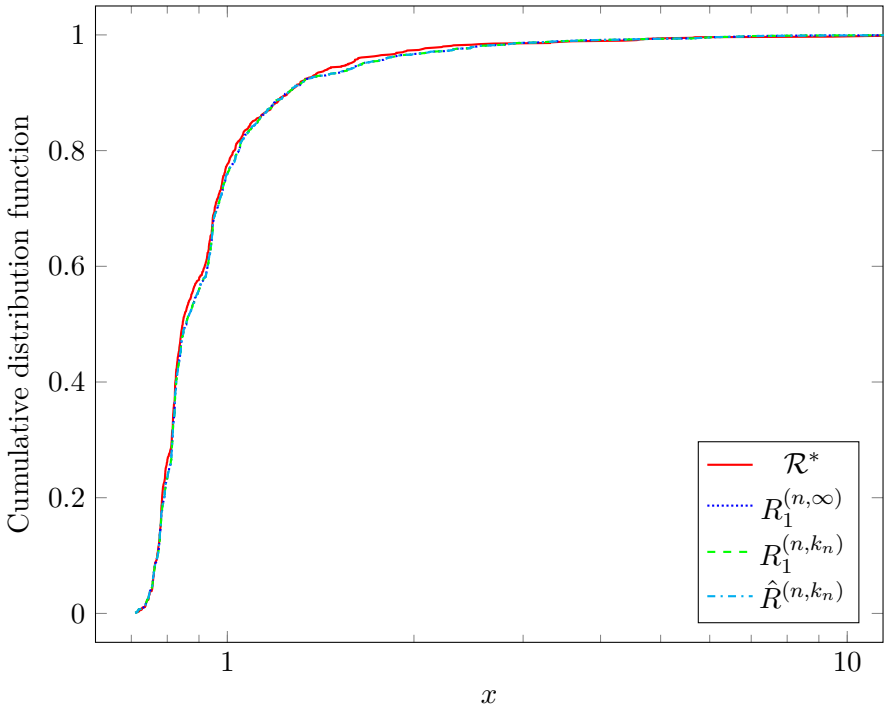

Figure 3.3: The empirical CDFs of 1000 samples of $\mathcal{R}^*$, $R_1^{(n,\infty)}$, $R_1^{(n,k_n)}$ and $\hat{R}^{(n,k_n)}$ for $n = 10000$ and $k_n = 9$.

convergence of PageRank on the graph, $R_1^{(n,\infty)}$, to its limit $\mathcal{R}^*$ as the size of the graph grows.

To quantify the distance between the CDFs, we sort the samples in ascending order and compute the mean squared error (MSE) $\sum_{i=1}^{1000}(x_i^{(n)} - y_i)/1000$, where y_i is the sorted ith sample of $\mathcal{R}^*$ and $x_i^{(n)}$ is the sorted ith sample of $R_1^{(n,\infty)}$. For robustness, we discard the squared error of the maximal value. As a result, the MSEs are 0.2950, 0.1813 and 0.0406 respectively for $n = 10$, 100 and 10000. It is clear that the approximation improves as n increases.

We also test our algorithm on real networks. We use the dataset of all Wikipedia pages in English, German, Italian, Chinese, French, and Dutch. Hence we have the information of six huge networks. Table 3.1 gives some summary statistics of these graphs.

We then compute the PageRank of each graph. This boils down to running the PageRank algorithm (3.3) on the graph for large enough times until the ranks converge. We

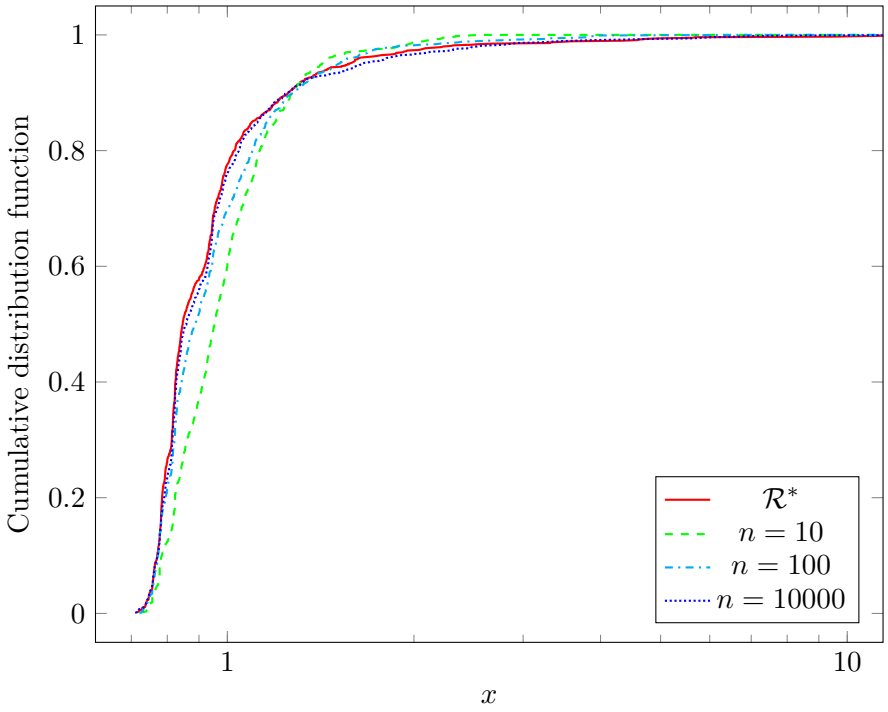

Figure 3.4: The empirical CDFs of 1000 samples of $\mathcal{R}^*$ and $R_1^{(n,\infty)}$ for $n = 10$, 100 and 10000.

compare the PageRank to the WBP approximation. The figures are shown below. We plot the tail empirical distributions in log-log scale. We can see that, even though the network of wikipages contains much more information than the in- and out-degrees, the WBP approximation of the PageRank turns out to be quite accurate.

Language	Node	Avg Degree	In-degree STD	Out-degree STD	Correlation
English	4212493	24.1	413.1	47.9	15%
German	1532978	24.0	247.6	43.8	30%
French	1352825	25.5	334.1	46.2	23%
Italian	1017953	25.2	343.2	49.4	17%
Dutch	1144615	14.0	273.5	36.1	14%
Chinese	663485	16.2	244.3	32.7	22%

Table 3.1: The number of nodes, average in(out)-degree, the standard deviation of in- and out-degrees, and the correlation between in- and out-degrees.

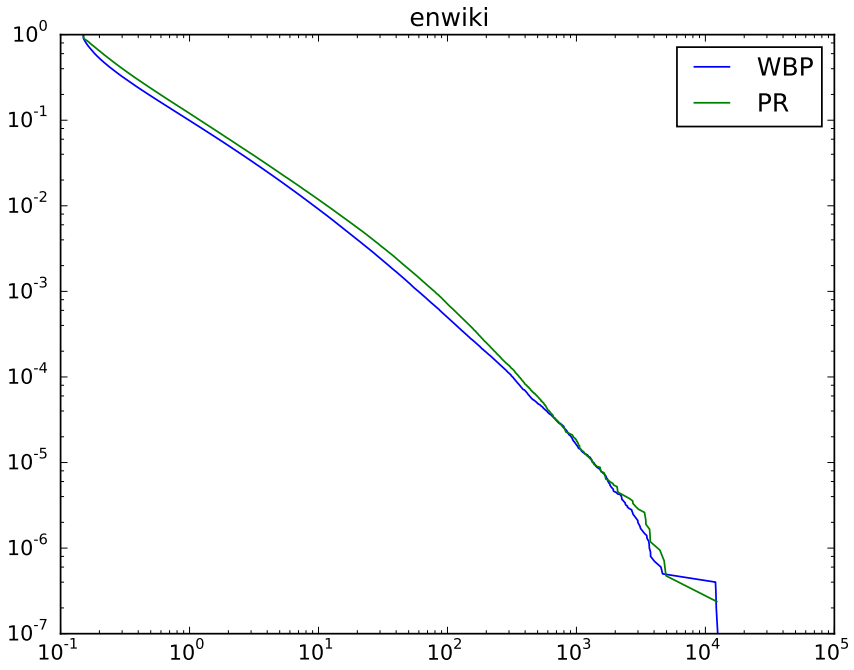

Figure 3.5: English wikipages.

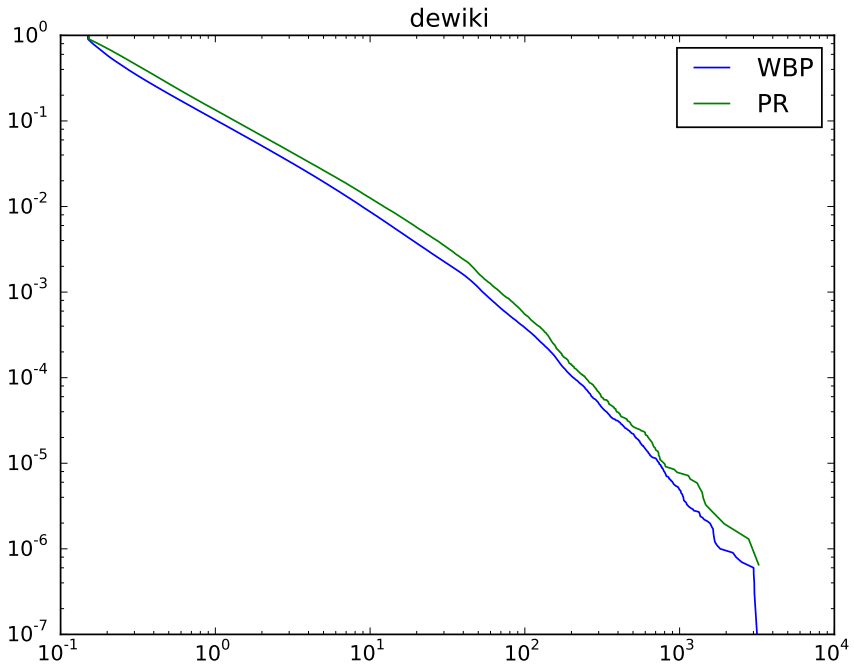

Figure 3.6: German wikipages.

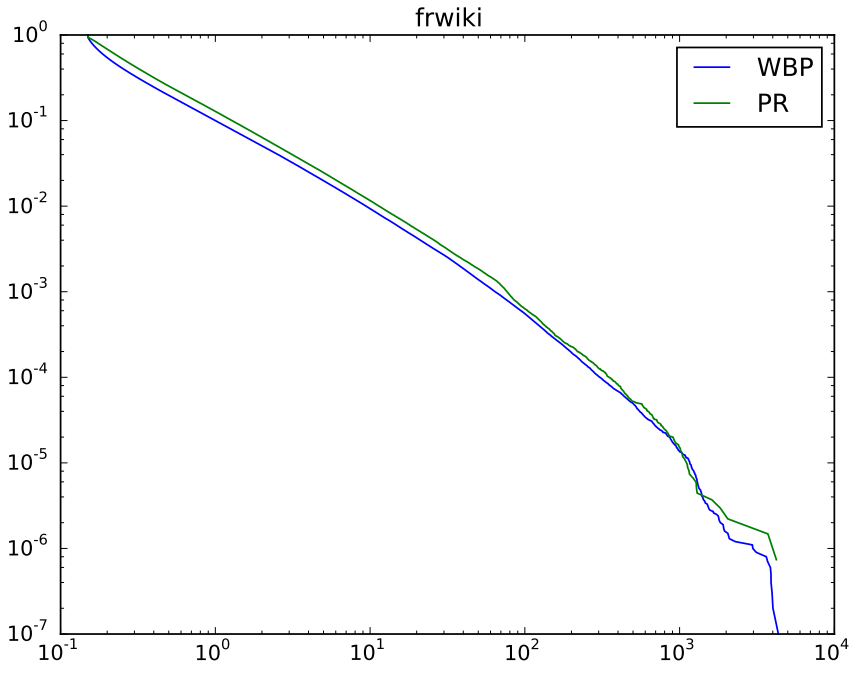

Figure 3.7: French wikipages.

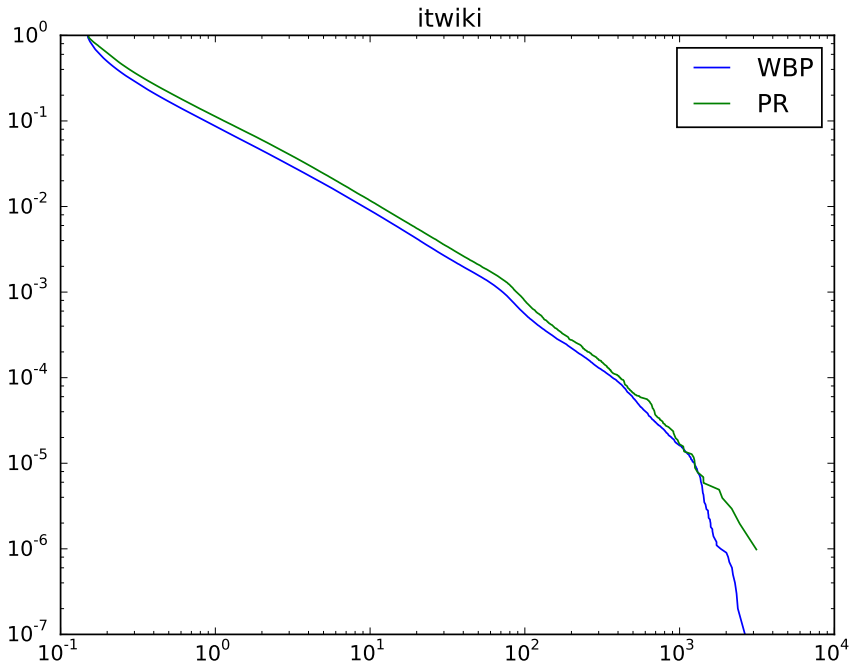

Figure 3.8: Italian wikipages.

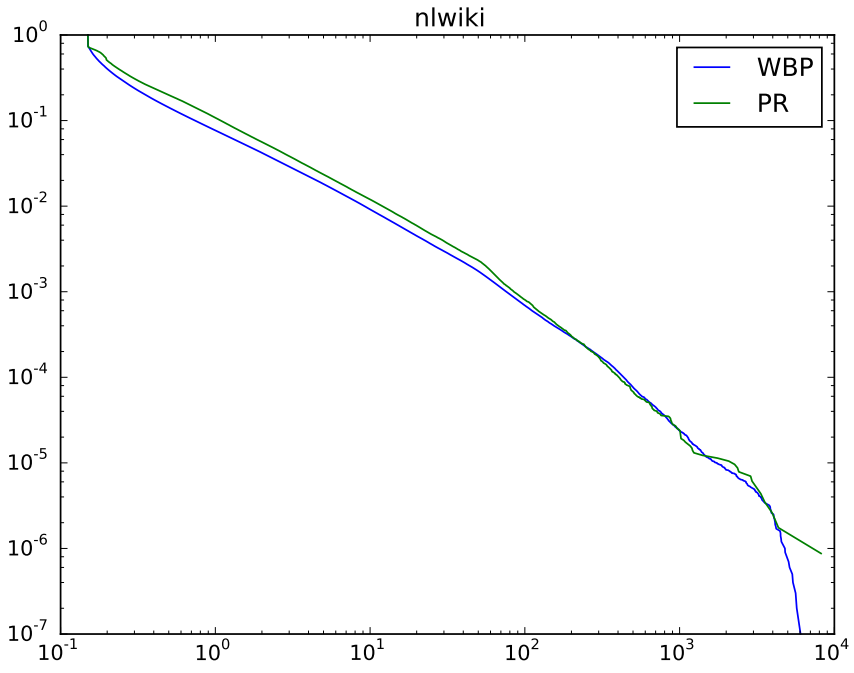

Figure 3.9: Dutch wikipages.

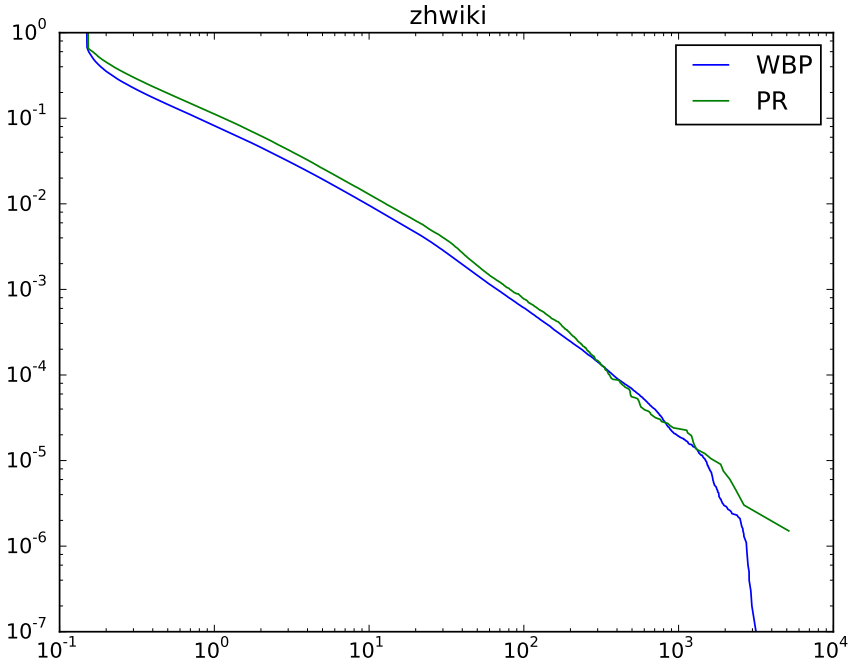

Figure 3.10: Chinese wikipages.

Part III

Simulation of branching recursions

Chapter 4

Efficient simulation for branching linear recursions

This chapter is organized as follows. Section 4.1 describes the weighted branching process and the linear recursion. The algorithm itself is given in Section 4.2 . Section 4.3 introduces the Kantorovich-Rubinstein distance and proves the convergence properties of our proposed algorithm. Numerical examples to illustrate the precision of the algorithm are presented in Section 4.4.

4.1 Linear recursions on weighted branching processes

The endogenous solution to

$$R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q, \tag{4.1}$$

can be explicitly constructed on a weighted branching process. To describe the structure of a weighted branching process, let $\mathbb{N}_+ = \{1, 2, 3, \dots\}$ be the set of positive integers and let $U = \bigcup_{k=0}^{\infty} (\mathbb{N}_+)^k$ be the set of all finite sequences $\mathbf{i} = (i_1, i_2, \dots, i_n)$, $n \geq 0$, where by convention $\mathbb{N}_+^0 = \{\emptyset\}$ contains the null sequence $\emptyset$. To ease the exposition, we will use $(\mathbf{i}, j) = (i_1, \dots, i_n, j)$ to denote the index concatenation operation.

Next, let $(Q, N, C_1, C_2, \dots)$ be a real-valued vector with $N \in \mathbb{N}$. We will refer to this vector as the generic branching vector. Now let $\{(Q_{\mathbf{i}}, N_{\mathbf{i}}, C_{(\mathbf{i},1)}, C_{(\mathbf{i},2)}, \dots)\}_{\mathbf{i} \in U}$ be a sequence

of i.i.d. copies of the generic branching vector. To construct a weighted branching process we start by defining a tree as follows: let $A_0 = \{\emptyset\}$ denote the root of the tree, and define the nth generation according to the recursion

$$A_n = \{(\mathbf{i}, i_n) \in U : \mathbf{i} \in A_{n-1}, 1 \le i_n \le N_{\mathbf{i}}\}, \quad n \ge 1.$$

Now, assign to each node $\mathbf{i}$ in the tree a weight $\Pi_{\mathbf{i}}$ according to the recursion

$$\Pi_\emptyset \equiv 1, \qquad \Pi_{(\mathbf{i},i_n)} = C_{(\mathbf{i},i_n)}\Pi_{\mathbf{i}}, \qquad n \ge 1,$$

see Figure 3.2. If $P(N < \infty) = 1$ and $C_i \equiv 1$ for all $i \ge 1$, the weighted branching process reduces to a Galton-Watson process.

For a weighted branching process with generic branching vector $(Q, N, C_1, C_2, \dots)$, define the process $\{R^{(k)} : k \ge 0\}$ as follows:

$$R^{(k)} = \sum_{j=0}^{k} \sum_{\mathbf{i} \in A_j} Q_{\mathbf{i}}\Pi_{\mathbf{i}}, \quad k \ge 0. \tag{4.2}$$

By focusing on the branching vector belonging to the root node, i.e., $(Q_\emptyset, N_\emptyset, C_1, C_2, \dots)$ we can see that the process $\{R^{(k)}\}$ satisfies the distributional equations

$$\begin{aligned} R^{(0)} &= Q_\emptyset \stackrel{\mathcal{D}}{=} Q \\ R^{(k)} &= \sum_{r=1}^{N_\emptyset} C_r \left(\sum_{j=1}^{k} \sum_{(r,\mathbf{i}) \in A_j} Q_{(r,\mathbf{i})}\Pi_{(r,\mathbf{i})}/C_r \right) + Q_\emptyset \stackrel{\mathcal{D}}{=} \sum_{r=1}^{N} C_r R_r^{(k-1)} + Q, \qquad k \ge 1, \end{aligned} \tag{4.3}$$

where $R_r^{(k-1)}$ are i.i.d. copies of $R^{(k-1)}$, all independent of $(Q, N, C_1, C_2, \dots)$. Here and throughout the paper the convention is that $XY/Y \equiv 1$ if $Y = 0$. Moreover, if we define

$$R = \sum_{j=0}^{\infty} \sum_{\mathbf{i} \in A_j} Q_{\mathbf{i}}\Pi_{\mathbf{i}}, \tag{4.4}$$

we have the following result. We use $x \vee y$ to denote the maximum of x and y.

Proposition 22. *Let $\beta \ge 1$ be such that $E[|Q|^\beta] < \infty$ and $E\left[\left(\sum_{i=1}^{N} |C_i|\right)^\beta\right] < \infty$. In addition, assume either (i) $(\rho_1 \vee \rho_\beta) < 1$, or (ii) $\beta = 2$, $\rho_1 = 1$, $\rho_\beta < 1$ and $E[Q] = 0$. Then, there exist constants $K_\beta > 0$ and $0 < c_\beta < 1$ such that for $R^{(k)}$ and R defined according to* (4.2) *and* (4.4)*, respectively, we have*

$$\sup_{k \ge 0} E\left[|R^{(k)}|^\beta\right] \le K_\beta < \infty \qquad \text{and} \qquad E\left[|R^{(k)} - R|^\beta\right] \le K_\beta c_\beta^{k+1}.$$

Proof. For the case $\rho_1 \vee \rho_\beta < 1$, Lemma 4.4 in [Jelenković and Olvera-Cravioto, 2012a] gives that for $W_n = \sum_{\mathbf{i}\in A_n} Q_{\mathbf{i}}\Pi_{\mathbf{i}}$ and some finite constant H_β we have

$$E\left[|W_n|^\beta\right] \leq H_\beta(\rho_1 \vee \rho_\beta)^n.$$

Let $c_\beta = \rho_1 \vee \rho_\beta$. Minkowski's inequality then gives

$$\left|\left|R^{(k)}\right|\right|_\beta \leq \sum_{n=0}^{k} ||W_n||_\beta \leq \sum_{n=0}^{\infty}\left(H_\beta c_\beta^n\right)^{1/\beta} = \left(\frac{H_\beta}{1-c_\beta^{1/\beta}}\right)^{1/\beta} \triangleq (K_\beta)^{1/\beta} < \infty.$$

Similarly,

$$\left|\left|R^{(k)} - R\right|\right|_\beta \leq \sum_{n=k+1}^{\infty} ||W_n||_\beta \leq \sum_{n=k+1}^{\infty}\left(H_\beta c_\beta^n\right)^{1/\beta} = c_\beta^{(k+1)/\beta}\left(\frac{H_\beta}{1-(\rho_1\vee\rho_\beta)^{1/\beta}}\right)^{1/\beta} = \left(K_\beta c_\beta^{k+1}\right)^{1/\beta}.$$

For the case $\beta = 2$, $\rho_1 = 1$, $\rho_\beta < 1$ and $E[Q] = 0$ we have that

$$E\left[W_n^2\right] = E\left[\left(\sum_{r=1}^{N_\emptyset} C_r W_{n-1,r}\right)^2\right] = E\left[\sum_{r=1}^{N_\emptyset} C_r^2 (W_{n-1,r})^2 + \sum_{1\leq r\neq s\leq N_\emptyset} C_r C_s W_{n-1,r}W_{n-1,s}\right],$$

where $W_{n-1,r} = \sum_{(r,\mathbf{i})\in A_n} Q_{(r,\mathbf{i})}\Pi_{(r,\mathbf{i})}/C_r$, and the $\{W_{n-1,r}\}_{r\geq 1}$ are i.i.d. copies of W_{n-1}, independent of $(N_\emptyset, C_1, C_2, \dots)$. Since $E[W_n] = 0$ for all $n \geq 0$, it follows that

$$E[W_n^2] = \rho_2 E[W_{n-1}^2] = \rho_2^n E[W_0^2] = \text{Var }(Q)\rho_2^n.$$

The two results now follow from the same arguments used above with $H_2 = \text{Var }(Q)$ and $c_2 = \rho_2$. □

It follows from the previous result that under the conditions of Proposition 22, $R^{(k)}$ converges to R both almost surely and in L^β-norm. Similarly, if we ignore the Q in the generic branching vector, assume that $C_i \geq 0$ for all i, and define the process

$$W^{(k)} = \sum_{\mathbf{i}\in A_k}\Pi_{\mathbf{i}} = \sum_{r=1}^{N_\emptyset} C_r\left(\sum_{(r,\mathbf{i})\in A_k}\Pi_{(r,\mathbf{i})}/C_r\right) \overset{\mathcal{D}}{=} \sum_{r=1}^{N} C_r W_r^{(k-1)},$$

where the $\{W_r^{(k-1)}\}_{r\geq 1}$ are i.i.d. copies of $W^{(k-1)}$ independent of $(N, C_1, C_2, \dots)$, then it can be shown that $\{W^{(k)}/\rho_1^k : k \geq 0\}$ defines a nonnegative martingale which converges almost surely to the endogenous solution of the stochastic fixed-point equation

$$W \overset{\mathcal{D}}{=} \sum_{i=1}^{N}\frac{C_i}{\rho_1}W_i,$$

where the $\{W_i\}_{i\geq 1}$ are i.i.d. copies of W, independent of $(N, C_1, C_2, \dots)$. We refer to this equation as the homogeneous case.

As mentioned in the introduction, our objective is to generate a sample of $R^{(k)}$ for values of k sufficiently large to suitably approximate R. Our proposed algorithm can also be used to simulate $W^{(k)}$, but due to space limitations we will omit the details.

4.2 The algorithm

Note that based on (4.2), one way to simulate $R^{(k)}$ would be to simulate a weighted branching process starting from the root and up to the k generation and then add all the weights $Q_{\mathbf{i}}\Pi_{\mathbf{i}}$ for $\mathbf{i} \in \bigcup_{j=0}^{k} A_j$. Alternatively, we could generate a large enough pool of i.i.d. copies of Q which would represent the $Q_{\mathbf{i}}$ for $\mathbf{i} \in A_k$, and use them to generate a pool of i.i.d. observations of $R^{(1)}$ by setting

$$R_i^{(1)} = \sum_{r=1}^{N_i} C_{(i,r)} R_r^{(0)} + Q_i,$$

where $\{(Q_i, N_i, C_{(i,1)}, C_{(i,2)}, \dots)\}_{i\geq 1}$ are i.i.d. copies of the generic branching vector, independent of everything else, and the $R_r^{(0)}$ are the Q's generated in the previous step. We can continue this process until we get to the root node. On average, we would need $(E[N])^k$ i.i.d. copies of Q for the first pool of observations, $(E[N])^{k-1}$ copies of the generic branching vector for the second pool, and in general, $(E[N])^{k-j}$ for the jth step. This approach is equivalent to simulating the weighted branching process starting from the kth generation and going up to the root, and is the result of iterating

$$R^{(k+1)} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i^{(k)} + Q.$$

Our proposed algorithm is based on this "leaves to root" approach, but to avoid the need for a geometric number of "leaves", we will resample from the initial pool to obtain a pool of the same size of observations of $R^{(1)}$. In general, for the jth generation we will sample from the pool obtained in the previous step of (approximate) observations of $R^{(j-1)}$ to obtain conditionally independent (approximate) copies of $R^{(j)}$. In other words, to obtain a pool of approximate copies of $R^{(j)}$ we bootstrap from the pool previously obtained of

approximate copies of $R^{(j-1)}$. The approximation lies in the fact that we are not sampling from $R^{(j-1)}$ itself, but from a finite sample of conditionally independent observations that are only approximately distributed as $R^{(j-1)}$. The algorithm is described below.

Let $(Q, N, C_1, C_2, \dots)$ denote the generic branching vector defining the weighted branching process. Let k be the depth of the recursion that we want to simulate, i.e., the algorithm will produce a sample of random variables approximately distributed as $R^{(k)}$. Choose $m \in \mathbb{N}_+$ to be the bootstrap sample size. For each $0 \le j \le k$, the algorithm outputs $\mathcal{P}^{(j,m)} \triangleq \left(\hat{R}_1^{(j,m)}, \hat{R}_2^{(j,m)}, \dots, \hat{R}_m^{(j,m)}\right)$, which we refer to as the sample pool at level j.

1. *Initialize*: Set $j = 0$. Simulate a sequence $\{Q_i\}_{i=1}^m$ of i.i.d. copies of Q and let $\hat{R}_i^{(0,m)} = Q_i$ for $i = 1, \dots, m$. Output $\mathcal{P}^{(0,m)} = \left(\hat{R}_1^{(0,m)}, \hat{R}_2^{(0,m)}, \dots, \hat{R}_m^{(0,m)}\right)$ and update $j = 1$.

2. While $j \le k$:

   i) Simulate a sequence $\{(Q_i, N_i, C_{(i,1)}, C_{(i,2)}, \dots)\}_{i=1}^m$ of i.i.d. copies of the generic branching vector, independent of everything else.

   ii) Let
   $$\hat{R}_i^{(j,m)} = \sum_{r=1}^{N_i} C_{(i,r)} \hat{R}_{(i,r)}^{(j-1,m)} + Q_i, \qquad i = 1, \dots, m, \tag{4.5}$$
   where the $\hat{R}_{(i,r)}^{(j-1,m)}$ are sampled uniformly with replacement from the pool $\mathcal{P}^{(j-1,m)}$.

   iii) Output $\mathcal{P}^{(j,m)} = \left(\hat{R}_1^{(j,m)}, \hat{R}_2^{(j,m)}, \dots, \hat{R}_m^{(j,m)}\right)$ and update $j = j + 1$.

Remark 9. *To simulate an approximation for the endogenous solution to the maximum equation* (1.5)*, given by* $R = \bigvee_{j=0}^{\infty} \bigvee_{\mathbf{i} \in A_j} Q_{\mathbf{i}} \Pi_{\mathbf{i}}$*, simply replace* (4.5) *with*

$$\hat{R}_i^{(j,m)} = Q_i \vee \bigvee_{r=1}^{N_i} C_{(i,r)} \hat{R}_{(i,r)}^{(j-1,m)}, \qquad i = 1, \dots, m.$$

Bootstrapping refers broadly to any method that relies on random sampling with replacement [Efron and Tibshirani, 1993]. For example, bootstrapping can be used to estimate the variance of an estimator, by constructing samples of the estimator from a number of resamples of the original dataset with replacement. With the same idea, our algorithm

draws samples uniformly with replacement from the previous bootstrap sample pool. Therefore, the $\hat{R}_{(i,r)}^{(j-1,m)}$ on the right-hand side of (4.5) are only conditionally independent given $\mathcal{P}^{(j-1,m)}$. Hence, the samples in $\mathcal{P}^{(j,m)}$ are identically distributed but not independent for $j \geq 1$.

As we mentioned earlier, the distribution of the $\{\hat{R}_i^{(j,m)}\}$ in $\mathcal{P}^{(j,m)}$ are only approximately distributed as $R^{(j)}$, with the exception of the $\{\hat{R}_i^{(0,m)}\}$ which are exact. The first thing that we need to prove is that the distribution of the observations in $\mathcal{P}^{(j,m)}$ does indeed converge to that of $R^{(j)}$. Intuitively, this should be the case since the empirical distribution of the $\{\hat{R}_i^{(0,m)}\}$ is the empirical distribution of m i.i.d. observations of $R^{(0)}$, and therefore should be close to the true distribution of $R^{(0)}$ for suitably large m. Similarly, since the $\{\hat{R}_i^{(1,m)}\}$ are constructed by sampling from the empirical distribution of $\mathcal{P}^{(0,m)}$, which is close to the true distribution of $R^{(0)}$, then their empirical distribution should be close to the empirical distribution of $R^{(1)}$, which in turn should be close to the true distribution of $R^{(1)}$. Inductively, provided the approximation is good in step $j-1$, we can expect the empirical distribution of $\mathcal{P}^{(j,m)}$ to be close to the true distribution of $R^{(j)}$. In the following section we make the mode of the convergence precise by considering the Kantorovich-Rubinstein distance between the empirical distribution of $\mathcal{P}^{(j,m)}$ and the true distribution of $R^{(j)}$.

The second technical aspect of our proposed algorithm is the lack of independence among the observations in $\mathcal{P}^{(k,m)}$, since a natural estimator for quantities of the form $E[h(R^{(k)})]$ would be to use

$$\frac{1}{m}\sum_{i=1}^{m} h(\hat{R}_i^{(k,m)}). \tag{4.6}$$

Hence, we also provide a result establishing the consistency of estimators of the form in (4.6) for a suitable family of functions h.

We conclude this section by pointing out that the complexity of the algorithm described above is of order km, while the naive Monte Carlo approach has order $(E[N])^k m$. This is a huge gain in efficiency.

4.3 Convergence and consistency

In order to show that our proposed algorithm does indeed produce observations that are approximately distributed as $R^{(k)}$ for any fixed k, we will show that the empirical distribution function of the observations in $\mathcal{P}^{(k,m)}$, i.e.,

$$\hat{F}_{k,m}(x) = \frac{1}{m}\sum_{i=1}^{m} 1(\hat{R}_i^{(k,m)} \leq x)$$

converges as $m \to \infty$ to the true distribution function of $R^{(k)}$, which we will denote by F_k. Recall the definition of the Kantorovich-Rubinstein distance d_1, which is a metric on the space of probability measures. In particular, convergence in this sense is equivalent to weak convergence plus convergence of the first absolute moments.

We point out that d_1 is only strictly speaking a distance when both μ and ν have finite first absolute moments. Moreover, it is well known that

$$d_1(\mu,\nu) = \int_0^1 |F^{-1}(u) - G^{-1}(u)|du = \int_{-\infty}^{\infty} |F(x) - G(x)|dx, \tag{4.7}$$

where F and G are the cumulative distribution functions of μ and ν, respectively, and $f^{-1}(t) = \inf\{x \in \mathbb{R} : f(x) \geq t\}$ denotes the pseudo-inverse of f. It follows that the optimal coupling of two real random variables X and Y is given by $(X,Y) = (F^{-1}(U), G^{-1}(U))$, where U is uniformly distributed in $[0,1]$.

Remark 10. *The Kantorovich-Rubinstein distance is also known as the Wasserstein metric of order 1. In general, both the Kantorovich-Rubinstein distance and the more general Wasserstein metric of order p can be defined in any metric space; we restrict our definition in this paper to the real line since that is all we need. We refer the interested reader to [Villani, 2009] for more details.*

With some abuse of notation, for two distribution functions F and G we use $d_1(F,G)$ to denote the Kantorovich-Rubinstein distance between their corresponding probability measures.

The following proposition shows that for i.i.d. samples, the expected value of the Kantorovich-Rubinstein distance between the empirical distribution function and the true distribution converges to zero.

Proposition 23. *Let $\{X_i\}_{i\geq 1}$ be a sequence of i.i.d. random variables with common distribution F. Let F_n denote the empirical distribution function of a sample of size n. Then, provided there exists $\alpha \in (1,2)$ such that $E\left[|X_1|^\alpha\right] < \infty$, we have that*

$$E\left[d_1(F_n, F)\right] \leq n^{-1+1/\alpha}\left(\frac{2\alpha}{\alpha-1} + \frac{2}{2-\alpha}\right) E[|X_1|^\alpha].$$

Proposition 23 can be proved following the same arguments used in the proof of Theorem 2.2 in [del Barrio *et al.*, 1999] by setting $M = 1$, and thus we omit it.

We now give the main theorem of the paper, which establishes the convergence of the expected Kantorovich-Rubinstein distance between $\hat{F}_{k,m}$ and F_k. Its proof is based on induction and the explicit representation (4.7). Recall that $\rho_\beta = E\left[\sum_{i=1}^N |C_i|^\beta\right]$.

Theorem 24. *Suppose that the conditions of Proposition 22 are satisfied for some $\beta > 1$. Then, for any $\alpha \in (1,2)$ with $\alpha \leq \beta$, there exists a constant $K_\alpha < \infty$ such that*

$$E\left[d_1(\hat{F}_{k,m}, F_k)\right] \leq K_\alpha m^{-1+1/\alpha} \sum_{i=0}^{k} \rho_1^i. \tag{4.8}$$

Proof. By Proposition 22 there exists a constant H_α such that

$$H_\alpha = \sup_{k\geq 0} E\left[|R^{(k)}|^\alpha\right] \leq \sup_{k\geq 0}\left(E\left[|R^{(k)}|^\beta\right]\right)^{\alpha/\beta} < \infty.$$

Set $K_\alpha = H_\alpha\left(\frac{2\alpha}{\alpha-1} + \frac{2}{2-\alpha}\right)$. We will give a proof by induction.

For $j = 0$, we have that

$$\hat{F}_{0,m}(x) = \frac{1}{m}\sum_{i=1}^{m} 1(Q_i \leq x),$$

where $\{Q_i\}_{i\geq 1}$ is a sequence of i.i.d. copies of Q. It follows that $\hat{F}_{0,m}$ is the empirical distribution function of $R^{(0)}$, and by Proposition 23 we have that

$$E\left[d_1(\hat{F}_{0,m}, F_0)\right] \leq K_\alpha m^{-1+1/\alpha}.$$

Now suppose that (4.8) holds for $j-1$. Let $\{U_r^i\}_{i,r\geq 1}$ be a sequence of i.i.d. Uniform$(0,1)$ random variables, independent of everything else. Let $\{(Q_i, N_i, C_{(i,1)}, C_{(i,2)}, \dots)\}_{i\geq 1}$ be a sequence of i.i.d. copies of the generic branching vector, also independent of everything else. Recall that F_{j-1} is the distribution function of $R^{(j-1)}$ and define the random variables

$$\hat{R}_i^{(j,m)} = \sum_{r=1}^{N_i} C_{(i,r)} \hat{F}_{j-1,m}^{-1}(U_r^i) + Q_i \qquad \text{and} \qquad R_i^{(j)} = \sum_{r=1}^{N_i} C_{(i,r)} F_{j-1}^{-1}(U_r^i) + Q_i$$

for each $i = 1, 2, \ldots, m$. Now use these random variables to define

$$\hat{F}_{j,m}(x) = \frac{1}{m}\sum_{i=1}^{m} 1(\hat{R}_i^{(j,m)} \leq x) \qquad \text{and} \qquad F_{j,m}(x) = \frac{1}{m}\sum_{i=1}^{m} 1(R_i^{(j)} \leq x).$$

Note that $F_{j,m}$ is an empirical distribution function of i.i.d. copies of $R^{(j)}$, which has been carefully coupled with the function $\hat{F}_{j,m}$ produced by the algorithm.

By the triangle inequality and Proposition 23 we have that

$$E\left[d_1(\hat{F}_{j,m}, F_j)\right] \leq E\left[d_1(\hat{F}_{j,m}, F_{j,m})\right] + E\left[d_1(F_{j,m}, F_j)\right] \leq E\left[d_1(\hat{F}_{j,m}, F_{j,m})\right] + K_\alpha m^{-1+1/\alpha}.$$

To analyze the remaining expectation note that

$$\begin{aligned}
E\left[d_1(\hat{F}_{j,m}, F_{j,m})\right] &= E\left[\int_{-\infty}^{\infty} |\hat{F}_{j,m}(x) - F_{j,m}(x)| dx\right] \\
&\leq \frac{1}{m}\sum_{i=1}^{m} E\left[\int_{-\infty}^{\infty} \left|1(\hat{R}_i^{(j,m)} \leq x) - 1(R_i^{(j)} \leq x)\right| dx\right] \\
&= \frac{1}{m}\sum_{i=1}^{m} E\left[\left|\hat{R}_i^{(j,m)} - R_i^{(j)}\right|\right] \\
&= \frac{1}{m}\sum_{i=1}^{m} E\left[\left|\sum_{r=1}^{N_i} C_{(i,r)}(\hat{F}_{j-1,m}^{-1}(U_r^i) - F_{j-1}^{-1}(U_r^i))\right|\right] \\
&\leq E\left[\sum_{r=1}^{N} |C_r|\right] E\left[d_1(\hat{F}_{j-1,m}, F_{j-1})\right],
\end{aligned}$$

where in the last step we used the fact that $(N_i, C_{(i,1)}, C_{(i,2)}, \ldots)$ is independent of $\{U_r^i\}_{r\geq 1}$ and of $\hat{F}_{j-1,m}$, combined with the explicit representation of the Kantorovich-Rubinstein distance given in (4.7). The induction hypothesis now gives

$$\begin{aligned}
E\left[d_1(\hat{F}_{j,m}, F_j)\right] &\leq \rho_1 E\left[d_1(\hat{F}_{j-1,m}, F_{j-1})\right] + K_\alpha m^{-1+1/\alpha} \\
&\leq K_\alpha m^{-1+1/\alpha}\rho_1 \sum_{i=0}^{j-1} \rho_1^i + K_\alpha m^{-1+1/\alpha} \\
&= K_\alpha m^{-1+1/\alpha} \sum_{i=0}^{j} \rho_1^i.
\end{aligned}$$

This completes the proof. □

Note that the proof of Theorem 24 implies that $\hat{R}_i^{(j,m)} \to R_i^{(j)} = \sum_{r=1}^{N_i} C_{(i,r)} F_{j-1}^{-1}(U_r^i) + Q_i \stackrel{\mathcal{D}}{=} R^{(j)}$ in L^1-norm for all fixed $j \in \mathbb{N}$, and hence in distribution. In other words,

$$P\left(\hat{R}_i^{(k,m)} \leq x\right) \to F_k(x) \qquad \text{as } m \to \infty, \tag{4.9}$$

for all $i = 1, 2, \ldots, m$, and for any continuity point of F_k. This also implies that

$$E\left[\hat{F}_{k,m}(x)\right] = P\left(\hat{R}_1^{(k,m)} \leq x\right) \to F_k(x) \qquad \text{as } m \to \infty, \tag{4.10}$$

for all continuity points of F_k.

Since our algorithm produces a pool $\mathcal{P}^{(k,m)}$ of m random variables approximately distributed according to F_k, it makes sense to use it for estimating expectations related to $R^{(k)}$. In particular, we are interested in estimators of the form in (4.6). The problem with this kind of estimators is that the random variables in $\mathcal{P}^{(k,m)}$ are only conditionally independent given $\hat{F}_{k-1,m}$.

Definition 11. *We say that Θ_n is a consistent estimator for θ if $\Theta_n \xrightarrow{P} \theta$ as $n \to \infty$, where $\xrightarrow{P}$ denotes convergence in probability.*

Our second theorem shows the consistency of estimators of the form in (4.6) for a broad class of functions.

Theorem 25. *Suppose that the conditions of Proposition 22 are satisfied for some $\beta > 1$. Suppose $h : \mathbb{R} \to \mathbb{R}$ is continuous and $|h(x)| \leq C(1 + |x|)$ for all $x \in \mathbb{R}$ and some constant $C > 0$. Then, the estimator*

$$\frac{1}{m}\sum_{i=1}^{m} h(\hat{R}_i^{(k,m)}) = \int_{\mathbb{R}} h(x) d\hat{F}_{k,m}(x),$$

where $\mathcal{P}^{(k,m)} = \left(\hat{R}_1^{(k,m)}, \hat{R}_2^{(k,m)}, \ldots, \hat{R}_m^{(k,m)}\right)$, is a consistent estimator for $E[h(R^{(k)})]$.

Proof. For any $M > 0$, define $h_M(x)$ as

$$h_M(x) = h(-M)1(x \leq -M) + h(x)1(-M < x \leq M) + h(M)1(x > M),$$

and note that h_M is uniformly continuous. We then have

$$\begin{aligned}\left|\int_{\mathbb{R}} h(x) d\hat{F}_{k,m}(x) - \int_{\mathbb{R}} h(x) dF_k(x)\right| &\leq 2C\int_{|x|>M}(1+|x|)dF_k(x) + 2C\int_{|x|>M}(1+|x|)d\hat{F}_{k,m}(x) \\ &\quad + \left|\int_{\mathbb{R}} h_M(x) d\hat{F}_{k,m}(x) - \int_{\mathbb{R}} h_M(x) dF_k(x)\right|. \end{aligned} \tag{4.11}$$

Fix $\epsilon > 0$ and choose $M_\epsilon > 0$ such that $E\left[(|R^{(k)}| + 1)1(|R^{(k)}| > M_\epsilon)\right] \leq \epsilon/(4C)$ and such that $-M_\epsilon$ and M_ϵ are continuity points of F_k. Define $(\hat{R}^{(k,m)}, R^{(k)}) = (\hat{F}_{k,m}^{-1}(U), F_k^{-1}(U))$,

where U is a uniform $[0,1]$ random variable independent of $\mathcal{P}^{(k,m)}$. Next, note that $g(x) = 1 + |x|$ is Lipschitz continuous with Lipschitz constant one and therefore

$$\begin{aligned}
\int_{|x|>M_\epsilon} (1+|x|)d\hat{F}_{k,m}(x) &= (1+M_\epsilon)\left(\hat{F}_{k,m}(-M_\epsilon) + 1 - \hat{F}_{k,m}(M_\epsilon)\right) \\
&\quad + \int_{x<-M_\epsilon} \hat{F}_{k,m}(x)\,dx + \int_{x>M_\epsilon} (1-\hat{F}_{k,m}(x))dx \\
&\le (1+M_\epsilon)\left(\hat{F}_{k,m}(-M_\epsilon) + 1 - \hat{F}_{k,m}(M_\epsilon)\right) + d_1(\hat{F}_{k,m}, F_k) \\
&\quad + \int_{x<-M_\epsilon} F_k(x)\,dx + \int_{x>M_\epsilon} (1-F_k(x))dx \\
&= (1+M_\epsilon)\left(\hat{F}_{k,m}(-M_\epsilon) - F_k(-M_\epsilon) + F_k(M_\epsilon) - \hat{F}_{k,m}(M_\epsilon)\right) + d_1(\hat{F}_{k,m}, F_k) \\
&\quad + E\left[(|R^{(k)}|+1)1(|R^{(k)}|>M_\epsilon)\right].
\end{aligned}$$

Finally, since h_{M_ϵ} is bounded and uniformly continuous, then $\omega(\delta) = \sup\{|h_{M_\epsilon}(x) - h_{M_\epsilon}(y)| : |x-y| \le \delta\}$ converges to zero as $\delta \to 0$. Hence, for any $\gamma > 0$,

$$\begin{aligned}
\left|\int_{\mathbb{R}} h_{M_\epsilon}(x)d\hat{F}_{k,m}(x) - \int_{\mathbb{R}} h_{M_\epsilon}(x)dF_k(x)\right| &\le E\left[\left|h_{M_\epsilon}(\hat{R}^{(k,m)}) - h_{M_\epsilon}(R^{(k)})\right| \hat{F}_{k,m}\right] \\
&\le \omega(m^{-\gamma}) + K_\epsilon E\left[1\left(|\hat{R}^{(k,m)} - R^{(k)}| > m^{-\gamma}\right)\Big| \hat{F}_{k,m}\right] \\
&\le \omega(m^{-\gamma}) + K_\epsilon m^\gamma d_1(\hat{F}_{k,m}, F_k),
\end{aligned}$$

where $2K_\epsilon = \sup\{|h_{M_\epsilon}(x)| : x \in \mathbb{R}\}$. Choose $0 < \gamma < 1 - 1/\alpha$ for the $\alpha \in (1,2)$ in Theorem 24 and combine the previous estimates to obtain

$$\begin{aligned}
&E\left[\left|\int_{\mathbb{R}} h(x)d\hat{F}_{k,m}(dx) - \int_{\mathbb{R}} h(x)dF_k(dx)\right|\right] \\
&\le 2C(1+M_\epsilon)\left(E[\hat{F}_{k,m}(-M_\epsilon)] - F_k(-M_\epsilon) + F_k(M_\epsilon) - E[\hat{F}_{k,m}(M_\epsilon)]\right) \\
&\quad + \epsilon + \omega(m^{-\gamma}) + (2C + K_\epsilon m^\gamma)E\left[d_1(\hat{F}_{k,m}, F_k)\right].
\end{aligned}$$

Since

$$E[\hat{F}_{k,m}(-M_\epsilon)] \to F_k(-M_\epsilon) \qquad \text{and} \qquad E[\hat{F}_{k,m}(M_\epsilon)] \to F_k(M_\epsilon)$$

by (4.10), and $m^\gamma E\left[d_1(\hat{F}_{k,m}, F_k)\right] \to 0$ by Theorem 24, it follows that

$$\limsup_{m\to\infty} E\left[\left|\int_{\mathbb{R}} h(x)d\hat{F}_{k,m}(dx) - \int_{\mathbb{R}} h(x)dF_k(dx)\right|\right] \le \epsilon.$$

Since $\epsilon > 0$ was arbitrary, the convergence in L^1, and therefore in probability, follows. $\square$

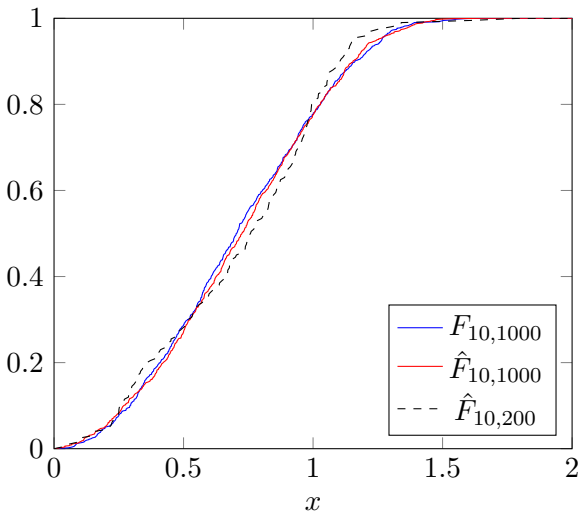

Figure 4.1: The functions $F_{10,1000}(x)$, $\hat{F}_{10,200}(x)$ and $\hat{F}_{10,1000}(x)$.

4.4 Numerical examples

This last section of the paper gives a numerical example to illustrate the performance of our algorithm. Consider a generic branching vector $(Q, N, C_1, C_2, \dots)$ where the $\{C_i\}_{i\geq 1}$ are i.i.d. and independent of N and Q, with N also independent of Q.

Figure 4.1 plots the empirical cumulative distribution function of 1000 samples of $R^{(10}$, i.e., $F_{10,1000}$ in our notation, versus the functions $\hat{F}_{10,200}$ and $\hat{F}_{10,1000}$ produced by our algorithm, for the case where the C_i are uniformly distributed in $[0, 0.2]$, Q uniformly distributed in $[0, 1]$ and N is a Poisson random variable with mean 3. Note that we cannot compare our results with the true distribution F_{10} since it is not available in closed form. Computing $F_{10,1000}$ required 883.3 seconds using Python with an Intel i7-4700MQ 2.40 GHz processor and 8 GB of memory, while computing $\hat{F}_{10,1000}$ required only 2.1 seconds. We point out that in applications to information ranking algorithms $E[N]$ can be in the thirties range, which would make the difference in computation time even more impressive.

Our second example plots the tail distribution of the empirical cumulative distribution function of $R^{(10)}$ for 10,000 samples versus the tail of $\hat{F}_{10,10000}$ for an example where N is a zeta random varialbe with a probability mass function $P(N = k) \propto k^{-2.5}$, Q is an exponential random variable with mean 1, and the C_i have a uniform distribution in $[0, 0.5]$.

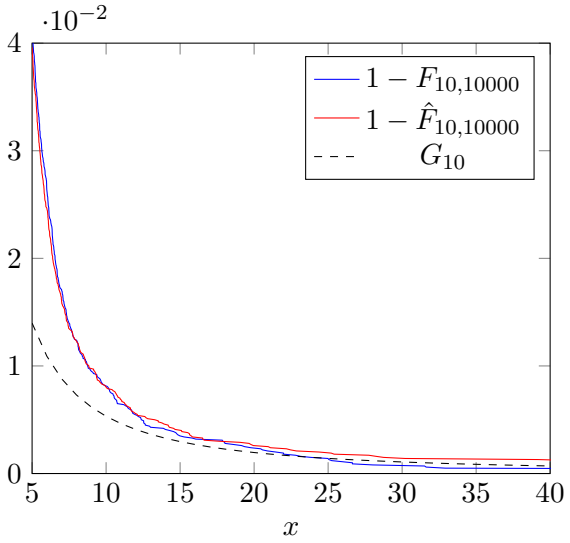

Figure 4.2: The functions $1-F_{10,10000}(x)$, $1-\hat{F}_{10,10000}(x)$ and $G_{10}(x)$, where G_{10} is evaluated only at integer values of x and linearly interpolated in between.

In this case the exact asymptotics for $P(R^{(k)} > x)$ as $x \to \infty$ are given by

$$P(R^{(k)} > x) \sim \frac{(E[C_1]E[Q])^\alpha}{(1-\rho_1)^\alpha} \sum_{j=0}^{k} \rho_\alpha^j (1-\rho_1^{k-j})^\alpha P(N > x),$$

where $P(N > x) = x^{-\alpha}L(x)$ is regularly varying (see Lemma 5.1 in [Jelenković and Olvera-Cravioto, 2010]), which reduces for the specific distributions we have chosen to

$$G_{10}(x) \triangleq \frac{(0.25)^{2.5}}{(1-(0.49))^{2.5}} \sum_{j=0}^{10} (0.07)^j (1-(0.49)^{10-j})^{2.5} P(N > x) = (0.365)\, P(N > x).$$

Figure 4.2 plots the complementary distributions of $F_{10,10000}$, $\hat{F}_{10,10000}$ and compares them to G. We can see that the tails of both $F_{10,10000}$ and $\hat{F}_{10,10000}$ approach the asymptotic roughly at the same time.

Part IV

Bibliography

Bibliography

[Abou-Chacra *et al.*, 1973] R. Abou-Chacra, D.J. Thouless, and P.W. Anderson. A self-consistent theory of localization. *Journal of Physics C: Solid State Physics*, 6(10):1734, 1973.

[Albert and Barabási, 1999] R. Albert and A. L. Barabási. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[Albert and Barabási, 2002] R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[Aldous and Bandyopadhyay, 2005] D.J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equation. *Ann. Appl. Probab.*, 15(2):1047–1110, 2005.

[Alsmeyer and Meiners, 2012] G. Alsmeyer and M. Meiners. Fixed points of inhomogeneous smoothing transforms. *Journal of Difference Equations and Applications*, 18(8):1287–1304, 2012.

[Alsmeyer and Meiners, 2013] G. Alsmeyer and M. Meiners. Fixed points of the smoothing transform: Two-sided solutions. *Probab. Theory Rel.*, 155(1-2):165–199, 2013.

[Alsmeyer *et al.*, 2012] G. Alsmeyer, J.D. Biggins, and M. Meiners. The functional equation of the smoothing transform. *Ann. Probab.*, 40(5):2069–2105, 2012.

[Amini *et al.*, 2013] H. Amini, R. Cont, and A. Minca. Resilience to contagion in financial networks. *Mathematical finance*, 2013.

[Andersen *et al.*, 2006] R. Andersen, F. Chung, and K. Lang. Local graph partitioning using PageRank vectors. In *Proceedings of FOCS2006*, pages 475–486, 2006.

[Arratia and Liggett, 2005] R. Arratia and T.M. Liggett. How likely is an i.i.d. degree sequence to be graphical? *Ann. Appl. Probab.*, 15(1B):652–670, 2005.

[Bender and Canfield, 1978] E.A. Bender and E.R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Comb. Theory A*, 24(3):296–307, 1978.

[Berge, 1976] C. Berge. *Graphs and hypergraphs*, volume 6. Elsevier, 1976.

[Bhamidi *et al.*, 2012] S. Bhamidi, J.M. Steele, and T. Zaman. Twitter event networks and the superstar model. *Preprint*, 2012. arXiv: 0902.0885.

[Biggins, 1998] J.D. Biggins. Lindley-type equations in the branching random walk. *Stochastic Process. Appl.*, 75:105–133, 1998.

[Bingham *et al.*, 1987] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular variation.* Cambridge University Press, Cambridge, 1987.

[Blanchet and Stauffer, 2013] Jose Blanchet and Alexandre Stauffer. Characterizing optimal sampling of binary contingency tables via the configuration model. *Random Structures & Algorithms*, 42(2):159–184, 2013.

[Blitzstein and Diaconis, 2011] J. Blitzstein and P. Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4):489–522, 2011.

[Boccaletti *et al.*, 2006] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.

[Boldi and Vigna, 2014] P. Boldi and S. Vigna. Axioms for centrality. *To Appear in Internet Mathematics*, 2014.

[Bollobás *et al.*, 2001] B. Bollobás, O. Riordan, J. Spencer, G. Tusnády, et al. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[Bollobás, 1980] 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.

[Brin and Page, 1998] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Comput. Networks ISDN Systems*, 30(1-7):107–117, 1998.

[Britton *et al.*, 2006] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, 124(6):1377–1397, 2006.

[Broder *et al.*, 2000] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Comput. Netw.*, 33:309–320, 2000.

[Chen and Olvera-Cravioto, 2013] N. Chen and M. Olvera-Cravioto. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013.

[Chen and Olvera-Cravioto, 2014] N. Chen and M. Olvera-Cravioto. Coupling on weighted branching trees. *Submitted*, 2014.

[Chen and Olvera-Cravioto, 2015] N. Chen and M. Olvera-Cravioto. Efficient simulation for branching linear recursions. *Submitted*, 2015.

[Chen *et al.*, 2007] 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.

[Chen *et al.*, 2014] N. Chen, N. Litvak, and M. Olvera-Cravioto. Ranking algorithms on directed configuration networks. *Submitted*, 2014.

[Chung and Lu, 2002a] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci.*, 99:15879–15882, 2002.

[Chung and Lu, 2002b] F. Chung and L. Lu. Connected components in random graphs with given degree sequences. *Ann. Comb.*, 6:125–145, 2002.

[Constantine and Gleich, 2009] P.G. Constantine and D.F. Gleich. Random alpha pagerank. *Internet Mathematics*, 6(2):189–236, 2009.

[del Barrio *et al.*, 1999] Eustasio del Barrio, Evarist Giné, and Carlos Matrán. Central limit theorems for the wasserstein distance between the empirical and the true distributions. *Annals of Probability*, pages 1009–1071, 1999.

[Durrett, 2007] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistics and Probabilistic Mathematics. Cambridge university press Cambridge, 2007.

[Durrett, 2010] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2010.

[Efron and Tibshirani, 1993] B. Efron and R. J. Tibshirani. *An introductin to the bootstrap*. 1993.

[Erdös and Gallai, 1960] P. Erdös and T. Gallai. Graphs with given degree of vertices. *Mat. Lapok.*, 11:264–274, 1960.

[Erdős and Rényi, 1959] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[Erdös *et al.*, 2010] P. L. Erdös, I. Miklós, and Z. Toroczkai. A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs. *Electron J. Comb.*, 17:1–10, 2010.

[Fill and Janson, 2001] J.A. Fill and S. Janson. Approximating the limiting Quicksort distribution. *Random Structures Algorithms*, 19(3-4):376–406, 2001.

[Fortunato *et al.*, 2008] S. Fortunato, M. Boguñá, A. Flammini, and F. Menczer. Approximating PageRank from in-degree. In *Algorithms and Models for the Web-Graph*, pages 59–71. Springer, 2008.

[Gyöngyi *et al.*, 2004] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating Web spam with TrustRank. In *Proceeding of VLDB2004*, pages 576–587, 2004.

[Haveliwala, 2002] T. H. Haveliwala. Topic-sensitive PageRank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526. ACM, 2002.

[Jelenković and Olvera-Cravioto, 2010] P.R. Jelenković and M. Olvera-Cravioto. Information ranking and power laws on trees. *Adv. Appl. Prob.*, 42(4):1057–1093, 2010.

[Jelenković and Olvera-Cravioto, 2012a] P.R. Jelenković and M. Olvera-Cravioto. Implicit renewal theorem for trees with general weights. *Stoch. Proc. Appl.*, 122(9):3209–3238, 2012.

[Jelenković and Olvera-Cravioto, 2012b] P.R. Jelenković and M. Olvera-Cravioto. Implicit renewal theory and power tails on trees. *Adv. Appl. Prob.*, 44(2):528–561, 2012.

[Jelenković and Olvera-Cravioto, 2015] P.R. Jelenković and M. Olvera-Cravioto. Maximums on trees. *Stochastic Process. Appl.*, 125:217–232, 2015.

[Jessen and Mikosch, 2006] A.H. Jessen and T. Mikosch. Regularly varying functions. *Publications de L'Institut Mathematique, Nouvelle Serie*, 80(94):171–192, 2006.

[K. Avrachenkov, 2006] and D. Lebedev K. Avrachenkov. PageRank of scale-free growing networks. *Internet Mathematics*, 3(2):207–231, 2006.

[Karpelevich *et al.*, 1994] F. I. Karpelevich, M. Y. Kelbert, and Y. M. Suhov. Higher-order lindley equations. *Stochastic Processes and their Applications*, 53(1):65–96, 1994.

[Kleinberg *et al.*, 1999] J. Kleinberg, R. Kumar, R. Raghavan, S. Rajagopalan, and A. Tomkins. *Proceedings of the International Conference on Combinatorics and Computing*, volume 1627 of *Lecture Notes in Computer Science*, chapter The Web as a Graph: Measurements, Models, and Methods, pages 1–17. Springer-Verlag, 1999.

[Krapivsky and Redner, 2002] P.L. Krapivsky and S. Redner. A statistical physics perspective on web growth. *Comput. Netw.*, 39:261–276, 2002.

[Krapivsky *et al.*, 2001] P.L. Krapivsky, G.J. Rodgers, and S. Redner. Degree distributions of growing networks. *Phys. Rev. Lett.*, 86(23):5401–5404, 2001.

[Langville and Meyer, 2011] A.N. Langville and C.D. Meyer. *Google's PageRank and beyond: The science of search engine rankings.* Princeton University Press, 2011.

[Litvak *et al.*, 2007] N. Litvak, W.R.W. Scheinhardt, and Y. Volkovich. In-degree and PageRank: Why do they follow similar power laws? *Internet Math.*, 4(2-3):175–198, 2007.

[Liu, 1998] Q. Liu. Fixed points of a generalized smoothing transformation and applications to the branching random walk. *Adv. Appl. Prob.*, 30:85–112, 1998.

[McKay and Wormald, 1990a] B.D. McKay and N.C. Wormald. Asymptotic enumeration by degree sequence of graphs of high degree. *Eur. J. Combin.*, 11:565–580, 1990.

[McKay and Wormald, 1990b] B.D. McKay and N.C. Wormald. Uniform generation of random regular graphs of moderate degree. *J. Algorithms*, 11(1):52–67, 1990.

[McKay and Wormald, 1991] B.D. McKay and N.C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11:369–382, 1991.

[Mezard and Montanari, 2009] M. Mezard and A. Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[Mézard and Parisi, 2001] M. Mézard and G. Parisi. The bethe lattice spin glass revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(2):217–233, 2001.

[Molloy and Reed, 1995] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algor.*, 6(2-3):161–180, 1995.

[Newman *et al.*, 2001] M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev.*, 64(2):1–17, 2001.

[Newman, 2010] Mark Newman. *Networks: an introduction.* Oxford University Press, 2010.

[Olvera-Cravioto and Ruiz-Lacedelli, 2014] M. Olvera-Cravioto and O. Ruiz-Lacedelli. Parallel queues with synchronization. *arXiv:1501.00186*, 2014.

[Olvera-Cravioto, 2012a] M. Olvera-Cravioto. Asymptotics for weighted random sums. *Adv. Appl. Prob.*, 44(4):1142–1172, 2012.

[Olvera-Cravioto, 2012b] M. Olvera-Cravioto. Tail behavior of solutions of linear recursions on trees. *Stochastic Processes and their Applications*, 122(4):1777–1807, 2012.

[Pandurangan *et al.*, 2002] G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. In *Computing and Combinatorics*, pages 330–339. Springer, 2002.

[Rösler and Rüschendorf, 2001] U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001.

[Rösler, 1991] U. Rösler. A limit theorem for "Quicksort". *RAIRO Theor. Inform. Appl.*, 25:85–100, 1991.

[van der Hofstad *et al.*, 2005] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures and Algorithms*, 27(1):76–123, 2005.

[van der Hofstad, 2014] R. van der Hofstad. Random graphs and complex networks. `http://www.win.tue.nl/rhofstad/NotesRGCN.pdf`, 2014.

[Vännman, 1976] K. Vännman. Estimators based on order statistics from a Pareto distribution. *Journal of the American Statistical Association*, 71(355):704–708, 1976.

[Vigna, 2014] S. Vigna. A weighted correlation index for rankings with ties. *Preprint*, 2014. arXiv: 1404.3325.

[Villani, 2009] C. Villani. *Optimal transport, old and new.* Springer, New York, 2009.

[Volkovich and Litvak, 2010] Y. Volkovich and N. Litvak. Asymptotic analysis for personalized web search. *Adv. Appl. Prob.*, 42(2):577–604, 2010.

[Volkovich *et al.*, 2007] Y. Volkovich, N. Litvak, and D. Donato. Determining factors behind the PageRank log-log plot. In *Proceedings of the 5th International Conference on Algorithms and Models for the Web-graph*, pages 108–123, 2007.

[Volkovich *et al.*, 2009] Y. Volkovich, N. Litvak, and B. Zwart. Extremal dependencies and rank correlations in power law networks. In *Complex Sciences*, pages 1642–1653. Springer, 2009.

[Waltman and van Eck, 2010] L. Waltman and N.J. van Eck. The relation between eigenfactor, audience factor, and influence weight. *J. Am. Soc. Inf. Sci.*, 61(7):1476–1486, 2010.

[Watts and Strogatz, 1998] D. J. Watts and S. H. Strogatz. Collective dynamics of âĂŸsmall-worldâĂŹnetworks. *Nature*, 393(6684):440–442, 1998.

[Wendel, 1948] J.G. Wendel. Note on the gamma function. *Amer. Math. Monthly*, 55(9):563–564, 1948.

[Williams, 1991] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[Wormald, 1978] N.C. Wormald. *Some problems in the enumeration of labelled graphs*. PhD thesis, Newcastle University, 1978.

[Wormald, 1999] N.C. Wormald. Models of random regular graphs. *London Mathematical Society Lecture Notes Series*, pages 239–298, 1999.

Part V

Appendices

Appendix A

Proofs of Chapter 2

We give the proofs of all the results in the paper. We divide the proofs into two subsections, one containing those belonging to Section 2.1 and those belonging to Section 2.3. Throughout the remainder of the paper we use the following notation: $g(x) \sim f(x)$ if $\lim_{x\to\infty} g(x)/f(x) = 1$, $g(x) = O(f(x))$ if $\limsup_{x\to\infty} g(x)/f(x) < \infty$, and $g(x) = o(f(x))$ if $\lim_{x\to\infty} g(x)/f(x) = 0$.

A.1 Degree sequences

This section contains the proofs of Lemma 1, Theorems 3 and 4, and Proposition 5.

Proof of Lemma 1. Let $Z_i = \gamma_i - \xi_i$ and note that the $\{Z_i\}$ are i.i.d. mean zero random variables. If $E[Z_1^2] < \infty$, then Chebyshev's inequality gives

$$P(\mathcal{D}_n^c) = P\left(\left|\sum_{i=1}^{n} Z_i\right| > n^{1/2+\delta_0}\right) \leq \frac{n\text{Var }(Z_1)}{n^{1+2\delta_0}} = O(n^{-2\delta_0}) = o(1)$$

as $n \to \infty$.

Suppose now that $E[Z_1^2] = \infty$, which implies that $\kappa = 1 - \max\{\alpha^{-1}, \beta^{-1}\} \in (0, 1/2]$. Let $\theta = \max\{\alpha^{-1}, \beta^{-1}\}$, define $t_n = n^{\theta+\epsilon}$, $0 < \epsilon < \min\{\delta_0, \theta^{-1} - \theta\}$, and let $\{\tilde{Z}_i\}$ be a sequence of i.i.d. random variables having distribution $P(\tilde{Z}_1 \leq x) = P(Z_1 \leq x||Z_1| \leq t_n)$.

Then,

$$\begin{aligned}
&P\left(\left|\sum_{i=1}^{n} Z_i\right| > n^{1-\kappa+\delta_0}\right) \\
&= P\left(\left|\sum_{i=1}^{n} \tilde{Z}_i\right| > n^{1-\kappa+\delta_0}\right) P(|Z_1| \le t_n)^n + P\left(\left|\sum_{i=1}^{n} Z_i\right| > n^{1-\kappa+\delta_0},\ \max_{1\le i\le n} |Z_i| > t_n\right) \\
&\le P\left(\left|\sum_{i=1}^{n} \tilde{Z}_i - nE[\tilde{Z}_1]\right| + n|E[\tilde{Z}_1]| > n^{1-\kappa+\delta_0}\right) + P\left(\max_{1\le i\le n} |Z_i| > t_n\right).
\end{aligned}$$

By the union bound,

$$\begin{aligned}
P\left(\max_{1\le i\le n} |Z_i| > t_n\right) &\le nP(|Z_1| > t_n) \le nP(\gamma_1 + \xi_1 > t_n) \le nP(\gamma_1 > t_n/2) + nP(\xi_1 > t_n/2) \\
&\le n(t_n/2)^{-\alpha} L_F(t_n/2) + n(t_n/2)^{-\beta} L_G(t_n/2) \\
&= O\left(n^{1-\alpha(\theta+\epsilon)} L_F(t_n) + n^{1-\beta(\theta+\epsilon)} L_G(t_n)\right) \\
&= O\left(n^{-\alpha\epsilon} L_F(t_n) + n^{-\beta\epsilon} L_G(t_n)\right)
\end{aligned}$$

as $n \to \infty$, which converges to zero by basic properties of slowly varying functions (see, e.g., Proposition 1.3.6 in [Bingham *et al.*, 1987]). Next, note that since $E[Z_1] = 0$,

$$\begin{aligned}
|E[\tilde{Z}_1]| &= \frac{|E[Z_1 1(|Z_1| > t_n)]|}{P(|Z_1| \le t_n)} \\
&\le \frac{E[|Z_1| 1(|Z_1| > t_n)]}{P(|Z_1| \le t_n)} \le (1 + o(1)) \left(t_n P(|Z_1| > t_n) + \int_{t_n}^{\infty} P(|Z_1| > z) dz\right),
\end{aligned}$$

where in the last inequality we used integration by parts for the numerator and the fact that $P(|Z_1| \le t_n) = 1 + o(1)$ as $n \to \infty$. To estimate the integral note that

$$\begin{aligned}
\int_{t_n}^{\infty} P(|Z_1| > z) dz &\le \int_{t_n}^{\infty} (P(\gamma_1 > z/2) + P(\xi_1 > z/2)) dz \\
&\le 2 \int_{t_n/2}^{\infty} \left(u^{-\alpha} L_F(u) + u^{-\beta} L_G(u)\right) du \\
&\sim 2 \left((\alpha-1)^{-1} (t_n/2)^{-\alpha+1} L_F(t_n/2) + (\beta-1)^{-1} (t_n/2)^{-\beta+1} L_G(t_n/2)\right) \\
&= O\left(n^{-(\alpha-1)(\theta+\epsilon)} L_F(t_n) + n^{-(\beta-1)(\theta+\epsilon)} L_G(t_n)\right),
\end{aligned}$$

where in the third step we used Proposition 1.5.10 in [Bingham *et al.*, 1987]. Now note that

$$\min\{(\alpha-1)(\theta+\epsilon), (\beta-1)(\theta+\epsilon)\} = (\theta^{-1}-1)(\theta+\epsilon) = \kappa + \epsilon(\theta^{-1}-1),$$

from where it follows that

$$|E[\tilde{Z}_1]| = O\left(n^{-\kappa-\epsilon(\theta^{-1}-1)} (L_F(t_n) + L_G(t_n))\right) = o\left(n^{-\kappa+\delta_0}\right)$$

as $n \to \infty$. In view of this, we can use Chebyshev's inequality to obtain

$$P\left(\left|\sum_{i=1}^{n} \tilde{Z}_i - nE[\tilde{Z}_1]\right| + n|E[\tilde{Z}_1]| > n^{1-\kappa+\delta_0}\right) \leq \frac{\text{Var }(\tilde{Z}_1)}{n^{1-2(\kappa-\delta_0)}(1+o(1))}. \tag{A.1}$$

Finally, to see that this last bound converges to zero note that

$$\text{Var }(\tilde{Z}_1) \leq E[\tilde{Z}_1^2] = \frac{1}{P(|Z_1| \leq t_n)} E[Z_1^2 1(|Z_1| \leq t_n)] \leq (1+o(1)) E\left[|Z_1|^{\theta^{-1}-\epsilon}\right] t_n^{2-\theta^{-1}+\epsilon},$$

where we used again the observation that $P(|Z_1| \leq t_n) = 1 + o(1)$ and the inequality

$$|Z_1|^2 = |Z_1|^{\theta^{-1}-\epsilon} |Z_1|^{2-\theta^{-1}+\epsilon} \leq |Z_1|^{\theta^{-1}-\epsilon} t_n^{2-\theta^{-1}+\epsilon}$$

for $|Z_1| \leq t_n$. Next note that by the remark following (2.1), $E[|Z_1|^{\theta^{-1}-\epsilon}] < \infty$. Hence, we conclude that (A.1) is of order

$$O\left(t_n^{2-\theta^{-1}+\epsilon} n^{2(\kappa-\delta_0)-1}\right) = O\left(n^{(\theta+\epsilon)(2-\theta^{-1}+\epsilon)+2(\kappa-\delta_0)-1}\right) = o\left(n^{-2(\delta_0-\epsilon)}\right) = o(1)$$

as $n \to \infty$. This completes the proof. □

Before giving the proof of Theorem 3 we will need the following preliminary lemma.

Lemma 26. *Let $\{X_1, \ldots, X_n\}$ be an i.i.d. sequence of nonnegative random variables having distribution function V, and let $X^{(i)}$ denote the ith order statistic. Then, for any $k \leq n$,*

$$\sum_{i=n-k+1}^{n} E\left[X^{(i)}\right] \leq \int_0^\infty \min\left\{n\overline{V}(x), k\right\} dx.$$

Proof. Note that

$$E\left[X^{(i)}\right] = \int_0^\infty P(X^{(i)} > x)\, dx = \int_0^\infty \sum_{j=n-i+1}^{n} \binom{n}{j} \overline{V}(x)^j V(x)^{n-j} dx,$$

from where it follows

$$\begin{aligned}
\sum_{i=n-k+1}^{n} E\left[X^{(i)}\right] &= \sum_{i=n-k+1}^{n} \sum_{j=n-i+1}^{n} \binom{n}{j} \int_0^\infty \overline{V}(x)^j V(x)^{n-j} dx \\
&= \sum_{j=1}^{n} \min\{j,k\} \binom{n}{j} \int_0^\infty \overline{V}(x)^j V(x)^{n-j} dx \\
&= \int_0^\infty E\left[\min\{B(n, \overline{V}(x)), k\}\right] dx,
\end{aligned}$$

where $B(n,p)$ is a Binomial(n,p) random variable. Since the function $u(t) = \min\{t,k\}$ is concave, Jensen's inequality gives

$$E\left[\min\{B(n,\overline{V}(x)),k\}\right] \leq \min\left\{E[B(n,\overline{V}(x))],k\right\} = \min\left\{n\overline{V}(x),k\right\}.$$

□

Proof of Theorem 3. Since by construction $\sum_{i=1}^n M_i = \sum_{i=1}^n D_i$, it follows from Theorem 2 that it suffices to show that

$$\lim_{n\to\infty} P\left(\max_{A\subseteq V}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^n \min\{D_i, |A-\{v_i\}|\}\right) > 0\right) = 0.$$

Fix $0 < \epsilon < \min\{\beta-1, \alpha-1, 1/2\}$ and use the union bound to obtain

$$\begin{aligned}
&P\left(\max_{A\subseteq V}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^n \min\{D_i, |A-\{v_i\}|\}\right) > 0\right) \\
&\leq P\left(\max_{A\subseteq V, |A|\leq n^{(1+\epsilon)/\beta}}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^n \min\{D_i, |A-\{v_i\}|\}\right) > 0\right) \qquad \text{(A.2)} \\
&\quad + P\left(\max_{A\subseteq V, |A|> n^{(1+\epsilon)/\beta}}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^n \min\{D_i, |A-\{v_i\}|\}\right) > 0\right). \qquad \text{(A.3)}
\end{aligned}$$

By conditioning on how many of the D_i are larger than $n^{(1+\epsilon)/\beta}$ we obtain that (A.3) is bounded by

$$\begin{aligned}
&P\left(\max_{A\subseteq V, |A|> n^{(1+\epsilon)/\beta}}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^n \min\{D_i, |A-\{v_i\}|\}\right) > 0, \max_{1\leq i\leq n} D_i \leq n^{(1+\epsilon)/\beta}\right) \\
&\quad + P\left(\max_{1\leq i\leq n} D_i > n^{(1+\epsilon)/\beta}\right) \\
&\leq P\left(\max_{A\subseteq V, |A|> n^{(1+\epsilon)/\beta}}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^n D_i\right) > 0\right) + P\left(\max_{1\leq i\leq n} D_i > n^{(1+\epsilon)/\beta}\right) \\
&= P\left(\max_{1\leq i\leq n}(\xi_i+\chi_i) > n^{(1+\epsilon)/\beta}\middle|\, \mathcal{D}_n\right),
\end{aligned}$$

where $\mathcal{D}_n$ was defined in Lemma 1 and we use the fact that, by construction, D_i has the same distribution as $\xi_i + \chi_i$ conditional on the event $\mathcal{D}_n$. (We use this observation several

times throughout the paper.) Now note that by the union bound we have

$$\begin{aligned} P\left(\max_{1\le i\le n}(\xi_i+\chi_i) > n^{(1+\epsilon)/\beta}\middle|\mathcal{D}_n\right) &\le \frac{1}{P(\mathcal{D}_n)}\cdot P\left(\max_{1\le i\le n}(\xi_i+\chi_i) > n^{(1+\epsilon)/\beta}\right) \\ &\le \frac{1}{P(\mathcal{D}_n)}\sum_{i=1}^{n} P\left(\xi_i+\chi_i > n^{(1+\epsilon)/\beta}\right) \\ &\le \frac{1}{P(\mathcal{D}_n)}\cdot n\left(n^{(1+\epsilon)/\beta}-1\right)^{-\beta} L_G\left(n^{(1+\epsilon)/\beta}-1\right) \\ &= O\left(n^{-\epsilon}L_G\left(n^{(1+\epsilon)/\beta}\right)\right) = o(1), \end{aligned}$$

as $n\to\infty$, where the last step follows from Lemma 1 and basic properties of slowly varying functions (see, e.g., Chapter 1 in [Bingham *et al.*, 1987]).

Next, to analyze (A.2) let $k_n = \lfloor n^{(1+\epsilon)/\beta}\rfloor$ and note that we can write it as

$$\begin{aligned} &P\left(\max_{A\subseteq V,|A|\le k_n}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^{n}\min\{D_i,|A-\{v_i\}|\}\right) > 0\right) \\ &\le P\Bigg(\max\left\{\max_{A\subseteq V,\,2\le|A|\le k_n}\left(\sum_{v_i\in A} M_i - \sum_{i=1}^{n}\min\{D_i,1\}\right),\right. \\ &\qquad \left.\max_{1\le j\le n}\left(M_j - \sum_{i=1}^{n}\min\{D_i,|\{v_j\}-\{v_i\}|\}\right)\right\} > 0\Bigg) \\ &= P\left(\max\left\{\sum_{i=n-k_n+1}^{n} M^{(i)},\,(M+D)^{(n)}\right\} - \sum_{i=1}^{n}\min\{D_i,1\} > 0\right), \end{aligned}$$

where $x^{(i)}$ is the ith smallest of $\{x_1,\ldots,x_n\}$. Now let $a_0 = E[\min\{\xi_1,1\}] = \overline{G}(0) > 0$ and split the last probability as follows

$$\begin{aligned} &P\left(\max\left\{\sum_{i=n-k_n+1}^{n} M^{(i)},\,(M+D)^{(n)}\right\} - \sum_{i=1}^{n}\min\{D_i,1\} > 0\right) \\ &\le P\left(\max\left\{\sum_{i=n-k_n+1}^{n} M^{(i)},\,(M+D)^{(n)}\right\} > a_0 n - n^{1/2+\epsilon},\ \sum_{i=1}^{n}\min\{D_i,1\} \ge a_0 n - n^{1/2+\epsilon}\right) \qquad \text{(A.4)} \\ &\quad + P\left(\sum_{i=1}^{n}\min\{D_i,1\} < a_0 n - n^{1/2+\epsilon}\right). \qquad \text{(A.5)} \end{aligned}$$

To bound (A.5) use $D_i \geq \xi_i$ for all $i = 1, \ldots, n$ and Chebyshev's inequality to obtain

$$P\left(\sum_{i=1}^{n} \min\{D_i, 1\} < a_0 n - n^{1/2+\epsilon}\right) \leq \frac{1}{P(\mathcal{D}_n)} P\left(\sum_{i=1}^{n} (a_0 - \min\{\xi_i, 1\}) > n^{1/2+\epsilon}\right)$$
$$\leq \frac{n\text{Var}\left(\min\{\xi_1, 1\}\right)}{P(\mathcal{D}_n) n^{1+2\epsilon}} = O\left(n^{-2\epsilon}\right),$$

while the union bound gives that (A.4) is bounded by

$$P\left(\max\left\{\sum_{i=n-k_n+1}^{n} M^{(i)}, (M+D)^{(n)}\right\} > b_n\right)$$
$$\leq P\left(\sum_{i=n-k_n+1}^{n} M^{(i)} > b_n\right) + P\left((M+D)^{(n)} > b_n\right),$$

where $b_n = a_0 n - n^{1/2+\epsilon}$. For the second probability the union bound again gives

$$\begin{aligned}
P\left((M+D)^{(n)} > b_n\right) &\leq P\left(M^{(n)} > b_n/2\right) + P\left(D^{(n)} > b_n/2\right) \\
&\leq \frac{1}{P(\mathcal{D}_n)}\left(P\left(\max_{1\leq i\leq n}(\gamma_i + \tau_i) > b_n/2\right) + P\left(\max_{1\leq i\leq n}(\xi_i + \chi_i) > b_n/2\right)\right) \\
&\leq \frac{n}{P(\mathcal{D}_n)}\left(P(\gamma_1 + \tau_1 > b_n/2) + P(\xi_1 + \chi_1 > b_n/2)\right) \\
&\leq \frac{n}{P(\mathcal{D}_n)}\left((b_n/2-1)^{-\alpha} L_F(b_n/2-1) + (b_n/2-1)^{-\beta} L_G(b_n/2-1)\right) \\
&= O\left(n^{-\alpha+1} L_F(n) + n^{-\beta+1} L_G(n)\right) = o(1)
\end{aligned}$$

as $n \to \infty$. Finally, by Markov's inequality and Lemma 26,

$$\begin{aligned}
&P\left(\sum_{i=n-k_n+1}^{n} M^{(i)} > b_n\right) \leq \frac{1}{b_n} \sum_{i=n-k_n+1}^{n} E\left[M^{(i)}\right] \leq \frac{1}{b_n P(\mathcal{D}_n)} \sum_{i=n-k_n+1}^{n} E[\gamma^{(i)} + 1] \\
&\leq \frac{1}{b_n P(\mathcal{D}_n)}\left(\int_0^\infty \min\left\{n\overline{F}(x), k_n\right\} dx + k_n\right) \\
&= a_0^{-1}(1+o(1)) \int_0^\infty \min\left\{\overline{F}(x), n^{(1+\epsilon)/\beta - 1}\right\} dx + o(1) \\
&\leq a_0^{-1}(1+o(1))\left(n^{(1+\epsilon)/\beta-1} + \int_1^\infty \min\left\{Kx^{-\alpha+\epsilon}, n^{(1+\epsilon)/\beta-1}\right\} dx\right) + o(1) \\
&= o(1) + O\left(\int_1^\infty \min\left\{x^{-\alpha+\epsilon}, n^{(1+\epsilon)/\beta-1}\right\} dx\right)
\end{aligned}$$

as $n \to \infty$, where $K = \sup_{t\geq 1} t^{-\epsilon} L_F(t) < \infty$. Since

$$\begin{aligned}
&\int_1^\infty \min\left\{x^{-\alpha+\epsilon}, n^{(1+\epsilon)/\beta-1}\right\} dx \\
&= n^{(1+\epsilon)/\beta-1}(n^{(\beta-1-\epsilon)/(\beta(\alpha-\epsilon))} - 1) + \int_{n^{(\beta-1-\epsilon)/(\beta(\alpha-\epsilon))}}^\infty x^{-\alpha+\epsilon} dx \\
&= O\left(n^{-(\beta-1-\epsilon)(\alpha-1-\epsilon)/(\beta(\alpha-\epsilon))}\right) = o(1),
\end{aligned}$$

the proof is complete. □

The last two proofs of this section are those of Theorem 4 and Proposition 5.

Proof of Theorem 4. Let $u : \mathbb{N}^{r+s} \to [-M, M]$, $M > 0$, be a continuous bounded function, and let $\Delta_n, \mathcal{D}_n$ be defined as in Lemma 1. Then,

$$\begin{aligned}
&|E\left[u(M_{i_1}, \ldots, M_{i_r}, D_{j_1}, \ldots, D_{j_s})\right] - E\left[u(\gamma_1, \ldots, \gamma_r, \xi_1, \ldots, \xi_s)\right]| \\
&= |E\left[u(\gamma_{i_1} + \tau_{i_1}, \ldots, \gamma_{i_r} + \tau_{i_r}, \xi_{j_1} + \chi_{j_1}, \ldots, \xi_{j_s} + \chi_{j_s})|\mathcal{D}_n\right] - E\left[u(\gamma_{i_1}, \ldots, \gamma_{i_r}, \xi_{j_1}, \ldots, \xi_{j_s})\right]| \\
&\leq |E\left[u(\gamma_{i_1} + \tau_{i_1}, \ldots, \gamma_{i_r} + \tau_{i_r}, \xi_{j_1} + \chi_{j_1}, \ldots, \xi_{j_s} + \chi_{j_s}) - u(\gamma_{i_1}, \ldots, \gamma_{i_r}, \xi_{j_1}, \ldots, \xi_{j_s})|\mathcal{D}_n\right]| \qquad \text{(A.6)} \\
&\quad + |E\left[u(\gamma_{i_1}, \ldots, \gamma_{i_r}, \xi_{j_1}, \ldots, \xi_{j_s})|\mathcal{D}_n\right] - E\left[u(\gamma_1, \ldots, \gamma_r, \xi_1, \ldots, \xi_s)\right]|. \qquad \text{(A.7)}
\end{aligned}$$

Let $T = \sum_{t=1}^r \tau_{i_t} + \sum_{t=1}^s \chi_{j_s}$. Since u is bounded then (A.6) is smaller than or equal to

$$\begin{aligned}
&E\left[|u(\gamma_{i_1} + \tau_{i_1}, \ldots, \gamma_{i_r} + \tau_{i_r}, \xi_{j_1} + \chi_{j_1}, \ldots, \xi_{j_s} + \chi_{j_s}) - u(\gamma_{i_1}, \ldots, \gamma_{i_r}, \xi_{j_1}, \ldots, \xi_{j_s})|\, 1\left(T \geq 1\right)|\, \mathcal{D}_n\right] \\
&\leq 2MP\left(T \geq 1|\, \mathcal{D}_n\right) \leq 2M\left(\sum_{t=1}^r P(\tau_{i_t} = 1|\mathcal{D}_n) + \sum_{t=1}^s P(\chi_{j_t} = 1|\mathcal{D}_n)\right) \\
&= \frac{2M}{P(\mathcal{D}_n)}\left(\sum_{t=1}^r E[1(\tau_{i_t} = 1, \mathcal{D}_n)] + \sum_{t=1}^s E[1(\chi_{j_t} = 1, \mathcal{D}_n)]\right).
\end{aligned}$$

To compute the last expectations let $\mathcal{F}_n = \sigma(\gamma_1, \ldots, \gamma_n, \xi_1, \ldots, \xi_n)$ be the σ-algebra generated by the γ_i's and ξ_i's and note that

$$\begin{aligned}
E[1(\chi_{j_t} = 1, \mathcal{D}_n)] = E[1(\mathcal{D}_n)E[1(\chi_{j_t} = 1)|\mathcal{F}_n]] &= E\left[1(\mathcal{D}_n, \Delta_n \geq 0)\frac{\binom{n-1}{\Delta_n - 1}}{\binom{n}{\Delta_n}}\right] \\
&= E\left[1(\mathcal{D}_n, \Delta_n \geq 0)\frac{\Delta_n}{n}\right],
\end{aligned}$$

and symmetrically,

$$E[1(\tau_{i_t} = 1, \mathcal{D}_n)] = E\left[1(\mathcal{D}_n, \Delta_n < 0)\frac{|\Delta_n|}{n}\right],$$

from where it follows that (A.6) is bounded by

$$2M\left(\sum_{t=1}^{r} E\left[\frac{\Delta_n}{n}1(\Delta_n \geq 0)\middle| \mathcal{D}_n\right] + \sum_{t=1}^{s} E\left[\frac{|\Delta_n|}{n}1(\Delta_n < 0)\middle| \mathcal{D}_n\right]\right) \leq 2M(r+s)n^{-\kappa+\delta_0} = o(1)$$

as $n \to \infty$. To analyze (A.7) we first note that by Lemma 1, $P(\mathcal{D}_n) \to 1$ as $n \to \infty$, hence

$$\begin{aligned} E\left[u(\gamma_{i_1}, \ldots, \gamma_{i_r}, \xi_{j_1}, \ldots, \xi_{j_s})|\mathcal{D}_n\right] &= \frac{1}{P(\mathcal{D}_n)} E\left[u(\gamma_1, \ldots, \gamma_r, \xi_1, \ldots, \xi_s)1(\mathcal{D}_n)\right] \\ &= E\left[u(\gamma_1, \ldots, \gamma_r, \xi_1, \ldots, \xi_s)1(\mathcal{D}_n)\right] + o(1). \end{aligned}$$

Therefore, (A.7) is equal to

$$\left|E\left[u(\gamma_1, \ldots, \gamma_r, \xi_1, \ldots, \xi_s)1(\mathcal{D}_n^c)\right] + o(1)\right| \leq MP(\mathcal{D}_n^c) + o(1) \to 0$$

as $n \to \infty$, which completes the proof. □

Proof of Proposition 5. Fix $\epsilon > 0$ and let $\mathcal{D}_n = \{|\Delta_n| \leq n^{1-\kappa+\delta_0}\}$. For the first limit fix $i, j = 0, 1, 2, \ldots$ and note that by the union bound,

$$\begin{aligned} &P\left(\left|\frac{1}{n}\sum_{k=1}^{n} 1(M_k = i, D_k = j) - f_i g_j\right| > \epsilon\right) \\ &\leq P\left(\left|\frac{1}{n}\sum_{k=1}^{n}(1(\gamma_k + \tau_k = i, \xi_k + \chi_k = j) - 1(\gamma_k = i, \xi_k = j))\right| > \epsilon/2\middle| \mathcal{D}_n\right) \\ &\quad + P\left(\left|\frac{1}{n}\sum_{k=1}^{n} 1(\gamma_k = i, \xi_k = j) - f_i g_j\right| > \epsilon/2\middle| \mathcal{D}_n\right) \\ &\leq P\left(\frac{1}{n}\sum_{k=1}^{n}|1(\gamma_k + \tau_k = i, \xi_k + \chi_k = j) - 1(\gamma_k = i, \xi_k = j))| > \epsilon/2\middle| \mathcal{D}_n\right) \\ &\quad + \frac{1}{P(\mathcal{D}_n)n(\epsilon/2)^2}\text{Var}\ (1(\gamma_1 = i, \xi_1 = j)), \end{aligned}$$

where in the last step we used Chebyshev's inequality. Clearly, Var $(1(\gamma_1 = i, \xi_1 = j)) = f_i g_j(1 - f_i g_j)$, and since by Lemma 1 $P(\mathcal{D}_n) \to 1$ as $n \to \infty$, then the second term converges

to zero. To analyze the first term note that at most one of χ_k or τ_k can be one, hence,

$$\begin{aligned}
&P\left(\frac{1}{n}\sum_{k=1}^{n}|1(\gamma_k+\tau_k=i,\xi_k+\chi_k=j)-1(\gamma_k=i,\xi_k=j))|>\epsilon/2\middle|\mathcal{D}_n\right)\\
&\leq P\left(\frac{1}{n}\sum_{k=1}^{n}(|1(\xi_k+\chi_k=j)-1(\xi_k=j)|+|1(\gamma_k+\tau_k=i)-1(\gamma_k=i)|)>\epsilon/2\middle|\mathcal{D}_n\right)\\
&\leq P\left(\frac{1}{n}\sum_{k=1}^{n}(1(\chi_k=1)+1(\tau_k=1))>\epsilon/2\middle|\mathcal{D}_n\right)\\
&= P\left(\frac{|\Delta_n|}{n}>\epsilon/2\middle|\mathcal{D}_n\right)\\
&\leq 1(n^{-\kappa+\delta_0}>\epsilon/2)\to 0
\end{aligned}$$

as $n\to\infty$.

Next, for the average degrees we have

$$\begin{aligned}
P\left(\left|\frac{1}{n}\sum_{i=1}^{n}M_i-E[\gamma_1]\right|>\epsilon\right)&=P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(\gamma_i+\tau_i)-E[\gamma_1]\right|>\epsilon\middle|\mathcal{D}_n\right)\\
&\leq P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i-E[\gamma_1]\right|+\frac{|\Delta_n|}{n}>\epsilon\middle|\mathcal{D}_n\right)\\
&\leq \frac{1}{P(\mathcal{D}_n)}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i-E[\gamma_1]\right|+n^{-\kappa+\delta_0}>\epsilon\right), \qquad \text{(A.8)}
\end{aligned}$$

symmetrically,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}D_i-E[\xi_1]\right|>\epsilon\right)\leq\frac{1}{P(\mathcal{D}_n)}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-E[\xi_1]\right|+n^{-\kappa+\delta_0}>\epsilon\right), \qquad \text{(A.9)}$$

and since $\tau_i\chi_i=0$ for all $1\leq i\leq n$,

$$\begin{aligned}
P\left(\left|\frac{1}{n}\sum_{i=1}^{n}M_iD_i-E[\gamma_1\xi_1]\right|>\epsilon\right)&=P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(\gamma_i\xi_i+\tau_i\xi_i+\gamma_i\chi_i-E[\gamma_1\xi_1]\right|>\epsilon\middle|\mathcal{D}_n\right)\\
&\leq P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\xi_i-E[\gamma_1\xi_1]\right|+\sum_{i=1}^{n}(\tau_i\xi_i+\gamma_i\chi_i)>\epsilon\middle|\mathcal{D}_n\right)\\
&\leq \frac{1}{P(\mathcal{D}_n)}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\xi_i-E[\gamma_1\xi_1]\right|+n^{-\kappa+\delta}>\epsilon\right) \qquad \text{(A.10)}\\
&\quad+P\left(\frac{1}{n}\sum_{i=1}^{n}(\tau_i\xi_i+\gamma_i\chi_i)>n^{-\kappa+\delta}\middle|\mathcal{D}_n\right), \qquad \text{(A.11)}
\end{aligned}$$

for any $\delta_0<\delta<\kappa$. By Lemma 1, $P(\mathcal{D}_n)$ converges to one, and by the Weak Law of Large Numbers (WLLN) we have that each of (A.8), (A.9) and (A.10) converges to zero as

$n \to \infty$, as required. To see that (A.11) converges to zero use Markov's inequality to obtain

$$P\left(\frac{1}{n}\sum_{i=1}^{n}(\tau_i\xi_i + \gamma_i\chi_i) > n^{-\kappa+\delta}\middle|\mathcal{D}_n\right) \leq \frac{E[\tau_1\xi_1 + \gamma_1\chi_1|\mathcal{D}_n]}{n^{-\kappa+\delta}} = \frac{E[(\tau_1\xi_1 + \gamma_1\chi_1)1(\mathcal{D}_n)]}{P(\mathcal{D}_n)n^{-\kappa+\delta}}. \tag{A.12}$$

Now let $\mathcal{F}_n = \sigma(\gamma_1, \ldots, \gamma_n, \xi_1, \ldots, \xi_n)$ to compute

$$\begin{aligned} E[(\tau_1\xi_1 + \gamma_1\chi_1)1(\mathcal{D}_n)] &= E[(\xi_1 E[\tau_1|\mathcal{F}_n] + \gamma_1 E[\chi_1|\mathcal{F}_n])1(\mathcal{D}_n)] \\ &\leq E\left[(\xi_1 + \gamma_1)\frac{|\Delta_n|}{n}1(\mathcal{D}_n)\right] \leq 2\mu n^{-\kappa+\delta_0}, \end{aligned}$$

which implies that (A.12) converges to zero.

Finally, provided that $E[\gamma_1^2 + \xi_1^2] < \infty$, the WLLN combined with the arguments used to bound (A.11) give

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}M_i^2 - E[\gamma_1^2]\right| > \epsilon\right) &\leq \frac{1}{P(\mathcal{D}_n)}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i^2 - E[\gamma_1^2]\right| + \frac{1}{n}\sum_{i=1}^{n}(2\gamma_i\tau_i + \tau_i^2) > \epsilon, \mathcal{D}_n\right) \\ &\leq \frac{1}{P(\mathcal{D}_n)}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i^2 - E[\gamma_1^2]\right| + n^{-\kappa+\delta} > \epsilon\right) \\ &\quad + P\left(\frac{1}{n}\sum_{i=1}^{n}(2\gamma_i\tau_i + \tau_i^2) > n^{-\kappa+\delta}\middle|\mathcal{D}_n\right) \\ &\leq o(1) + \frac{E[(2\gamma_1 + 1)\tau_1|\mathcal{D}_n]}{n^{-\kappa+\delta}} \\ &\leq o(1) + \frac{E[2\gamma_1 + 1]}{P(\mathcal{D}_n)n^{\delta-\delta_0}}, \end{aligned}$$

and symmetrically,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}D_i^2 - E[\xi_1^2]\right| > \epsilon\right) \to 0,$$

as $n \to \infty$. □

A.2 Configuration model

This section contains the proofs of Proposition 6, which establishes the uniformity of simple graphs, Propositions 7 and 9, which concern the repeated directed configuration model, and Proposition 10 which refers to the erased directed configuration model.

Proof of Proposition 6. Suppose $\mathbf{m}$ and $\mathbf{d}$ have equal sum l_n, and number the inbound and outbound half-edges by $1, 2, \ldots, l_n$. The process of matching half edges in the configuration

model is equivalent to a permutation $(p(1), p(2), \ldots, p(l_n))$ of the numbers $(1, 2, \ldots, l_n)$ where we pair the ith inbound half-edge to the $p(i)$th outbound half-edge, with all $l_n!$ permutations being equally likely. Note that different permutations can actually lead to the same graph, for example, if we switch the position of two outbound half-edges of the same node, so not all multigraphs have the same probability. Nevertheless, a simple graph can only be produced by $\prod_{i=1}^{n} d_i! m_i!$ different permutations; to see this note that for each node v_i, $i = 1, \ldots, n$, we can permute its m_i inbound half-edges and its d_i outbound half-edges without changing the graph. It follows that since the number of permutations leading to a simple graph is the same for all simple graphs, then conditional on the resulting graph being simple, it is uniformly chosen among all simple graphs having bi-degree-sequence $(\mathbf{m}, \mathbf{d})$. □

Next, we give the proofs of the results related to the repeated directed configuration model. Before proceeding with the proof of Proposition 7 we give the following preliminary lemma, which will be used to establish that under Condition 2 the maximum in- and out-degrees cannot grow too fast.

Lemma 27. *Let $\{a_{nk} : 1 \leq k \leq n, n \in \mathbb{N}\}$ be a triangular array of nonnegative integers, and suppose there exist nonnegative numbers $\{p_j : j \in \mathbb{N} \cup \{0\}\}$ such that $\sum_{j=0}^{\infty} p_j = 1$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1(a_{nk} = j) = p_j, \quad \textit{for all } j \in \mathbb{N}\cup\{0\} \qquad \textit{and} \qquad \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_{nk} = \sum_{j=0}^{\infty} j p_j < \infty.$$

Then,

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{a_{nk}}{n} = 0.$$

Proof. Define

$$F(x) = \sum_{j=0}^{\lfloor x \rfloor} p_j \qquad \text{and} \qquad F_n(x) = \frac{1}{n} \sum_{k=1}^{n} 1(a_{nk} \leq x)$$

and note that F and F_n are both distribution functions with support on the nonnegative integers. Define the pseudoinverse operator $h^{-1}(u) = \inf\{x \geq 0 : u \leq h(x)\}$ and let

$$X_n = F_n^{-1}(U) \qquad \text{and} \qquad X = F^{-1}(U),$$

where U is a Uniform(0,1) random variable. It is easy to verify that X_n and X have distributions F_n and F, respectively. Furthermore, the assumptions imply that

$$X_n \to X \qquad \text{a.s.}$$

as $n \to \infty$ and

$$E[X_n] = \sum_{j=0}^{\infty} j \frac{1}{n} \sum_{k=1}^{n} 1(a_{nk} = j) = \frac{1}{n} \sum_{k=1}^{n} \sum_{j=0}^{\infty} j 1(a_{nk} = j) = \frac{1}{n} \sum_{k=1}^{n} a_{nk} \to E[X]$$

as $n \to \infty$, where the exchange of sums is justified by Fubini's theorem. Now note that by Fatou's lemma,

$$\liminf_{n\to\infty} E[X_n 1(X_n \leq \sqrt{n})] \geq E\left[\liminf_{n\to\infty} X_n 1(X_n \leq \sqrt{n})\right] = E[X],$$

which implies that

$$\lim_{n\to\infty} E[X_n 1(X_n > \sqrt{n})] = 0.$$

Finally,

$$\begin{aligned} E[X_n 1(X_n \geq n)] &= \sum_{j=\lfloor\sqrt{n}\rfloor+1}^{\infty} j \frac{1}{n} \sum_{k=1}^{n} 1(a_{nk} = j) \\ &= \frac{1}{n} \sum_{k=1}^{n} \sum_{j=\lfloor\sqrt{n}\rfloor+1}^{\infty} j 1(a_{nk} = j) = \frac{1}{n} \sum_{k=1}^{n} a_{nk} 1(a_{nk} > \sqrt{n}), \end{aligned}$$

from where it follows that

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{a_{nk} 1(a_{nk} > \sqrt{n})}{n} = 0,$$

which in turn implies that

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{a_{nk}}{n} \leq \lim_{n\to\infty} \left(\frac{\sqrt{n}}{n} + \max_{1\leq k\leq n} \frac{a_{nk} 1(a_{nk} > \sqrt{n})}{n} \right) = 0.$$

□

Proof of Proposition 7. Following the proof of Proposition 7.9 in [van der Hofstad, 2014], we define the random variable $\tilde{T}_n$ to be the total number of pairs of multiple edges in the same direction, e.g., if from node v_i to node v_j there are $k \geq 2$ edges, their contribution to $\tilde{T}_n$ is $\binom{k}{2}$. Note that $T_n \leq \tilde{T}_n$, with strict inequality whenever there is at least one pair of nodes

having three or more multiple edges in the same direction. We claim that $\tilde{T}_n - T_n \xrightarrow{P} 0$ as $n \to \infty$, which implies that

$$\text{if} \qquad (S_n, \tilde{T}_n) \Rightarrow (S, T), \qquad \text{then} \qquad (S_n, T_n) \Rightarrow (S, T)$$

as $n \to \infty$. To prove the claim start by defining indicator random variables for each of the possible self-loops and multiple edges in the same direction that the multigraph can have. For the self-loops we use the notation $\mathbf{u} = (r, t, i)$ to define

$$I_{\mathbf{u}} := 1(\text{self-loop from the } r\text{th outbound stub to the } t\text{th inbound stub of node } v_i),$$

and for the pairs of multiple edges in the same direction we use $\mathbf{w} = (r_1, t_1, r_2, t_2, i, j)$ to define

$$J_{\mathbf{w}} := 1(r_s\text{th outbound stub of node } v_i \text{ paired to } t_s\text{th inbound stub of node } v_j,\ s = 1, 2).$$

The sets of possible vectors $\mathbf{u}$ and $\mathbf{w}$ are given by

$$\mathcal{I} = \{(r, t, i) : 1 \le i \le n, 1 \le r \le d_{ni}, 1 \le t \le m_{ni}\}, \quad \text{and}$$
$$\mathcal{J} = \{(r_1, t_1, r_2, t_2, i, j) : 1 \le i \ne j \le n, 1 \le r_1 < r_2 \le d_{ni}, 1 \le t_1 \ne t_2 \le m_{nj}\},$$

respectively. It follows from this notation that

$$S_n = \sum_{\mathbf{u} \in \mathcal{I}} I_{\mathbf{u}} \qquad \text{and} \qquad \tilde{T}_n = \sum_{\mathbf{w} \in \mathcal{J}} J_{\mathbf{w}}.$$

Next, note that by the union bound,

$$\begin{aligned}
P\left(\tilde{T}_n - T_n \ge 1\right) &\le P\left(\text{at least two nodes with three or more edges in the same direction}\right) \\
&\le \sum_{1 \le i \ne j \le n} P\left(\text{three or more edges from node } v_i \text{ to node } v_j\right) \\
&\le \sum_{1 \le i \ne j \le n} \frac{d_{ni}(d_{ni}-1)(d_{ni}-2) m_{nj}(m_{nj}-1)(m_{nj}-2)}{l_n(l_n-1)(l_n-2)} \\
&\le \left(\frac{1}{\sqrt{n}} \max_{1 \le i \le n} d_{ni}\right)\left(\frac{1}{\sqrt{n}} \max_{1 \le j \le n} m_{nj}\right)\left(\frac{n}{l_n - 2}\right)^3 \cdot \frac{1}{n}\sum_{i=1}^{n} d_{ni}^2 \cdot \frac{1}{n}\sum_{j=1}^{n} m_{nj}^2 \\
&= o(1)
\end{aligned}$$

as $n \to \infty$, where for the last step we used Condition 2 and Lemma 27. It follows that $\tilde{T}_n - T_n \xrightarrow{P} 0$ as claimed.

We now proceed to prove that $(S_n, \tilde{T}_n) \Rightarrow (S, T)$, where S and M are independent Poisson random variables with means λ_1 and λ_2, respectively. To do this we use Theorem 2.6 in [van der Hofstad, 2014] which says that if for any $p, q \in \mathbb{N}$

$$\lim_{n\to\infty} E\left[(S_n)_p(\tilde{T}_n)_q\right] = \lambda_1^p \lambda_2^q,$$

where $(X)_r = X(X-1)\cdots(X-r+1)$, then $(S_n, \tilde{T}_n) \Rightarrow (S, T)$ as $n \to \infty$. To compute the expectation we use Theorem 2.7 in [van der Hofstad, 2014], which gives

$$E\left[(S_n)_p(\tilde{T}_n)_q\right] = \sum_{\mathbf{u}_1,\ldots,\mathbf{u}_p \in \mathcal{I}} \sum_{\mathbf{w}_1,\ldots,\mathbf{w}_q \in \mathcal{J}} P\left(I_{\mathbf{u}_1} = \cdots = I_{\mathbf{u}_p} = J_{\mathbf{w}_1} = \cdots = J_{\mathbf{w}_q} = 1\right), \quad \text{(A.13)}$$

where the sums are taken over all the p-permutations, respectively q-permutations, of the distinct indices in $\mathcal{I}$, respectively $\mathcal{J}$.

Next, by the fact that all stubs are uniformly paired, we have that

$$P\left(I_{\mathbf{u}_1} = \cdots = I_{\mathbf{u}_p} = J_{\mathbf{w}_1} = \cdots = J_{\mathbf{w}_q} = 1\right), = \frac{1}{\prod_{i=0}^{p+2q-1}(l_n - i)}$$

unless there is a conflict in the attachment rules, i.e., one stub is required to pair with two or more different stubs within the indices $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_q\}$, in which case

$$P\left(I_{\mathbf{u}_1} = \cdots = I_{\mathbf{u}_p} = J_{\mathbf{w}_1} = \cdots = J_{\mathbf{w}_q} = 1\right) = 0. \quad \text{(A.14)}$$

Therefore, from (A.13) we obtain

$$\begin{aligned} E[(S_n)_p(\tilde{T}_n)_q] &\leq \sum_{\mathbf{u}_1,\ldots,\mathbf{u}_p \in \mathcal{I}} \sum_{\mathbf{w}_1,\ldots,\mathbf{w}_q \in \mathcal{J}} \frac{1}{\prod_{i=0}^{p+2q-1}(l_n - i)} \\ &= \frac{|\mathcal{I}|(|\mathcal{I}|-1)\cdots(|\mathcal{I}|-p+1)|\mathcal{J}|(|\mathcal{J}|-1)\cdots(|\mathcal{J}|-q+1)}{l_n(l_n-1)\cdots(l_n-(p+2q-1))}, \end{aligned} \quad \text{(A.15)}$$

where $|A|$ denotes the cardinality of set A. Now note that

$$\begin{aligned} |\mathcal{I}| &= \sum_{i=1}^{n} m_{ni} d_{ni}, \qquad \text{and} \\ |\mathcal{J}| &= \sum_{1\leq i\neq j\leq n} \frac{d_{ni}(d_{ni}-1)}{2} m_{nj}(m_{nj}-1) \\ &= \frac{1}{2}\left(\sum_{i=1}^{n} m_{ni}(m_{ni}-1)\right)\left(\sum_{i=1}^{n} d_{ni}(d_{ni}-1)\right) - \frac{1}{2}\sum_{i=1}^{n} m_{ni}(m_{ni}-1)d_{ni}(d_{ni}-1). \end{aligned}$$

By Lemma 27 and Condition 2 we have

$$\sum_{i=1}^{n} m_{ni}(m_{ni}-1)d_{ni}(d_{ni}-1) \le \left(\max_{1\le i\le n} m_{ni}\right)\left(\max_{1\le i\le n} d_{ni}\right)\sum_{i=1}^{n} m_{ni}d_{ni} = o(n^2)$$

as $n\to\infty$. Hence, it follows from Condition 2 that

$$\begin{aligned}\frac{|\mathcal{I}|}{n} &= E[\gamma\xi]+o(1),\\ \frac{|\mathcal{J}|}{n^2} &= \frac{1}{2}E[\gamma(\gamma-1)]E[\xi(\xi-1)]+o(1), \qquad \text{and}\\ \frac{n}{l_n} &= \frac{1}{\mu}+o(1)\end{aligned}$$

as $n\to\infty$. Since p and q remain fixed as $n\to\infty$, we have

$$\begin{aligned}\limsup_{n\to\infty} E[(S_n)_p(\tilde{T}_n)_q] &= \left(\lim_{n\to\infty}\frac{|\mathcal{I}|}{n}\right)^p\left(\lim_{n\to\infty}\frac{|\mathcal{J}|}{n^2}\right)^q\left(\lim_{n\to\infty}\frac{n}{l_n}\right)^{p+2q}\\ &= (E[\gamma\xi])^p\left(\frac{E[\gamma(\gamma-1)]E[\xi(\xi-1)]}{2}\right)^q\left(\frac{1}{\mu}\right)^{p+2q} = \lambda_1^p\lambda_2^q.\end{aligned}$$

To prove the matching lower bound, we note that (A.14) occurs exactly when there is a conflict in the attachment rules. Each time a conflict happens, the numerator of (A.15) decreases by one. Therefore,

$$\begin{aligned}E\left[(S_n)_p(\tilde{T}_n)_q\right] &= \frac{|\mathcal{I}|(|\mathcal{I}|-1)\cdots(|\mathcal{I}|-p+1)|\mathcal{J}|(|\mathcal{J}|-1)\cdots(|\mathcal{J}|-q+1)}{l_n(l_n-1)\cdots(l_n-(p+2q-1))}\\ &\quad-\sum_{\mathbf{u}_1,\ldots,\mathbf{u}_p\in\mathcal{I}}\sum_{\mathbf{w}_1,\ldots,\mathbf{w}_q\in\mathcal{J}}\frac{1(\mathbf{u}_1,\ldots,\mathbf{u}_p,\mathbf{w}_1,\ldots,\mathbf{w}_q \text{ have a conflict})}{\prod_{i=0}^{p+2q-1}(l_n-i)}\\ &= \lambda_1^p\lambda_2^q-\frac{1}{(\mu n)^{p+2q}}\sum_{\mathbf{u}_1,\ldots,\mathbf{u}_p\in\mathcal{I}}\sum_{\mathbf{w}_1,\ldots,\mathbf{w}_q\in\mathcal{J}}1(\mathbf{u}_1,\ldots,\mathbf{u}_p,\mathbf{w}_1,\ldots,\mathbf{w}_q \text{ have a conflict})+o(1)\end{aligned}$$

as $n\to\infty$. To bound the total number of conflicts note that there are three possibilities:

1. a stub is assigned to two different self-loops, or
2. a stub is assigned to a self-loop and a multiple edge, or
3. a stub is assigned to two different multiple edges.

We now discuss each of the cases separately. For conflicts of type (a) suppose there is a conflict between the self-loops $\mathbf{u}_a$ and $\mathbf{u}_b$; the remaining $p-2$ self-loops and q pairs of multiple edges can be chosen freely. Then the number of such conflicts is bounded by

$|\mathcal{I}|^{p-2}|\mathcal{J}|^q = O\left(n^{p+2q-2}\right)$, hence it suffices to show that the total number of conflicting pairs $(\mathbf{u}_a, \mathbf{u}_b)$ is $o(n^2)$ as $n \to \infty$. Now, to see that this is indeed the case, first choose the node v_i where the conflicting pair is; if the conflict is that an outbound stub is assigned to two different inbound stubs then we can choose the problematic outbound stub in d_{ni} ways and the two inbound stubs in $m_{ni}(m_{ni}-1)$ ways, whereas if the conflict is that an inbound stub is assigned to two different outbound stubs then we can choose the problematic inbound stub in m_{ni} ways and the two outbound stubs in $d_{ni}(d_{ni}-1)$ ways. Thus, the total number of conflicting pairs is bounded by

$$\begin{aligned}\sum_{i=1}^{n}(d_{ni}m_{ni}^2 + m_{ni}d_{ni}^2) &\leq \left(\max_{1\leq i\leq n} m_{ni} + \max_{1\leq i\leq n} d_{ni}\right) 2\sum_{i=1}^{n} m_{ni}d_{ni} \\ &= o(n^{3/2}) = o(n^2).\end{aligned}$$

For conflicts of type (b) suppose there is a conflict between the self-loop $\mathbf{u}_a$ and the pair of multiple edges $\mathbf{w}_b$; choose the remaining $p-1$ self-loops and $q-1$ multiple edges freely. Then, the number of such conflicts is bounded by $|\mathcal{I}|^{p-1}|\mathcal{J}|^{q-1} = O\left(n^{p+2q-3}\right)$, and it suffices to show that the number of conflicting pairs $(\mathbf{u}_a, \mathbf{w}_b)$ is $o(n^3)$ as $n \to \infty$. Similarly as in case (a), an outbound stub of node v_i can be paired to a self-loop and a multiple edge to node v_j in $d_{ni}m_{ni}m_{nj}(d_{ni}-1)(m_{nj}-1)$ ways, and an inbound stub of node v_i can be paired to a self-loop and a multiple edge from node v_j in $m_{ni}d_{ni}d_{nj}(m_{ni}-1)(d_{nj}-1)$ ways, and so the total number of conflicting pairs is bounded by

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}(d_{ni}^2 m_{ni}m_{nj}^2 + m_{ni}^2 d_{ni}d_{nj}^2) &\leq \left(\max_{1\leq i\leq n} m_{ni} + \max_{1\leq i\leq n} d_{ni}\right) 2\left(\sum_{i=1}^{n} m_{ni}^2\right)\left(\sum_{i=1}^{n} d_{ni}^2\right) \\ &= o(n^{5/2}) = o(n^3).\end{aligned}$$

Finally, for conflicts of type (c) we first fix $\mathbf{w}_a$ and $\mathbf{w}_b$ and choose freely the remaining p self-loops and $q-2$ multiple edges, which can be done in less than $|\mathcal{I}|^p|\mathcal{J}|^{q-2} = O\left(n^{p+2q-4}\right)$ ways. It then suffices to show that the number of conflicting pairs $(\mathbf{w}_a, \mathbf{w}_b)$ is $o(n^4)$ as $n \to \infty$. A similar reasoning to that used in the previous cases gives that the total number

of conflicting pairs is bounded by

$$\begin{aligned}
&2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(d_{ni}^3 m_{nj}^2 m_{nk}^2 + m_{ni}^3 d_{nj}^2 d_{nk}^2) \\
&\leq 2\left(\max_{1\leq i\leq n} m_{ni} + \max_{1\leq i\leq n} d_{ni}\right)\left(\sum_{i=1}^{n} d_{ni}^2\left(\sum_{i=1}^{n} m_{ni}^2\right)^2 + \sum_{i=1}^{n} m_{ni}^2\left(\sum_{i=1}^{n} d_{ni}^2\right)^2\right) \\
&= o(n^{7/2}) = o(n^4).
\end{aligned}$$

We conclude that in any of the three cases the number of conflicts is negligible, which completes the proof. □

Proof of Proposition 9. Let $\mathcal{S}_n$ be the event that the resulting graph is simple, and note that the bi-degree-sequence $(\mathbf{M}^{(r)}, \mathbf{D}^{(r)})$ is the same as $(\mathbf{M}, \mathbf{D})$ given $\mathcal{S}_n$.

To prove part (a) note that for any $i, j = 0, 1, 2, \ldots,$

$$h^{(n)}(i,j) = \frac{1}{n}\sum_{i=1}^{n} P(M_k = i, D_k = j|\mathcal{S}_n) = \frac{1}{P(\mathcal{S}_n)} P(M_1 = i, D_1 = j, \mathcal{S}_n),$$

since the $\{(M_k, D_k)\}_{k=1}^n$ are identically distributed. Now let $\mathcal{G}_n = \sigma(M_1, \ldots, M_n, D_1, \ldots, D_n)$ and condition on $\mathcal{G}_n$ to obtain

$$P(M_1 = i, D_1 = j, \mathcal{S}_n) = E[1(M_1 = i, D_1 = j)P(\mathcal{S}_n|\mathcal{G}_n)],$$

from where it follows that

$$\begin{aligned}
\left|h^{(n)}(i,j) - f_i g_j\right| &\leq \left|\frac{E[1(M_1 = i, D_1 = j)(P(\mathcal{S}_n|\mathcal{G}_n) - P(\mathcal{S}_n))]}{P(\mathcal{S}_n)}\right| + \left|P(M_1 = i, D_1 = j) - f_i g_j\right| \\
&\leq E\left[\left|\frac{P(\mathcal{S}_n|\mathcal{G}_n)}{P(\mathcal{S}_n)} - 1\right|\right] + \left|P(M_1 = i, D_1 = j) - f_i g_j\right|.
\end{aligned}$$

Theorem 4 gives that the second term converges to zero, and for the first term use Theorem 8 to obtain that both $P(\mathcal{S}_n)$ and $P(\mathcal{S}_n|\mathcal{G}_n)$ converge to the same positive limit, so by dominated convergence,

$$\lim_{n\to\infty} E\left[\left|\frac{P(\mathcal{S}_n|\mathcal{G}_n)}{P(\mathcal{S}_n)} - 1\right|\right] \leq E\left[\lim_{n\to\infty}\left|\frac{P(\mathcal{S}_n|\mathcal{G}_n)}{P(\mathcal{S}_n)} - 1\right|\right] = 0.$$

For part (b) we only show the proof for $\widehat{g_k}^{(n)}$ since the proof for $\widehat{f_k}^{(n)}$ is symmetrical. Note that $\widehat{g_k}^{(n)}$ is a quantity defined on $\mathcal{S}_n$; recall that $\mathcal{D}_n = \left\{|\Delta_n| \leq n^{1-\kappa+\delta_0}\right\}$ and that

D_i has the same distribution as $\xi_i + \chi_i$ conditional on the event $\mathcal{D}_n$. Fix $\epsilon > 0$ and use the union bound to obtain

$$P\left(\left|\widehat{g_k}^{(n)} - g_k\right| > \epsilon \middle| \mathcal{S}_n\right) \leq \frac{1}{P(\mathcal{S}_n)} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} 1(D_i = k) - g_k\right| > \epsilon\right)$$

$$\leq \frac{1}{P(\mathcal{S}_n)} P\left(\frac{1}{n}\sum_{i=1}^{n} |1(\xi_i + \chi_i = k) - 1(\xi_i = k)| > \epsilon/2 \middle| \mathcal{D}_n\right) \quad \text{(A.16)}$$

$$+ \frac{1}{P(\mathcal{S}_n)P(\mathcal{D}_n)} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} 1(\xi_i = k) - g_k\right| > \epsilon/2\right). \quad \text{(A.17)}$$

By Theorem 8 and Lemma 1, $P(\mathcal{S}_n)$ and $P(\mathcal{D}_n)$ are bounded away from zero, so we only need to show that the numerators converge to zero. The arguments are the same as those used in the proof of Proposition 5; for (A.17) use Chebyshev's inequality to obtain that

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} 1(\xi_i = k) - g_k\right| > \epsilon/2\right) \leq \frac{\text{Var}\ (1(\xi_1 = k))}{n(\epsilon/2)^2} = O(n^{-1}),$$

as $n \to \infty$, and for (A.16)

$$P\left(\frac{1}{n}\sum_{i=1}^{n} |1(\xi_i + \chi_i = k) - 1(\xi_i = k)| > \epsilon/2 \middle| \mathcal{D}_n\right) \leq P\left(\frac{1}{n}\sum_{i=1}^{n} 1(\chi_i = 1) > \epsilon/2 \middle| \mathcal{D}_n\right)$$

$$\leq P\left(\frac{|\Delta_n|}{n} > \epsilon/2 \middle| \mathcal{D}_n\right) \leq 1(n^{-\kappa+\delta_0} > \epsilon/2),$$

which also converges to zero. This completes the proof. □

Finally, the last result of the paper, which refers to the erased directed configuration model, is given below. Since the technical part of the proof is to show that the probability that no in-degrees or out-degrees of a fixed node are removed during the erasing procedure, we split the proof of Proposition 10 into two parts. The following lemma contains the more delicate step.

Lemma 28. *Consider the graph obtained through the erased directed configuration model using as bi-degree-sequence* $(\mathbf{M}, \mathbf{D})$*, as constructed in Section 2.1.1. Let* E^+ *and* E^- *be the number of inbound stubs and outbound stubs, respectively, that have been removed from node* v_1 *during the erasing procedure. Then,*

$$\lim_{n\to\infty} P(E^+ = 0) = 1 \qquad \textit{and} \qquad \lim_{n\to\infty} P(E^- = 0) = 1.$$

Proof. We only show the result for E^+ since the proof for E^- is symmetric. Define the set

$$\mathcal{P}_n^+ = \{(i_1, \ldots, i_t) : 2 \le i_1 \ne i_2 \cdots \ne i_t \le n,\, 1 \le t \le n\},$$

and note that in order for all the inbound stubs of node v_1 to survive the erasing procedure, it must have been that they were paired to outbound stubs of M_1 different nodes from $\{v_2, \ldots, v_n\}$. Before we proceed it is helpful to recall some definitions from Section 2.1, $L_n = \sum_{i=1}^n M_i = \sum_{i=1}^n D_i$, $\Gamma_n = \sum_{i=1}^n \gamma_i$, $\Xi_n = \sum_{i=1}^n \xi_i$, $\Delta_n = \Gamma_n - \Xi_n$, and $\mathcal{D}_n = \{|\Delta_n| \le n^s\}$, where $s = 1 - \kappa + \delta_0$; also, $\{\gamma_i\}$ and $\{\xi_i\}$ are independent sequences of i.i.d. random variables having distributions F and G, respectively. Now fix $0 < \epsilon < 1 - s$ and let $\mathcal{G}_n = \sigma(M_1, \ldots, M_n, D_1, \ldots, D_n)$. Then, since $D_i = \xi_i + \chi_i \ge \xi_i$,

$$\begin{aligned}
P\left(E^+ = 0\right) &= E\left[P\left(E^+ = 0 \middle| \mathcal{G}_n\right)\right] \ge E\left[P\left(E^+ = 0 \middle| \mathcal{G}_n\right) 1(1 \le M_1 \le n^\epsilon)\right] + P(M_1 = 0) \\
&= E\left[\frac{1(1 \le M_1 \le n^\epsilon)}{L_n!} \sum_{(i_1, i_2, \ldots, i_{M_1}) \in \mathcal{P}_n^+} D_{i_1} D_{i_2} \cdots D_{i_{M_1}} (L_n - M_1)!\right] + P(M_1 = 0) \\
&\ge E\left[\frac{1(1 \le \gamma_1 + \tau_1 \le n^\epsilon)}{L_n!} \sum_{(i_1, i_2, \ldots, i_{(\gamma_1+\tau_1)}) \in \mathcal{P}_n^+} \xi_{i_1} \xi_{i_2} \cdots \xi_{i_{(\gamma_1+\tau_1)}} (L_n - \gamma_1 - \tau_1)! \middle| \mathcal{D}_n\right] \\
&\quad + P(M_1 = 0) \\
&\ge E\left[\frac{1(1 \le \gamma_1 \le n^\epsilon) 1(\tau_1 = 0)}{(L_n)^{\gamma_1}} \sum_{(i_1, i_2, \ldots, i_{\gamma_1}) \in \mathcal{P}_n^+} \xi_{i_1} \xi_{i_2} \cdots \xi_{i_{\xi_1}} \middle| \mathcal{D}_n\right] + P(M_1 = 0).
\end{aligned} \tag{A.18}$$

Next, condition on $\mathcal{F}_n = \sigma(\gamma_1, \ldots, \gamma_n, \xi_1, \ldots, \xi_n)$ and note that

$$P(\tau_1 = 0 | \mathcal{F}_n) = 1(\Delta_n \ge 0) + \frac{\Gamma_n}{\Gamma_n + |\Delta_n|} 1(\Delta_n < 0) \ge \frac{\Gamma_n}{\Gamma_n + |\Delta_n|}.$$

It follows that the expectation in (A.18) is equal to

$$\begin{aligned}
&E\left[P(\tau_1 = 0 | \mathcal{F}_n) \frac{1(1 \le \gamma_1 \le n^\epsilon)}{(L_n)^{\gamma_1}} \sum_{(i_1, i_2, \ldots, i_{\gamma_1}) \in \mathcal{P}_n^+} \xi_{i_1} \xi_{i_2} \cdots \xi_{i_{\gamma_1}} \middle| \mathcal{D}_n\right] \\
&\ge E\left[\frac{\Gamma_n}{\Gamma_n + |\Delta_n|} \cdot \frac{1(1 \le \gamma_1 \le n^\epsilon)}{(\Gamma_n + |\Delta_n|)^{\gamma_1}} \sum_{(i_1, i_2, \ldots, i_{\gamma_1}) \in \mathcal{P}_n^+} \xi_{i_1} \xi_{i_2} \cdots \xi_{i_{\gamma_1}} \middle| \mathcal{D}_n\right] \\
&\ge E\left[\frac{1(1 \le \gamma_1 \le n^\epsilon) \Gamma_n}{(\Gamma_n + n^s)^{\gamma_1 + 1}} \sum_{(i_1, i_2, \ldots, i_{\gamma_1}) \in \mathcal{P}_n^+} \xi_{i_1} \xi_{i_2} \cdots \xi_{i_{\gamma_1}} \middle| \mathcal{D}_n\right]
\end{aligned}$$

$$= \frac{1}{P(\mathcal{D}_n)} E\left[1(1 \le \gamma_1 \le n^\epsilon) \sum_{(i_1,i_2,\ldots,i_{\gamma_1}) \in \mathcal{P}_n^+} E\left[\frac{1(\mathcal{D}_n)\Gamma_n}{(\Gamma_n + n^s)^{\gamma_1+1}} \cdot \xi_{i_1}\xi_{i_2}\cdots\xi_{i_{\gamma_1}} \middle| \gamma_1\right]\right]$$
$$= \frac{1}{P(\mathcal{D}_n)} E\left[1(1 \le \gamma_1 \le n^\epsilon) \frac{(n-1)!}{(n-1-\gamma_1)!n^{\gamma_1}} E\left[\frac{1(\mathcal{D}_n)\Gamma_n n^{\gamma_1}}{(\Gamma_n + n^s)^{\gamma_1+1}} \cdot \xi_1\xi_2\cdots\xi_{\gamma_1} \middle| \gamma_1\right]\right].$$

It follows by Fatou's lemma, Lemma 1 and Theorem 4 that

$$\liminf_{n\to\infty} P(E^+ = 0) \ge E\left[1(\gamma_1 \ge 1) \liminf_{n\to\infty} E\left[\frac{1(\mathcal{D}_n)\Gamma_n n^{\gamma_1}}{(\Gamma_n + n^s)^{\gamma_1+1}} \cdot \xi_1\xi_2\cdots\xi_{\gamma_1} \middle| \gamma_1\right]\right] + P(\gamma_1 = 0).$$

Next, define the function $u_n^+ : \mathbb{N} \to [0,\infty)$ as

$$u_n^+(t) = E\left[\frac{1(|\Gamma_{n-1} + t - \Xi_n| \le n^s)(\Gamma_{n-1}+t)n^t}{(\Gamma_{n-1}+t+n^s)^{t+1}} \cdot \xi_1\xi_2\cdots\xi_t\right],$$

and note that it only remains to prove that for all $t \in \mathbb{N}$, $\liminf_{n\to\infty} u_n^+(t) = 1$.

Now let $0 < a < \mu$ and note that

$$u_n^+(t) \ge E\left[\frac{1(|\Gamma_{n-1} + t - \Xi_n| \le n^s)}{\mu^t} \cdot \xi_1\xi_2\cdots\xi_t\right] - P(\Gamma_{n-1} < an)$$
$$- E\left[1(\Gamma_{n-1} \ge an)\left|\frac{(\Gamma_{n-1}+t)n^t}{(\Gamma_{n-1}+t+n^s)^{t+1}} - \frac{1}{\mu^t}\right| \xi_1\xi_2\cdots\xi_t\right].$$

The SLLN and bounded convergence give $\lim_{n\to\infty} P(\Gamma_{n-1} < an) = 0$ and

$$\limsup_{n\to\infty} E\left[1(\Gamma_{n-1} \ge an)\left|\frac{(\Gamma_{n-1}+t)n^t}{(\Gamma_{n-1}+t+n^s)^{t+1}} - \frac{1}{\mu^t}\right| \xi_1\xi_2\cdots\xi_t\right]$$
$$\le E\left[\xi_1\xi_2\cdots\xi_t \limsup_{n\to\infty}\left|\frac{(\Gamma_{n-1}+t)n^t}{(\Gamma_{n-1}+t+n^s)^{t+1}} - \frac{1}{\mu^t}\right|\right] = 0,$$

from where it follows that

$$\liminf_{n\to\infty} u_n^+(t) \ge \liminf_{n\to\infty} E\left[\frac{1(|\Gamma_{n-1} + t - \Xi_n| \le n^s)}{\mu^t} \cdot \xi_1\xi_2\cdots\xi_t\right].$$

The last step is to condition on $\xi_1, \xi_2 \ldots, \xi_t$ and use Fatou's Lemma again to obtain

$$\liminf_{n\to\infty} E\left[\frac{1(|\Gamma_{n-1} + t - \Xi_n| \le n^s)}{\mu^t} \cdot \xi_1\xi_2\cdots\xi_t\right]$$
$$= \liminf_{n\to\infty} E\left[\frac{\xi_1\xi_2\cdots\xi_t}{\mu^t} P(|\Gamma_{n-1} + t - \Xi_n| \le n^s | \xi_1, \ldots, \xi_t)\right]$$
$$\ge E\left[\frac{\xi_1\xi_2\cdots\xi_t}{\mu^t} \liminf_{n\to\infty} P(|\Gamma_{n-1} + t - \Xi_n| \le n^s | \xi_1, \ldots, \xi_t)\right].$$

Finally, by the same reasoning used in the proof of Lemma 1, we obtain

$$\lim_{n\to\infty} P(|\Gamma_{n-1} + t - \Xi_n| \leq n^s | \xi_1, \ldots, \xi_t) = 1 \qquad \text{a.s.}$$

Since $E[\xi_1 \xi_2 \cdots \xi_t]/\mu^t = 1$, this completes the proof. □

Proof of Proposition 10. To prove part (a) note that since the $\{(M_i^{(e)}, D_i^{(e)})\}_{i=1}^n$ are identically distributed, then $h^{(n)}(i,j) = P(M_1^{(e)} = i, D_1^{(e)} = j)$. It follows that

$$\left|h^{(n)}(i,j) - f_i g_j\right| \leq \left|P(M_1^{(e)} = i, D_1^{(e)} = j) - P(M_1 = i, D_1 = j)\right| + \left|P(M_1 = i, D_1 = j) - f_i g_j\right|.$$

By Theorem 4 we have that $|P(M_1 = i, D_1 = j) - f_i g_j| \to 0$, as $n \to \infty$, and for the remaining term note that

$$\begin{aligned}
&\left|P(M_1^{(e)} = i, D_1^{(e)} = j) - P(M_1 = i, D_1 = j)\right| \\
&\leq E\left[\left|1(M_1^{(e)} = i, D_1^{(e)} = j) - 1(M_1 = i, D_1 = j)\right|\right] \\
&\leq E\left[\left|1(D_1^{(e)} = j) - 1(D_1 = j)\right|\right] + E\left[\left|1(M_1^{(e)} = i) - 1(M_1 = i)\right|\right]. \qquad \text{(A.19)}
\end{aligned}$$

To bound the expectations in (A.19) let E^+ and E^- be the number of inbound stubs and outbound stubs, respectively, that have been removed from node v_1 during the erasing procedure. Then,

$$\begin{aligned}
E\left[\left|1(D_1^{(e)} = j) - 1(D_1 = j)\right|\right] &\leq P\left(E^- \geq 1\right) \quad \text{and} \\
E\left[\left|1(M_1^{(e)} = i) - 1(M_1 = i)\right|\right] &\leq P\left(E^+ \geq 1\right).
\end{aligned}$$

By Lemma 28,

$$\lim_{n\to\infty} P(E^- \geq 1) = 0 \qquad \text{and} \qquad \lim_{n\to\infty} P(E^+ \geq 1) = 0,$$

which completes the proof of part (a).

For part (b) we only show the proof for $\widehat{g_k}^{(n)}$, since the proof for $\widehat{f_k}^{(n)}$ is symmetrical. Fix $\epsilon > 0$ and use the triangle inequality and the union bound to obtain

$$\begin{aligned}
P\left(|\widehat{g_k}(k) - g_k| > \epsilon\right) \leq\ & P\left(\left|\widehat{g_k}(k) - \frac{1}{n}\sum_{i=1}^n 1(D_i = k)\right| > \epsilon/2\right) \\
&+ P\left(\left|\frac{1}{n}\sum_{i=1}^n 1(D_i = k) - g_k\right| > \epsilon/2\right).
\end{aligned}$$

From the proof of Proposition 9, we know that the second probability converges to zero as $n \to \infty$, and for the first one use Markov's inequality to obtain

$$\begin{aligned} P\left(\left|\widehat{g_k}(k) - \frac{1}{n}\sum_{i=1}^{n} 1(D_i = k)\right| > \epsilon/2\right) &\leq P\left(\frac{1}{n}\sum_{i=1}^{n}\left|1(D_i^{(e)} = k) - 1(D_i = k)\right| > \epsilon/2\right) \\ &\leq \frac{2}{\epsilon} E\left[\left|1(D_1^{(e)} = k) - 1(D_1 = k)\right|\right] \\ &\leq \frac{2}{\epsilon} P(E^- \geq 1) \to 0, \end{aligned}$$

as $n \to \infty$, by Lemma 28. □

Appendix B

Proofs of Chapter 3

The last section of the paper contains most of the proofs. For the reader's convenience we have organized them in subsections according to the order in which their corresponding statements appear in the paper.

B.1 Proof of the coupling lemma

Recall from Section 3.3 that $\hat{N}_\emptyset$ denotes the number of offspring of the root node in the TBT (chosen from distribution (3.7)) and $\hat{N}_1$ denotes the number of offspring of a node chosen from distribution (3.8). Throughout this section we will also need to define

$$\mu_n^* = \mathbb{E}_n\left[\hat{N}_\emptyset\right] = \sum_{i,j,s,t} i f_n^*(i,j,s,t) = \frac{1}{n}\sum_{k=1}^{n} N_k = \frac{L_n}{n},$$

and

$$\mu_n = \mathbb{E}_n\left[\hat{N}_1\right] = \sum_{i,j,s,t} i f_n(i,j,s,t) = \frac{1}{L_n}\sum_{k=1}^{n} N_k D_k.$$

Before we give the proof of the Coupling Lemma 12 we will need the following estimates for the growth of the process $\{\hat{Z}_k\}$.

Lemma 29. *Suppose* $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ *satisfies Assumption 1 and recall that* $\mu = \nu_2/\nu_1$. *Then, for any constants* $K > 0$, *any nonnegative sequence* $\{x_n\}$ *with* $x_n \to \infty$ *and any* $k = O(n^\gamma)$,

$$P\left(\max_{0\le r\le k} \frac{\hat{Z}_r}{\mu^r} > K x_n \middle| \Omega_n\right) = O\left(x_n^{-1}\right), \qquad n \to \infty.$$

Proof. Start by noting that for any $r = 0, 1, 2, \ldots,$

$$\mathbb{E}_n[\hat{Z}_r] = \mu_n^* \mu_n^r. \tag{B.1}$$

Moreover, on the event Ω_n,

$$\begin{aligned}
\mu_n &= \frac{n\nu_2(1 + O(n^{-\gamma}))}{n\nu_1(1 + O(n^{-\gamma}))} = \mu(1 + O(n^{-\gamma})), \qquad \text{and} \\
\mu_n^* &= \frac{n\nu_1(1 + O(n^{-\gamma}))}{n} = \nu_1(1 + O(n^{-\gamma})).
\end{aligned}$$

Next, note that conditionally on $\mathscr{F}_n$, the process

$$X_r = \frac{\hat{Z}_r}{\mu_n^* \mu_n^r} = \frac{1}{\mu_n^* \mu_n^r} \sum_{\mathbf{i} \in \hat{A}_{r-1}} \hat{N}_{\mathbf{i}}, \quad r \geq 1, \qquad X_0 = \frac{\hat{N}_\emptyset}{\mu_n^*}$$

is a nonnegative martingale with respect to the filtration $\sigma\left(\mathcal{F}_r \cup \mathscr{F}_n\right)$, where $\mathcal{F}_r = \sigma\big(\hat{N}_{\mathbf{i}} : \mathbf{i} \in \hat{A}_s,\, s \leq r\big)$. Therefore, we can apply Doob's inequality, conditionally on $\mathscr{F}_n$, to obtain

$$\begin{aligned}
P\left(\max_{0 \leq r \leq k} \frac{\hat{Z}_r}{\mu^r} > K x_n \middle| \Omega_n\right) &= P\left(\max_{0 \leq r \leq k} \frac{X_r \mu_n^* \mu_n^r}{\mu^r} > K x_n \middle| \Omega_n\right) \\
&= P\left(\max_{0 \leq r \leq k} X_r \nu_1 (1 + O(n^{-\gamma}))^{r+1} > K x_n \middle| \Omega_n\right) \\
&\leq \frac{1}{P(\Omega_n)} E\left[1(\Omega_n) \mathbb{E}_n\left[1\left(\max_{0 \leq r \leq k} X_r > \frac{K x_n}{\nu_1 (1 + O(n^{-\gamma}))^{k+1}}\right)\right]\right] \\
&\leq \frac{1}{P(\Omega_n)} E\left[1(\Omega_n) \frac{\mathbb{E}_n[X_k] \nu_1 (1 + O(n^{-\gamma}))^{k+1}}{K x_n}\right] \\
&= \frac{\nu_1 (1 + O(n^{-\gamma}))^{k+1}}{K x_n} \qquad (\text{since } \mathbb{E}_n[X_k] = 1).
\end{aligned}$$

Noting that $(1 + O(n^{-\gamma}))^k = e^{O(kn^{-\gamma})} = O(1)$ as $n \to \infty$ gives that this last term is $O(x_n^{-1})$. This completes the proof. □

We now give the proof of the coupling lemma.

Proof of Lemma 12. Start by defining

$$x_n = \begin{cases} (n/\mu^{2k})^{1/2}, & \mu > 1, \\ (n/k^2)^{1/2}, & \mu = 1, \\ n^{1/2}, & \mu < 1, \end{cases} \qquad \text{and} \qquad F_k = \left\{\max_{0 \leq r \leq k} \frac{\hat{Z}_r}{\mu^r} \leq x_n\right\}.$$

Note that $x_n \to \infty$ as $n \to \infty$ for all $1 \le k \le h \log n$ when $\mu > 1$ and for all $1 \le k \le n^b$, $b < \min\{1/2, \gamma\}$, when $\mu \le 1$. The constraint $b < \gamma$ will allow us to use Lemma 29.

Next, note that the jth inbound stub of node $i \in A_s$ (where the label i refers to the order in which the node was added to the graph during the exploration process) will be the first one to be paired with an outbound stub having label 2 or 3 with probability

$$\frac{1}{L_n}\left(\sum_{r=0}^{s-1}\hat{V}_r + \sum_{t=1}^{i-1} D_t + (j-1)\right) \le \frac{1}{L_n}\sum_{r=0}^{s}\hat{V}_r =: P_s.$$

It follows that,

$$\begin{aligned} P(\tau = s|\Omega_n) &\le P(\tau = s, F_k|\Omega_n) + P(\tau = s, F_k^c|\Omega_n) \\ &\le P(\mathrm{Bin}(\hat{Z}_s, P_s) \ge 1,\, F_k|\Omega_n) + P(\tau = s, F_k^c|\Omega_n), \end{aligned}$$

where $\mathrm{Bin}(n,p)$ is a Binomial random variable with parameters (n,p). It follows that if we let $\mathcal{F}_k = \sigma(\hat{Z}_r, \hat{V}_r : 1 \le r \le k)$, then

$$\begin{aligned} P(\tau \le k|\Omega_n) &= \sum_{s=0}^{k} P(\tau = s|\Omega_n) \\ &\le \sum_{s=0}^{k}\left\{P\left(\mathrm{Bin}(\hat{Z}_s, P_s) \ge 1,\, F_k\middle|\,\Omega_n\right) + P\left(\tau = s,\, F_k^c|\,\Omega_n\right)\right\} \\ &\le \sum_{s=0}^{k} E\left[1(F_k)P(\mathrm{Bin}(\hat{Z}_s, P_s) \ge 1|\mathcal{F}_k)\middle|\,\Omega_n\right] + P\left(F_k^c|\,\Omega_n\right) \\ &\le \sum_{s=0}^{k} E\left[1(F_k)\hat{Z}_s P_s\middle|\,\Omega_n\right] + P\left(F_k^c|\,\Omega_n\right), \end{aligned}$$

where in the last step we used Markov's inequality. Now, use the bound for $\hat{Z}_s$ implied by F_k and recall that $|\hat{A}_r| = \hat{Z}_{r-1}$ to obtain

$$\begin{aligned} E\left[1(F_k)\hat{Z}_s P_s\middle|\,\Omega_n\right] &\le E\left[\mu^s x_n P_s|\,\Omega_n\right] \qquad \text{(B.2)} \\ &= \frac{\mu^s x_n}{\nu_1 n}\sum_{r=0}^{s} E\left[\hat{V}_r\middle|\,\Omega_n\right](1 + O(n^{-\gamma})) \\ &= \frac{\mu^s x_n}{\nu_1 n}\left\{E\left[\hat{V}_0\middle|\,\Omega_n\right] + \sum_{r=1}^{s} E\left[\mathbb{E}_n\left[\hat{V}_r|\hat{Z}_{r-1}\right]\middle|\,\Omega_n\right]\right\}(1 + O(n^{-\gamma})) \\ &= \frac{\mu^s x_n}{\nu_1 n}\left\{E\left[\mu_n^*|\,\Omega_n\right] + \sum_{r=1}^{s} E\left[\hat{Z}_{r-1}\lambda_n\middle|\,\Omega_n\right]\right\}(1 + O(n^{-\gamma})), \end{aligned}$$

where in the first equality we used that on the set Ω_n we have $L_n = \nu_1 n(1 + O(n^{-\gamma}))$, and on the second equality we used the observation that

$$\mathbb{E}_n\left[\hat{V}_0\right] = \mathbb{E}_n\left[\hat{D}_\emptyset\right] = \mu_n^*, \qquad \mathbb{E}_n\left[\hat{V}_r \middle| \hat{Z}_{r-1}\right] = \hat{Z}_{r-1}\lambda_n, \quad r \geq 1,$$

where $\lambda_n = \mathbb{E}_n[\hat{D}_1]$. Moreover, on the set Ω_n we have that

$$\lambda_n = \frac{1}{L_n}\sum_{k=1}^{n} D_k^2 = \frac{n\nu_3(1 + O(n^{-\gamma}))}{n\nu_1(1 + O(n^{-\gamma}))} = \lambda(1 + O(n^{-\gamma})),$$

so we obtain

$$\begin{aligned} E\left[1(F_k)\hat{Z}_s P_s \middle| \Omega_n\right] &\leq \frac{\mu^s x_n}{\nu_1 n}\left\{\nu_1 + \sum_{r=1}^{s}\lambda E\left[\hat{Z}_{r-1}\middle|\Omega_n\right]\right\}(1 + O(n^{-\gamma})) \\ &= \frac{\mu^s x_n}{\nu_1 n}\left\{\nu_1 + \sum_{r=1}^{s}\lambda E\left[\mu_n^*\mu_n^{r-1}\middle|\Omega_n\right]\right\}(1 + O(n^{-\gamma})) \qquad \text{(by (B.1))}. \end{aligned}$$

Using the observation that $E\left[\mu_n^*\mu_n^{r-1}\middle|\Omega_n\right] = \nu_1\mu^{r-1}(1 + O(n^{-\gamma}))^{r-1}$ (see the proof of Lemma 29), and the condition $r - 1 < s \leq k = O(n^\gamma)$, gives

$$P(\tau \leq k|\Omega_n) \leq (1 + O(1))\frac{(\lambda + 1)x_n}{n}\sum_{s=0}^{k}\sum_{r=0}^{s}\mu^{s+r} + P(F_k^c|\Omega_n).$$

Note that we did not compute $E\left[\hat{Z}_s P_s\middle|\Omega_n\right]$ in (B.2) directly, since that would have led to having to compute $\mathbb{E}_n\left[\hat{Z}_{s-1}^2\right]$ and neither $\hat{N}_0$ nor $\hat{N}_1$ are required to have finite second moments in the limit. Now, since by Lemma 29 we have that $P(F_k^c|\Omega_n) = O\left(x_n^{-1}\right)$, and

$$\sum_{s=0}^{k}\sum_{r=0}^{s}\mu^{s+r} \leq \begin{cases} \mu^{2(k+1)}/(\mu-1)^2, & \mu > 1, \\ (k+1)(k+2)/2, & \mu = 1, \\ 1/(1-\mu), & \mu < 1, \end{cases}$$

we conclude that

$$P(\tau \leq k|\Omega_n) = \begin{cases} O\left(x_n\mu^{2k}n^{-1} + x_n^{-1}\right) = O\left((n/\mu^{2k})^{-1/2}\right), & \mu > 1, \\ O\left(x_n k^2 n^{-1} + x_n^{-1}\right) = O\left((n/k^2)^{-1/2}\right), & \mu = 1, \\ O\left(x_n n^{-1} + x_n^{-1}\right) = O\left(n^{-1/2}\right), & \mu < 1, \end{cases}$$

as $n \to \infty$. This completes the proof. □

B.2 Coupling of weighted branching processes

We first prove Proposition 13, which bounds the Kantorovich-Rubinstein distance of the linear processes on two coupled weighted branching processes, by the same distance of their generic branching vectors.

Proof of Proposition 13. Define $\mathcal{E} = E_\pi\left[|\hat{Q} - Q| + \sum_{i=1}^{\infty} |\hat{B}_i - B_i|\right]$, where the vector $(Q, B_1, B_2, \dots, \hat{Q}, \hat{B}_1, \hat{B}_2, \dots)$ is distributed according to π. Recall that the weights $\Pi_{\mathbf{i}}$ and $\hat{\Pi}_{\mathbf{i}}$ follow the recursions

$$\Pi_{(\mathbf{i},j)} = \Pi_{\mathbf{i}} B_{(\mathbf{i},j)} \qquad \text{and} \qquad \hat{\Pi}_{(\mathbf{i},j)} = \hat{\Pi}_{\mathbf{i}} \hat{B}_{(\mathbf{i},j)},$$

with $\Pi_\emptyset = \hat{\Pi}_\emptyset = 1$. Now note that for $j = 0$ we have

$$E\left[\left|\hat{W}^{(0)} - W^{(0)}\right|\right] = E\left[\left|\hat{Q} - Q\right|\right] \le \mathcal{E}.$$

To analyze the expression for $j \ge 1$, define for $r \ge 1$, $W_r^{(j-1)} = \sum_{(r,\mathbf{i})\in\mathbb{N}_+^j} Q_{(r,\mathbf{i})}\Pi_{(r,\mathbf{i})}/B_r$ and $\hat{W}_r^{(j-1)} = \sum_{(r,\mathbf{i})\in\mathbb{N}_+^j} \hat{Q}_{(j,\mathbf{i})}\hat{\Pi}_{(r,\mathbf{i})}/\hat{B}_r$. We then have

$$\hat{W}^{(j)} = \sum_{r=1}^{\infty} \hat{B}_r \hat{W}_r^{(j-1)} \qquad \text{and} \qquad W^{(j)} = \sum_{r=1}^{\infty} B_r W_r^{(j-1)}.$$

Next, note that

$$\begin{aligned}
E\left[\left|\hat{W}^{(j)} - W^{(j)}\right|\right] &\le \sum_{r=1}^{\infty} E\left[\left|\hat{B}_r\hat{W}_r^{(j-1)} - B_r W_r^{(j-1)}\right|\right] \\
&\le \sum_{r=1}^{\infty} \left\{E\left[\left|W_r^{(j-1)}(\hat{B}_r - B_r)\right| + \left|\hat{B}_r\left(\hat{W}_r^{(j-1)} - W_r^{(j-1)}\right)\right|\right]\right\} \\
&\le \sum_{r=1}^{\infty} E\left[\left|\hat{B}_r - B_r\right|\right] E\left[\left|W_r^{(j)}\right|\right] + \sum_{r=1}^{\infty} E\left[|\hat{B}_r|\right] E\left[\left|\hat{W}_r^{(j)} - W_r^{(j)}\right|\right] \\
&\le E\left[\left|W^{(j-1)}\right|\right]\mathcal{E} + \hat{\rho} E\left[\left|\hat{W}^{(j-1)} - W^{(j-1)}\right|\right],
\end{aligned}$$

where we used the independence of the root vectors and their offspring, the observation that the random variables $\{W_r^{(j-1)}\}_{r\ge1}$ are i.i.d. with the same distribution as $W^{(j-1)}$ and $\{(\hat{W}_r^{(j-1)} - W_r^{(j-1)})\}_{r\ge1}$ are i.i.d. with the same distribution as $\hat{W}^{(j-1)} - W^{(j-1)}$. Moreover,

$$E\left[\left|W^{(j-1)}\right|\right] \le E\left[|Q|\right] \sum_{\mathbf{i}\in\mathbb{N}_+^{j-1}} E\left[|\Pi_{\mathbf{i}}|\right] = E\left[|Q|\right]\rho^{j-1}.$$

It follows that

$$E\left[\left|\hat{W}^{(j)} - W^{(j)}\right|\right] \le E\left[|Q|\right]\rho^{j-1}\mathcal{E} + \hat{\rho}E\left[\left|\hat{W}^{(j-1)} - W^{(j-1)}\right|\right] \quad \text{after } (j-1) \text{ iterations}$$
$$\le \left(\hat{\rho}^j + E\left[|Q|\right]\sum_{t=0}^{j-1}\rho^t\hat{\rho}^{j-1-t}\right)\mathcal{E}.$$

This completes the proof. □

Similarly we can prove an upper bound for weighted branching trees.

Proof of Proposition 14. We construct the processes $\hat{W}^{(j)}$ and $W^{(j)}$ on two weighted branching trees using a coupled vector $(Q_\emptyset, N_\emptyset, \hat{Q}_\emptyset, \hat{N}_\emptyset)$ for the root nodes $\emptyset$, distributed according to π^*, and a sequence of i.i.d. random vectors $\{(Q_\mathbf{i}, N_\mathbf{i}, C_\mathbf{i}, \hat{Q}_\mathbf{i}, \hat{N}_\mathbf{i}, \hat{C}_\mathbf{i}))\}_{\mathbf{i}\in U, \mathbf{i}\neq\emptyset}$, independent of $(Q_\emptyset, N_\emptyset, \hat{Q}_\emptyset, \hat{N}_\emptyset)$, distributed according to π for all other nodes.

Next, for $\mathbf{i} \in \mathbb{N}_+^k$, $k \ge 1$, let $B_\mathbf{i}^{(0)} = C_\mathbf{i}Q_\mathbf{i}$, $\hat{B}_\mathbf{i}^{(0)} = \hat{C}_\mathbf{i}\hat{Q}_\mathbf{i}$, $B_\mathbf{i}^{(j)} = C_\mathbf{i}1(N_\mathbf{i} \ge i)$, and $\hat{B}_\mathbf{i}^{(j)} = \hat{C}_\mathbf{i}1(\hat{N}_\mathbf{i} \ge i)$, for $j \ge 1$, and note that

$$\Pi_\mathbf{i}Q_\mathbf{i} = Q_\mathbf{i}\prod_{r=1}^{k} C_{\mathbf{i}|r}1(i_r \le N_{\mathbf{i}|r-1}) = 1(i_1 \le N_\emptyset)\prod_{r=1}^{k-1} B_{\mathbf{i}|r}^{(i_{r+1})}B_\mathbf{i}^{(0)},$$

and similarly,

$$\hat{\Pi}_\mathbf{i}\hat{Q}_\mathbf{i} = 1(i_1 \le \hat{N}_\emptyset)\prod_{r=1}^{k-1} \hat{B}_{\mathbf{i}|r}^{(i_{r+1})}\hat{B}_\mathbf{i}^{(0)},$$

with the convention that $\prod_{i=a}^{b} x_i \equiv 1$ if $b < a$.

Let $\mathcal{E}^* = E_{\pi^*}[|\hat{Q}-Q| + |\hat{N}-N|]$, where $(Q, N, \hat{Q}, \hat{N})$ is distributed according to π^*, and $\mathcal{E} = E_\pi[\sum_{i=0}^{\infty}|\hat{B}^{(i)} - B^{(i)}|]$, where $(B^{(0)}, B^{(1)}, B^{(2)}, \dots, \hat{B}^{(0)}, \hat{B}^{(1)}, \hat{B}^{(2)}, \dots)$ is distributed according to π. It follows that

$$E\left[\left|\hat{W}^{(0)} - W^{(0)}\right|\right] = E_{\pi^*}\left[\left|\hat{Q} - Q\right|\right] \le \mathcal{E}^*,$$

and for $j \geq 1$,

$$
\begin{aligned}
E\left[\left|\hat{W}^{(j)} - W^{(j)}\right|\right] &= E\left[\left|\sum_{\mathbf{i}\in\mathbb{N}_+^j} 1(i_1 \leq \hat{N}_\emptyset) \prod_{r=1}^{j-1} \hat{B}_{\mathbf{i}|r}^{(i_{r+1})} \hat{B}_{\mathbf{i}}^{(0)} - \sum_{\mathbf{i}\in\mathbb{N}_+^j} 1(i_1 \leq N_\emptyset) \prod_{r=1}^{j-1} B_{\mathbf{i}|r}^{(i_{r+1})} B_{\mathbf{i}}^{(0)}\right|\right] \\
&\leq \sum_{\mathbf{i}\in\mathbb{N}_+^j} E\left[\left|1(i_1 \leq \hat{N}_\emptyset) \prod_{r=1}^{j-1} \hat{B}_{\mathbf{i}|r}^{(i_{r+1})} \left(\hat{B}_{\mathbf{i}}^{(0)} - B_{\mathbf{i}}^{(0)}\right)\right|\right] \\
&\quad + \sum_{\mathbf{i}\in\mathbb{N}_+^j} E\left[\left|\left(1(i_1 \leq \hat{N}_\emptyset) - 1(i_1 \leq N_\emptyset)\right) \prod_{r=1}^{j-1} \hat{B}_{\mathbf{i}|r}^{(i_{r+1})} B_{\mathbf{i}}^{(0)}\right|\right] \\
&\quad + \sum_{\mathbf{i}\in\mathbb{N}_+^j} E\left[\left|1(i_1 \leq N_\emptyset) \left(\prod_{r=1}^{j-1} \hat{B}_{\mathbf{i}|r}^{(i_{r+1})} - \prod_{r=1}^{j-1} B_{\mathbf{i}|r}^{(i_{r+1})}\right) B_{\mathbf{i}}^{(0)}\right|\right] \\
&= \sum_{\mathbf{i}\in\mathbb{N}_+^j} P(\hat{N} \geq i_1) \prod_{r=1}^{j-1} E\left[|\hat{C}|1(\hat{N} \geq i_{r+1})\right] E_\pi\left[|\hat{B}^{(0)} - B^{(0)}|\right] \\
&\quad + \sum_{\mathbf{i}\in\mathbb{N}_+^j} E_{\pi^*}\left[\left|1(i_1 \leq \hat{N}) - 1(i_1 \leq N)\right|\right] \prod_{r=1}^{j-1} E\left[|\hat{C}|1(\hat{N} \geq i_{r+1})\right] E\left[|CQ|\right] \\
&\quad + \sum_{\mathbf{i}\in\mathbb{N}_+^j} P(N \geq i_1) E\left[\left|\prod_{r=1}^{j-1} \hat{B}_{\mathbf{i}|r}^{(i_{r+1})} - \prod_{r=1}^{j-1} B_{\mathbf{i}|r}^{(i_{r+1})}\right|\right] E[|CQ|],
\end{aligned}
$$

where we have used the independence among the generic branching vectors of the weighted branching trees. Moreover,

$$
\sum_{\mathbf{i}\in\mathbb{N}_+^j} P(\hat{N} \geq i_1) \prod_{r=1}^{j-1} E\left[|\hat{C}|1(\hat{N} \geq i_{r+1})\right] = \sum_{i=1}^{\infty} P(\hat{N} \geq i) \left(\sum_{k=1}^{\infty} E\left[|\hat{C}|1(\hat{N} \geq k)\right]\right)^{j-1} = E[\hat{N}]\hat{\rho}^{j-1},
$$

where $\hat{\rho} = E[\hat{N}|\hat{C}|]$. Similarly,

$$
\sum_{i=1}^{\infty} E_{\pi^*}\left[\left|1(i \leq \hat{N}) - 1(i \leq N)\right|\right] = \sum_{i=1}^{\infty} E_{\pi^*}\left[1(N < i \leq \hat{N}) + 1(\hat{N} < i \leq N)\right] = E_{\pi^*}\left[|\hat{N} - N|\right].
$$

It follows that

$$
\begin{aligned}
E\left[\left|\hat{W}^{(j)} - W^{(j)}\right|\right] &\leq E[\hat{N}]\hat{\rho}^{j-1} E_\pi\left[|\hat{B}^{(0)} - B^{(0)}|\right] + E[|CQ|]\hat{\rho}^{j-1} E_{\pi^*}\left[|\hat{N} - N|\right] \\
&\quad + E[N]E[|CQ|] \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} E\left[\left|\prod_{r=1}^{j-1} \hat{B}_{(1,i_2,\ldots,i_r)}^{(i_{r+1})} - \prod_{r=1}^{j-1} B_{(1,i_2,\ldots,i_r)}^{(i_{r+1})}\right|\right].
\end{aligned}
$$

To analyze the last expectation let $a_j = \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} E\left[\left|\prod_{r=1}^{j-1}\hat{B}^{(i_{r+1})}_{(1,i_2,\ldots,i_r)} - \prod_{r=1}^{j-1}B^{(i_{r+1})}_{(1,i_2,\ldots,i_r)}\right|\right]$ for $j \geq 2$, and $a_1 = 0$. It follows that for $j \geq 2$,

$$\begin{aligned}
a_j \leq & \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} E\left[\left|\prod_{r=1}^{j-2}\hat{B}^{(i_{r+1})}_{(1,i_2,\ldots,i_r)} - \prod_{r=1}^{j-2}B^{(i_{r+1})}_{(1,i_2,\ldots,i_r)}\right|\left|\hat{B}^{(i_j)}_{(1,i_2,\ldots,i_{j-1})}\right|\right] \\
& + \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} E\left[\left|\prod_{r=1}^{j-2}B^{(i_{r+1})}_{(1,i_2,\ldots,i_r)}\right|\left|\hat{B}^{(i_j)}_{(1,i_2,\ldots,i_{j-1})} - B^{(i_j)}_{(1,i_2,\ldots,i_{j-1})}\right|\right] \\
= & \, a_{j-1}\sum_{i_j=1}^{\infty} E\left[|\hat{C}|1(\hat{N}\geq i_j)\right] + \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}}\prod_{r=1}^{j-2} E\left[|C|1(N\geq i_{r+1})\right] E_\pi\left[\left|\hat{B}^{(i_j)} - B^{(i_j)}\right|\right] \\
= & \, \hat{\rho}\, a_{j-1} + \rho^{j-2}\sum_{i=1}^{\infty} E_\pi\left[\left|\hat{B}^{(i)} - B^{(i)}\right|\right],
\end{aligned}$$

where $\rho = E[N|C|]$. Iterating this recursion $j-2$ times gives

$$a_j \leq \hat{\rho}^{j-1}a_1 + \sum_{t=0}^{j-2}\hat{\rho}^t\rho^{j-2-t}\sum_{i=1}^{\infty}E_\pi\left[\left|\hat{B}^{(i)} - B^{(i)}\right|\right] = \sum_{t=0}^{j-2}\hat{\rho}^t\rho^{j-2-t}\sum_{i=1}^{\infty}E_\pi\left[\left|\hat{B}^{(i)} - B^{(i)}\right|\right].$$

We conclude that for $j \geq 1$,

$$\begin{aligned}
E\left[\left|\hat{W}^{(j)} - W^{(j)}\right|\right] \leq & \, E[\hat{N}]\hat{\rho}^{j-1}E_\pi\left[|\hat{B}^{(0)} - B^{(0)}|\right] + E[|CQ|]\hat{\rho}^{j-1}E_{\pi^*}\left[|\hat{N} - N|\right] \\
& + 1(j\geq 2)E[N]E[|CQ|]\sum_{t=0}^{j-2}\hat{\rho}^t\rho^{j-2-t}\sum_{i=1}^{\infty}E_\pi\left[\left|\hat{B}^{(i)} - B^{(i)}\right|\right] \\
\leq & \left(E[\hat{N}] \vee \frac{E[N]E[|CQ|]}{\rho}\right)\left(\sum_{t=0}^{j-1}\hat{\rho}^t\rho^{j-1-t}\right)\mathcal{E} + E[|CQ|]\hat{\rho}^{j-1}\mathcal{E}^*.
\end{aligned}$$

□

We now proceed to prove the two main theorems of the paper, Theorems 15 and 16.

Proof of Theorem 15. Case 1: Weighted branching processes.

Choose a coupling π of μ_n and μ such that $E_\pi\left[\left|Q - Q^{(n)}\right| + \sum_{j=1}^{\infty}\left|B_j - B_j^{(n)}\right|\right] = d_1(\mu,\mu_n)$. If we construct both WBPs based on this optimal coupling, then by Proposition 13,

$$\begin{aligned}
E\left[\left|W^{(n,j)} - W^{(j)}\right|\right] \leq & \left(\rho_n^j + E\left[|Q|\right]\sum_{t=0}^{j-1}\rho^t\rho_n^{j-1-t}\right)d_1(\mu,\mu_n) \\
& \leq (E[|Q|\vee\rho)(j+1)\rho^{j-1}\left(1\vee\frac{\rho_n}{\rho}\right)^j d_1(\mu,\mu_n).
\end{aligned}$$

For fixed $j \geq 1$ note that $|\rho_n - \rho| \leq d_1(\mu, \mu_n)$, and hence $(1 \vee (\rho_n/\rho))^j \to 1$ as $n \to \infty$, which in turn implies that $E\left[\left|W^{(n,j)} - W^{(j)}\right|\right] \to 0$.

Assume now $Q^{(n)} = Q \equiv 1$, and $\{C_j^{(n)}, C_j\}$ are nonnegative for all n, j; suppose $j_n \to \infty$ and $j_n d_1(\mu, \mu_n) \to 0$ as $n \to \infty$. First note that $\{W^{(j)}/\rho^j\}$ is a mean one nonnegative martingale with respect to the filtration generated by $\mathcal{G}_j = \sigma((B_{(\mathbf{i},1)}, B_{(\mathbf{i},2)}, \dots) : \mathbf{i} \in \mathbb{N}_+^r, 0 \leq r < j)$, $\mathcal{G}_0 = \sigma(\varnothing)$. Therefore,

$$\begin{aligned} E\left[\left|\frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}}\right|\right] &\leq E\left[\frac{1}{\rho_n^{j_n}}\left|W^{(n,j_n)} - W^{(j_n)}\right|\right] + E\left[\frac{W^{(j_n)}}{\rho^{j_n}}\left|\left(\frac{\rho}{\rho_n}\right)^{j_n} - 1\right|\right] \\ &\leq \frac{(1 \vee \rho)}{\rho}(j_n + 1)\left(\frac{\rho}{\rho_n}\right)^{j_n}\left(1 \vee \frac{\rho_n}{\rho}\right)^{j_n} d_1(\mu, \mu_n) + \left|\left(\frac{\rho}{\rho_n}\right)^{j_n} - 1\right| \\ &\leq \frac{(1 \vee \rho)}{\rho}(j_n + 1)e^{j_n(\rho/\rho_n - 1)^+} d_1(\mu, \mu_n) + j_n\left|\frac{\rho}{\rho_n} - 1\right| e^{(j_n - 1)(\rho/\rho_n - 1)^+}, \end{aligned}$$

where in the last step we used the inequalities

$$(x \vee 1)^j \leq e^{j(x-1)^+} \qquad \text{and} \qquad |x^j - 1| \leq j|x - 1|e^{(j-1)(x-1)^+} \qquad \text{for all } x > 0, j \in \mathbb{N}. \tag{B.3}$$

Since $j_n d_1(\mu, \mu_n) \to 0$ as $n \to \infty$, then so does $j_n|\rho/\rho_n - 1| \to 0$ as $n \to \infty$, and we conclude that the expected value converges to zero. Since by the martingale convergence theorem $W^{(j_n)}/\rho^{j_n} \to \mathcal{W}$ almost surely, then

$$\frac{W^{(n,j_n)}}{\rho_n^{j_n}} \Rightarrow \mathcal{W}, \qquad n \to \infty.$$

If $E[\mathcal{W}] = 1$ then $E\left[|W^{(j_n)}/\rho^{j_n} - \mathcal{W}|\right] \to 0$ and we can replace the convergence in distribution to convergence in the Kantorovich-Rubinstein distance.

The last statement of the theorem for weighted branching processes follows from noting that

$$\begin{aligned} \frac{1}{\rho^j} E\left[\left|W^{(n,j)} - W^{(j)}\right|\right] &\leq \left|\frac{1}{\rho^j} - \frac{1}{\rho_n^j}\right| E\left[|W^{(n,j)}|\right] + E\left[\left|\frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}}\right|\right] \\ &= \left|\left(\frac{\rho}{\rho_n}\right)^j - 1\right| + E\left[\left|\frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}}\right|\right], \end{aligned}$$

which were already shown to converge to zero for all $0 \leq j \leq j_n$. This completes the proof for this case.

Case 2: Weighted branching trees.

Construct versions of the processes $\{W^{(n,j)} : j \geq 0\}$ and $\{W^{(j)} : j \geq 0\}$ using a sequence of coupled vectors $\{(Q_{\mathbf{i}}^{(n)}, N_{\mathbf{i}}^{(n)}, C_{\mathbf{i}}^{(n)}, Q_{\mathbf{i}}, N_{\mathbf{i}}, C_{\mathbf{i}})\}_{\mathbf{i}\in U,\mathbf{i}\neq\emptyset}$ according to the coupling π satisfying $d_1(\mu_n,\mu) = E_\pi\left[|C^{(n)}Q^{(n)} - CQ| + \sum_{i=1}^{\infty}|C^{(n)}1(N^{(n)}\geq i) - C1(N\geq i)|\right]$. Let the root vector $(Q_\emptyset^{(n)}, N_\emptyset^{(n)}, Q_\emptyset, N_\emptyset)$ be distributed according to π^*, where $d_1(\nu_n^*,\nu^*) = E_{\pi^*}\left[|Q^{(n)} - Q| + |N^{(n)} - N|\right]$, and be independent of all other nodes.

By Proposition 14 we have $E\left[\left|W^{(n,0)} - W^{(0)}\right|\right] \leq d_1(\nu_n^*,\nu^*)$ and

$$E\left[\left|W^{(n,j)} - W^{(j)}\right|\right] \leq Kj(\rho_n\vee\rho)^{j-1}d_1(\mu_n,\mu) + K\rho_n^{j-1}d_1(\nu_n^*,\nu^*), \qquad j\geq 1,$$

with $K = \max\{E[N^{(n)}], E[|Q|]\}$. Note that

$$|\rho_n - \rho| = \left|\sum_{i=1}^{\infty} E\left[C^{(n)}1(N^{(n)}\geq i) - C1(N\geq i)\right]\right| \leq d_1(\mu_n,\mu).$$

The result for fixed j follows immediately.

Assume now that $Q^{(n)} = Q = 1$, and $\{C^{(n)}, C\}$ are nonnegative, and recall that C is independent of (Q,N), and therefore defines a weighted branching process. This in turn implies that $\{W^{(j)}/\rho^j\}$ is a nonnegative martingale with respect to the filtration generated by $\mathcal{H}_j = \sigma((N_{\mathbf{i}}, C_{(\mathbf{i},1)},\ldots,C_{(\mathbf{i},N_{\mathbf{i}})}) : \mathbf{i}\in A_r, 0\leq r<j)$, $\mathcal{H}_0 = \sigma(\varnothing)$. It follows that

$$\begin{aligned}
E\left[\left|\frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}}\right|\right] &\leq E\left[\left|W^{(n,j_n)} - W^{(j_n)}\right|\right]\frac{1}{\rho_n^{j_n}} + \left|\left(\frac{\rho}{\rho_n}\right)^{j_n} - 1\right| \\
&\leq Kj_n\left(\frac{\rho\vee\rho_n}{\rho_n}\right)^{j_n} d_1(\mu_n,\mu) + \frac{K}{\rho_n}d_1(\nu_n^*,\nu^*) + \left|\left(\frac{\rho}{\rho_n}\right)^{j_n} - 1\right| \\
&\leq Kj_n e^{j_n(\rho/\rho_n-1)^+}d_1(\mu_n,\mu) + \frac{K}{\rho_n}d_1(\nu_n^*,\nu^*) + j_n\left|\frac{\rho}{\rho_n} - 1\right| e^{(j_n-1)(\rho/\rho_n-1)^+},
\end{aligned}$$

where in the last step we used the inequalities (B.3). This last expression converges to zero since $j_n d_1(\mu_n,\mu)\to 0$ as $n\to\infty$.

The proof of the last statement is identical to that of *Case 1* and is therefore omitted. □

We now proceed to the non homogeneous case.

Proof of Theorem 16. Case 1: Weighted branching processes.

The result for fixed k follows from Theorem 16, since

$$\left|R^{(n,k)} - R^{(k)}\right| = \left|\sum_{j=0}^{k}\left(W^{(n,j)} - W^{(j)}\right)\right| \leq \sum_{j=0}^{k}\left|W^{(n,j)} - W^{(j)}\right|.$$

If in addition we have $\rho < 1$, then, by Proposition 13 (using the optimal coupling),

$$E\left[\left|R^{(n,k_n)} - R^{(k_n)}\right|\right] \leq \sum_{j=0}^{k_n} E\left[\left|W^{(n,j)} - W^{(j)}\right|\right] \leq \sum_{j=0}^{k_n}\left(\rho_n^j + E\left[|Q|\right]\sum_{t=0}^{j-1}\rho^t\rho_n^{j-1-t}\right) d_1(\mu_n,\mu).$$

Now note that since $|\rho_n - \rho| \leq d_1(\mu_n,\mu)$, then for any $0 < \varepsilon < 1-\rho$ we have that $\rho_n < 1-\varepsilon$ for all n sufficiently large. In this case,

$$E\left[\left|R^{(n,k_n)} - R^{(k_n)}\right|\right] \leq \sum_{j=0}^{\infty}\left((1-\varepsilon)^j + E\left[|Q|\right]j(1-\varepsilon)^{j-1}\right) d_1(\mu_n,\mu) \to 0,$$

as $n \to \infty$ for any $k_n \geq 1$. Since we also have that

$$E\left[\left|R^{(k_n)} - R\right|\right] = E\left[\left|\sum_{j=k_n+1}^{\infty} W^{(j)}\right|\right] \leq \sum_{j=k_n+1}^{\infty} E[|Q|]\rho^j = \frac{E[|Q|]\rho^{k_n+1}}{1-\rho},$$

then for any $k_n \to \infty$,

$$R^{(n,k_n)} \xrightarrow{d_1} R, \qquad n \to \infty.$$

Case 2: Weighted branching trees.

The proof of the result for fixed k follows from Theorem 15 as before. For k_n and $\rho < 1$ we use Proposition 14 (using the optimal couplings π^* and π) to obtain

$$\begin{aligned} E\left[\left|R^{(n,k_n)} - R^{(k_n)}\right|\right] &\leq \sum_{j=0}^{k_n} E\left[\left|W^{(n,j)} - W^{(j)}\right|\right] \\ &\leq d_1(\nu_n^*,\nu^*) + K\sum_{j=1}^{k_n}\left(\sum_{t=0}^{j-1}\rho_n^t\rho^{j-1-t}d_1(\mu_n,\mu) + \rho_n^{j-1}d_1(\nu_n^*,\nu^*)\right), \end{aligned}$$

with $K = \max\{E[N^{(n)}], E[|Q|]\}$. Using the same arguments from *Case 1* note that for any $0 < \varepsilon < 1-\rho$ and n sufficiently large,

$$E\left[\left|R^{(n,k_n)} - R^{(k_n)}\right|\right] \leq d_1(\nu_n^*,\nu^*) + K\sum_{j=1}^{\infty}\left(j(1-\varepsilon)^{j-1}d_1(\mu_n,\mu) + (1-\varepsilon)^{j-1}d_1(\nu_n^*,\nu^*)\right) \to 0,$$

as $n \to \infty$ for any $k_n \geq 1$. The rest of proof is the same as that of *Case 1* and is therefore omitted. □

The last result we need to prove in this section is Lemma 17

Proof of Lemma 17. From the definition of the Kantorovich-Rubinstein metric and the fact that the infimum is always attained (see, e.g., [Villani, 2009], Theorem 4.1), there exists a coupling π of $(N^{(n)}, Q^{(n)}, C^{(n)}, N, Q, C)$ such that

$$d_1(\nu_n, \nu) = E_\pi \left[|Q^{(n)} - Q| + |N^{(n)} - N| + |C^{(n)} - C| \right]. \tag{B.4}$$

Next, define the vectors

$$\mathbf{Y}_n = C^{(n)}(Q^{(n)}, 1(N^{(n)} \geq 1), 1(N^{(n)} \geq 2), \dots) \qquad \text{and} \qquad \mathbf{Y} = C(Q, 1(N \geq 1), 1(N \geq 2), \dots).$$

We will first show that $\|\mathbf{Y}_n - \mathbf{Y}\|_1 \xrightarrow{P} 0$ as $n \to \infty$. To this end, let $(\hat{Q}, \hat{N}, \hat{C}) = (Q^{(n)}, N^{(n)}, C^{(n)})$ to simplify the notation and define $X_n = \|(N^{(n)}, Q^{(n)}, C^{(n)}) - (N, Q, C)\|_1$. Note that (B.4) implies that $X_n \to 0$ in mean, and therefore in probability. Now note that

$$\begin{aligned}
\|\mathbf{Y}_n - \mathbf{Y}\|_1 &= |\hat{Q}\hat{C} - QC| + \sum_{i=1}^{\infty} |\hat{C}1(\hat{N} \geq i) - C1(N \geq i)| \\
&= |\hat{Q}\hat{C} - QC| + \sum_{i=1}^{\infty} \Big(|\hat{C} - C|1(i \leq \hat{N} \wedge N) + |\hat{C}|1(N < i \leq \hat{N}) + |C|1(\hat{N} < i \leq N) \Big) \\
&= |\hat{Q}\hat{C} - QC| + |\hat{C} - C|(\hat{N} \wedge N) + |\hat{C}|(\hat{N} - N)^+ + |C|(N - \hat{N})^+ \\
&\leq |\hat{C}||\hat{Q} - Q| + |Q||\hat{C} - C| + |\hat{C} - C|(\hat{N} \wedge N) + |\hat{C}|(\hat{N} - N)^+ + |C|(N - \hat{N})^+ \\
&\leq \Big(2|C^{(n)}| + |Q| + N + |C| \Big) X_n \xrightarrow{P} 0, \qquad n \to \infty,
\end{aligned}$$

by the converging together lemma. It remains to show that $\|\mathbf{Y}_n - \mathbf{Y}\|_1 \to 0$ in mean.

By the triangle's inequality we have that

$$Q_n \triangleq |\|\mathbf{Y}_n\|_1 - \|\mathbf{Y}\|_1| \leq \|\mathbf{Y}_n - \mathbf{Y}\|_1 \xrightarrow{P} 0, \qquad n \to \infty.$$

Also, by assumption,

$$E\left[\|\mathbf{Y}_n\|_1\right] = E\left[|\hat{Q}\hat{C}| + \sum_{i=1}^{\infty} |\hat{C}|1(\hat{N} \geq i) \right] = E\left[|\hat{Q}\hat{C}| + |\hat{C}|\hat{N} \right] \to E[|CQ| + |C|N] = E\left[\|\mathbf{Y}\|_1\right]$$

as $n \to \infty$, and therefore $E[Q_n] \to 0$ (see, e.g., Theorem 5.5.2 in [Durrett, 2010]). Now note that since $\|\mathbf{Y}_n - \mathbf{Y}\|_1 \leq \|\mathbf{Y}_n\|_1 + \|\mathbf{Y}\|_1 \leq Q_n + 2\|\mathbf{Y}\|_1$, we have

$$E\left[\|\mathbf{Y}_n - \mathbf{Y}\|_1\right] \leq E\left[\|\mathbf{Y}_n - \mathbf{Y}\|_1 1(Q_n \leq 1)\right] + E[Q_n] + 2E\left[\|\mathbf{Y}\|_1 1(Q_n > 1)\right],$$

where $\|\mathbf{Y}_n - \mathbf{Y}\|_1 1(Q_n \leq 1)$ and $\|\mathbf{Y}\|_1 1(Q_n > 1)$ are uniformly integrable by Theorem 13.3 in [Williams, 1991], and hence

$$\lim_{n\to\infty} E\left[\|\mathbf{Y}_n - \mathbf{Y}\|_1 1(Q_n \leq 1)\right] = \lim_{n\to\infty} E\left[\|\mathbf{Y}\|_1 1(Q_n > 1)\right] = 0.$$

□

B.3 Proof of the asymptotic behavior of $\mathcal{R}^*$

We give in this section the proof of Theorem 19 which describes the asymptotic behavior of the limit $\mathcal{R}^*$, which is essentially determined by the asymptotic behavior of the endogenous solution $\mathcal{R}$ given in (**??**). The tail behavior of $\mathcal{R}$ is the main focus of the work in [Volkovich and Litvak, 2010; Jelenković and Olvera-Cravioto, 2010; Jelenković and Olvera-Cravioto, 2012b; Jelenković and Olvera-Cravioto, 2012a; Olvera-Cravioto, 2012b].

Proof of Theorem 19. We consider the case when $\mathcal{N}$ is regularly varying first. By Theorem 3.4 in [Olvera-Cravioto, 2012b] and the remarks that follow it (see also Theorem 4.1 in [Volkovich and Litvak, 2010]),

$$P(\mathcal{R} > x) \sim \frac{(E[\mathcal{Q}]E[\mathcal{C}_1])^\alpha}{(1-\rho)^\alpha(1-\rho_\alpha)} P(\mathcal{N} > x), \qquad x \to \infty,$$

and therefore, $P(\mathcal{R} > x) \in \mathscr{R}_{-\alpha}$. Next, since the $\{\mathcal{C}_i\}$ are i.i.d. and independent of $\mathcal{N}$, Minkowski's inequality gives for any $\beta \geq 1$,

$$E\left[\left(\sum_{i=1}^{\mathcal{N}} \mathcal{C}_i\right)^\beta\right] = E\left[E\left[\left(\sum_{i=1}^{\mathcal{N}} \mathcal{C}_i\right)^\beta \middle| \mathcal{N}\right]\right] \leq E\left[\mathcal{N}^\beta E[\mathcal{C}_1^\beta]\right]. \tag{B.5}$$

Applying Lemma 2.3 in [Olvera-Cravioto, 2012b] with $\beta = 1+\delta$ gives that $E[|\mathcal{R}|^{1+\delta}] < \infty$ for all $0 < \delta < \alpha - 1$. By conditioning on the filtration $\mathcal{F}_k = \sigma\left((\mathcal{N}_{\mathbf{i}}, \mathcal{C}_{(\mathbf{i},1)}, \mathcal{C}_{(\mathbf{i},2)}, \dots) : \mathbf{i} \in \mathcal{A}_s, s < k\right)$ it can be shown that $E\left[\sum_{\mathbf{i}\in\mathcal{A}_k} \Pi_{\mathbf{i}} \mathcal{Q}_{\mathbf{i}}\right] = \rho^k E[\mathcal{Q}]$, which implies that $E[\mathcal{R}] = (1-\rho)^{-1}E[\mathcal{Q}] > 0$. Also, by Lemma 3.7(2) in [Jessen and Mikosch, 2006] we have

$$P\left(\sum_{i=1}^{\mathcal{N}_0} \mathcal{C}_i > x\right) \sim (E[\mathcal{C}_1])^\alpha P(\mathcal{N}_0 > x) \sim \kappa \frac{(1-\rho)^\alpha(1-\rho_\alpha)}{(E[\mathcal{Q}])^\alpha} P(\mathcal{R} > x).$$

Using Theorem A.1 in [Olvera-Cravioto, 2012b] we conclude that

$$P(\mathcal{R}^* > x) \sim \left(E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa \frac{(1-\rho)^\alpha(1-\rho_\alpha)}{(E[\mathcal{Q}])^\alpha} (E[\mathcal{R}])^\alpha \right) P(\mathcal{R} > x)$$
$$\sim (E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa(1-\rho_\alpha)) \frac{(E[\mathcal{Q}]E[\mathcal{C}_1])^\alpha}{(1-\rho)^\alpha(1-\rho_\alpha)} P(\mathcal{N} > x)$$

as $x \to \infty$.

Now, for the case when $\mathcal{Q}$ is regularly varying, note that $E\left[\left(\sum_{i=1}^{\mathcal{N}} \mathcal{C}_i\right)^{\alpha+\epsilon}\right] < \infty$ by (B.5) and the theorem's assumptions. Then, by Theorem 4.4 in [Olvera-Cravioto, 2012b] (see also Theorem 4.1 in [Volkovich and Litvak, 2010]) we have

$$P(\mathcal{R} > x) \sim (1-\rho_\alpha)^{-1} P(\mathcal{Q} > x), \qquad x \to \infty.$$

The same observations made for the previous case give $E[|\mathcal{R}|^{1+\delta}] < \infty$ for all $0 < \delta < \alpha - 1$. In addition, note that the same argument used above gives $E\left[\left(\sum_{i=1}^{\mathcal{N}_0} \mathcal{C}_i\right)^{\alpha+\epsilon}\right] < \infty$. Also,

$$P(\mathcal{Q}_0 > x) \sim \kappa P(\mathcal{Q} > x) \sim \kappa(1-\rho_\alpha) P(\mathcal{R} > x).$$

It follows, by Theorem A.2 in [Olvera-Cravioto, 2012b], that

$$P(\mathcal{R}^* > x) \sim (E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa(1-\rho_\alpha)) P(\mathcal{R} > x)$$
$$\sim (E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa(1-\rho_\alpha)) (1-\rho_\alpha)^{-1} P(\mathcal{Q} > x)$$

as $x \to \infty$. □

B.4 Proofs of properties of the *IID Algorithm*

Before giving the proofs of Propositions 20 and 21 we will need some general results for sequences of i.i.d. random variables, which may be of independent interest. The first result establishes a bound for the sum of the largest order statistics in a sample. The second result is essentially an explicit version of the Weak Law of Large Numbers.

Lemma 30. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. nonnegative random variables satisfying $E[X_1^{1+\kappa}] < \infty$ for some $\kappa > 0$, and let $X_{(i)}$ denote the ith smallest observation from the set $\{X_1, X_2, \ldots, X_n\}$. Let $\{\pi_1, \pi_2, \ldots, \pi_n\}$ be any permutation of the set $\{1, 2, \ldots, n\}$. Then, for any $k_n \in \{1, 2, 3, 4, \ldots, n\}$ we have*

$$P\left(\sum_{i=n-k_n+1}^{n} X_{(i)} > n^{1-\gamma} \right) = O\left(k_n^{\kappa/(1+\kappa)} n^{-(\kappa/(1+\kappa)-\gamma)} \right)$$

as $n \to \infty$.

Proof. Note that, by Markov's inequality,

$$P(X_1 > x) \leq E[X_1^{1+\kappa}]x^{-1-\kappa},$$

and therefore,

$$P(X_i > x) \leq P(Y_i > x),$$

where $\{Y_1, Y_2, \ldots, Y_n\}$ are i.i.d. Pareto random variables having distribution $G(x) = 1 - (x/b)^{-1-\kappa}$ for $x > b := \left(E[X_1^{1+\kappa}]\right)^{-1/(1+\kappa)}$. We then have that

$$\begin{aligned} P\left(\sum_{i=n-k_n+1}^{n} X_{(i)} > n^{1-\gamma}\right) &\leq P\left(\sum_{i=n-k_n+1}^{n} Y_{(i)} > n^{1-\gamma}\right) \\ &\leq \frac{1}{n^{1-\gamma}} \sum_{i=n-k_n+1}^{n} E[Y_{(i)}], \end{aligned}$$

where $Y_{(i)}$ is the ith smallest from the set $\{Y_1, Y_2, \ldots, Y_n\}$. Moreover, it is known (see [Vännman, 1976], for example) that

$$E[Y_{(i)}] = b \cdot \frac{n!}{(n-i)!} \cdot \frac{\Gamma(n-i+1-(1+\kappa)^{-1})}{\Gamma(n+1-(1+\kappa)^{-1})},$$

where $\Gamma(\cdot)$ is the Gamma function. By Wendel's inequality [Wendel, 1948], for any $0 < s < 1$ and $x > 0$,

$$\left(\frac{x}{x+s}\right)^{1-s} \leq \frac{\Gamma(x+s)}{x^s\Gamma(x)} \leq 1,$$

and therefore, for $i < n$, and $\vartheta = (1+\kappa)^{-1}$,

$$E[Y_{(i)}] \leq b \cdot \frac{n!}{\Gamma(n+1-\vartheta)} \cdot \frac{1}{(n-i)^{\vartheta}} \leq b\left(\frac{n+1-\vartheta}{n-i}\right)^{\vartheta}.$$

We conclude that

$$\begin{aligned}
\frac{1}{n^{1-\gamma}} \sum_{i=n-k_n+1}^{n} E[Y_{(i)}] &\leq \frac{b}{n^{1-\gamma}} \left(\sum_{i=n-k_n+1}^{n-1} \left(\frac{n+1-\vartheta}{n-i} \right)^{\vartheta} + \frac{n!\Gamma(1-\vartheta)}{\Gamma(n+1-\vartheta)} \right) \\
&\leq \frac{b(n+1-\vartheta)^{\vartheta}}{n^{1-\gamma}} \left(\sum_{i=n-k_n+1}^{n-1} \left(\frac{1}{n-i} \right)^{\vartheta} + \Gamma(1-\vartheta) \right) \\
&\leq \frac{b(n+1)^{\vartheta}}{n^{1-\gamma}} \left(\sum_{j=1}^{k_n-1} \int_{j-1}^{j} \frac{1}{t^{\vartheta}}\, dt + \Gamma(1-\vartheta) \right) \\
&= \frac{b(n+1)^{\vartheta}}{n^{1-\gamma}} \left(\frac{(k_n-1)^{1-\vartheta}}{1-\vartheta} + \Gamma(1-\vartheta) \right) \\
&= O\left(\frac{k_n^{1-\vartheta}}{n^{1-\vartheta-\gamma}} \right),
\end{aligned}$$

where in the second inequality we used Wendel's inequality. This completes the proof. □

Lemma 31. *Let $\{X_1, X_2, \ldots, X_n\}$ be i.i.d. random variables satisfying $E[|X_1|^{1+\kappa}] < \infty$ for some $\kappa > 0$ and $\mu = E[X_1]$. Set $S_m = X_1 + \cdots + X_m$ and $\theta = \min\{1+\kappa, 2\}$. Then, for any $K > 0$, any nonnegative sequence $\{x_n\}$ such that $x_n \to \infty$ as $n \to \infty$, and all $m = o\left(x_n^{1+\kappa}\right)$, there exists an $n_0 \geq 1$ such that for all $n \geq n_0$,*

$$P\left(|S_m - m\mu| > Kx_n\right) \leq E[|X_1|^{\theta}] \left(\frac{2}{K^2} + 1 \right) \frac{m}{x_n^{\theta}}.$$

Proof. If $\kappa \geq 1$, then Chebyshev's inequality gives, for all $m \geq 1$,

$$P\left(|S_m - m\mu| > Kx_n\right) \leq \frac{m\text{Var}\,(X_1)}{K^2 x_n^2} \leq \frac{mE[|X_1|^2]}{K^2 x_n^2} = \frac{mE[|X_1|^{\theta}]}{K^2 x_n^{\theta}}.$$

Suppose now that $0 < \kappa < 1$ and let $G(t) = P(|X_1| \leq t)$. Set $t = x_n$ and define $P(\tilde{X}_i \leq x) = P(X_i \leq x | X_i \leq t)$, and note that

$$\begin{aligned}
\left| E[\tilde{X}_1] - \mu \right| &= |E[X_1 1(|X_1| \leq t)]/G(t) - \mu| \\
&\leq \frac{1}{G(t)} |E[X_1 1(|X_1| \leq t)] - \mu| + \frac{|\mu|\overline{G}(t)}{G(t)} \\
&= \frac{1}{G(t)} \left(|E[X_1 1(|X_1| > t)]| + |\mu|\overline{G}(t) \right) \\
&\leq \frac{1}{G(t)} \left(t\overline{G}(t) + \int_t^{\infty} \overline{G}(x) dx + |\mu|\overline{G}(t) \right) \\
&\leq \frac{E[|X_1|^{1+\kappa}]}{G(t)} \left(t^{-\kappa} + \int_t^{\infty} x^{-1-\kappa}\, dx + |\mu| t^{-1-\kappa} \right) \qquad \text{(by Markov's inequality)} \\
&= \frac{E[|X_1|^{1+\kappa}]}{G(t)} \left(\frac{1+\kappa}{\kappa} + |\mu| t^{-1} \right) t^{-\kappa}.
\end{aligned}$$

Then, for sufficiently large n, we obtain that

$$\left|E[\tilde{X}_1] - \mu\right| \leq 2E[|X_1|^{1+\kappa}] \left(\frac{1+\kappa}{\kappa} + |\mu|\right) t^{-\kappa} \triangleq K' t^{-\kappa} = K' x_n^{-\kappa}.$$

It follows that for sufficiently large n and $m = o(x_n^{1+\kappa})$,

$$\begin{aligned}
&P\left(|S_m - m\mu| > Kx_n\right) \\
&= P\left(\left|\sum_{i=1}^{m}(\tilde{X}_i - \mu)\right| > Kx_n\right) G(t)^m + P\left(\left|\sum_{i=1}^{m}(X_i - \mu)\right| > Kx_n, \max_{1\leq i\leq m}|X_i| > t\right) \\
&\leq P\left(\left|\sum_{i=1}^{m}(\tilde{X}_i - E[\tilde{X}_1])\right| + m\left|E[\tilde{X}_1] - \mu\right| > Kx_n\right) G(t)^m + P\left(\max_{1\leq i\leq m}|X_i| > t\right) \\
&\leq \frac{G(t)^m}{\left(Kx_n - K'mt^{-\kappa}\right)^2} \cdot m\text{Var}\left(\tilde{X}_1\right) + 1 - G(t)^m \qquad \text{(by Chebyshev's inequality)} \\
&\leq \frac{G(t)^m m\text{Var}\left(\tilde{X}_1\right)}{K^2 x_n^2 (1 - mx_n^{-1-\kappa}K'/K)^2} + m\overline{G}(t).
\end{aligned}$$

To estimate Var $(\tilde{X}_1)$ note that

$$\text{Var}\left(\tilde{X}_1\right) \leq E[\tilde{X}_1^2] = \frac{E[X_1^2 1(|X_1| \leq t)]}{G(t)} \leq \frac{E[|X_1|^{1+\kappa}]t^{1-\kappa}}{G(t)},$$

so using Markov's inequality again to estimate $\overline{G}(t)$ gives us

$$\begin{aligned}
P\left(|S_m - m\mu| > Kx_n\right) &\leq \frac{E[|X_1|^{1+\kappa}]}{K^2(1 - mx_n^{-1-\kappa}K'/K)^2} \cdot \frac{mt^{1-\kappa}}{x_n^2} + \frac{E[|X_1|^{1+\kappa}]m}{t^{1+\kappa}} \\
&= E[|X_1|^{1+\kappa}] \left(\frac{1}{K^2(1 - mx_n^{-1-\kappa}K'/K)^2} + 1\right) \frac{m}{x_n^{1+\kappa}} \\
&= E[|X_1|^{\theta}] \left(\frac{1}{K^2(1 - mx_n^{-1-\kappa}K'/K)^2} + 1\right) \frac{m}{x_n^{\theta}}.
\end{aligned}$$

This completes the proof. □

By setting $m = n$ and $x_n = n^{1-\gamma}$ we immediately obtain the following corollary.

Corollary 32. *Let $\{X_1, X_2, \ldots, X_n\}$ be i.i.d. random variables satisfying $E[|X_1|^{1+\kappa}] < \infty$ for some $\kappa > 0$ and $\mu = E[X_1]$. Set $S_n = X_1 + \cdots + X_n$. Then, for any $0 \leq \gamma < 1 - 1/\theta$, $\theta = \min\{1+\kappa, 2\}$ and any constant $K > 0$, there exists an $n_0 \geq 1$ such that for all $n \geq n_0$*

$$P\left(|S_n - n\mu| > Kn^{1-\gamma}\right) \leq E[|X_1|^{\theta}] \left(\frac{2}{K^2} + 1\right) n^{-\theta(1-1/\theta-\gamma)}.$$

We now proceed to prove that the extended bi-degree sequence generated by the IID Algorithm satisfies Assumptions 1 and 2.

Proof of Proposition 20. It suffices to show that $P\left(\Omega_{n,i}^c\right) = O(n^{-\varepsilon})$ for some $\varepsilon > 0$ and $i = 1, \ldots, 6$. Throughout the proof let $E_n = \{|\Delta_n| \leq n^{1-\kappa_0+\delta_0}\}$ and recall that by (3.21) $P(E_n^c) = O\left(n^{-\delta_0\eta}\right)$, where $\eta = (\kappa_0 - \delta_0)/(1-\kappa_0)$.

We start with $\Omega_{n,2}$. Let $\nu_2 = (E[\mathscr{D}])^2$ and define $\chi_i = D_i - \mathscr{D}_i$, $\tau_i = N_i - \mathscr{N}_i$. Note that $\chi_i, \tau_i \in \{0,1\}$ for all $i = 1, \ldots, n$; moreover, either all the $\{\chi_i\}$ or all the $\{\tau_i\}$ are zero, and therefore $\chi_i\tau_j = 0$ for all $1 \leq i, j \leq n$. We now have

$$\begin{aligned}\left|\sum_{i=1}^n D_iN_i - n\nu_2\right| &= \left|\sum_{i=1}^n \mathscr{D}_i\mathscr{N}_i - n\nu_2 + \sum_{i=1}^n(\mathscr{D}_i\tau_i + \chi_i\mathscr{N}_i)\right| \\ &\leq \left|\sum_{i=1}^n \mathscr{D}_i\mathscr{N}_i - n\nu_2\right| + \max\left\{\sum_{i=n-\Delta_n+1}^n \mathscr{D}_{(i)}, \sum_{i=n-\Delta_n+1}^n \mathscr{N}_{(i)}\right\},\end{aligned}$$

where $\mathscr{D}_{(i)}$ (respectively, $\mathscr{N}_{(i)}$) is the ith smallest value from the set $\{\mathscr{D}_1, \ldots, \mathscr{D}_n\}$ (respectively, $\{\mathscr{N}_1, \ldots, \mathscr{N}_n\}$). Since $|\Delta_n| \leq n^{1-\kappa_0+\delta_0}$ on E_n, we have

$$\begin{aligned}P(\Omega_{n,2}^c) &= P\left(\left|\sum_{i=1}^n D_iN_i - n\nu_2\right| > n^{1-\gamma}\,\middle|\, E_n\right) \\ &\leq \frac{1}{P(E_n)}\left\{P\left(\left|\sum_{i=1}^n \mathscr{D}_i\mathscr{N}_i - n\nu_2\right| > \frac{n^{1-\gamma}}{2}\right)\right. \\ &\quad \left. + P\left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)}\rfloor+1}^n \mathscr{D}_{(i)} > \frac{n^{1-\gamma}}{2}\right) + P\left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)}\rfloor+1}^n \mathscr{N}_{(i)} > \frac{n^{1-\gamma}}{2}\right)\right\}.\end{aligned}$$

Now apply Corollary 32 to $X_i = \mathscr{D}_i\mathscr{N}_i$, which satisfies $E[(\mathscr{D}_1\mathscr{N}_1)^{1+\eta}] = E[\mathscr{N}_1^{1+\eta}]E[\mathscr{D}_1^{1+\eta}] < \infty$, to obtain

$$P\left(\left|\sum_{i=1}^n \mathscr{D}_i\mathscr{N}_i - n\nu_2\right| > \frac{n^{1-\gamma}}{2}\right) = O\left(n^{-\eta+(1+\eta)\gamma}\right).$$

For the remaining two probabilities use Lemma 30 to see that

$$\begin{aligned}&P\left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)}\rfloor+1}^n \mathscr{D}_{(i)} > \frac{n^{1-\gamma}}{2}\right) + P\left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)}\rfloor+1}^n \mathscr{N}_{(i)} > \frac{n^{1-\gamma}}{2}\right) \\ &= O\left(n^{(1-\eta(1-\kappa_0))\eta/(1+\eta)-(\eta/(1+\eta)-\gamma)}\right) \\ &= O\left(n^{-\eta(\kappa_0-\delta_0)/(1+\eta)+\gamma}\right).\end{aligned}$$

It follows from these estimates that

$$P(\Omega_{n,2}^c) = O\left(n^{-\eta(\kappa_0-\delta_0)/(1+\eta)+\gamma}\right). \tag{B.6}$$

Next, we can analyze $\Omega_{n,1}, \Omega_{n,3}$ and $\Omega_{n,4}$ by considering the sequence $\{D_i^\vartheta\}$ where ϑ can be taken to be $1, 2$ or $2+\kappa$. Correspondingly, we have $\nu_1 = E[\mathscr{D}]$, $\nu_3 = E[\mathscr{D}^2]$ and $\nu_4 = E[\mathscr{D}^{2+\kappa}]$. Similarly as what was done for $\Omega_{n,2}$, note that

$$\begin{aligned}\left|\sum_{i=1}^n D_i^\vartheta - nE[\mathscr{D}^\vartheta]\right| &\le \left|\sum_{i=1}^n \mathscr{D}_i^\vartheta - nE[\mathscr{D}^\vartheta]\right| + \sum_{i=1}^n \left((\mathscr{D}_i+\chi_i)^\vartheta - \mathscr{D}_i^\vartheta\right)\\ &\le \left|\sum_{i=1}^n \mathscr{D}_i^\vartheta - nE[\mathscr{D}^\vartheta]\right| + \sum_{i=1}^n \vartheta(\mathscr{D}_i+1)^{\vartheta-1}\chi_i,\end{aligned}$$

where we used the inequality $(d+x)^\vartheta - d^\vartheta \le \vartheta(d+1)^{\vartheta-1}x$ for $d \ge 0$, $x \in [0,1]$ and $\vartheta \ge 1$. Now note that $E[(\mathscr{D}^\vartheta)^{1+\sigma}] < \infty$ for any $0 < \sigma < (\beta-2-\kappa)/(2+\kappa)$; in particular, since $\gamma < (\beta-2-\kappa)/\beta$, we can choose $\gamma/(1-\gamma) < \sigma < (\beta-2-\kappa)/(2+\kappa)$. For such σ, Corollary 32 gives

$$P\left(\left|\sum_{i=1}^n \mathscr{D}_i^\vartheta - nE[\mathscr{D}^\vartheta]\right| > \frac{n^{1-\gamma}}{2}\right) = O\left(n^{-\sigma+(1+\sigma)\gamma}\right).$$

For the term involving the $\{\chi_i\}$ we use again Lemma 30 to obtain

$$\begin{aligned}P\left(\sum_{i=1}^n \vartheta(\mathscr{D}_i+1)^{\vartheta-1}\chi_i > \frac{n^{1-\gamma}}{2}\right) &\le P\left(\sum_{i=n-\lceil n^{1-\eta}\rceil+1}^n \vartheta(\mathscr{D}_{(i)}+1)^{\vartheta-1} > \frac{n^{1-\gamma}}{2}\right)\\ &= O\left(n^{(1-\eta)(1-1/2)-(1-\gamma-1/2)}\right)\\ &= O\left(n^{-\eta/2+\gamma}\right).\end{aligned}$$

It follows that

$$P(\Omega_{n,i}^c) \le \frac{1}{P(E_n)} \cdot O\left(n^{-\sigma+(1+\sigma)\gamma} + n^{-\eta/2+\gamma}\right), \qquad i = 1,3,4. \tag{B.7}$$

Now note that since $|\zeta| \le c < 1$ a.s., then $E[|\zeta|^2] < \infty$ and Corollary 32 gives

$$\begin{aligned}P(\Omega_{n,5}^c) &= P\left(\left|\sum_{r=1}^n |\zeta_r| 1(D_r \ge 1) - n\nu_5\right| > n^{1-\gamma}\right)\\ &= P\left(\left|\sum_{r=1}^n |\zeta_r| 1(\mathscr{D}_r \ge 1) - n\nu_5\right| + c|\Delta_n| > n^{1-\gamma}\right) = O\left(n^{-1+2\gamma}\right).\end{aligned} \tag{B.8}$$

Finally, by Corollary 32 and (3.21),

$$P(\Omega_{n,6}^c) \leq P\left(\left|\sum_{r=1}^{n}|Q_r| - nE[|Q|]\right| > n\middle| E_n\right) = O\left(n^{-\epsilon_Q} + n^{-\delta_0\eta}\right). \tag{B.9}$$

Our choice of $0 < \gamma < \min\{\eta(\kappa_0-\delta_0)(1+\eta),\, \sigma/(1+\sigma)\}$ guarantees that all the exponents of n in expressions (B.6) - (B.8) are strictly negative, which completes the proof. □

Proof of Proposition 21. We will show that $d_1(F_n^*, F^*)$ and $d_1(F_n, F)$ converge to zero a.s. by using the duality formula for the Kantorovich-Rubinstein distance. To this end, let $S_n = \sum_{i=1}^n \mathscr{D}_i$, $\mathscr{C}_k = \zeta_k/\mathscr{D}_k 1(\mathscr{D}_k \geq 1) + c\,\mathrm{sgn}(\zeta_k)1(\mathscr{D}_k = 0)$, and fix $\psi^* : \mathbb{R}^2 \to \mathbb{R}$ and $\psi : \mathbb{R}^3 \to \mathbb{R}$ to be Lipschitz continuous functions with Lipschitz constant one. Then,

$$\begin{aligned}
\mathcal{E}_0 &:= \left|\frac{1}{n}\sum_{k=1}^{n}\psi^*(N_k, Q_k) - \frac{1}{n}\sum_{k=1}^{n}\psi^*(\mathscr{N}_k, Q_k)\right| \\
&\leq \frac{1}{n}\sum_{k=1}^{n}\left|\psi^*(\mathscr{N}_k+1, Q_k) - \psi^*(\mathscr{N}_k, Q_k)\right| 1(N_k = \mathscr{N}_k + 1) \\
&\leq \frac{1}{n}\sum_{k=1}^{n} 1(N_k = \mathscr{N}_k + 1) \leq \frac{|\Delta_n|}{n},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{E}_1 &:= \left|\sum_{k=1}^{n}\psi(N_k, Q_k, C_k)\frac{D_k}{L_n} - \sum_{k=1}^{n}\psi(\mathscr{N}_k, Q_k, \mathscr{C}_k)\frac{\mathscr{D}_k}{S_n}\right| \\
&\leq \sum_{k=1}^{n}\frac{\mathscr{D}_k}{S_n}\left|\psi(N_k, Q_k, \mathscr{C}_k) - \psi(\mathscr{N}_k, Q_k, \mathscr{C}_k)\right| 1(\Delta_n \leq 0) \\
&\quad + \sum_{k=1}^{n}\frac{D_k}{L_n}\left|\psi(\mathscr{N}_k, Q_k, C_k) - \psi(\mathscr{N}_k, Q_k, \mathscr{C}_k)\right| 1(\Delta_n > 0) \\
&\quad + \sum_{k=1}^{n}\left|\psi(\mathscr{N}_k, Q_k, \zeta_k/\mathscr{D}_k)\left(\frac{D_k}{L_n} - \frac{\mathscr{D}_k}{S_n}\right)\right| 1(\Delta_n > 0) \\
&\leq \sum_{k=1}^{n}\frac{\mathscr{D}_k}{S_n}1(N_k = \mathscr{N}_k + 1) + \sum_{k=1}^{n}\frac{D_k}{L_n}\left|\zeta_k/(\mathscr{D}_k+1) - \mathscr{C}_k\right| 1(D_k = \mathscr{D}_k + 1) \\
&\quad + \sum_{k=1}^{n}\left|\psi(\mathscr{N}_k, Q_k, \mathscr{C}_k)\right|\left|\frac{(D_k - \mathscr{D}_k)S_n - \mathscr{D}_k\Delta_n}{L_n S_n}\right| 1(\Delta_n > 0),
\end{aligned}$$

where we used the fact that ψ^* and ψ have Lipschitz constant one. To bound further $\mathcal{E}_1$ use the Cauchy-Schwarz inequality to obtain

$$\sum_{k=1}^{n}\frac{\mathscr{D}_k}{S_n}1(N_k = \mathscr{N}_k + 1) \leq \frac{n}{S_n}\left(\frac{1}{n}\sum_{k=1}^{n}\mathscr{D}_k^2\right)^{1/2}\left(\frac{|\Delta_n|}{n}\right)^{1/2}.$$

Now, use the observation that $|\zeta_k| \leq c$ to obtain

$$\begin{aligned}
&\sum_{k=1}^{n} \frac{D_k}{L_n} \left|\zeta_k/(\mathscr{D}_k+1) - \mathscr{C}_k\right| 1(D_k = \mathscr{D}_k + 1) \\
&\leq c \sum_{k=1}^{n} \frac{1}{L_n \mathscr{D}_k} 1(D_k = \mathscr{D}_k + 1, \mathscr{D}_k \geq 1) + \sum_{k=1}^{n} \frac{1}{L_n} \left|\zeta_k - c\,\mathrm{sgn}(\zeta_k)\right| 1(D_k = \mathscr{D}_k + 1, \mathscr{D}_k = 0) \\
&\leq \frac{c}{L_n} \sum_{k=1}^{n} 1(D_k = \mathscr{D}_k + 1) \leq \frac{c|\Delta_n|}{S_n}.
\end{aligned}$$

Next, use the bound $|\psi(m,q,x)| \leq ||(m,q,x)||_1 + |\psi(0,0,0)|$ and Hölder's inequality to obtain

$$\begin{aligned}
&\sum_{k=1}^{n} |\psi(\mathscr{N}_k, Q_k, \mathscr{C}_k)| \left|\frac{(D_k - \mathscr{D}_k)S_n - \mathscr{D}_k \Delta_n}{L_n S_n}\right| 1(\Delta_n > 0) \\
&\leq \sum_{k=1}^{n} |\psi(\mathscr{N}_k, Q_k, \mathscr{C}_k)| \frac{1(D_k = \mathscr{D}_k + 1)}{S_n} + \sum_{k=1}^{n} |\psi(\mathscr{N}_k, Q_k, \mathscr{C}_k)| \frac{\mathscr{D}_k |\Delta_n|}{S_n^2} \\
&\leq \frac{1}{S_n} \sum_{k=1}^{n} ||(\mathscr{N}_k, Q_k, c)||_1 \, 1(D_k = \mathscr{D}_k + 1) + \frac{|\Delta_n|}{S_n^2} \sum_{k=1}^{n} (\mathscr{N}_k \mathscr{D}_k + |Q_k| \mathscr{D}_k + c) + \frac{2|\psi(0,0,0)\Delta_n|}{S_n} \\
&\leq \frac{n}{S_n} \left\{ \left(\frac{1}{n} \sum_{k=1}^{n} \mathscr{N}_k^{1+\delta} \right)^{1/(1+\delta)} + \left(\frac{1}{n} \sum_{k=1}^{n} |Q_k|^{1+\delta} \right)^{1/(1+\delta)} \right\} \left(\frac{|\Delta_n|}{n} \right)^{\delta/(1+\delta)} \\
&\quad + \frac{|\Delta_n|}{S_n^2} \sum_{k=1}^{n} (\mathscr{N}_k \mathscr{D}_k + |Q_k| \mathscr{D}_k) + \frac{H|\Delta_n|}{S_n},
\end{aligned}$$

where $0 < \delta < \min\{\alpha - 1, \epsilon_Q\}$ and $H = 2|\psi(0,0,0)| + 2c$. Now note that since the bi-degree sequence is constructed on the event $|\Delta_n| \leq n^{1-\kappa_0+\delta_0}$, we have that $\mathcal{E}_0 \leq n^{-\kappa_0+\delta_0}$ a.s. To show that $\mathcal{E}_1$ converges to zero a.s. use the Strong Law of Large Numbers (SLLN) (recall that $E[\mathscr{D}^2] < \infty$ and that $\mathscr{N}, \mathscr{D}, Q$ are mutually independent) and the bounds derived above.

Finally, by the SLLN again and the fact that $E[||(\mathscr{N}, Q, \mathscr{C})||_1] < \infty$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \psi^*(N_k, Q_k) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \psi^*(\mathscr{N}_k, Q_k) = E[\psi^*(\mathscr{N}, Q)] \quad \text{a.s.}$$

and

$$\lim_{n\to\infty} \sum_{i=1}^{n} \psi(N_k, Q_k, C_k) \frac{\mathscr{D}_i}{S_n} = \lim_{n\to\infty} \sum_{k=1}^{n} \psi(\mathscr{N}_k, Q_k, \mathscr{C}_k) \frac{\mathscr{D}_k}{S_n} = \frac{1}{\mu} E[\psi(\mathscr{N}, Q, \mathscr{C}) \mathscr{D}] \quad \text{a.s.}$$

The first limit combined with the duality formula gives that $d_1(F_n^*, F^*) \to 0$ a.s. For the

second limit we still need to identify the limiting distribution, for which we note that

$$\frac{1}{\mu}E[\psi(\mathscr{N},Q,\mathscr{C})\mathscr{D}] = \frac{1}{\mu}E\left[E[\psi(\mathscr{N},Q,\mathscr{C})\mathscr{D}|\mathscr{N},Q]\right] = \frac{1}{\mu}E\left[\sum_{i=1}^{\infty}\int_{-\infty}^{\infty}\psi(\mathscr{N},Q,z/i)i\,dF^{\zeta}(z)P(\mathscr{D}=i)\right]$$
$$= \frac{1}{\mu}E\left[\sum_{i=1}^{\infty}\int_{-\infty}^{\infty}\psi(\mathscr{N},Q,y)i\,dF^{\zeta}(yi)P(\mathscr{D}=i)\right] =: E\left[\psi(\mathscr{N},Q,Y)\right],$$

where Y has distribution function

$$P(Y\le x) = \frac{1}{\mu}E\left[\sum_{i=1}^{\infty}\int_{-\infty}^{\infty}1(y\le x)i\,dF^{\zeta}(yi)P(\mathscr{D}=i)\right] = \frac{1}{\mu}E\left[\sum_{i=1}^{\infty}iF^{\zeta}(ix)P(\mathscr{D}=i)\right]$$
$$= \frac{1}{\mu}E[\mathscr{D}F^{\zeta}(\mathscr{D}x)] = \frac{1}{\mu}E[\mathscr{D}1(\zeta/\mathscr{D}\le x)] = P(\mathcal{C}\le x).$$

It follows that $E[\psi(\mathscr{N},Q,\mathscr{C})\mathscr{D}]/\mu = E[\psi(\mathscr{N},Q,\mathcal{C})]$, which combined with the duality formula gives that $d_1(F_n,F)\to 0$ a.s. □