

Ranking algorithms on directed configuration networks

Ningyuan Chen
Columbia University

Nelly Litvak
University of Twente

Mariana Olvera-Cravioto
Columbia University

October 14, 2014

Abstract

This paper studies the distribution of a family of rankings, which includes Google's PageRank, on a directed configuration model. In particular, it is shown that the distribution of the rank of a randomly chosen node in the graph converges in distribution to a finite random variable \mathcal{R}^* that can be written as a linear combination of i.i.d. copies of the endogenous solution to a stochastic fixed point equation of the form

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

where $(\mathcal{Q}, \mathcal{N}, \{\mathcal{C}_i\})$ is a real-valued vector with $\mathcal{N} \in \{0, 1, 2, \dots\}$, $P(|\mathcal{Q}| > 0) > 0$, and the $\{\mathcal{R}_i\}$ are i.i.d. copies of \mathcal{R} , independent of $(\mathcal{Q}, \mathcal{N}, \{\mathcal{C}_i\})$. Moreover, we provide precise asymptotics for the limit \mathcal{R}^* , which when the in-degree distribution in the directed configuration model has a power law imply a power law distribution for \mathcal{R}^* with the same exponent.

Keywords: PageRank, ranking algorithms, directed configuration model, complex networks, stochastic fixed-point equations, weighted branching processes, power laws.

2000 MSC: Primary: 05C80, 60J80, 68P20. Secondary: 41A60, 37A30, 60B10.

1 Introduction

Ranking of nodes according to their centrality, or importance, in a complex network such as the Internet, the World Wide Web, and other social and biological networks, has been a hot research topic for several years in physics, mathematics, and computer science. For a comprehensive overview of the vast literature on rankings in networks we refer the reader to [27], and more recently to [7] for a thorough up-to-date mathematical classification of centrality measures.

In this paper we analyze a family of ranking algorithms which includes Google's PageRank, the algorithm proposed by Brin and Page [10], 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:

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

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 [20] or penalize spam pages [19].

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 . 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:

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

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

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

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

The basic definition (1.1) has many modifications and generalizations. The analysis in this paper 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 [15]. Numerous applications of PageRank and its modifications include graph clustering [5], spam detection [19], and citation analysis [13, 43].

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 $\alpha \in (1, 3)$, see e.g., [10, 29]. Thus, a lot of research has been devoted to the study of random graph models with highly skewed, or scale-free, degree distributions. By now, classical examples are the Chung-Lu model [14], the Preferential Attachment model [9], and the Configuration Model [35, 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 [6]. We refer to [35, 17, 29] for a more detailed discussion of random graph models for complex networks. In this paper we focus on the Directed Configuration Model as studied in [11]. Originally, an (undirected) Configuration Model is defined as a graph, randomly sampled from the set of graphs with a given degree sequence [8]. We emphasize that, to the best of our knowledge, [11] is the only paper that formally addresses the directed version of the Configuration Model and obtains its exact mathematical properties. We will provide more details in Section 3.

From the work of Pandurangan et al. [31], 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 [18]. 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 [13, 42, 38]. Exact mathematical evidence supporting the power law hypothesis is surprisingly scarce. As one of the few examples, [26] 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. [28, 41, 40] 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):

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

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 [40] 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 [40] 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 [22] using a sample-path approach.

The properties of equation (1.3) and the study of its multiple solutions has itself been an interesting topic in the recent literature [4, 22, 24, 23, 30, 2], and is related to the broader study of weighted branching processes (WBPs) [32, 33, 34]. The tail behavior of the endogenous solution, the one relevant to PageRank, was given in [22, 24, 23, 30]. In particular, in [22] it was discovered that when the C_i ’s are not bounded by one and there exists a positive root to the equation $E \left[\sum_{i=1}^N |C_i|^\alpha \right] = 1$ with $0 < E \left[\sum_{i=1}^N |C_i|^\alpha \log |C_i| \right] < \infty$, then R will have a power law tail with exponent α ; the main tool for this type of analysis is the implicit renewal theory on trees developed there and later extended in [24, 23] to study (1.3) in its full generality.

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. In this work we complete the argument when the underlying network is a Directed Configuration Model 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.

The essential theoretical contribution of this work is two-fold. First, we prove that the PageRank in the Directed Configuration Model is well approximated by the endogenous solution to a specific SFPE of the same type as (1.3). Second, we develop a methodology to analyze processes on graphs based on a coupling with a new type of stochastic process: a *weighted* branching process. Due to the presence of weights, couplings with weighted branching processes are more complex compared to traditional couplings with standard branching processes, and therefore, our approach may be of independent interest.

In Section 2 we describe our main results, outline the methodology, and provide an overview of the rest of the paper.

2 Overview of the paper

Although a rigorous presentation of the main result in the paper requires a significant amount of notation, we provide here a somewhat imprecise version that still captures the essence of our work. The paper is written according to the different steps needed in the proof of the main result, outlined in Section 2.2, and the precise statement can be found in Section 6.2.

2.1 An overview of the main result

Let $\mathcal{G}_n = (V_n, E_n)$ be a directed graph. We number the nodes $V_n = \{1, 2, \dots, n\}$ 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 i th 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 already mentioned, we will assume throughout the paper that \mathcal{G}_n is constructed according to the Directed Configuration Model (DCM). To briefly explain the construction of the DCM 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 (see Section 3 for more details). 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 5.1) 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 6.2) 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 [35]. 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 7 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 6.2. Let \mathcal{R} denote the endogenous solution to the following SFPE:

$$\mathcal{R} \stackrel{D}{=} \sum_{j=1}^{\mathcal{N}} \mathcal{C}_j \mathcal{R}_j + \mathcal{Q}, \quad (2.1)$$

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

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

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

$$\mathcal{R}^* := \sum_{j=1}^{\mathcal{N}_0} \mathcal{C}_j \mathcal{R}_j + \mathcal{Q}_0, \quad (2.2)$$

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 6. 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.

2.2 Methodology

As mentioned earlier, 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 by describing the DCM, which on its own does not require any assumptions on the bi-degree sequence. Then, in Section 4 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 4.2). 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 5). 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 5.4, 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 5.1) on the bi-degree sequence are important for this step.

3. *Convergence to a weighted branching process* (Section 6). Show that the rank of the root node of the TBT converges weakly to (2.2). This last step requires the weak convergence of the random distributions that define the TBT in the previous step (Assumption 6.2).

Finally, Section 7 gives an algorithm to construct an extended bi-degree sequence satisfying the two main assumptions. The technical proofs are postponed to Section 8.

3 The directed configuration model

The Configuration Model (CM) was originally defined as an undirected graph sampled uniformly at random from the collection of graphs with a given degree sequence [8]. In order to ensure a desired degree distribution, one may generate an i.i.d. degree sequence sampled from this distribution, see [35, Section 7.6]. In this case each node receives a random number of half-edges, or stubs, and then the stubs are paired uniformly at random. The resulting graph is, in general, a multi-graph, because two stubs of the same node may form an edge (self-loop), or a node may have two or more stubs connected to the same other node (multiple edges). There are two ways to create a simple graph. In the *repeated* CM, the pairing is repeated until a simple graph is obtained. This will occur with positive probability if the degrees have finite variance, see [35, Section 7.6]. In the *erased* CM self-loops and double-edges are removed. In the erased CM, the degree sequence is altered because of edge removal, but the distribution of the original degree sequence is preserved asymptotically under very general conditions, see again [35, Section 7.6]. A literature review and discussion of the undirected CM is provided in [35, Section 7.9].

While the undirected CM has been thoroughly studied, a formal analysis of the Directed Configuration Model (DCM) with given in- and out-degree distributions has only been recently presented by Chen and Olvera-Cravioto [11]. The crucial difference compared to the undirected case is that now we have a *bi-degree* sequence, i.e., a pair of sequences of nonnegative integers determining the in- and out-degrees of the nodes. Note that the sums of the in-degrees must be equal to that of the out-degrees for one to be able to draw a graph. The difficulty and originality of the DCM is that sums of i.i.d. in- and out-degrees will only be equal with a probability converging to zero as the size of the graph grows. To circumvent this problem, the algorithm given in [11], and included in Section 7 in this paper, forces the sums to match by adding the necessary half-edges in such a way that the degree distributions are essentially unchanged.

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 3.1 *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 3.2 Let $(\mathbf{N}_n, \mathbf{D}_n)$ be a bi-degree sequence and let $V_n = \{1, 2, \dots, n\}$ 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 $\{1, 2, \dots, L_n\}$, and let $\mathbf{x}_n = (x_1, x_2, \dots, x_{L_n})$ 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 5 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 [11] 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 paper, 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. [36] when analyzing distances in the undirected CM.

Throughout the paper, we will use $\mathcal{F}_n = \sigma((\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n))$ 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 $\mathbb{P}_n(\cdot) = P(\cdot | \mathcal{F}_n)$ and $\mathbb{E}_n[\cdot] = E[\cdot | \mathcal{F}_n]$ to denote the conditional probability and conditional expectation, respectively, given \mathcal{F}_n .

4 Spectral ranking algorithms

In this section we introduce the class of ranking algorithms that we analyze in this paper. Following the terminology from [7], 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 4.1 is not particular to the DCM.

4.1 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:

$$M_{i,j} = \begin{cases} s_{ij}C_i, & \text{if there are } s_{ij} \text{ edges from } i \text{ to } j, \\ 0, & \text{otherwise.} \end{cases} \quad (4.1)$$

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

$$\mathbf{R} = \mathbf{R}M + \mathbf{Q}. \quad (4.2)$$

Remark 4.1 *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.*

4.2 Finitely many iterations

To solve the system of equations given in (4.2) we proceed via matrix iterations [27]. 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 $k \geq 1$,

$$\mathbf{R}^{(n,k)} = \mathbf{r}_0M^k + \sum_{i=0}^{k-1} \mathbf{Q}M^i.$$

With this notation, we have that the solution \mathbf{R} to (4.2), 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, as described in Section 2.2, is to show that we can do so by using only finitely many matrix iterations. To this end note that

$$\mathbf{R}^{(n,k)} - \mathbf{R}^{(n,\infty)} = \mathbf{r}_0M^k - \sum_{i=k}^{\infty} \mathbf{Q}M^i = \left(\mathbf{r}_0 - \sum_{i=0}^{\infty} \mathbf{Q}M^i \right) M^k.$$

Moreover,

$$\left\| \mathbf{R}^{(n,k)} - \mathbf{R}^{(n,\infty)} \right\|_1 \leq \left\| \mathbf{r}_0M^k \right\|_1 + \sum_{i=0}^{\infty} \left\| \mathbf{Q}M^{k+i} \right\|_1.$$

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

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

where $A_{i\bullet}$ and $A_{\bullet j}$ are the i th row and j th 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

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

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

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

It follows that if we let

$$B_n = \left\{ \max_{1 \leq i \leq n} |C_i|D_i \leq c, \frac{1}{n} \sum_{i=1}^n |Q_i| \leq H \right\} \quad (4.3)$$

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

$$\begin{aligned}
& P\left(\left|R_1^{(n,k)} - R_1^{(n,\infty)}\right| > n^{-\epsilon} \mid B_n\right) \\
&= \frac{1}{P(B_n)} E\left[1(B_n) \mathbb{E}_n\left[1\left(\left|R_1^{(n,k)} - R_1^{(n,\infty)}\right| > n^{-\epsilon}\right)\right]\right] \\
&\leq \frac{1}{P(B_n)} E\left[1(B_n) n^\epsilon \mathbb{E}_n\left[\left|R_1^{(n,k)} - R_1^{(n,\infty)}\right|\right]\right] \\
&\leq \left(|r_0| + \frac{1}{1-c} E\left[\frac{1}{n} \sum_{i=1}^n |Q_i| \mid B_n\right]\right) n^\epsilon c^k \\
&\leq \left(|r_0| + \frac{H}{1-c}\right) n^\epsilon c^k. \tag{4.4}
\end{aligned}$$

We have thus derived the following result.

Proposition 4.2 *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 (4.3). Then, for any $x_n \rightarrow \infty$ and any $k \geq 1$, we have*

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

as $n \rightarrow \infty$.

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.

5 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 5.1 *Let $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ be an extended bi-degree sequence for which there exists constants $H, \nu_i > 0$, $i = 1, \dots, 5$, with*

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

$0 < \kappa \leq 1$, 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 $n \rightarrow \infty$,

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

It is clear from (4.3) that $\Omega_n \subseteq B_n$, hence Proposition 4.2 holds under Assumption 5.1. 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 7 an algorithm based on sequences of i.i.d. random variables that satisfies Assumption 5.1.

In Sections 5.1–5.4 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 5.1, followed by the construction itself in Section 5.2. Then, in Section 5.3 we present the Coupling Lemma 5.4, which is the main result of Section 5. Finally, Section 5.4 explains how to compute the rank of the root node in the coupled tree.

5.1 Terminology and notation

Throughout the remainder of the paper 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 1. We point out that the structure of the tree (i.e., parent-offspring relations) is solely determined by the number of offspring.

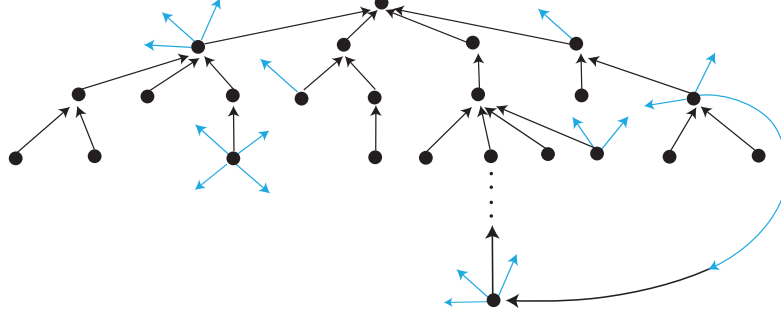


Figure 1: Graph construction process. Unpaired outbound links are in blue.

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 $\mathbf{i} = (i_1, i_2, \dots, i_k) \in \mathcal{U}$, 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 individual in the tree.

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

Definition 5.2 *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 r th 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:

$$\begin{aligned} f_n^*(i, j, s, t) &= \sum_{k=1}^n \mathbb{1}(N_k = i, D_k = j, C_k = s, Q_k = t) \mathbb{P}_n(\text{node } k \text{ is sampled}) \\ &= \frac{1}{n} \sum_{k=1}^n \mathbb{1}(N_k = i, D_k = j, C_k = s, Q_k = t). \end{aligned} \quad (5.1)$$

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:

$$\begin{aligned} f_n(i, j, s, t) &= \sum_{k=1}^n \mathbb{1}(N_k = i, D_k = j, C_k = s, Q_k = t) \mathbb{P}_n(\text{an outbound stub from node } k \text{ is sampled}) \\ &= \sum_{k=1}^n \mathbb{1}(N_k = i, D_k = j, C_k = s, Q_k = t) \frac{D_k}{L_n}. \end{aligned} \quad (5.2)$$

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.

5.2 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 [36] 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 (5.1).

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 (5.2). 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 (5.2), 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 i th node to be added to the graph.

In general, the k th 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 $(\hat{N}_i, \hat{D}_i, \hat{C}_i, \hat{Q}_i)$, $\mathbf{i} \in \mathcal{U}$. 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 (5.1). Next, for $k \geq 1$, each of the \hat{Z}_{k-1} individuals in the k th generation will independently have offspring, outbound stubs, weight and personalization value according to the joint distribution $f_n(i, j, s, t)$ given by (5.2).

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 j th node to be added to the graph, where:

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

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

Definition 5.3 *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 5.1, the coupling breaks only after a number of generations that is of order $\log n$, which combined with Proposition 4.2 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.

5.3 The coupling lemma

It follows from the construction in Section 5.2 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 4.2 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 4.2. Its proof is rather technical and is therefore postponed to Section 8.1.

Lemma 5.4 *Suppose $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ satisfies Assumption 5.1. 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

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

as $n \rightarrow \infty$.

Remark 5.5 *The constant μ was defined in Assumption 5.1, 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 5.4 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 $\delta > 0$,

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

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 (5.1) and (5.2), respectively).

This observation, combined with Proposition 4.2, 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

$$\begin{aligned} P\left(\left|R_1^{(n,\infty)} - \hat{R}^{(n,k_n)}\right| > n^{-\delta} \mid \Omega_n\right) &\leq P\left(\left|R_1^{(n,\infty)} - R_1^{(n,k_n)}\right| > n^{-\delta}/2 \mid \Omega_n\right) \\ &\quad + P\left(\left|R_1^{(n,k_n)} - \hat{R}^{(n,k_n)}\right| > n^{-\delta}/2 \mid \Omega_n\right) \\ &= O\left(n^\delta c^{k_n} + \varphi(k_n, n)\right) \\ &= O\left(n^{\delta - h|\log c|} + n^{-\varepsilon/2}\right). \end{aligned} \tag{5.3}$$

In view of (5.3), 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.

5.4 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 $\mathbf{i} = (i_1, \dots, i_k)$ we will use $(\mathbf{i}, j) = (i_1, \dots, i_k, j)$ 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

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

Note that the $\hat{\Pi}_\mathbf{i}$'s are the products of all the weights \hat{C}_j along the path leading to node \mathbf{i} , as depicted in Figure 2.

Next, for each fixed $k \in \mathbb{N}$ and each node \mathbf{i} in the TBT define $\hat{R}_\mathbf{i}^{(n,k)}$ 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,

$$\hat{R}_\mathbf{i}^{(n,k)} = \sum_{j=1}^{\hat{N}_\mathbf{i}} \hat{C}_{(\mathbf{i},j)} \hat{R}_{(\mathbf{i},j)}^{(n,k-1)} + \hat{Q}_\mathbf{i}, \quad k \geq 1, \quad \hat{R}_\mathbf{j}^{(n,0)} = r_0. \tag{5.4}$$

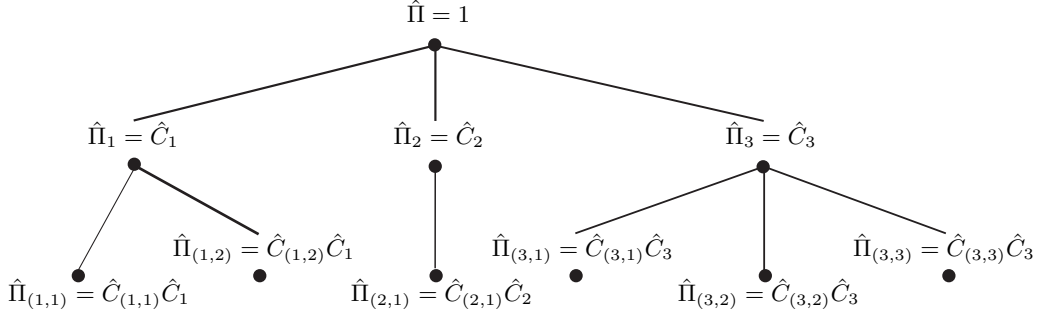


Figure 2: Weighted tree.

Iterating (5.4) gives

$$\begin{aligned}
\hat{R}^{(n,k)} &= \sum_{\mathbf{i} \in \hat{A}_1} \hat{\Pi}_{\mathbf{i}} \hat{R}_{\mathbf{i}}^{(n,k-1)} + \hat{Q}_{\emptyset} = \sum_{\mathbf{i} \in \hat{A}_1} \hat{\Pi}_{\mathbf{i}} \left(\sum_{j=1}^{\hat{N}_{\mathbf{i}}} \hat{C}_{(\mathbf{i},j)} \hat{R}_{(\mathbf{i},j)}^{(n,k-2)} + \hat{Q}_{\mathbf{i}} \right) + \hat{Q}_{\emptyset} \\
&= \sum_{\mathbf{i} \in \hat{A}_2} \hat{\Pi}_{\mathbf{i}} \hat{R}_{\mathbf{i}}^{(n,k-2)} + \sum_{\mathbf{i} \in \hat{A}_1} \hat{\Pi}_{\mathbf{i}} \hat{Q}_{\mathbf{i}} + \hat{Q}_{\emptyset} = \dots = \sum_{\mathbf{i} \in \hat{A}_k} \hat{\Pi}_{\mathbf{i}} r_0 + \sum_{s=0}^{k-1} \sum_{\mathbf{i} \in \hat{A}_s} \hat{\Pi}_{\mathbf{i}} \hat{Q}_{\mathbf{i}}. \quad (5.5)
\end{aligned}$$

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

6 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 [22, 23, 30]. We will elaborate more on this point after we state our main result.

As already mentioned in Section 2, our main result shows that

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

as $n \rightarrow \infty$, 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.

6.1 The linear branching stochastic fixed-point equation

We define the linear branching SFPE according to:

$$\mathcal{R} \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\mathcal{N}} \mathcal{C}_j \mathcal{R}_j + \mathcal{Q}, \quad (6.1)$$

where $(\mathcal{N}, \mathcal{Q}, \mathcal{C}_1, \mathcal{C}_2, \dots)$ is a real-valued random vector with $\mathcal{N} \in \mathbb{N} \cup \{\infty\}$, $P(|\mathcal{Q}| > 0) > 0$, and the $\{\mathcal{R}_i\}$ are i.i.d. copies of \mathcal{R} , independent of the vector $(\mathcal{N}, \mathcal{Q}, \mathcal{C}_1, \mathcal{C}_2, \dots)$. The vector $(\mathcal{N}, \mathcal{Q}, \mathcal{C}_1, \mathcal{C}_2, \dots)$ is often referred to as the generic branching vector, and in the general setting is allowed to be arbitrarily dependent with the weights $\{\mathcal{C}_i\}$ not necessarily identically distributed. This equation is also known as the “smoothing transform” [21, 16, 1, 3].

In the context of ranking algorithms, we can identify \mathcal{N} with the in-degree of a node, \mathcal{Q} with its personalization value, and the $\{\mathcal{C}_i\}$ with the weights of the neighboring nodes pointing to it. We now explain how to construct a solution to (6.1).

Similarly as what we did in Section 5.4 and using the same notation introduced there, we construct a weighted tree using a sequence $\{(\mathcal{N}_i, \mathcal{Q}_i, \mathcal{C}_{(i,1)}, \mathcal{C}_{(i,2)}, \dots)\}_{i \in \mathcal{U}}$ of i.i.d. copies of the vector $(\mathcal{N}, \mathcal{Q}, \mathcal{C}_1, \mathcal{C}_2, \dots)$ to define its structure and its node attributes. This construction is known in the literature as a WBP [32]. Next, let \mathcal{A}_k denote the number of individuals in the k th generation of the tree, and to each node \mathbf{i} in the tree assign a weight $\Pi_{\mathbf{i}}$ according to the recursion

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

Then, the random variable formally defined as

$$\mathcal{R} := \sum_{k=0}^{\infty} \sum_{\mathbf{i} \in \mathcal{A}_k} \Pi_{\mathbf{i}} \mathcal{Q}_{\mathbf{i}} \quad (6.2)$$

is called the endogenous solution to (6.1), and provided $E \left[\sum_{i=1}^{\mathcal{N}} |\mathcal{C}_i|^{\beta} \right] < 1$ for some $0 < \beta \leq 1$, it is well defined (see [23], Lemma 4.1). The name “endogenous” comes from its explicit construction in terms of the weighted tree. We point out that equation (6.1) has in general multiple solutions [3, 4], so it is important to emphasize that the one considered here is the endogenous one.

Comparing (5.5) and (6.2) suggests that $\hat{R}^{(n,k_n)}$ should converge to \mathcal{R} provided the distribution of the attribute vectors in the TBT converges to the distribution of the generic branching vector in the WBP, but in order to formalize this heuristic there are two difficulties that we need to overcome. The first one is that the TBT was defined using a sequence of (conditionally) independent vectors of the form $\{(\hat{N}_{\mathbf{i}}, \hat{Q}_{\mathbf{i}}, \hat{C}_{\mathbf{i}})\}_{\mathbf{i} \in \mathcal{U}}$, where by construction (see Assumption 5.1 and (5.2)) the generic attribute vector $(\hat{N}_1, \hat{Q}_1, \hat{C}_1)$ is dependent. Note that this implies that the vectors $(\hat{N}_{\mathbf{i}}, \hat{Q}_{\mathbf{i}}, \hat{C}_{(\mathbf{i},1)}, \hat{C}_{(\mathbf{i},2)}, \dots)$ and $\{(\hat{N}_{(\mathbf{i},j)}, \hat{Q}_{(\mathbf{i},j)}, \hat{C}_{(\mathbf{i},j,1)}, \hat{C}_{(\mathbf{i},j,2)}, \dots)\}_{j \geq 1}$ are dependent through the dependence between $\hat{N}_{(\mathbf{i},j)}$ and $\hat{C}_{(\mathbf{i},j)}$, which destroys the branching property of the WBP. The second problem is that the root node of the TBT has a different distribution from the rest of the nodes in the tree.

It is therefore to be expected that we will need something more than weak convergence of the node attributes to obtain the convergence of $\hat{R}^{(n,k_n)}$ we seek. To solve the first problem we will require that $(\hat{N}_1, \hat{Q}_1, \hat{C}_1)$ converges to $(\mathcal{N}, \mathcal{Q}, \mathcal{C})$ with \mathcal{C} independent of $(\mathcal{N}, \mathcal{Q})$. Note that this will naturally

lead to the $\{\mathcal{C}_i\}$ being i.i.d. in (6.1). To solve the second problem we will allow the attributes of the root node in the TBT to converge to their own limit $(\mathcal{N}_0, \mathcal{Q}_0)$. In view of these observations we can now identify the limit of $\hat{R}^{(n, k_n)}$ to be:

$$\mathcal{R}^* := \sum_{i=1}^{\mathcal{N}_0} \mathcal{C}_i \mathcal{R}_i + \mathcal{Q}_0, \quad (6.3)$$

where the $\{\mathcal{R}_i\}$ are i.i.d. copies of \mathcal{R} , as given by (6.2), independent of the vector $(\mathcal{N}_0, \mathcal{Q}_0, \{\mathcal{C}_i\})$ with $\{\mathcal{C}_i\}$ i.i.d. and independent of $(\mathcal{N}_0, \mathcal{Q}_0)$. The appropriate condition ensuring that \mathcal{R}^* is the correct limit is given in terms of the Kantorovich-Rubinstein distance (also known as the minimal l_1 distance or the Wasserstein distance).

Definition 6.1 Consider the metric space $(\mathbb{R}^d, \|\cdot\|_1)$, where $\|\mathbf{x}\|_1$ is the l_1 norm in \mathbb{R}^d . Let $M(\mu, \nu)$ denote the set of joint probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν . Then, the Kantorovich-Rubinstein distance between μ and ν is given by

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

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

$$\mathcal{P}_1(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|\mathbf{x} - \mathbf{x}_0\|_1 d\mu(\mathbf{x}) < \infty \right\},$$

for some $\mathbf{x}_0 \in \mathbb{R}^d$, where $\mathcal{P}(\mathbb{R}^d)$ is the set of Borel probability measures on \mathbb{R}^d . We refer the interested reader to [39] for a thorough treatment of this distance, since Definition 6.1 gives only a special case.

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

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

for all $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, where the supremum is taken over all Lipschitz continuous functions $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ with Lipschitz constant one (see Remark 6.5 in [39]).

We now 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 6.2 Given the extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ define

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

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

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

as $n \rightarrow \infty$, where

$$F^*(m, q) := P(\mathcal{N}_0 \leq m, \mathcal{Q}_0 \leq q) \quad \text{and} \quad F(m, q, x) := P(\mathcal{N} \leq m, \mathcal{Q} \leq q)P(\mathcal{C} \leq x).$$

Remark 6.3 Note that Assumption 6.2 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,

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

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, $(\hat{N}_0, \hat{Q}_0) \Rightarrow (\mathcal{N}_0, \mathcal{Q}_0)$. The duality formula, combined with Assumption 5.1, also implies that $E[\mathcal{N}_0] = \nu_1$, $E[\mathcal{N}] = \mu$ and $E[\mathcal{C}] = \nu_5/\nu_1$.

6.2 Main Result

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

Theorem 6.4 Suppose the extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ satisfies Assumptions 5.1 and 6.2. Then,

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

as $n \rightarrow \infty$, where \mathcal{R}^* is defined as in (6.3) 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 (2.1).

Proof. Define Ω_n according to Assumption 5.1 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 5.1, $\rho = E[\mathcal{N}]E[\mathcal{C}] = \nu_5\mu/\nu_1 < 1$, which is a sufficient condition for \mathcal{R} to be well defined (see Lemma 4.1 in [23]). 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 (5.3),

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

as $n \rightarrow \infty$. When $\mu \leq 1$ we can take $k_n = n^\epsilon$, with $\epsilon < \min\{1/2, \gamma\}$, 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 4.8 in [12] and Assumption 6.2 once we verify that, as $n \rightarrow \infty$,

$$\mathbb{E}_n \left[\hat{N}_1 | \hat{C}_1 \right] \xrightarrow{P} E[\mathcal{N}]E[\mathcal{C}] \quad \text{and} \quad \mathbb{E}_n \left[|\hat{Q}_1 \hat{C}_1| \right] \xrightarrow{P} E[|\mathcal{Q}|]E[\mathcal{C}]. \quad (6.4)$$

To show that (6.4) 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 6.2 and Remark 6.3 imply that

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

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

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

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

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

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 $n \rightarrow \infty$, 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. ■

6.3 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 Section 8.2. 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 [30, 40].

Definition 6.5 *We say that a function f is regularly varying at infinity with index $-\alpha$, denoted $f \in \mathcal{R}_{-\alpha}$, if $f(x) = x^{-\alpha}L(x)$ for some slowly varying function L ; and $L : [0, \infty) \rightarrow (0, \infty)$ is slowly varying if $\lim_{x \rightarrow \infty} L(\lambda x)/L(x) = 1$ for any $\lambda > 0$.*

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

Theorem 6.6 *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 (6.2).*

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

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

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

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

Remark 6.7 (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

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

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

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

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 6.6 requires the weights $\{C_i\}$ to be nonnegative, which is not a condition in Theorem 6.4. The tail asymptotics of \mathcal{R} , and therefore of \mathcal{R}^* , in the real-valued case are unknown.

7 Algorithm to generate bi-degree sequences

As an example of an extended bi-degree sequence satisfying Assumptions 5.1 and 6.2, 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 [11], 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 $f_k^{\text{in}} = P(\mathcal{N} = k)$, and $f_k^{\text{out}} = P(\mathcal{D} = k)$, $k \geq 0$, respectively. Furthermore, we will assume that these target distributions satisfy $E[\mathcal{N}] = E[\mathcal{D}]$,

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

for some slowly varying functions L_{in} and L_{out} , and $\alpha > 1, \beta > 2$. To the original construction given in [11] 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 $\{\mathcal{N}_1, \dots, \mathcal{N}_n\}$ from distribution F^{in} ; let $\overline{\mathcal{N}}_n = \sum_{i=1}^n \mathcal{N}_i$.
3. Sample an i.i.d. sequence $\{\mathcal{D}_1, \dots, \mathcal{D}_n\}$ from distribution F^{out} , independent of $\{\mathcal{N}_i\}$; let $\overline{\mathcal{D}}_n = \sum_{i=1}^n \mathcal{D}_i$.
4. Define $\Delta_n = \overline{\mathcal{N}}_n - \overline{\mathcal{D}}_n$. 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 $\{i_1, i_2, \dots, i_{|\Delta_n|}\}$ without replacement and let

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

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

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

Remark 7.1 *Note that since $E[|\mathcal{N} - \mathcal{D}|^{1+a}] < \infty$ for any $0 < a < \min\{\alpha - 1, \beta - 1\}$, then $E[|\mathcal{N} - \mathcal{D}|^{1+(\kappa_0-\delta_0)/(1-\kappa_0)}] < \infty$, and Corollary 8.4 in Section 8 gives*

$$P\left(|\Delta_n| > n^{1-\kappa_0+\delta_0}\right) = O\left(n^{-\delta_0(\kappa_0-\delta_0)/(1-\kappa_0)}\right) \quad (7.1)$$

as $n \rightarrow \infty$.

The two propositions below give the desired properties. Their proofs are given in Section 8.3.

Proposition 7.2 *The extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ generated by the IID Algorithm satisfies Assumption 5.1 for any $0 < \kappa < \beta - 2$, any $0 < \gamma < \min\{(\kappa_0 - \delta_0)^2/(1 - \delta_0), (\beta - 2 - \kappa)/\beta\}$, $\mu = \nu_1 = E[\mathcal{N}] = E[\mathcal{D}]$, $\nu_2 = (E[\mathcal{D}])^2$, $\nu_3 = E[\mathcal{D}^2]$, $\nu_4 = E[\mathcal{D}^{2+\kappa}]$, $\nu_5 = E[|\zeta|]P(\mathcal{D} \geq 1)$, $H = E[|Q|] + 1$, and some $\varepsilon > 0$.*

Proposition 7.3 *The extended bi-degree sequence $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ generated by the IID Algorithm satisfies Assumption 6.2 with*

$$F^*(m, q) = P(\mathcal{N} \leq m)P(Q \leq q) \quad \text{and}$$

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

7.1 Numerical examples

To complement the theoretical contribution of the paper, 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

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

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

Once the sequences $\{\mathcal{N}_i\}$ and $\{\mathcal{D}_i\}$ 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 5. 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

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

To draw a sample from \mathcal{R}^* , note that by Proposition 7.3, \mathcal{R}^* in the IID Algorithm has the same distribution as \mathcal{R} , i.e., the endogenous solution to the SFPE

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

where $P(\mathcal{C} \leq x) = E[1(c/\mathcal{D} \leq x)\mathcal{D}]/\mu$. To sample \mathcal{R} we construct a WBP with generic branching vector $(\mathcal{N}, 1-c, \{\mathcal{C}_i\})$, with the $\{\mathcal{C}_i\}$ i.i.d. and independent of \mathcal{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 $\alpha = 1.5, \beta = 2.5, E[\mathcal{N}] = E[\mathcal{D}] = 2, c = 0.3$. For the TBT, we simulate up to $k_n = \lceil \log n \rceil$ 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 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 4 illustrates the weak 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 i th sample of \mathcal{R}^* and $x_i^{(n)}$ is the sorted i th 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.

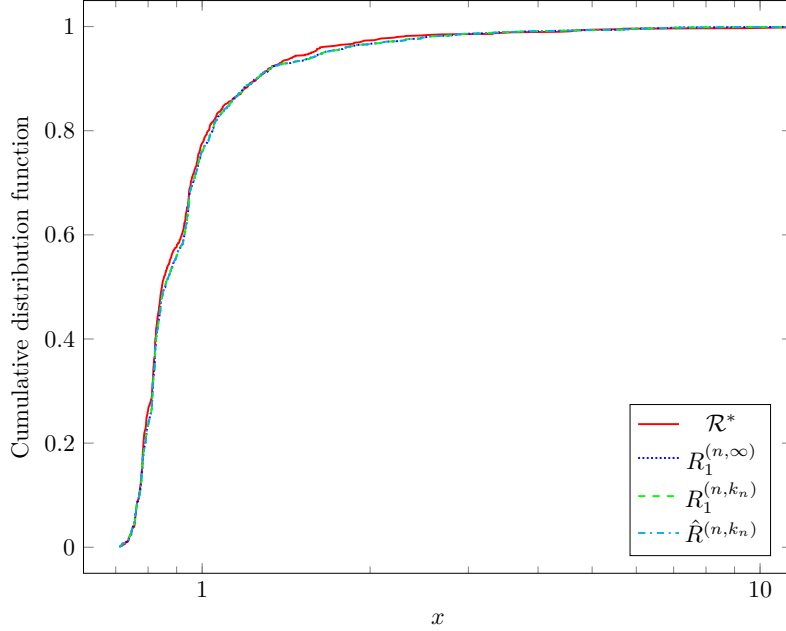


Figure 3: The empirical CDFs of 1000 samples of \mathcal{R}^* , $R_1^{(n,\infty)}$, $R_1^{(n,k_n)}$ and $\hat{R}^{(n,k_n)}$ for $n = 10000$ and $k_n = 9$.

8 Proofs

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.

8.1 Proof of the coupling lemma

Recall from Section 5 that \hat{N}_\emptyset denotes the number of offspring of the root node in the TBT (chosen from distribution (5.1)) and \hat{N}_1 denotes the number of offspring of a node chosen from distribution (5.2). Throughout this section we will also need to define

$$\mu_n^* = \mathbb{E}_n [\hat{N}_\emptyset] = \sum_{i,j,s,t} i f_n^*(i, j, s, t) = \frac{1}{n} \sum_{k=1}^n N_k = \frac{L_n}{n},$$

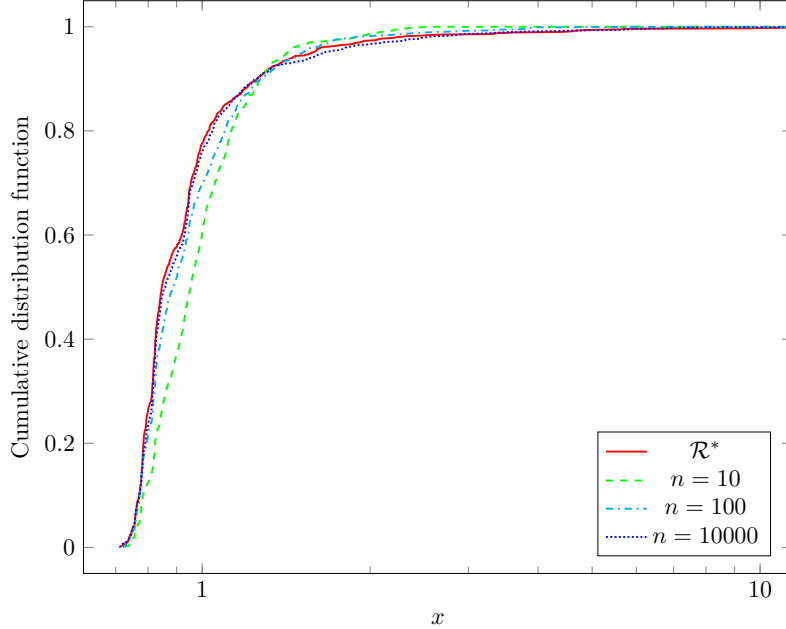


Figure 4: The empirical CDFs of 1000 samples of \mathcal{R}^* and $R_1^{(n,\infty)}$ for $n = 10, 100$ and 10000 .

and

$$\mu_n = \mathbb{E}_n [\hat{N}_1] = \sum_{i,j,s,t} if_n(i,j,s,t) = \frac{1}{L_n} \sum_{k=1}^n N_k D_k.$$

Before we give the proof of the Coupling Lemma 5.4 we will need the following estimates for the growth of the process $\{\hat{Z}_k\}$.

Lemma 8.1 *Suppose $(\mathbf{N}_n, \mathbf{D}_n, \mathbf{C}_n, \mathbf{Q}_n)$ satisfies Assumption 5.1 and recall that $\mu = \nu_2/\nu_1$. Then, for any constants $K > 0$, any nonnegative sequence $\{x_n\}$ with $x_n \rightarrow \infty$ and any $k = O(n^\gamma)$,*

$$P \left(\max_{0 \leq r \leq k} \frac{\hat{Z}_r}{\mu^r} > Kx_n \mid \Omega_n \right) = O(x_n^{-1}), \quad n \rightarrow \infty.$$

Proof. Start by noting that for any $r = 0, 1, 2, \dots$,

$$\mathbb{E}_n[\hat{Z}_r] = \mu_n^* \mu_n^r. \tag{8.1}$$

Moreover, on the event Ω_n ,

$$\begin{aligned} \mu_n &= \frac{n\nu_2(1 + O(n^{-\gamma}))}{n\nu_1(1 + O(n^{-\gamma}))} = \mu(1 + O(n^{-\gamma})), \quad \text{and} \\ \mu_n^* &= \frac{n\nu_1(1 + O(n^{-\gamma}))}{n} = \nu_1(1 + O(n^{-\gamma})). \end{aligned}$$

Next, note that conditionally on \mathcal{F}_n , the process

$$X_r = \frac{\hat{Z}_r}{\mu_n^* \mu_n^r} = \frac{1}{\mu_n^* \mu_n^r} \sum_{\mathbf{i} \in \hat{A}_{r-1}} \hat{N}_{\mathbf{i}}, \quad r \geq 1, \quad X_0 = \frac{\hat{N}_\emptyset}{\mu_n^*}$$

is a nonnegative martingale with respect to the filtration $\sigma(\mathcal{F}_r \cup \mathcal{F}_n)$, where $\mathcal{F}_r = \sigma(\hat{N}_{\mathbf{i}} : \mathbf{i} \in \hat{A}_s, s \leq r)$. Therefore, we can apply Doob's inequality, conditionally on \mathcal{F}_n , to obtain

$$\begin{aligned} P\left(\max_{0 \leq r \leq k} \frac{\hat{Z}_r}{\mu^r} > Kx_n \mid \Omega_n\right) &= P\left(\max_{0 \leq r \leq k} \frac{X_r \mu_n^* \mu_n^r}{\mu^r} > Kx_n \mid \Omega_n\right) \\ &= P\left(\max_{0 \leq r \leq k} X_r \nu_1 (1 + O(n^{-\gamma}))^{r+1} > Kx_n \mid \Omega_n\right) \\ &\leq \frac{1}{P(\Omega_n)} E \left[1(\Omega_n) \mathbb{E}_n \left[1\left(\max_{0 \leq r \leq k} X_r > \frac{Kx_n}{\nu_1 (1 + O(n^{-\gamma}))^{k+1}}\right) \right] \right] \\ &\leq \frac{1}{P(\Omega_n)} E \left[1(\Omega_n) \frac{\mathbb{E}_n[X_k] \nu_1 (1 + O(n^{-\gamma}))^{k+1}}{Kx_n} \right] \\ &= \frac{\nu_1 (1 + O(n^{-\gamma}))^{k+1}}{Kx_n} \quad (\text{since } \mathbb{E}_n[X_k] = 1). \end{aligned}$$

Noting that $(1 + O(n^{-\gamma}))^k = e^{O(kn^{-\gamma})} = O(1)$ as $n \rightarrow \infty$ 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 5.4. Start by defining

$$x_n = \begin{cases} (n/\mu^{2k})^{1/2}, & \mu > 1, \\ (n/k^2)^{1/2}, & \mu = 1, \\ n^{1/2}, & \mu < 1, \end{cases} \quad \text{and} \quad F_k = \left\{ \max_{0 \leq r \leq k} \frac{\hat{Z}_r}{\mu^r} \leq x_n \right\}.$$

Note that $x_n \rightarrow \infty$ as $n \rightarrow \infty$ for all $1 \leq k \leq h \log n$ when $\mu > 1$ and for all $1 \leq k \leq n^b$, $b < \min\{1/2, \gamma\}$, when $\mu \leq 1$. The constraint $b < \gamma$ will allow us to use Lemma 8.1.

Next, note that the j th 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

$$\frac{1}{L_n} \left(\sum_{r=0}^{s-1} \hat{V}_r + \sum_{t=1}^{i-1} D_t + (j-1) \right) \leq \frac{1}{L_n} \sum_{r=0}^s \hat{V}_r =: P_s.$$

It follows that,

$$\begin{aligned} P(\tau = s \mid \Omega_n) &\leq P(\tau = s, F_k \mid \Omega_n) + P(\tau = s, F_k^c \mid \Omega_n) \\ &\leq P(\text{Bin}(\hat{Z}_s, P_s) \geq 1, F_k \mid \Omega_n) + P(\tau = s, F_k^c \mid \Omega_n), \end{aligned}$$

where $\text{Bin}(n, p)$ is a Binomial random variable with parameters (n, p) . It follows that if we let $\mathcal{F}_k = \sigma(\hat{Z}_r, \hat{V}_r : 1 \leq r \leq k)$, then

$$\begin{aligned}
P(\tau \leq k | \Omega_n) &= \sum_{s=0}^k P(\tau = s | \Omega_n) \\
&\leq \sum_{s=0}^k \left\{ P\left(\text{Bin}(\hat{Z}_s, P_s) \geq 1, F_k \mid \Omega_n\right) + P(\tau = s, F_k^c | \Omega_n) \right\} \\
&\leq \sum_{s=0}^k E \left[1(F_k) P(\text{Bin}(\hat{Z}_s, P_s) \geq 1 | \mathcal{F}_k) \mid \Omega_n \right] + P(F_k^c | \Omega_n) \\
&\leq \sum_{s=0}^k E \left[1(F_k) \hat{Z}_s P_s \mid \Omega_n \right] + P(F_k^c | \Omega_n),
\end{aligned}$$

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

$$\begin{aligned}
E \left[1(F_k) \hat{Z}_s P_s \mid \Omega_n \right] &\leq E \left[\mu^s x_n P_s \mid \Omega_n \right] \tag{8.2} \\
&= \frac{\mu^s x_n}{\nu_1 n} \sum_{r=0}^s E \left[\hat{V}_r \mid \Omega_n \right] (1 + O(n^{-\gamma})) \\
&= \frac{\mu^s x_n}{\nu_1 n} \left\{ E \left[\hat{V}_0 \mid \Omega_n \right] + \sum_{r=1}^s E \left[\mathbb{E}_n \left[\hat{V}_r | \hat{Z}_{r-1} \right] \mid \Omega_n \right] \right\} (1 + O(n^{-\gamma})) \\
&= \frac{\mu^s x_n}{\nu_1 n} \left\{ E \left[\mu_n^* \mid \Omega_n \right] + \sum_{r=1}^s E \left[\hat{Z}_{r-1} \lambda_n \mid \Omega_n \right] \right\} (1 + O(n^{-\gamma})),
\end{aligned}$$

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

$$\mathbb{E}_n \left[\hat{V}_0 \right] = \mathbb{E}_n \left[\hat{D}_\emptyset \right] = \mu_n^*, \quad \mathbb{E}_n \left[\hat{V}_r | \hat{Z}_{r-1} \right] = \hat{Z}_{r-1} \lambda_n, \quad r \geq 1,$$

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

$$\begin{aligned}
E \left[1(F_k) \hat{Z}_s P_s \mid \Omega_n \right] &\leq \frac{\mu^s x_n}{\nu_1 n} \left\{ \nu_1 + \sum_{r=1}^s \lambda E \left[\hat{Z}_{r-1} \mid \Omega_n \right] \right\} (1 + O(n^{-\gamma})) \\
&= \frac{\mu^s x_n}{\nu_1 n} \left\{ \nu_1 + \sum_{r=1}^s \lambda E \left[\mu_n^* \mu_n^{r-1} \mid \Omega_n \right] \right\} (1 + O(n^{-\gamma})) \quad (\text{by (8.1)}).
\end{aligned}$$

Using the observation that $E \left[\mu_n^* \mu_n^{r-1} \mid \Omega_n \right] = \nu_1 \mu^{r-1} (1 + O(n^{-\gamma}))^{r-1}$ (see the proof of Lemma 8.1), 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 $E[\hat{Z}_s P_s | \Omega_n]$ in (8.2) directly, since that would have led to having to compute $\mathbb{E}_n[\hat{Z}_{s-1}^2]$ and neither \hat{N}_0 nor \hat{N}_1 are required to have finite second moments in the limit. Now, since by Lemma 8.1 we have that $P(F_k^c | \Omega_n) = O(x_n^{-1})$, 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

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

as $n \rightarrow \infty$. This completes the proof. ■

8.2 Proof of the asymptotic behavior of \mathcal{R}^*

We give in this section the proof of Theorem 6.6 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 (6.2). The tail behavior of \mathcal{R} is the main focus of the work in [40, 22, 24, 23, 30].

Proof of Theorem 6.6. We consider the case when \mathcal{N} is regularly varying first. By Theorem 3.4 in [30] and the remarks that follow it (see also Theorem 4.1 in [40]),

$$P(\mathcal{R} > x) \sim \frac{(E[\mathcal{Q}]E[\mathcal{C}_1])^\alpha}{(1-\rho)^\alpha(1-\rho_\alpha)} P(\mathcal{N} > x), \quad x \rightarrow \infty,$$

and therefore, $P(\mathcal{R} > x) \in \mathcal{R}_{-\alpha}$. Next, since the $\{\mathcal{C}_i\}$ are i.i.d. and independent of \mathcal{N} , Minkowski's inequality gives for any $\beta \geq 1$,

$$E \left[\left(\sum_{i=1}^{\mathcal{N}} \mathcal{C}_i \right)^\beta \right] = E \left[E \left[\left(\sum_{i=1}^{\mathcal{N}} \mathcal{C}_i \right)^\beta \middle| \mathcal{N} \right] \right] \leq E \left[\mathcal{N}^\beta E[\mathcal{C}_1^\beta] \right]. \quad (8.3)$$

Applying Lemma 2.3 in [30] with $\beta = 1 + \delta$ gives that $E[|\mathcal{R}|^{1+\delta}] < \infty$ for all $0 < \delta < \alpha - 1$. By conditioning on the filtration $\mathcal{F}_k = \sigma(\mathcal{N}_i, \mathcal{C}_{(i,1)}, \mathcal{C}_{(i,2)}, \dots) : \mathbf{i} \in \mathcal{A}_s, s < k$ 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 [25] 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 [30] we conclude that

$$\begin{aligned} P(\mathcal{R}^* > x) &\sim \left(E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa \frac{(1-\rho)^\alpha(1-\rho_\alpha)}{(E[\mathcal{Q}])^\alpha} (E[\mathcal{R}])^\alpha \right) P(\mathcal{R} > x) \\ &\sim (E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa(1-\rho_\alpha)) \frac{(E[\mathcal{Q}]E[\mathcal{C}_1])^\alpha}{(1-\rho)^\alpha(1-\rho_\alpha)} P(\mathcal{N} > x) \end{aligned}$$

as $x \rightarrow \infty$.

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 (8.3) and the theorem's assumptions. Then, by Theorem 4.4 in [30] (see also Theorem 4.1 in [40]) we have

$$P(\mathcal{R} > x) \sim (1-\rho_\alpha)^{-1}P(\mathcal{Q} > x), \quad x \rightarrow \infty.$$

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 [30], that

$$\begin{aligned} P(\mathcal{R}^* > x) &\sim (E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa(1-\rho_\alpha)) P(\mathcal{R} > x) \\ &\sim (E[\mathcal{N}_0]E[\mathcal{C}_1^\alpha] + \kappa(1-\rho_\alpha)) (1-\rho_\alpha)^{-1}P(\mathcal{Q} > x) \end{aligned}$$

as $x \rightarrow \infty$. ■

8.3 Proofs of properties of the IID Algorithm

Before giving the proofs of Propositions 7.2 and 7.3 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 8.2 *Let X_1, X_2, \dots, 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 i th smallest observation from the set $\{X_1, X_2, \dots, X_n\}$. Let $\{\pi_1, \pi_2, \dots, \pi_n\}$ be any permutation of the set $\{1, 2, \dots, n\}$. Then, for any $k_n \in \{1, 2, 3, 4, \dots, n\}$ 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 $n \rightarrow \infty$.

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 $\{Y_1, Y_2, \dots, Y_n\}$ 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 i th smallest from the set $\{Y_1, Y_2, \dots, Y_n\}$. Moreover, it is known (see [37], 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 [44], 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. \blacksquare

Lemma 8.3 *Let $\{X_1, X_2, \dots, X_n\}$ be i.i.d. random variables satisfying $E[|X_1|^{1+\kappa}] < \infty$ for some $\kappa > 0$ and $\mu = E[X_1]$. Set $S_m = X_1 + \dots + X_m$ and $\theta = \min\{1+\kappa, 2\}$. Then, for any $K > 0$,*

any nonnegative sequence $\{x_n\}$ such that $x_n \rightarrow \infty$ as $n \rightarrow \infty$, and all $m = o(x_n^{1+\kappa})$, there exists an $n_0 \geq 1$ such that for all $n \geq n_0$,

$$P(|S_m - m\mu| > Kx_n) \leq E[|X_1|^\theta] \left(\frac{2}{K^2} + 1 \right) \frac{m}{x_n^\theta}.$$

Proof. If $\kappa \geq 1$, then Chebyshev's inequality gives, for all $m \geq 1$,

$$P(|S_m - m\mu| > Kx_n) \leq \frac{m\text{Var}(X_1)}{K^2x_n^2} \leq \frac{mE[|X_1|^2]}{K^2x_n^2} = \frac{mE[|X_1|^\theta]}{K^2x_n^\theta}.$$

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

$$\begin{aligned} |E[\tilde{X}_1] - \mu| &= |E[X_1 1(|X_1| \leq t)]/G(t) - \mu| \\ &\leq \frac{1}{G(t)} |E[X_1 1(|X_1| \leq t)] - \mu| + \frac{|\mu|\bar{G}(t)}{G(t)} \\ &= \frac{1}{G(t)} (|E[X_1 1(|X_1| > t)]| + |\mu|\bar{G}(t)) \\ &\leq \frac{1}{G(t)} \left(t\bar{G}(t) + \int_t^\infty \bar{G}(x)dx + |\mu|\bar{G}(t) \right) \\ &\leq \frac{E[|X_1|^{1+\kappa}]}{G(t)} \left(t^{-\kappa} + \int_t^\infty x^{-1-\kappa} dx + |\mu|t^{-1-\kappa} \right) \quad (\text{by Markov's inequality}) \\ &= \frac{E[|X_1|^{1+\kappa}]}{G(t)} \left(\frac{1+\kappa}{\kappa} + |\mu|t^{-1} \right) t^{-\kappa}. \end{aligned}$$

Then, for sufficiently large n , we obtain that

$$|E[\tilde{X}_1] - \mu| \leq 2E[|X_1|^{1+\kappa}] \left(\frac{1+\kappa}{\kappa} + |\mu| \right) t^{-\kappa} \triangleq K't^{-\kappa} = K'x_n^{-\kappa}.$$

It follows that for sufficiently large n and $m = o(x_n^{1+\kappa})$,

$$\begin{aligned} &P(|S_m - m\mu| > Kx_n) \\ &= P\left(\left| \sum_{i=1}^m (\tilde{X}_i - \mu) \right| > Kx_n \right) G(t)^m + P\left(\left| \sum_{i=1}^m (X_i - \mu) \right| > Kx_n, \max_{1 \leq i \leq m} |X_i| > t \right) \\ &\leq P\left(\left| \sum_{i=1}^m (\tilde{X}_i - E[\tilde{X}_1]) \right| + m |E[\tilde{X}_1] - \mu| > Kx_n \right) G(t)^m + P\left(\max_{1 \leq i \leq m} |X_i| > t \right) \\ &\leq \frac{G(t)^m}{(Kx_n - K'mt^{-\kappa})^2} \cdot m\text{Var}(\tilde{X}_1) + 1 - G(t)^m \quad (\text{by Chebyshev's inequality}) \\ &\leq \frac{G(t)^m m\text{Var}(\tilde{X}_1)}{K^2x_n^2(1 - mx_n^{-1-\kappa}K'/K)^2} + m\bar{G}(t). \end{aligned}$$

To estimate $\text{Var}(\tilde{X}_1)$ note that

$$\text{Var}(\tilde{X}_1) \leq E[\tilde{X}_1^2] = \frac{E[X_1^2 1(|X_1| \leq t)]}{G(t)} \leq \frac{E[|X_1|^{1+\kappa}]t^{1-\kappa}}{G(t)},$$

so using Markov's inequality again to estimate $\overline{G}(t)$ gives us

$$\begin{aligned} P(|S_m - m\mu| > Kx_n) &\leq \frac{E[|X_1|^{1+\kappa}]}{K^2(1 - mx_n^{-1-\kappa}K'/K)^2} \cdot \frac{mt^{1-\kappa}}{x_n^2} + \frac{E[|X_1|^{1+\kappa}]m}{t^{1+\kappa}} \\ &= E[|X_1|^{1+\kappa}] \left(\frac{1}{K^2(1 - mx_n^{-1-\kappa}K'/K)^2} + 1 \right) \frac{m}{x_n^{1+\kappa}} \\ &= E[|X_1|^\theta] \left(\frac{1}{K^2(1 - mx_n^{-1-\kappa}K'/K)^2} + 1 \right) \frac{m}{x_n^\theta}. \end{aligned}$$

This completes the proof. ■

By setting $m = n$ and $x_n = n^{1-\gamma}$ we immediately obtain the following corollary.

Corollary 8.4 *Let $\{X_1, X_2, \dots, X_n\}$ be i.i.d. random variables satisfying $E[|X_1|^{1+\kappa}] < \infty$ for some $\kappa > 0$ and $\mu = E[X_1]$. Set $S_n = X_1 + \dots + X_n$. 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 $n \geq n_0$*

$$P(|S_n - n\mu| > Kn^{1-\gamma}) \leq E[|X_1|^\theta] \left(\frac{2}{K^2} + 1 \right) n^{-\theta(1-1/\theta-\gamma)}.$$

We now proceed to prove that the extended bi-degree sequence generated by the IID Algorithm satisfies Assumptions 5.1 and 6.2.

Proof of Proposition 7.2. It suffices to show that $P(\Omega_{n,i}^c) = O(n^{-\varepsilon})$ for some $\varepsilon > 0$ and $i = 1, \dots, 6$. Throughout the proof let $E_n = \{|\Delta_n| \leq n^{1-\kappa_0+\delta_0}\}$ and recall that by (7.1) $P(E_n^c) = O(n^{-\delta_0\eta})$, where $\eta = (\kappa_0 - \delta_0)/(1 - \kappa_0)$.

We start with $\Omega_{n,2}$. Let $\nu_2 = (E[\mathcal{D}])^2$ and define $\chi_i = D_i - \mathcal{D}_i$, $\tau_i = N_i - \mathcal{N}_i$. Note that $\chi_i, \tau_i \in \{0, 1\}$ for all $i = 1, \dots, n$; 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

$$\begin{aligned} \left| \sum_{i=1}^n D_i N_i - n\nu_2 \right| &= \left| \sum_{i=1}^n \mathcal{D}_i \mathcal{N}_i - n\nu_2 + \sum_{i=1}^n (\mathcal{D}_i \tau_i + \chi_i \mathcal{N}_i) \right| \\ &\leq \left| \sum_{i=1}^n \mathcal{D}_i \mathcal{N}_i - n\nu_2 \right| + \max \left\{ \sum_{i=n-\Delta_n+1}^n \mathcal{D}_{(i)}, \sum_{i=n-\Delta_n+1}^n \mathcal{N}_{(i)} \right\}, \end{aligned}$$

where $\mathcal{D}_{(i)}$ (respectively, $\mathcal{N}_{(i)}$) is the i th smallest value from the set $\{\mathcal{D}_1, \dots, \mathcal{D}_n\}$ (respectively, $\{\mathcal{N}_1, \dots, \mathcal{N}_n\}$). Since $|\Delta_n| \leq n^{1-\kappa_0+\delta_0}$ on E_n , we have

$$\begin{aligned} P(\Omega_{n,2}^c) &= P \left(\left| \sum_{i=1}^n D_i N_i - n\nu_2 \right| > n^{1-\gamma} \middle| E_n \right) \\ &\leq \frac{1}{P(E_n)} \left\{ P \left(\left| \sum_{i=1}^n \mathcal{D}_i \mathcal{N}_i - n\nu_2 \right| > \frac{n^{1-\gamma}}{2} \right) \right. \\ &\quad \left. + P \left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)} \rfloor + 1}^n \mathcal{D}_{(i)} > \frac{n^{1-\gamma}}{2} \right) + P \left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)} \rfloor + 1}^n \mathcal{N}_{(i)} > \frac{n^{1-\gamma}}{2} \right) \right\}. \end{aligned}$$

Now apply Corollary 8.4 to $X_i = \mathcal{D}_i \mathcal{N}_i$, which satisfies $E[(\mathcal{D}_1 \mathcal{N}_1)^{1+\eta}] = E[\mathcal{N}_1^{1+\eta}]E[\mathcal{D}_1^{1+\eta}] < \infty$, to obtain

$$P\left(\left|\sum_{i=1}^n \mathcal{D}_i \mathcal{N}_i - n\nu_2\right| > \frac{n^{1-\gamma}}{2}\right) = O\left(n^{-\eta+(1+\eta)\gamma}\right).$$

For the remaining two probabilities use Lemma 8.2 to see that

$$\begin{aligned} & P\left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)} \rfloor + 1}^n \mathcal{D}_{(i)} > \frac{n^{1-\gamma}}{2}\right) + P\left(\sum_{i=n-\lfloor n^{1-\eta(1-\kappa_0)} \rfloor + 1}^n \mathcal{N}_{(i)} > \frac{n^{1-\gamma}}{2}\right) \\ &= O\left(n^{(1-\eta(1-\kappa_0))\eta/(1+\eta) - (\eta/(1+\eta) - \gamma)}\right) \\ &= O\left(n^{-\eta(\kappa_0 - \delta_0)/(1+\eta) + \gamma}\right). \end{aligned}$$

It follows from these estimates that

$$P(\Omega_{n,2}^c) = O\left(n^{-\eta(\kappa_0 - \delta_0)/(1+\eta) + \gamma}\right). \quad (8.4)$$

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 $\nu_1 = E[\mathcal{D}]$, $\nu_3 = E[\mathcal{D}^2]$ and $\nu_4 = E[\mathcal{D}^{2+\kappa}]$. Similarly as what was done for $\Omega_{n,2}$, note that

$$\begin{aligned} \left|\sum_{i=1}^n D_i^\vartheta - nE[\mathcal{D}^\vartheta]\right| &\leq \left|\sum_{i=1}^n \mathcal{D}_i^\vartheta - nE[\mathcal{D}^\vartheta]\right| + \sum_{i=1}^n \left((\mathcal{D}_i + \chi_i)^\vartheta - \mathcal{D}_i^\vartheta\right) \\ &\leq \left|\sum_{i=1}^n \mathcal{D}_i^\vartheta - nE[\mathcal{D}^\vartheta]\right| + \sum_{i=1}^n \vartheta(\mathcal{D}_i + 1)^{\vartheta-1} \chi_i, \end{aligned}$$

where we used the inequality $(d+x)^\vartheta - d^\vartheta \leq \vartheta(d+1)^{\vartheta-1}x$ for $d \geq 0$, $x \in [0, 1]$ and $\vartheta \geq 1$. Now note that $E[(\mathcal{D}^\vartheta)^{1+\sigma}] < \infty$ 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 8.4 gives

$$P\left(\left|\sum_{i=1}^n \mathcal{D}_i^\vartheta - nE[\mathcal{D}^\vartheta]\right| > \frac{n^{1-\gamma}}{2}\right) = O\left(n^{-\sigma+(1+\sigma)\gamma}\right).$$

For the term involving the $\{\chi_i\}$ we use again Lemma 8.2 to obtain

$$\begin{aligned} P\left(\sum_{i=1}^n \vartheta(\mathcal{D}_i + 1)^{\vartheta-1} \chi_i > \frac{n^{1-\gamma}}{2}\right) &\leq P\left(\sum_{i=n-\lfloor n^{1-\eta} \rfloor + 1}^n \vartheta(\mathcal{D}_{(i)} + 1)^{\vartheta-1} > \frac{n^{1-\gamma}}{2}\right) \\ &= O\left(n^{(1-\eta)(1-1/2) - (1-\gamma-1/2)}\right) \\ &= O\left(n^{-\eta/2 + \gamma}\right). \end{aligned}$$

It follows that

$$P(\Omega_{n,i}^c) \leq \frac{1}{P(E_n)} \cdot O\left(n^{-\sigma+(1+\sigma)\gamma} + n^{-\eta/2 + \gamma}\right), \quad i = 1, 3, 4. \quad (8.5)$$

Now note that since $|\zeta| \leq c < 1$ a.s., then $E[|\zeta|^2] < \infty$ and Corollary 8.4 gives

$$\begin{aligned} P(\Omega_{n,5}^c) &= P\left(\left|\sum_{r=1}^n |\zeta_r| 1(D_r \geq 1) - n\nu_5\right| > n^{1-\gamma}\right) \\ &= P\left(\left|\sum_{r=1}^n |\zeta_r| 1(\mathcal{D}_r \geq 1) - n\nu_5\right| + c|\Delta_n| > n^{1-\gamma}\right) = O\left(n^{-1+2\gamma}\right). \end{aligned} \quad (8.6)$$

Finally, by Corollary 8.4 and (7.1),

$$P(\Omega_{n,6}^c) \leq P\left(\left|\sum_{r=1}^n |Q_r| - nE[|Q|]\right| > n \left|E_n\right.\right) = O\left(n^{-\epsilon_Q} + n^{-\delta_0\eta}\right). \quad (8.7)$$

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 (8.4) - (8.6) are strictly negative, which completes the proof. ■

Proof of Proposition 7.3. 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 $S_n = \sum_{i=1}^n \mathcal{D}_i$, $\mathcal{C}_k = \zeta_k/\mathcal{D}_k 1(\mathcal{D}_k \geq 1) + c \operatorname{sgn}(\zeta_k) 1(\mathcal{D}_k = 0)$, and fix $\psi^* : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $\psi : \mathbb{R}^3 \rightarrow \mathbb{R}$ to be Lipschitz continuous functions with Lipschitz constant one. Then,

$$\begin{aligned} \mathcal{E}_0 &:= \left| \frac{1}{n} \sum_{k=1}^n \psi^*(N_k, Q_k) - \frac{1}{n} \sum_{k=1}^n \psi^*(\mathcal{N}_k, Q_k) \right| \\ &\leq \frac{1}{n} \sum_{k=1}^n |\psi^*(\mathcal{N}_k + 1, Q_k) - \psi^*(\mathcal{N}_k, Q_k)| 1(N_k = \mathcal{N}_k + 1) \\ &\leq \frac{1}{n} \sum_{k=1}^n 1(N_k = \mathcal{N}_k + 1) \leq \frac{|\Delta_n|}{n}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{E}_1 &:= \left| \sum_{k=1}^n \psi(N_k, Q_k, C_k) \frac{D_k}{L_n} - \sum_{k=1}^n \psi(\mathcal{N}_k, Q_k, \mathcal{C}_k) \frac{\mathcal{D}_k}{S_n} \right| \\ &\leq \sum_{k=1}^n \frac{\mathcal{D}_k}{S_n} |\psi(N_k, Q_k, \mathcal{C}_k) - \psi(\mathcal{N}_k, Q_k, \mathcal{C}_k)| 1(\Delta_n \leq 0) \\ &\quad + \sum_{k=1}^n \frac{D_k}{L_n} |\psi(\mathcal{N}_k, Q_k, C_k) - \psi(\mathcal{N}_k, Q_k, \mathcal{C}_k)| 1(\Delta_n > 0) \\ &\quad + \sum_{k=1}^n \left| \psi(\mathcal{N}_k, Q_k, \zeta_k/\mathcal{D}_k) \left(\frac{D_k}{L_n} - \frac{\mathcal{D}_k}{S_n} \right) \right| 1(\Delta_n > 0) \\ &\leq \sum_{k=1}^n \frac{\mathcal{D}_k}{S_n} 1(N_k = \mathcal{N}_k + 1) + \sum_{k=1}^n \frac{D_k}{L_n} |\zeta_k/(\mathcal{D}_k + 1) - \mathcal{C}_k| 1(D_k = \mathcal{D}_k + 1) \\ &\quad + \sum_{k=1}^n |\psi(\mathcal{N}_k, Q_k, \mathcal{C}_k)| \left| \frac{(D_k - \mathcal{D}_k)S_n - \mathcal{D}_k\Delta_n}{L_n S_n} \right| 1(\Delta_n > 0), \end{aligned}$$

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

$$\sum_{k=1}^n \frac{\mathcal{D}_k}{S_n} 1(N_k = \mathcal{N}_k + 1) \leq \frac{n}{S_n} \left(\frac{1}{n} \sum_{k=1}^n \mathcal{D}_k^2 \right)^{1/2} \left(\frac{|\Delta_n|}{n} \right)^{1/2}.$$

Now, use the observation that $|\zeta_k| \leq c$ to obtain

$$\begin{aligned} & \sum_{k=1}^n \frac{D_k}{L_n} |\zeta_k / (\mathcal{D}_k + 1) - \mathcal{C}_k| 1(D_k = \mathcal{D}_k + 1) \\ & \leq c \sum_{k=1}^n \frac{1}{L_n \mathcal{D}_k} 1(D_k = \mathcal{D}_k + 1, \mathcal{D}_k \geq 1) + \sum_{k=1}^n \frac{1}{L_n} |\zeta_k - c \operatorname{sgn}(\zeta_k)| 1(D_k = \mathcal{D}_k + 1, \mathcal{D}_k = 0) \\ & \leq \frac{c}{L_n} \sum_{k=1}^n 1(D_k = \mathcal{D}_k + 1) \leq \frac{c|\Delta_n|}{S_n}. \end{aligned}$$

Next, use the bound $|\psi(m, q, x)| \leq \|(m, q, x)\|_1 + |\psi(0, 0, 0)|$ and Hölder's inequality to obtain

$$\begin{aligned} & \sum_{k=1}^n |\psi(\mathcal{N}_k, Q_k, \mathcal{C}_k)| \left| \frac{(D_k - \mathcal{D}_k)S_n - \mathcal{D}_k \Delta_n}{L_n S_n} \right| 1(\Delta_n > 0) \\ & \leq \sum_{k=1}^n |\psi(\mathcal{N}_k, Q_k, \mathcal{C}_k)| \frac{1(D_k = \mathcal{D}_k + 1)}{S_n} + \sum_{k=1}^n |\psi(\mathcal{N}_k, Q_k, \mathcal{C}_k)| \frac{\mathcal{D}_k |\Delta_n|}{S_n^2} \\ & \leq \frac{1}{S_n} \sum_{k=1}^n \|(\mathcal{N}_k, Q_k, c)\|_1 1(D_k = \mathcal{D}_k + 1) + \frac{|\Delta_n|}{S_n^2} \sum_{k=1}^n (\mathcal{N}_k \mathcal{D}_k + |Q_k| \mathcal{D}_k + c) + \frac{2|\psi(0, 0, 0)\Delta_n|}{S_n} \\ & \leq \frac{n}{S_n} \left\{ \left(\frac{1}{n} \sum_{k=1}^n \mathcal{N}_k^{1+\delta} \right)^{1/(1+\delta)} + \left(\frac{1}{n} \sum_{k=1}^n |Q_k|^{1+\delta} \right)^{1/(1+\delta)} \right\} \left(\frac{|\Delta_n|}{n} \right)^{\delta/(1+\delta)} \\ & \quad + \frac{|\Delta_n|}{S_n^2} \sum_{k=1}^n (\mathcal{N}_k \mathcal{D}_k + |Q_k| \mathcal{D}_k) + \frac{H|\Delta_n|}{S_n}, \end{aligned}$$

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 $E[\mathcal{D}^2] < \infty$ and that $\mathcal{N}, \mathcal{D}, Q$ are mutually independent) and the bounds derived above.

Finally, by the SLLN again and the fact that $E[\|(\mathcal{N}, Q, \mathcal{C})\|_1] < \infty$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \psi^*(N_k, Q_k) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \psi^*(\mathcal{N}_k, Q_k) = E[\psi^*(\mathcal{N}, Q)] \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \psi(N_k, Q_k, C_k) \frac{\mathcal{D}_i}{S_n} = \lim_{n \rightarrow \infty} \sum_{k=1}^n \psi(\mathcal{N}_k, Q_k, \mathcal{C}_k) \frac{\mathcal{D}_k}{S_n} = \frac{1}{\mu} E[\psi(\mathcal{N}, Q, \mathcal{C}) \mathcal{D}] \quad \text{a.s.}$$

The first limit combined with the duality formula gives that $d_1(F_n^*, F^*) \rightarrow 0$ a.s. For the second limit we still need to identify the limiting distribution, for which we note that

$$\begin{aligned} \frac{1}{\mu} E[\psi(\mathcal{N}, Q, \mathcal{C}) \mathcal{D}] &= \frac{1}{\mu} E[E[\psi(\mathcal{N}, Q, \mathcal{C}) \mathcal{D} | \mathcal{N}, Q]] = \frac{1}{\mu} E \left[\sum_{i=1}^{\infty} \int_{-\infty}^{\infty} \psi(\mathcal{N}, Q, z/i) i dF^{\zeta}(z) P(\mathcal{D} = i) \right] \\ &= \frac{1}{\mu} E \left[\sum_{i=1}^{\infty} \int_{-\infty}^{\infty} \psi(\mathcal{N}, Q, y) i dF^{\zeta}(yi) P(\mathcal{D} = i) \right] =: E[\psi(\mathcal{N}, Q, Y)], \end{aligned}$$

where Y has distribution function

$$\begin{aligned} P(Y \leq x) &= \frac{1}{\mu} E \left[\sum_{i=1}^{\infty} \int_{-\infty}^{\infty} 1(y \leq x) i dF^{\zeta}(yi) P(\mathcal{D} = i) \right] = \frac{1}{\mu} E \left[\sum_{i=1}^{\infty} i F^{\zeta}(ix) P(\mathcal{D} = i) \right] \\ &= \frac{1}{\mu} E[\mathcal{D} F^{\zeta}(\mathcal{D}x)] = \frac{1}{\mu} E[\mathcal{D} 1(\zeta/\mathcal{D} \leq x)] = P(\mathcal{C} \leq x). \end{aligned}$$

It follows that $E[\psi(\mathcal{N}, Q, \mathcal{C}) \mathcal{D}] / \mu = E[\psi(\mathcal{N}, Q, \mathcal{C})]$, which combined with the duality formula gives that $d_1(F_n, F) \rightarrow 0$ a.s. ■

References

- [1] G. Alsmeyer, J.D. Biggins, and M. Meiners. The functional equation of the smoothing transform. *Ann. Probab.*, 40(5):2069–2105, 2012.
- [2] G. Alsmeyer, E. Damek, and S. Mentemeier. Tails of fixed points of the two-sided smoothing transform. In *Springer Proceedings in Mathematics & Statistics: Random Matrices and Iterated Random Functions*, 2012.
- [3] G. Alsmeyer and M. Meiners. Fixed points of inhomogeneous smoothing transforms. *Journal of Difference Equations and Applications*, 18(8):1287–1304, 2012.
- [4] G. Alsmeyer and M. Meiners. Fixed points of the smoothing transform: Two-sided solutions. *Probab. Theory Rel.*, 155(1-2):165–199, 2013.
- [5] R. Andersen, F. Chung, and K. Lang. Local graph partitioning using PageRank vectors. In *Proceedings of FOCS2006*, pages 475–486, 2006.
- [6] S. Bhamidi, J.M. Steele, and T. Zaman. Twitter event networks and the superstar model. *Preprint*, 2012. arXiv: 0902.0885.
- [7] P. Boldi and S. Vigna. Axioms for centrality. *To Appear in Internet Mathematics*, 2014.
- [8] 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.
- [9] 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.

- [10] 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.
- [11] N. Chen and M. Olvera-Cravioto. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013.
- [12] N. Chen and M. Olvera-Cravioto. Coupling on weighted branching trees. http://www.columbia.edu/~mo2291/Coupling_Chen_Olv.pdf, 2014. Preprint.
- [13] 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.
- [14] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of combinatorics*, 6(2):125–145, 2002.
- [15] P.G. Constantine and D.F. Gleich. Random alpha pagerank. *Internet Mathematics*, 6(2):189–236, 2009.
- [16] R. Durrett and T. Liggett. Fixed points of the smoothing transformation. *Z. Wahrsch. verw. Gebiete*, 64:275–301, 1983.
- [17] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistics and Probabilistic Mathematics. Cambridge university press Cambridge, 2007.
- [18] 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.
- [19] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating Web spam with TrustRank. In *Proceeding of VLDB2004*, pages 576–587, 2004.
- [20] T. H. Haveliwala. Topic-sensitive PageRank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526. ACM, 2002.
- [21] R. Holley and T. Liggett. Generalized potlatch and smoothing processes. *Z. Wahrsch. verw. Gebiete*, 55:165–195, 1981.
- [22] P.R. Jelenković and M. Olvera-Cravioto. Information ranking and power laws on trees. *Adv. Appl. Prob.*, 42(4):1057–1093, 2010.
- [23] P.R. Jelenković and M. Olvera-Cravioto. Implicit renewal theorem for trees with general weights. *Stoch. Proc. Appl.*, 122(9):3209–3238, 2012.
- [24] P.R. Jelenković and M. Olvera-Cravioto. Implicit renewal theory and power tails on trees. *Adv. Appl. Prob.*, 44(2):528–561, 2012.
- [25] A.H. Jessen and T. Mikosch. Regularly varying functions. *Publications de L’Institut Mathématique, Nouvelle Serie*, 80(94):171–192, 2006.
- [26] and D. Lebedev K. Avrachenkov. PageRank of scale-free growing networks. *Internet Mathematics*, 3(2):207–231, 2006.

- [27] A.N. Langville and C.D. Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.
- [28] 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.
- [29] Mark Newman. *Networks: an introduction*. Oxford University Press, 2010.
- [30] M. Olvera-Cravioto. Tail behavior of solutions of linear recursions on trees. *Stochastic Processes and their Applications*, 122(4):1777–1807, 2012.
- [31] G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. In *Computing and Combinatorics*, pages 330–339. Springer, 2002.
- [32] U. Röslér. The weighted branching process. *Dynamics of complex and irregular systems (Bielefeld, 1991)*, pages 154–165, 1993. Bielefeld Encounters in Mathematics and Physics VIII, World Science Publishing, River Edge, NJ.
- [33] U. Röslér, V.A. Topchii, and V.A. Vatutin. Convergence conditions for the weighted branching process. *Discrete Math. Appl.*, 10(1):5–21, 2000.
- [34] U. Röslér, V.A. Topchii, and V.A. Vatutin. The rate of convergence for weighted branching processes. *Siberian Adv. Math.*, 12(4):57–82, 2002.
- [35] R. van der Hofstad. Random graphs and complex networks. <http://www.win.tue.nl/rhofstad/NotesRGCN.pdf>, 2014.
- [36] 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.
- [37] K. Vännman. Estimators based on order statistics from a Pareto distribution. *Journal of the American Statistical Association*, 71(355):704–708, 1976.
- [38] S. Vigna. A weighted correlation index for rankings with ties. *Preprint*, 2014. arXiv: 1404.3325.
- [39] C. Villani. *Optimal transport, old and new*. Springer, New York, 2009.
- [40] Y. Volkovich and N. Litvak. Asymptotic analysis for personalized web search. *Adv. Appl. Prob.*, 42(2):577–604, 2010.
- [41] 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.
- [42] Y. Volkovich, N. Litvak, and B. Zwart. Extremal dependencies and rank correlations in power law networks. In *Complex Sciences*, pages 1642–1653. Springer, 2009.
- [43] 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.
- [44] J.G. Wendel. Note on the gamma function. *Amer. Math. Monthly*, 55(9):563–564, 1948.