Adv. Appl. Prob. **48,** 499–524 (2016)
*doi:*10.1017/apr.2016.12
© *Applied Probability Trust* 2016

COUPLING ON WEIGHTED BRANCHING TREES

NINGYUAN CHEN * AND
MARIANA OLVERA-CRAVIOTO,* ** *Columbia University*

Abstract

In this paper we consider linear functions constructed on two different weighted branching processes and provide explicit bounds for their Kantorovich–Rubinstein distance in terms of couplings of their corresponding generic branching vectors. Motivated by applications to the analysis of random graphs, we also consider a variation of the weighted branching process where the generic branching vector has a different dependence structure from the usual one. By applying the bounds to sequences of weighted branching processes, we derive sufficient conditions for the convergence in the Kantorovich–Rubinstein distance of linear functions. We focus on the case where the limits are endogenous fixed points of suitable smoothing transformations.

Keywords: Weighted branching processes; smoothing transform; coupling; Kantorovich–Rubinstein distance; Wasserstein distance; weak convergence

2010 Mathematics Subject Classification: Primary 60J80
Secondary 60B10; 60H25

1. Introduction

In this paper we study one particular solution of the linear stochastic fixed-point equation (SFPE)

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

where $(Q, N, C_1, C_2, \dots)$ is a real-valued vector with $N \in \mathbb{N} \cup \{\infty\}$, and $\{R_i\}_{i\in\mathbb{N}}$ are independent and identically distributed (i.i.d.) random variables having the same distribution as R. This distributional equation appears in the probabilistic analysis of algorithms and has been studied in [9], [11], [14], [21], [22], and its homogeneous version ($Q \equiv 0$) has been studied extensively in the literature of weighted branching processes and multiplicative cascades; see, e.g. [4], [6], [12], [16], [17], and the references therein.

Although it is well known that (1) has multiple solutions (see [2]–[4], [13]), it is often the case that in applications (e.g. [9], [11], [21]) only one of them is relevant. More precisely, we are usually interested in the solution obtained by iterating the SFPE starting from a well-behaved initial condition, which can be explicitly constructed on a weighted branching process (as explained in Section 2) and hence is often referred to as a special endogenous solution. We refer the interested reader to [2]–[4], and [13] for a more thorough discussion on the existence of multiple solutions to (1) and the role of the endogenous solution(s) in their characterization; in particular, the recent work in [13] treats the case of real-valued weights $\{C_i\}$ where there may

Received 17 October 2014; revision received 25 June 2015.
* Postal address: Industrial Engineering and Operations Research Department, Columbia University, 321 S. W. Mudd Building, 500 W. 120th Street, New York, NY 10027, USA.
** Email address: mo2291@columbia.edu

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

be multiple endogenous solutions. The focus of this paper is to analyze the 'most attractive' endogenous solution mentioned above both for $Q \equiv 0$ and $\mathbb{P}(Q \neq 0) > 0$. In particular, we consider two different weighted branching processes and compare their corresponding special endogenous solutions in the Kantorovich–Rubinstein distance d_1, also known as the Wasserstein distance of order 1; see, e.g. [25]. Although convergence to the (unique) endogenous solution to (1) for nonnegative weights has been previously studied in the context of the analysis of divide-and-conquer algorithms using the Wasserstein distance of order 2 (see [19], [21], and [22]), recent applications to the analysis of information ranking algorithms [9], [14], [26], where the variance of R might be infinite, suggest the use of the weaker Kantorovich–Rubinstein distance.

Moreover, motivated by the same applications to the analysis of information ranking algorithms mentioned above, we analyze a variation of the weighted branching process constructed using a generic branching vector of the form (Q, N, C), i.e. where the weight C of a node in the tree is allowed to depend on the node's copy of (Q, N), rather than on its *parent's* copy as is the case with a generic branching vector of the form $(Q, N, C_1, C_2, \dots)$. To avoid confusion, we will refer to this variation as a weighted branching tree. Using this weighted branching tree we mimic the construction of the special endogenous solution and provide conditions under which it will be close, in the Kantorovich–Rubinstein distance, to the special endogenous solution to (1) constructed on a usual weighted branching process.

The main results in this paper provide explicit bounds for the Kantorovich–Rubinstein distance between two random variables constructed according to the representation for the special endogenous solution to (1); these bounds are given in terms of the Kantorovich–Rubinstein distance between their generic branching vectors. We then use these bounds to obtain the convergence of a sequence of such random variables in the same distance. We illustrate the main results with applications to the analysis of random graphs and information ranking algorithms.

The remainder of the paper is organized as follows. In Section 2 we describe the weighted branching process and its variation, the weighted branching tree. Section 3 contains a brief exposition of the Kantorovich–Rubinstein distance and some of its main properties. Section 4 contains our main results, with the explicit bounds for the Kantorovich–Rubinstein distance given in Section 4.1. Finally, Section 5 contains applications of the main results to the analysis of random graphs and information ranking algorithms. All the proofs in the paper are postponed until Section 6.

2. Weighted branching processes

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

Next, let $(Q, N, C_1, C_2, \dots)$ be a real-valued vector with $N \in \mathbb{N} \cup \{\infty\}$. We will refer to this vector as the generic branching vector. Now let $\{(Q_{\boldsymbol{i}}, N_{\boldsymbol{i}}, C_{(\boldsymbol{i},1)}, C_{(\boldsymbol{i},2)}, \dots)\}_{\boldsymbol{i} \in U}$ be a sequence of i.i.d. copies of the generic branching vector. To construct a weighted branching process we start by defining a tree as follows: let $A_0 = \{\varnothing\}$ denote the root of the tree, and

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

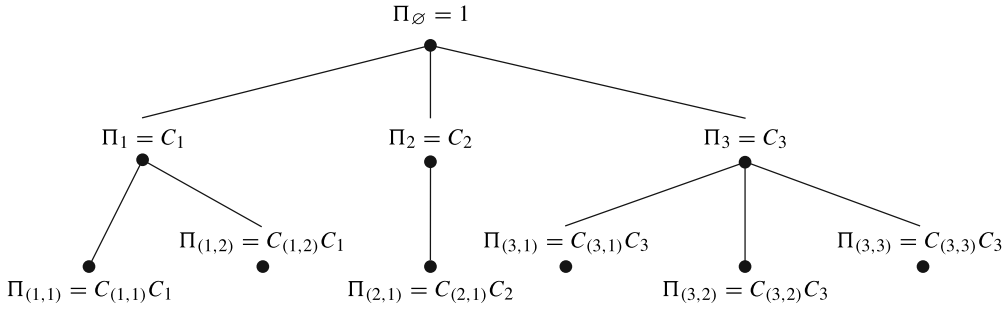


FIGURE 1: Weighted branching process.

define the nth generation according to the recursion

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

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

$$\Pi_\varnothing \equiv 1, \qquad \Pi_{(\boldsymbol{i},i_n)} = C_{(\boldsymbol{i},i_n)}\Pi_{\boldsymbol{i}}, \quad n \ge 1;$$

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

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

We will now explain how to construct the special endogenous solution to the linear SFPE (1) using a weighted branching process.

2.1. Special endogenous solution to the linear SFPE

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

$$W^{(0)} = Q_0, \qquad W^{(j)} = \sum_{\boldsymbol{i}\in A_j} Q_{\boldsymbol{i}}\Pi_{\boldsymbol{i}}, \quad j \ge 1, \tag{2}$$

$$R^{(k)} = \sum_{j=0}^{k} W^{(j)} = \sum_{j=0}^{k}\sum_{\boldsymbol{i}\in A_j} Q_{\boldsymbol{i}}\Pi_{\boldsymbol{i}}, \qquad k \ge 0. \tag{3}$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

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

$$W^{(j)} = \sum_{r=1}^{N_\varnothing} C_r \left(\frac{\sum_{(r,\boldsymbol{i})\in A_j} Q_{(r,\boldsymbol{i})}\Pi_{(r,\boldsymbol{i})}}{C_r} \right) \stackrel{\mathrm{D}}{=} \sum_{r=1}^{N} C_r W_r^{(j-1)}, \qquad j \geq 1 \tag{4}$$

and

$$R^{(k)} = \sum_{r=1}^{N_\varnothing} C_r \left(\frac{\sum_{j=1}^{k} \sum_{(r,\boldsymbol{i})\in A_j} Q_{(r,\boldsymbol{i})}\Pi_{(r,\boldsymbol{i})}}{C_r} \right) + Q_\varnothing \stackrel{\mathrm{D}}{=} \sum_{r=1}^{N} C_r R_r^{(k-1)} + Q, \qquad k \geq 1, \tag{5}$$

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

For the homogeneous case ($Q \equiv 0$ in (1)), assume the weights $\{C_i\}$ are nonnegative and redefine the $\{W^{(j)}\}$ process as

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

In this case, and provided $\rho = \mathbb{E}[\sum_{i=1}^{N} C_i] < \infty$, the process $M^{(j)} = W^{(j)}/\rho^j$, $j \geq 0$, defines a nonnegative martingale. It follows that $M^{(j)}$ converges almost surely (a.s.), as $j \to \infty$, to a finite limit W with $\mathbb{E}[W] \leq 1$. Taking the limit as $j \to \infty$ in (4), it follows that W satisfies

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

where the $\{W_r\}$ are i.i.d. copies of W, independent of $(N, C_1, C_2, \dots)$. Hence, W is a solution to the homogeneous version of (1) with the generic branching vector $(N, C_1', C_2', \dots)$.

For the nonhomogeneous case ($\mathbb{P}(Q \neq 0) > 0$), one can argue, as was done in [15], that provided $\mathbb{E}[\sum_{i=1}^{N} |C_i|^\beta] < 1$ and $\mathbb{E}[|Q|^\beta] < \infty$ for some $0 < \beta \leq 1$, then the random variable $R^{(k)}$ converges a.s. as $k \to \infty$, to a finite limit R. Taking the limit as $k \to \infty$ in (5), it follows that R is a solution to (1). We refer to the random variables W and R described above as the special endogenous solutions to (1) in the homogeneous and nonhomogeneous cases, respectively. We point out that in the case of nonnegative weights, W and R are the unique endogenous solutions to (1), whereas in the real-valued case there can be other endogenous solutions, i.e., that can be explicitly constructed using a weighted branching process; see [13] for more details.

3. Kantorovich–Rubinstein distance

Before proceeding to the main results in this paper we give a brief description of the Kantorovich–Rubinstein distance. This distance on the space of probability measures is also known as the minimal l_1 metric or the Wasserstein distance of order 1. For the purposes of this paper, we consider the vector space of infinite real sequences $\mathbb{R}^\infty$ having finite l_1 norm, i.e., $\boldsymbol{x} \in \mathbb{R}^\infty$ such that

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

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

On some occasions, which will become clear from the context, we will work with elements of $\mathbb{R}^d$ instead. The norm $\|\boldsymbol{x}\|_1$ will always refer to the corresponding l_1 norm.

Definition 1. Let $M(\mu, \nu)$ denote the set of joint probability measures on $\mathcal{S} \times \mathcal{S}$ ($\mathcal{S} = \mathbb{R}^d$ or $\mathbb{R}^\infty$) with marginals μ and ν. Then, the Kantorovich–Rubinstein distance between μ and ν is given by

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

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

$$\mathcal{P}_1(\mathcal{S}) \triangleq \left\{ \mu \in \mathcal{P}(\mathcal{S}) \colon \int_{\mathcal{S}} \|\boldsymbol{x}\|_1 \, \mathrm{d}\mu(\boldsymbol{x}) < \infty \right\},$$

where $\mathcal{P}(\mathcal{S})$ is the set of Borel probability measures on $\mathcal{S}$. We refer the interested reader to [25] for a thorough treatment of this distance, since Definition 1 is only a special case.

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

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

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

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

$$d_1(\mu, \nu) = \mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1]$$

always exists; see, e.g. [25, Theorem 4.1] or [5, Lemma 8.1]. This optimal coupling, nonetheless, is not in general explicitly available. One noteworthy exception is when μ and ν are probability measures on the real line, in which case we have

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

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

Another important property of the Kantorovich–Rubinstein distance is that the convergence in d_1 to a limit $\mu \in \mathcal{P}_1(\mathcal{S})$ is equivalent to weak convergence plus convergence of the first moments. Furthermore, it satisfies the useful *duality formula*

$$d_1(\mu, \nu) = \sup_{\|\psi\|_{\mathrm{Lip}} \leq 1} \left\{ \int_{\mathcal{S}} \psi(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) - \int_{\mathcal{S}} \psi(\boldsymbol{x}) \, \mathrm{d}\nu(\boldsymbol{x}) \right\} \quad \text{for all } \mu, \nu \in \mathcal{P}_1(\mathcal{S}),$$

where the supremum is taken over all Lipschitz continuous functions $\psi \colon \mathcal{S} \to \mathbb{R}$ with Lipschitz constant 1; see [25, Remark 6.5].

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

4. Main results

This paper contains two sets of results; the first one provides explicit bounds for the Kantorovich–Rubinstein distance between two versions of the processes $\{W^{(j)}\colon j \geq 0\}$ (as defined by (2)) constructed on weighted branching processes, respectively weighted branching trees, using different generic branching vectors. These bounds are given in terms of the Kantorovich–Rubinstein distance between the two generic branching vectors. The second set of results apply the explicit bounds to a sequence of processes $\{W^{(n,j)}\colon j \geq 0\}$ and $\{R^{(n,k)}\colon k \geq 0\}$ for $n \geq 1$, in order to obtain the convergence in the Kantorovich–Rubinstein distance to the special endogenous solution to (1) constructed on a limiting weighted branching process.

4.1. Bounds for the Kantorovich–Rubinstein distance

Let $\{W^{(j)}\colon j \geq 0\}$ and $\{\hat{W}^{(j)}\colon j \geq 0\}$ be defined according to (2) on two different weighted branching processes using the generic branching vectors $(Q, N, C_1, C_2, \dots)$ and $(\hat{Q}, \hat{N}, \hat{C}_1, \hat{C}_2, \dots)$, respectively. As our result will show, it is enough to consider generic branching vectors of the form $(Q, B_1, B_2, \dots)$ and $(\hat{Q}, \hat{B}_1, \hat{B}_2, \dots)$, where $B_i = C_i \mathbf{1}_{\{N \geq i\}}$ and $\hat{B}_i = \hat{C}_i \mathbf{1}_{\{\hat{N} \geq i\}}$ for all $i \in \mathbb{N}_+$. Let μ denote the probability measure of $(Q, B_1, B_2, \dots)$ and let $\hat{\mu}$ denote the probability measure of $(\hat{Q}, \hat{B}_1, \hat{B}_2, \dots)$. Using $\mathcal{S} = \mathbb{R}^\infty$, we assume throughout the paper that

$$\begin{aligned} \int_{\mathbb{R}^\infty} \|\boldsymbol{x}\|_1 \, \mathrm{d}\mu(\boldsymbol{x}) &= \mathbb{E}\left[|Q| + \sum_{i=1}^{\infty} |B_i| \right] < \infty, \\ \int_{\mathbb{R}^\infty} \|\boldsymbol{x}\|_1 \, \mathrm{d}\hat{\mu}(\boldsymbol{x}) &= \mathbb{E}\left[|\hat{Q}| + \sum_{i=1}^{\infty} |\hat{B}_i| \right] < \infty. \end{aligned} \tag{6}$$

To construct the two processes on the same probability space, let π denote any coupling of μ and $\hat{\mu}$, and let $\{(Q_{\boldsymbol{i}}, B_{(\boldsymbol{i},1)}, B_{(\boldsymbol{i},2)}, \dots, \hat{Q}_{\boldsymbol{i}}, \hat{B}_{(\boldsymbol{i},1)}, \hat{B}_{(\boldsymbol{i},2)}, \dots)\}_{\boldsymbol{i} \in U}$ be a sequence of i.i.d. random vectors distributed according to π. Then, use the vectors $\{(Q_{\boldsymbol{i}}, B_{(\boldsymbol{i},1)}, B_{(\boldsymbol{i},2)}, \dots)\}_{\boldsymbol{i} \in U}$ to construct $\{W^{(j)}\colon j \geq 0\}$, as described in Section 2, and the vectors $\{(\hat{Q}_{\boldsymbol{i}}, \hat{B}_{(\boldsymbol{i},1)}, \hat{B}_{(\boldsymbol{i},2)}, \dots)\}_{\boldsymbol{i} \in U}$ to construct $\{\hat{W}^{(j)}\colon j \geq 0\}$. Our first result is stated below. We use the convention that $\sum_{i=a}^{b} x_i \equiv 0$ if $b < a$, and the notation $\mathbb{E}_\pi[\cdot]$ to denote the expectation taken with respect to the coupling π; we also use $x \wedge y$ and $x \vee y$ to denote the minimum and the maximum of x and y, respectively, and $x^+ = \max\{0, x\}$.

Proposition 1. *For any coupling π of μ and $\hat{\mu}$, and any $j \geq 0$,*

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

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

We point out that the bound provided by Proposition 1 is also a bound for the Kantorovich–Rubinstein distance between $\hat{W}^{(j)}$ and $W^{(j)}$, and if we take π to be the optimal coupling of μ and $\hat{\mu}$ then we have $\mathcal{E} = d_1(\hat{\mu}, \mu)$. It is also worth mentioning that if we let ν and $\hat{\nu}$ be the probability measures of $(Q, N, C_1, C_2, \dots)$ and $(\hat{Q}, \hat{N}, \hat{C}_1, \hat{C}_2, \dots)$, respectively, and assume that $\mathbb{E}[N + \hat{N}] < \infty$, then $d_1(\mu, \hat{\mu})$ can be small even if $d_1(\nu, \hat{\nu})$ is not. This is due to the observation that, in general, large disagreements between C_r and $\hat{C}_r$ for values of r for which $\mathbb{P}(N > r)$ and $\mathbb{P}(\hat{N} > r)$ are negligible do not affect $d_1(\mu, \hat{\mu})$, whereas they do adversely affect $d_1(\nu, \hat{\nu})$.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

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

Proposition 2. *For any coupling π of ν and $\hat{\nu}$ and any coupling π^* of ν^* and $\hat{\nu}^*$,*

$$\mathbb{E}[|\hat{W}^{(0)} - W^{(0)}|] \le \mathcal{E}^*$$

and for $j \ge 1$,

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

where $\rho = \mathbb{E}[N|C|]$, $\hat{\rho} = \mathbb{E}[\hat{N}|\hat{C}|]$, $\mathcal{E}^* = \mathbb{E}_{\pi^*}[|\hat{Q} - Q| + |\hat{N} - N|]$, and

$$\mathcal{E} = \mathbb{E}_\pi\left[|\hat{C}\hat{Q} - CQ| + \sum_{i=1}^{\infty} |\hat{C}\,\mathbf{1}_{\{\hat{N}\ge i\}} - C\,\mathbf{1}_{\{N\ge i\}}|\right].$$

4.2. Convergence to the special endogenous solution

Our second set of results considers a sequence of weighted branching processes (respectively, weighted branching trees), each constructed using a generic branching vector having probability measure ν_n, $n \ge 1$. In other words, for weighted branching processes, ν_n is the probability measure of a vector of the form $(Q^{(n)}, N^{(n)}, C_1^{(n)}, C_2^{(n)}, \dots)$, while for weighted branching trees it corresponds to a vector of the form $(Q^{(n)}, N^{(n)}, C^{(n)})$. On each of them we define the processes $\{W^{(n,j)} : j \ge 0\}$ and $\{R^{(n,k)} : k \ge 0\}$ according to (2) and (3), and we are interested in providing conditions under which $W^{(n,j)}$ (suitably scaled) and $R^{(n,k)}$ will converge, as n, j, k go to ∞, to the special endogenous solution of a linear SFPE of the form in (1).

The main conditions for the convergence we seek will be in terms of the sequence of probability measures $\{\mu_n\}_{n\ge1}$, where μ_n is the probability measure of the vector

$$(Q^{(n)}, C_1^{(n)}\,\mathbf{1}_{\{N^{(n)}\ge1\}}, C_2^{(n)}\,\mathbf{1}_{\{N^{(n)}\ge2\}}, \dots)$$

for weighted branching processes, and of

$$(C^{(n)}Q^{(n)}, C^{(n)}\,\mathbf{1}_{\{N^{(n)}\ge1\}}, C^{(n)}\,\mathbf{1}_{\{N^{(n)}\ge2\}}, \dots)$$

for weighted branching trees.

In both cases, we assume that there exists a probability measure μ such that $d_1(\mu_n, \mu) \to 0$. We point out that for a weighted branching process, μ is always the probability measure of a generic branching vector, since each of the μ_n is. However, this is not necessarily the case for a weighted branching tree, and we need to further assume that there exist probability measures η_1 on $\mathbb{R}$ and η_2 on $\mathbb{R} \times \{0, 1\}^\infty$ such that

$$\int_{\mathbb{R}^\infty} h(\boldsymbol{x})\mu(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}\times\{0,1\}^\infty}\int_{\mathbb{R}} h(yx_1, yx_2, yx_3, \dots)\eta_1(\mathrm{d}y)\eta_2(\mathrm{d}\boldsymbol{x})$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

for all functions $h\colon \mathbb{R}^\infty \to \mathbb{R}$. We can then identify η_1 with the probability distribution of C and η_2 with the probability distribution of the vector $(Q, \mathbf{1}_{\{N\geq 1\}}, \mathbf{1}_{\{N\geq 2\}}, \dots)$, which fully determines (Q, N). With this interpretation, the limiting measure μ defines a weighted branching process with a generic branching vector of the form $(Q, N, C_1, C_2, \dots)$ with the $\{C_i\}_{i\geq 1}$ i.i.d. and independent of (Q, N); condition (6) implies that $\mathbb{E}[N] < \infty$.

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

Theorem 1. *Define the processes* $\{W^{(n,j)}\colon j \geq 0\}, n \geq 1$, *and* $\{W^{(j)}\colon j \geq 0\}$ *according to (2). Suppose that as* $n \to \infty$,

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

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

$$W^{(n,j)} \xrightarrow{d_1} W^{(j)} \quad \text{as } n \to \infty.$$

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

$$j_n d_1(\mu_n, \mu) \to 0 \quad \textit{(case 1)} \quad \textit{or} \quad d_1(\nu_n^*, \nu^*) + j_n d_1(\mu_n, \mu) \to 0 \quad \textit{(case 2)} \quad \text{as } n \to \infty,$$

we have

$$\frac{W^{(n,j_n)}}{\rho_n^{j_n}} \xrightarrow{\mathrm{D}} \mathcal{W}, \qquad \frac{W^{(n,j_n)}}{\rho^{j_n}} \xrightarrow{\mathrm{D}} \mathcal{W},$$

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

As pointed out in Section 2.1, $\mathcal{W}$ is the unique endogenous solution to the SFPE

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

where the $\{\mathcal{W}_i\}$ are i.i.d. copies of $\mathcal{W}$, independent of $(N, C_1, C_2, \dots)$. See [1], [16], and [18] for conditions on when the random variable $\mathcal{W}$, which satisfies $\mathbb{E}[\mathcal{W}] \leq 1$, is nontrivial, as well as characterizations of its tail behavior. Furthermore, when $\mathbb{E}[\mathcal{W}] = 1$ we can replace the convergence in distribution with convergence in the Kantorovich–Rubinstein distance, i.e.

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

We now give a similar result for the nonhomogeneous equation.

Theorem 2. *Define the processes* $\{R^{(n,k)}\colon k \geq 0\}, n \geq 1$, *and* $\{R^{(k)}\colon k \geq 0\}$ *according to (3). Suppose that as* $n \to \infty$,

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

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

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

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

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

$$R^{(n,k_n)} \xrightarrow{d_1} R \quad \text{as } n \to \infty,$$

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

In the statement of the theorem, provided $\rho < 1$, R is the unique endogenous solution to the SFPE

$$R \overset{\mathrm{D}}{=} \sum_{i=1}^{N} C_i R_i + Q, \tag{7}$$

where the $\{R_i\}$ are i.i.d. copies of R, independent of $(Q, N, C_1, C_2, \dots)$. Moreover, the asymptotic behavior of $\mathbb{P}(R > x)$ as $x \to \infty$ can be described for several different assumptions on the generic vector $(Q, N, C_1, C_2, \dots)$. We refer the reader to [15] and [20] for the precise set of theorems.

Note that in case 1, the convergence of $R^{(n,k)}$ as $k \to \infty$ for a fixed n is guaranteed whenever $\mathbb{E}[\sum_{i=1}^{N^{(n)}} |C_i^{(n)}|^{\beta}] < 1$ for some $0 < \beta \le 1$ (see [15, Lemma 4.1]), and its limit $R^{(n)}$ would be the unique endogenous solution to

$$R^{(n)} \overset{\mathrm{D}}{=} \sum_{i=1}^{N^{(n)}} C_i^{(n)} R_i^{(n)} + Q^{(n)}. \tag{8}$$

For case 2, on the other hand, an adaptation of the proof of [15, Lemma 4.1] would show that $R^{(n,k)}$ converges a.s. to

$$R^{(n)} = \sum_{j=0}^{\infty} W^{(n,j)} \quad \text{as } k \to \infty$$

with $R^{(n)}$ finite a.s. provided $\mathbb{E}[N^{(n)}|C^{(n)}|^{\beta}] < 1$ for some $0 < \beta \le 1$. However, this random variable $R^{(n)}$ would not necessarily have the interpretation of being a solution to (8).

We end this section with a result for the weighted branching tree setting that states that $d_1(\mu_n, \mu)$ converges to 0 whenever $d_1(\nu_n, \nu)$ and the moments of $Q^{(n)}C^{(n)}$ and $N^{(n)}C^{(n)}$ do.

Lemma 1. *For case 2, suppose that as $n \to \infty$, $d_1(\nu_n, \nu) \to 0$, $\mathbb{E}[|C^{(n)}Q^{(n)}|] \to \mathbb{E}[|CQ|]$, and $\mathbb{E}[|C^{(n)}|N^{(n)}] \to \mathbb{E}[|C|N]$. Then*

$$d_1(\mu_n, \mu) \to 0 \quad \text{as } n \to \infty.$$

5. Applications

As mentioned in the introduction, our interest in analyzing the processes $\{\hat{W}^{(j)} : j \ge 0\}$ and $\{\hat{R}^{(k)} : k \ge 0\}$ when they are constructed on a weighted branching tree, rather than a weighted branching process, comes from applications to the analysis of random graphs and information ranking algorithms. In this section we provide two examples in which an application of the main results in Section 4 lead to the special endogenous solution to (1).

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

5.1. Ranking algorithms on a directed configuration network

Our first example is related to the analysis of ranking algorithms on directed graphs. In particular, we are interested in studying the distribution of the ranks produced by spectral ranking algorithms, e.g. Google's Pagerank™, used to rank webpages on the World Wide Web. More precisely, the recent work in [9] considers a sequence of random graphs constructed according to the directed configuration model [8] and shows that the rank of a randomly chosen node can be coupled with a random variable $R^{(n,k_n)}$ (of the form in (3) and built on a weighted branching tree) where the n refers to the number of nodes in the graph. An application of a version of Theorem 2 then leads to the rank of a randomly chosen node having a representation in terms of the special endogenous solution to (1), as defined in Section 2.1, as the number of nodes in the graph grows to ∞.

Before stating the precise version of Theorem 2 that is needed in this application it will be helpful to give some details about the configuration model. The configuration or pairing model (see, e.g. [7] and [23]), produces a random graph from a given degree sequence by assigning to each node a number of half-edges equal to its degree and then randomly pairing the half-edges to form a graph. Similarly, the directed configuration model generates a directed graph from a given bi-degree sequence (sequence of in-degrees and out-degrees). In both cases, the resulting graph, conditional on it not having self-loops or multiple edges, is a graph uniformly chosen at random from all simple graphs having that degree (bi-degree) sequence. We observe that given the degree sequence(s), the randomness in the graph comes from the pairing process, so it makes sense that in applications one often works on the conditional probability space, given the degree sequence(s).

It is due to this last observation that in order to obtain the convergence of the rank of a randomly chosen node on a directed configuration model one needs to apply Theorem 2 conditionally on the sigma-algebra generated by the bi-degree sequence. Other random graph models, e.g. the generalized random graph [23], require conditioning on a 'weight' sequence. Moreover, the analysis of problems related to the configuration model in general (directed or undirected) often relies on a coupling with a weighted branching tree where the root node has a different distribution, hence the need to further tailor the theorem.

In order to state a suitable theorem for the analysis of spectral algorithms on random graphs, we first need to introduce some additional notation. We consider a sequence of sigma-algebras generated by a finite set of random variables/vectors (e.g. the degree sequences). Next, for each $n \geq 1$, and conditionally on $\mathcal{F}_n$, we construct a weighted branching tree using the generic branching vector $(Q^{(n)}, N^{(n)}, C^{(n)})$, whose (conditional) probability measure we denote by ν_n. Moreover, we allow the root branching vector $(Q_\varnothing^{(n)}, N_\varnothing^{(n)})$ to have a different (conditional) distribution, say having a probability measure ν_n^*. In other words, the nth weighted branching tree is constructed, conditionally on $\mathcal{F}_n$, using the sequence of conditionally i.i.d. vectors $\{(Q_{\boldsymbol{i}}^{(n)}, N_{\boldsymbol{i}}^{(n)}, C_{\boldsymbol{i}}^{(n)})\}_{\boldsymbol{i}\in U, \boldsymbol{i}\neq\varnothing}$ distributed according to ν_n, and $(Q_\varnothing^{(n)}, N_\varnothing^{(n)})$ is conditionally independent of this sequence and is distributed according to ν_n^*. Note that unconditionally, the ν_n and ν_n^* are random elements of $\mathcal{P}_1(\mathbb{R}^3)$ and $\mathcal{P}_1(\mathbb{R}^2)$, respectively (e.g., the empirical measures constructed from the degree sequences).

In the statement of Theorem 4 below, we assume that ν and ν^* are the probability measures of the vectors (Q, N, C) and $(Q_\varnothing, N_\varnothing)$, respectively, with C independent of (Q, N), i.e. the limiting weighted branching tree is a delayed weighted branching process. The measures ν and ν^* are fixed elements of $\mathcal{P}_1(\mathbb{R}^3)$ and $\mathcal{P}_1(\mathbb{R}^2)$, respectively, and hence are independent of $\mathcal{F}_n$. The symbol $\xrightarrow{\mathbb{P}}$ denotes convergence in probability.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

Theorem 3. *Conditionally on $\mathcal{F}_n$, define the processes $\{R^{(n,k)}: k \geq 0\}, n \geq 1$, according to (3). Similarly, define $\{R^{(k)}: k \geq 0\}$. Suppose that as $n \to \infty$,*

$$d_1(\nu_n^*, \nu^*) + d_1(\mu_n, \mu) \xrightarrow{\mathbb{P}} 0.$$

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

$$R^{(n,k)} \xrightarrow{\mathrm{D}} R^{(k)} \quad \text{as } n \to \infty.$$

Moreover, if $\rho = \mathbb{E}[N]\mathbb{E}[|C|] < 1$ then for any $k_n \in \mathbb{N}$ such that $k_n \to \infty$ as $n \to \infty$, we have

$$R^{(n,k_n)} \xrightarrow{\mathrm{D}} \mathcal{R} \quad \text{as } n \to \infty,$$

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

Remark 1. (i) Because we allow ν^* to be different from ν, the random variable $\mathcal{R}$ appearing in the limit can be written as

$$\mathcal{R} = \sum_{i=1}^{N_\varnothing} C_i R_i + Q_\varnothing,$$

where the $\{R_i\}$ are i.i.d. copies of the special endogenous solution R to (7), independent of $(Q_\varnothing, N_\varnothing, C_1, C_2, \dots)$, and with the $\{C_i\}$ i.i.d. and independent of $(Q_\varnothing, N_\varnothing)$. In other words, $\mathcal{R}$ is a linear combination of i.i.d. copies of the special endogenous solution to (7).

(ii) No restrictions need to be imposed on ν^* besides $\int \|\boldsymbol{x}\|_1 \, \mathrm{d}\nu^*(\boldsymbol{x}) < \infty$, since $C_\varnothing$ does not appear in the definitions of $W^{(j)}$ and $R^{(k)}$.

(iii) An important observation is that we have replaced the convergence in the Kantorovich–Rubinstein distance in Theorem 2 with weak convergence, this is due to the fact that the proof of Theorem 3 requires that we apply Proposition 2 conditionally on $\mathcal{F}_n$, which if done directly starting with $\mathbb{E}[|W^{(n,j)} - W^{(j)}|]$ would lead to having to compute all the moments of $\rho_n = \mathbb{E}[N^{(n)}|C^{(n)}| \mid \mathcal{F}_n]$, which in general may not even be finite. Therefore, Proposition 2 needs to be used after having guaranteed that ρ_n is sufficiently close to ρ, hence the weaker mode of convergence.

(iv) The previous remark also implies that if $\rho_n \leq c < 1$ a.s. then the weak convergence in Theorem 3 can be replaced by convergence in the Kantorovich–Rubinstein sense.

5.2. Analyzing the configuration model

Our second example is also related to the analysis of random graphs. As mentioned earlier, when analyzing the properties of random graphs, e.g. connectivity, existence of a giant component, typical distances, or phase transitions, one of the basic techniques consists in coupling a 'graph exploration process' with a branching process. This is true for the configuration model (directed or undirected) as well as for other random graph models such as the Erdős–Rényi graph or the generalized random graph. In line with our previous example for the analysis of ranking algorithms, we show an application of Theorem 1 that can be used for analyzing the properties of the configuration model.

Consider an undirected graph with n nodes generated according to the configuration model. In [24] the authors provide an asymptotic characterization, as $n \to \infty$, of the hop count (length of the shortest path) between two randomly chosen nodes in the graph. In particular, they show that conditional on the two nodes belonging to the same component, this distance

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

grows logarithmically in n. The main step of the proof consists in coupling a breadth-first exploration process, where starting from a randomly chosen node we sequentially uncover all nodes at distance 1, then those at distance 2, etc. with a delayed branching process (Galton–Watson process). This delayed branching process, as in our previous example, is constructed conditionally on the sigma-algebra $\mathcal{F}_n$ generated by the degree sequence.

If we denote by $\{Z^{(n,j)}\colon j \geq 0\}$ the number of individuals in the jth generation of the coupled branching process (obtained by setting $Q_{\boldsymbol{i}}^{(n)} \equiv C_{\boldsymbol{i}}^{(n)} \equiv 1$ for all $\boldsymbol{i} \in U$ in (2)) then the goal is to show that $Z^{(n,j)}/(\mathbb{E}[N^{(n)} \mid \mathcal{F}_n])^j$ converges to a limit as $n, j \to \infty$. The following version of Theorem 1 provides such a limit; here ν_n denotes the random probability measure of the conditionally i.i.d. random variables $\{N_{\boldsymbol{i}}^{(n)}\}_{\boldsymbol{i}\in U, \boldsymbol{i}\neq\varnothing}$ and ν_n^* denotes that of $N_{\varnothing}^{(n)}$.

Theorem 4. *Suppose there exist probability measures ν and ν^* such that*

$$d_1(\nu_n, \nu) \xrightarrow{\mathbb{P}} 0, \qquad d_1(\nu_n^*, \nu^*) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty.$$

Let $\{Z^{(n,j)}\colon j \geq 0\}, n \geq 1$, be the (delayed) branching process defined by the sequence $\{N_{\boldsymbol{i}}^{(n)}\}_{\boldsymbol{i}\in U}$, and let $\{Z^{(j)}\colon j \geq 0\}$ be the one defined using $\{N_{\boldsymbol{i}}\}_{\boldsymbol{i}\in U}$. Then, there exists a coupling of $\{Z^{(n,j)}\colon j \geq 0\}$ and $\{Z^{(j)}\colon j \geq 0\}$ such that for any $j_n \in \mathbb{N}$ satisfying

$$j_n\, d_1(\nu_n, \nu) \xrightarrow{\mathbb{P}} 0,$$

we have

$$\max_{1\leq j\leq j_n} \left| \frac{Z^{(n,j)}}{m_n^* m_n^{j-1}} - \frac{Z^{(j)}}{m^* m^{j-1}} \right| \xrightarrow{\mathbb{P}} 0, \qquad \max_{1\leq j\leq j_n} \frac{|Z^{(n,j)} - Z^{(j)}|}{m^{j-1}} \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty,$$

where $m_n = \mathbb{E}[N^{(n)} \mid \mathcal{F}_n]$, $m_n^* = \mathbb{E}[N_{\varnothing}^{(n)} \mid \mathcal{F}_n]$, $m = \mathbb{E}[N]$, and $m^* = \mathbb{E}[N_{\varnothing}]$.

In particular, Theorem 4 can be used to obtain

$$\frac{Z^{(n,j_n)}}{m_n^* m_n^{j_n-1}} \xrightarrow{\mathrm{D}} \mathcal{W} \quad \text{as } n \to \infty,$$

where $\mathcal{W}$ is the a.s. limit of the martingale $Z^{(j)}/(m^* m^{j-1})$ as $j \to \infty$.

Our last result in this paper shows how large we can take j_n in Theorem 4 when analyzing the configuration model using a degree sequence $\{D_1, D_2, \ldots, D_n\}$ of i.i.d. random variables having common probability mass function f, a model used in [24] to analyze the typical distance between nodes in the graph. In this context, $\mathcal{F}_n = \sigma(D_1, D_2, \ldots, D_n)$,

$$\nu_n^*(\{k\}) = \mathbb{P}(N_{\varnothing}^{(n)} = k \mid \mathcal{F}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{D_i=k\}}, \qquad k = 0, 1, 2, \ldots,$$

and

$$\nu_n(\{k\}) = \mathbb{P}(N_1^{(n)} = k \mid \mathcal{F}_n) = \sum_{i=1}^{n} \frac{D_i}{L_n} \mathbf{1}_{\{D_i=k+1\}}, \qquad k = 0, 1, 2, \ldots$$

with $L_n = \sum_{j=1}^{n} D_j$. The measure ν_n corresponds to the so-called *size-biased* empirical distribution. The limiting probability measures are given by

$$\nu^*(\{k\}) = f(k), \qquad \nu(\{k\}) = \frac{\mathbb{E}[D\,\mathbf{1}_{\{D=k+1\}}]}{\mathbb{E}[D]}, \quad k = 0, 1, 2, \ldots,$$

where D is distributed according to f.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

Lemma 2. *Suppose that* $\mathbb{E}[D^{2+\epsilon}] < \infty$ *for some* $\epsilon > 0$*, then*

$$n^{\delta'} d_1(\nu_n^*, \nu^*) \xrightarrow{\mathbb{P}} 0, \qquad n^{\delta} d_1(\nu_n, \nu) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty$$

for any $0 < \delta' < \frac{1}{2}$ *and* $0 < \delta < \min\{\frac{1}{2}, \epsilon/(2+\epsilon)\}$.

6. Proofs

This last section contains the proofs of all the results presented throughout this paper. For the reader's convenience they are organized according to the section in which their statements appear.

6.1. Bounds for the Kantorovich–Rubinstein distance

We first prove Proposition 1, which bounds the Kantorovich–Rubinstein distance of the linear processes on two coupled weighted branching processes by the distance between their generic branching vectors.

Proof of Proposition 1. Define $\mathcal{E} = \mathbb{E}_\pi[|\hat{Q} - Q| + \sum_{i=1}^{\infty} |\hat{B}_i - B_i|]$, where the vector $(Q, B_1, B_2, \dots, \hat{Q}, \hat{B}_1, \hat{B}_2, \dots)$ is distributed according to π. Recall that the weights $\Pi_{\boldsymbol{i}}$ and $\hat{\Pi}_{\boldsymbol{i}}$ follow the recursions

$$\Pi_{(\boldsymbol{i},j)} = \Pi_{\boldsymbol{i}} B_{(\boldsymbol{i},j)}, \qquad \hat{\Pi}_{(\boldsymbol{i},j)} = \hat{\Pi}_{\boldsymbol{i}} \hat{B}_{(\boldsymbol{i},j)}$$

with $\Pi_\varnothing = \hat{\Pi}_\varnothing = 1$. Now note that for $j = 0$, we have

$$\mathbb{E}[|\hat{W}^{(0)} - W^{(0)}|] = \mathbb{E}[|\hat{Q} - Q|] \le \mathcal{E}.$$

To analyze the expression for $j \ge 1$, define for $r \ge 1$,

$$W_r^{(j-1)} = \frac{\sum_{(r,\boldsymbol{i}) \in \mathbb{N}_+^j} Q_{(r,\boldsymbol{i})} \Pi_{(r,\boldsymbol{i})}}{B_r}, \qquad \hat{W}_r^{(j-1)} = \frac{\sum_{(r,\boldsymbol{i}) \in \mathbb{N}_+^j} \hat{Q}_{(j,\boldsymbol{i})} \hat{\Pi}_{(r,\boldsymbol{i})}}{\hat{B}_r}.$$

We then have

$$\hat{W}^{(j)} = \sum_{r=1}^{\infty} \hat{B}_r \hat{W}_r^{(j-1)}, \qquad W^{(j)} = \sum_{r=1}^{\infty} B_r W_r^{(j-1)}.$$

Next, note that

$$\begin{aligned}
\mathbb{E}[|\hat{W}^{(j)} - W^{(j)}|] &\le \sum_{r=1}^{\infty} \mathbb{E}[|\hat{B}_r \hat{W}_r^{(j-1)} - B_r W_r^{(j-1)}|] \\
&\le \sum_{r=1}^{\infty} \{\mathbb{E}[|W_r^{(j-1)}(\hat{B}_r - B_r)| + |\hat{B}_r(\hat{W}_r^{(j-1)} - W_r^{(j-1)})|]\} \\
&\le \sum_{r=1}^{\infty} \mathbb{E}[|\hat{B}_r - B_r|]\mathbb{E}[|W_r^{(j)}|] + \sum_{r=1}^{\infty} \mathbb{E}[|\hat{B}_r|]\mathbb{E}[|\hat{W}_r^{(j)} - W_r^{(j)}|] \\
&\le \mathbb{E}[|W^{(j-1)}|]\mathcal{E} + \hat{\rho}\mathbb{E}[|\hat{W}^{(j-1)} - W^{(j-1)}|],
\end{aligned}$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

where we used the independence of the root vectors and their offspring, the observation that the random variables $\{W_r^{(j-1)}\}_{r\geq 1}$ are i.i.d. with the same distribution as $W^{(j-1)}$, and $\{(\hat{W}_r^{(j-1)} - W_r^{(j-1)})\}_{r\geq 1}$ are i.i.d. with the same distribution as $\hat{W}^{(j-1)} - W^{(j-1)}$. Moreover,

$$\mathbb{E}[|W^{(j-1)}|] \leq \mathbb{E}[|Q|] \sum_{\boldsymbol{i}\in\mathbb{N}_+^{j-1}} \mathbb{E}[|\Pi_{\boldsymbol{i}}|] = \mathbb{E}[|Q|]\rho^{j-1}.$$

It follows that

$$\begin{aligned}\mathbb{E}[|\hat{W}^{(j)} - W^{(j)}|] &\leq \mathbb{E}[|Q|]\rho^{j-1}\mathcal{E} + \hat{\rho}\mathbb{E}[|\hat{W}^{(j-1)} - W^{(j-1)}|] \quad \text{(after } (j-1) \text{ iterations)}\\ &\leq \left(\hat{\rho}^j + \mathbb{E}[|Q|]\sum_{t=0}^{j-1}\rho^t\hat{\rho}^{j-1-t}\right)\mathcal{E}.\end{aligned}$$

This completes the proof. □

Similarly we can prove an upper bound for weighted branching trees.

Proof of Proposition 2. We construct the processes $\hat{W}^{(j)}$ and $W^{(j)}$ on two weighted branching trees using a coupled vector $(Q_\varnothing, N_\varnothing, \hat{Q}_\varnothing, \hat{N}_\varnothing)$ for the root nodes $\varnothing$, distributed according to π^*, and a sequence of i.i.d. random vectors $\{(Q_{\boldsymbol{i}}, N_{\boldsymbol{i}}, C_{\boldsymbol{i}}, \hat{Q}_{\boldsymbol{i}}, \hat{N}_{\boldsymbol{i}}, \hat{C}_{\boldsymbol{i}}))\}_{\boldsymbol{i}\in U, \boldsymbol{i}\neq\varnothing}$, independent of $(Q_\varnothing, N_\varnothing, \hat{Q}_\varnothing, \hat{N}_\varnothing)$, distributed according to π for all other nodes.

Next, for $\boldsymbol{i} \in \mathbb{N}_+^k$, $k \geq 1$, let $B_{\boldsymbol{i}}^{(0)} = C_{\boldsymbol{i}}Q_{\boldsymbol{i}}$, $\hat{B}_{\boldsymbol{i}}^{(0)} = \hat{C}_{\boldsymbol{i}}\hat{Q}_{\boldsymbol{i}}$, $B_{\boldsymbol{i}}^{(j)} = C_{\boldsymbol{i}}\mathbf{1}_{\{N_{\boldsymbol{i}}\geq i\}}$, and $\hat{B}_{\boldsymbol{i}}^{(j)} = \hat{C}_{\boldsymbol{i}}\mathbf{1}_{\{\hat{N}_{\boldsymbol{i}}\geq i\}}$ for $j \geq 1$, and note that

$$\Pi_{\boldsymbol{i}}Q_{\boldsymbol{i}} = Q_{\boldsymbol{i}}\prod_{r=1}^{k} C_{\boldsymbol{i}|r}\mathbf{1}_{\{i_r\leq N_{\boldsymbol{i}|r-1}\}} = \mathbf{1}_{\{i_1\leq N_\varnothing\}}\prod_{r=1}^{k-1} B_{\boldsymbol{i}|r}^{(i_{r+1})}B_{\boldsymbol{i}}^{(0)}.$$

Similarly,

$$\hat{\Pi}_{\boldsymbol{i}}\hat{Q}_{\boldsymbol{i}} = \mathbf{1}_{\{i_1\leq \hat{N}_\varnothing\}}\prod_{r=1}^{k-1}\hat{B}_{\boldsymbol{i}|r}^{(i_{r+1})}\hat{B}_{\boldsymbol{i}}^{(0)},$$

with the convention that $\prod_{i=a}^{b} x_i \equiv 1$ if $b < a$.

Let

$$\mathcal{E}^* = \mathbb{E}_{\pi^*}[|\hat{Q} - Q| + |\hat{N} - N|],$$

where $(Q, N, \hat{Q}, \hat{N})$ is distributed according to π^*, and

$$\mathcal{E} = \mathbb{E}_\pi[\sum_{i=0}^{\infty}|\hat{B}^{(i)} - B^{(i)}|],$$

where $(B^{(0)}, B^{(1)}, B^{(2)}, \ldots, \hat{B}^{(0)}, \hat{B}^{(1)}, \hat{B}^{(2)}, \ldots)$ is distributed according to π. It follows that

$$\mathbb{E}[|\hat{W}^{(0)} - W^{(0)}|] = E_{\pi^*}[|\hat{Q} - Q|] \leq \mathcal{E}^*,$$

and for $j \geq 1$,

$$\begin{aligned}&\mathbb{E}[|\hat{W}^{(j)} - W^{(j)}|]\\ &\quad= \mathbb{E}\left[\left|\sum_{\boldsymbol{i}\in\mathbb{N}_+^j}\mathbf{1}_{\{i_1\leq\hat{N}_\varnothing\}}\prod_{r=1}^{j-1}\hat{B}_{\boldsymbol{i}|r}^{(i_{r+1})}\hat{B}_{\boldsymbol{i}}^{(0)} - \sum_{\boldsymbol{i}\in\mathbb{N}_+^j}\mathbf{1}_{\{i_1\leq N_\varnothing\}}\prod_{r=1}^{j-1}B_{\boldsymbol{i}|r}^{(i_{r+1})}B_{\boldsymbol{i}}^{(0)}\right|\right]\end{aligned}$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

$$
\begin{aligned}
&\le \sum_{\boldsymbol{i}\in\mathbb{N}_+^j} \mathbb{E}\Bigg[\Bigg|\mathbf{1}_{\{i_1\le \hat{N}_\varnothing\}} \prod_{r=1}^{j-1} \hat{B}^{(i_{r+1})}_{\boldsymbol{i}\,|\,r} (\hat{B}^{(0)}_{\boldsymbol{i}} - B^{(0)}_{\boldsymbol{i}})\Bigg|\Bigg] \\
&\quad + \sum_{\boldsymbol{i}\in\mathbb{N}_+^j} \mathbb{E}\Bigg[\Bigg|(\mathbf{1}_{\{i_1\le \hat{N}_\varnothing\}} - \mathbf{1}_{\{i_1\le N_\varnothing\}}) \prod_{r=1}^{j-1} \hat{B}^{(i_{r+1})}_{\boldsymbol{i}\,|\,r} B^{(0)}_{\boldsymbol{i}}\Bigg|\Bigg] \\
&\quad + \sum_{\boldsymbol{i}\in\mathbb{N}_+^j} \mathbb{E}\Bigg[\Bigg|\mathbf{1}_{\{i_1\le N_\varnothing\}} \Bigg(\prod_{r=1}^{j-1} \hat{B}^{(i_{r+1})}_{\boldsymbol{i}\,|\,r} - \prod_{r=1}^{j-1} B^{(i_{r+1})}_{\boldsymbol{i}\,|\,r}\Bigg) B^{(0)}_{\boldsymbol{i}}\Bigg|\Bigg] \\
&= \sum_{\boldsymbol{i}\in\mathbb{N}_+^j} \mathbb{P}(\hat{N}\ge i_1) \prod_{r=1}^{j-1} \mathbb{E}[|\hat{C}|\,\mathbf{1}_{\{\hat{N}\ge i_{r+1}\}}] \mathbb{E}_\pi[|\hat{B}^{(0)} - B^{(0)}|] \\
&\quad + \sum_{\boldsymbol{i}\in\mathbb{N}_+^j} \mathbb{E}_{\pi^*}[|\,\mathbf{1}_{\{i_1\le \hat{N}\}} - \mathbf{1}_{\{i_1\le N\}}\,|] \prod_{r=1}^{j-1} \mathbb{E}[|\hat{C}|\,\mathbf{1}_{\{\hat{N}\ge i_{r+1}\}}] \mathbb{E}[|CQ|] \\
&\quad + \sum_{\boldsymbol{i}\in\mathbb{N}_+^j} \mathbb{P}(N\ge i_1) \mathbb{E}\Bigg[\Bigg|\prod_{r=1}^{j-1} \hat{B}^{(i_{r+1})}_{\boldsymbol{i}\,|\,r} - \prod_{r=1}^{j-1} B^{(i_{r+1})}_{\boldsymbol{i}\,|\,r}\Bigg|\Bigg] \mathbb{E}[|CQ|],
\end{aligned}
$$

where we have used the independence among the generic branching vectors of the weighted branching trees. Moreover,

$$
\sum_{\boldsymbol{i}\in\mathbb{N}_+^j} \mathbb{P}(\hat{N}\ge i_1) \prod_{r=1}^{j-1} \mathbb{E}[|\hat{C}|\,\mathbf{1}_{\{\hat{N}\ge i_{r+1}\}}] = \sum_{i=1}^{\infty} \mathbb{P}(\hat{N}\ge i) \Bigg(\sum_{k=1}^{\infty} \mathbb{E}[|\hat{C}|\,\mathbf{1}_{\{\hat{N}\ge k\}}]\Bigg)^{j-1} = \mathbb{E}[\hat{N}]\hat{\rho}^{j-1},
$$

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

$$
\sum_{i=1}^{\infty} \mathbb{E}_{\pi^*}[|\,\mathbf{1}_{\{i\le \hat{N}\}} - \mathbf{1}_{\{i\le N\}}\,|] = \sum_{i=1}^{\infty} \mathbb{E}_{\pi^*}[\mathbf{1}_{\{N<i\le \hat{N}\}} + \mathbf{1}_{\{\hat{N}<i\le N\}}] = \mathbb{E}_{\pi^*}[|\hat{N} - N|].
$$

It follows that

$$
\begin{aligned}
&\mathbb{E}[|\hat{W}^{(j)} - W^{(j)}|] \\
&\quad \le \mathbb{E}[\hat{N}]\hat{\rho}^{j-1} \mathbb{E}_\pi[|\hat{B}^{(0)} - B^{(0)}|] + \mathbb{E}[|CQ|]\hat{\rho}^{j-1} \mathbb{E}_{\pi^*}[|\hat{N} - N|] \\
&\qquad + \mathbb{E}[N]\mathbb{E}[|CQ|] \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} \mathbb{E}\Bigg[\Bigg|\prod_{r=1}^{j-1} \hat{B}^{(i_{r+1})}_{(1,i_2,\ldots,i_r)} - \prod_{r=1}^{j-1} B^{(i_{r+1})}_{(1,i_2,\ldots,i_r)}\Bigg|\Bigg].
\end{aligned}
$$

Now define $a_j = \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} \mathbb{E}[|\prod_{r=1}^{j-1} \hat{B}^{(i_{r+1})}_{(1,i_2,\ldots,i_r)} - \prod_{r=1}^{j-1} B^{(i_{r+1})}_{(1,i_2,\ldots,i_r)}|]$ for $j \ge 2$,

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

and $a_1 = 0$. It follows that, for $j \geq 2$,

$$\begin{aligned} a_j &\leq \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} \mathbb{E}\left[\left|\prod_{r=1}^{j-2} \hat{B}^{(i_{r+1})}_{(1,i_2,\ldots,i_r)} - \prod_{r=1}^{j-2} B^{(i_{r+1})}_{(1,i_2,\ldots,i_r)}\right| |\hat{B}^{(i_j)}_{(1,i_2,\ldots,i_{j-1})}|\right] \\ &\quad + \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} \mathbb{E}\left[\left|\prod_{r=1}^{j-2} B^{(i_{r+1})}_{(1,i_2,\ldots,i_r)}\right| |\hat{B}^{(i_j)}_{(1,i_2,\ldots,i_{j-1})} - B^{(i_j)}_{(1,i_2,\ldots,i_{j-1})}|\right] \\ &= a_{j-1} \sum_{i_j=1}^{\infty} \mathbb{E}[|\hat{C}|\,\mathbf{1}_{\{\hat{N}\geq i_j\}}] \\ &\quad + \sum_{(i_2,i_3,\ldots,i_j)\in\mathbb{N}_+^{j-1}} \prod_{r=1}^{j-2} \mathbb{E}[|C|\,\mathbf{1}_{\{N\geq i_{r+1}\}}]\mathbb{E}_\pi[|\hat{B}^{(i_j)} - B^{(i_j)}|] \\ &= \hat{\rho} a_{j-1} + \rho^{j-2} \sum_{i=1}^{\infty} \mathbb{E}_\pi[|\hat{B}^{(i)} - B^{(i)}|], \end{aligned}$$

where $\rho = \mathbb{E}[N|C|]$. Iterating this recursion $j-2$ times, we obtain

$$a_j \leq \hat{\rho}^{j-1} a_1 + \sum_{t=0}^{j-2} \hat{\rho}^t \rho^{j-2-t} \sum_{i=1}^{\infty} \mathbb{E}_\pi[|\hat{B}^{(i)} - B^{(i)}|] = \sum_{t=0}^{j-2} \hat{\rho}^t \rho^{j-2-t} \sum_{i=1}^{\infty} \mathbb{E}_\pi[|\hat{B}^{(i)} - B^{(i)}|].$$

We conclude that, for $j \geq 1$,

$$\begin{aligned} \mathbb{E}[|\hat{W}^{(j)} - W^{(j)}|] &\leq \mathbb{E}[\hat{N}]\hat{\rho}^{j-1}\mathbb{E}_\pi[|\hat{B}^{(0)} - B^{(0)}|] + \mathbb{E}[|CQ|]\hat{\rho}^{j-1}E_{\pi^*}[|\hat{N} - N|] \\ &\quad + \mathbf{1}_{\{j\geq 2\}}\,\mathbb{E}[N]\mathbb{E}[|CQ|] \sum_{t=0}^{j-2} \hat{\rho}^t \rho^{j-2-t} \sum_{i=1}^{\infty} \mathbb{E}_\pi[|\hat{B}^{(i)} - B^{(i)}|] \\ &\leq \left(\mathbb{E}[\hat{N}] \vee \frac{\mathbb{E}[N]\mathbb{E}[|CQ|]}{\rho}\right)\left(\sum_{t=0}^{j-1} \hat{\rho}^t \rho^{j-1-t}\right)\mathcal{E} + \mathbb{E}[|CQ|]\hat{\rho}^{j-1}\mathcal{E}^*. \qquad \square \end{aligned}$$

6.2. Convergence to the special endogenous solution

We now proceed to prove the two main theorems of the paper, Theorems 1 and 2.

Proof of Theorem 1. We split the proof into two parts.

Case 1: weighted branching processes. Choose a coupling π of μ_n and μ such that $\mathbb{E}_\pi[|Q - Q^{(n)}| + \sum_{j=1}^{\infty} |B_j - B_j^{(n)}|] = d_1(\mu, \mu_n)$. If we construct both weighted branching processes based on this optimal coupling, then by Proposition 1,

$$\begin{aligned} \mathbb{E}[|W^{(n,j)} - W^{(j)}|] &\leq \left(\rho_n^j + \mathbb{E}[|Q|]\sum_{t=0}^{j-1} \rho^t \rho_n^{j-1-t}\right) d_1(\mu, \mu_n) \\ &\leq (\mathbb{E}[|Q| \vee \rho)(j+1)\rho^{j-1}\left(1 \vee \frac{\rho_n}{\rho}\right)^j d_1(\mu, \mu_n). \end{aligned}$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

For fixed $j \geq 1$ note that $|\rho_n - \rho| \leq d_1(\mu, \mu_n)$, and hence $(1 \vee (\rho_n/\rho))^j \to 1$ as $n \to \infty$, which in turn implies that $\mathbb{E}[|W^{(n,j)} - W^{(j)}|] \to 0$.

Assume now $Q^{(n)} = Q \equiv 1$, and $\{C_j^{(n)}, C_j\}$ are nonnegative for all n, j; suppose that $j_n \to \infty$ and $j_n d_1(\mu, \mu_n) \to 0$ as $n \to \infty$. First, note that $\{W^{(j)}/\rho^j\}$ is a mean 1 nonnegative martingale with respect to the filtration generated by $\mathcal{G}_j = \sigma((B_{(\boldsymbol{i},1)}, B_{(\boldsymbol{i},2)}, \dots) \colon \boldsymbol{i} \in \mathbb{N}_+^r, 0 \leq r < j)$, $\mathcal{G}_0 = \sigma(\varnothing)$. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}}\right|\right] \\
&\quad\leq \mathbb{E}\left[\frac{1}{\rho_n^{j_n}}|W^{(n,j_n)} - W^{(j_n)}|\right] + \mathbb{E}\left[\frac{W^{(j_n)}}{\rho^{j_n}}\left|\left(\frac{\rho}{\rho_n}\right)^{j_n} - 1\right|\right] \\
&\quad\leq \frac{(1 \vee \rho)}{\rho}(j_n + 1)\left(\frac{\rho}{\rho_n}\right)^{j_n}\left(1 \vee \frac{\rho_n}{\rho}\right)^{j_n} d_1(\mu, \mu_n) + \left|\left(\frac{\rho}{\rho_n}\right)^{j_n} - 1\right| \\
&\quad\leq \frac{(1 \vee \rho)}{\rho}(j_n + 1)\mathrm{e}^{j_n(\rho/\rho_n - 1)^+} d_1(\mu, \mu_n) + j_n\left|\frac{\rho}{\rho_n} - 1\right|\mathrm{e}^{(j_n-1)(\rho/\rho_n-1)^+},
\end{aligned}$$

where in the last step we used the inequalities

$$(x \vee 1)^j \leq \mathrm{e}^{j(x-1)^+}, \qquad |x^j - 1| \leq j|x - 1|\mathrm{e}^{(j-1)(x-1)^+} \quad \text{for all } x > 0,\ j \in \mathbb{N}. \tag{9}$$

Since $j_n d_1(\mu, \mu_n) \to 0$ as $n \to \infty$, then so does $j_n|\rho/\rho_n - 1| \to 0$ as $n \to \infty$, and we conclude that the expected value converges to 0. Since by the martingale convergence theorem $W^{(j_n)}/\rho^{j_n} \to \mathcal{W}$ a.s., then

$$\frac{W^{(n,j_n)}}{\rho_n^{j_n}} \xrightarrow{\mathrm{D}} \mathcal{W} \quad \text{as } n \to \infty.$$

If $\mathbb{E}[\mathcal{W}] = 1$ then $\mathbb{E}[|W^{(j_n)}/\rho^{j_n} - \mathcal{W}|] \to 0$ and we can replace the convergence in distribution to convergence in the Kantorovich–Rubinstein distance.

The last statement of the theorem for weighted branching processes follows from noting that

$$\begin{aligned}
\frac{1}{\rho^j}\mathbb{E}[|W^{(n,j)} - W^{(j)}|] &\leq \left|\frac{1}{\rho^j} - \frac{1}{\rho_n^j}\right|\mathbb{E}[|W^{(n,j)}|] + \mathbb{E}\left[\left|\frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}}\right|\right] \\
&= \left|\left(\frac{\rho}{\rho_n}\right)^j - 1\right| + \mathbb{E}\left[\left|\frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}}\right|\right],
\end{aligned}$$

which were already shown to converge to 0 for all $0 \leq j \leq j_n$. This completes the proof of case 1.

Case 2: weighted branching trees. Construct versions of the processes $\{W^{(n,j)} \colon j \geq 0\}$ and $\{W^{(j)} \colon j \geq 0\}$ using a sequence of coupled vectors $\{(Q_{\boldsymbol{i}}^{(n)}, N_{\boldsymbol{i}}^{(n)}, C_{\boldsymbol{i}}^{(n)}, Q_{\boldsymbol{i}}, N_{\boldsymbol{i}}, C_{\boldsymbol{i}})\}_{\boldsymbol{i} \in U, \boldsymbol{i} \neq \varnothing}$ according to the coupling π satisfying

$$d_1(\mu_n, \mu) = \mathbb{E}_\pi\left[|C^{(n)}Q^{(n)} - CQ| + \sum_{i=1}^{\infty}|C^{(n)}\mathbf{1}_{\{N^{(n)} \geq i\}} - C\mathbf{1}_{\{N \geq i\}}|\right].$$

Let the root vector $(Q_\varnothing^{(n)}, N_\varnothing^{(n)}, Q_\varnothing, N_\varnothing)$ be distributed according to π^*, where $d_1(\nu_n^*, \nu^*) = \mathbb{E}_{\pi^*}[|Q^{(n)} - Q| + |N^{(n)} - N|]$, and be independent of all other nodes.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

By Proposition 2, we have $\mathbb{E}[|W^{(n,0)} - W^{(0)}|] \le d_1(\nu_n^*, \nu^*)$ and

$$\mathbb{E}[|W^{(n,j)} - W^{(j)}|] \le Kj(\rho_n \vee \rho)^{j-1} d_1(\mu_n, \mu) + K\rho_n^{j-1} d_1(\nu_n^*, \nu^*), \qquad j \ge 1$$

with $K = \max\{\mathbb{E}[N^{(n)}], \mathbb{E}[|Q|]\}$. Note that

$$|\rho_n - \rho| = \left| \sum_{i=1}^{\infty} \mathbb{E}[C^{(n)} \mathbf{1}_{\{N^{(n)} \ge i\}} - C \mathbf{1}_{\{N \ge i\}}] \right| \le d_1(\mu_n, \mu).$$

The result for fixed j follows immediately.

Assume now that $Q^{(n)} = Q = 1$ and $\{C^{(n)}, C\}$ are nonnegative, and recall that C is independent of (Q, N), and therefore μ defines a weighted branching process. This in turn implies that $\{W^{(j)}/\rho^j\}$ is a nonnegative martingale with respect to the filtration generated by $\mathcal{H}_j = \sigma((N_{\boldsymbol{i}}, C_{(\boldsymbol{i},1)}, \dots, C_{(\boldsymbol{i},N_{\boldsymbol{i}})}) \colon \boldsymbol{i} \in A_r, 0 \le r < j)$, $\mathcal{H}_0 = \sigma(\varnothing)$. It follows that

$$\begin{aligned}
&\mathbb{E}\left[\left| \frac{W^{(n,j_n)}}{\rho_n^{j_n}} - \frac{W^{(j_n)}}{\rho^{j_n}} \right|\right] \\
&\quad \le \mathbb{E}[|W^{(n,j_n)} - W^{(j_n)}|] \frac{1}{\rho_n^{j_n}} + \left| \left(\frac{\rho}{\rho_n}\right)^{j_n} - 1 \right| \\
&\quad \le K j_n \left(\frac{\rho \vee \rho_n}{\rho_n} \right)^{j_n} d_1(\mu_n, \mu) + \frac{K}{\rho_n} d_1(\nu_n^*, \nu^*) + \left| \left(\frac{\rho}{\rho_n}\right)^{j_n} - 1 \right| \\
&\quad \le K j_n \mathrm{e}^{j_n(\rho/\rho_n - 1)^+} d_1(\mu_n, \mu) + \frac{K}{\rho_n} d_1(\nu_n^*, \nu^*) + j_n \left| \frac{\rho}{\rho_n} - 1 \right| \mathrm{e}^{(j_n - 1)(\rho/\rho_n - 1)^+},
\end{aligned}$$

where in the last step we used the inequalities in (9). This last expression converges to 0 since $j_n d_1(\mu_n, \mu) \to 0$ as $n \to \infty$.

The proof of the last statement is identical to that of case 1 and is therefore omitted. □

We now proceed to the nonhomogeneous case.

Proof of Theorem 2. Again, we split the proof into two parts.

Case 1: weighted branching processes. The result for fixed k follows from Theorem 2, since

$$|R^{(n,k)} - R^{(k)}| = \left| \sum_{j=0}^{k} (W^{(n,j)} - W^{(j)}) \right| \le \sum_{j=0}^{k} |W^{(n,j)} - W^{(j)}|.$$

If, in addition, we have $\rho < 1$ then, by Proposition 1 (using the optimal coupling),

$$\begin{aligned}
\mathbb{E}[|R^{(n,k_n)} - R^{(k_n)}|] &\le \sum_{j=0}^{k_n} \mathbb{E}[|W^{(n,j)} - W^{(j)}|] \\
&\le \sum_{j=0}^{k_n} \left(\rho_n^j + \mathbb{E}[|Q|] \sum_{t=0}^{j-1} \rho^t \rho_n^{j-1-t} \right) d_1(\mu_n, \mu).
\end{aligned}$$

Note that since $|\rho_n - \rho| \le d_1(\mu_n, \mu)$ then for any $\rho < a < 1$, we have $\rho_n < a$ for all sufficiently large n. In this case

$$\mathbb{E}[|R^{(n,k_n)} - R^{(k_n)}|] \le \sum_{j=0}^{\infty} (a^j + \mathbb{E}[|Q|] j a^{j-1}) d_1(\mu_n, \mu) \to 0 \quad \text{as } n \to \infty$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

for any $k_n \geq 1$. Since we also have

$$\mathbb{E}[|R^{(k_n)} - R|] = \mathbb{E}\left[\left|\sum_{j=k_n+1}^{\infty} W^{(j)}\right|\right] \leq \sum_{j=k_n+1}^{\infty} \mathbb{E}[|Q|]\rho^j = \frac{\mathbb{E}[|Q|]\rho^{k_n+1}}{1-\rho},$$

then for any $k_n \to \infty$,

$$R^{(n,k_n)} \xrightarrow{d_1} R \quad \text{as } n \to \infty.$$

Case 2: weighted branching trees. The proof of the result for fixed k follows from Theorem 1 as before. For k_n and $\rho < 1$, we use Proposition 2 (using the optimal couplings π^* and π) to obtain

$$\begin{aligned}\mathbb{E}[|R^{(n,k_n)} - R^{(k_n)}|] &\leq \sum_{j=0}^{k_n} \mathbb{E}[|W^{(n,j)} - W^{(j)}|] \\ &\leq d_1(\nu_n^*, \nu^*) + K \sum_{j=1}^{k_n} \left(\sum_{t=0}^{j-1} \rho_n^t \rho^{j-1-t} d_1(\mu_n, \mu) + \rho_n^{j-1} d_1(\nu_n^*, \nu^*) \right)\end{aligned}$$

with $K = \max\{\mathbb{E}[N^{(n)}], \mathbb{E}[|Q|]\}$. Using the same arguments from case 1 note that for any $\rho < a < 1$ and sufficiently large n,

$$\mathbb{E}[|R^{(n,k_n)} - R^{(k_n)}|] \leq d_1(\nu_n^*, \nu^*) + K \sum_{j=1}^{\infty} (j a^{j-1} d_1(\mu_n, \mu) + a^{j-1} d_1(\nu_n^*, \nu^*)),$$

which converges to 0 as $n \to \infty$ for any $k_n \geq 1$. The rest of proof is the same as that of case 1 and is therefore omitted. □

The last result we need to prove in this section is Lemma 1.

Proof of Lemma 1. From the definition of the Kantorovich–Rubinstein metric and the fact that the infimum is always attained (see, e.g. [25, Theorem 4.1]), there exists a coupling π of $(N^{(n)}, Q^{(n)}, C^{(n)}, N, Q, C)$ such that

$$d_1(\nu_n, \nu) = \mathbb{E}_\pi[|Q^{(n)} - Q| + |N^{(n)} - N| + |C^{(n)} - C|]. \tag{10}$$

Next, define the vectors

$$\boldsymbol{Y}_n = C^{(n)}(Q^{(n)}, \mathbf{1}_{\{N^{(n)} \geq 1\}}, \mathbf{1}_{\{N^{(n)} \geq 2\}}, \dots), \qquad \boldsymbol{Y} = C(Q, \mathbf{1}_{\{N \geq 1\}}, \mathbf{1}_{\{N \geq 2\}}, \dots).$$

We will first show that $\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1 \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. To this end, let $(\hat{Q}, \hat{N}, \hat{C}) = (Q^{(n)}, N^{(n)}, C^{(n)})$ to simplify the notation and define $X_n = \|(N^{(n)}, Q^{(n)}, C^{(n)}) - (N, Q, C)\|_1$.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

Note that (10) implies that $X_n \to 0$ in mean, and therefore in probability. Now note that

$$\begin{aligned}
&\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1 \\
&\quad = |\hat{Q}\hat{C} - QC| + \sum_{i=1}^{\infty} |\hat{C}\,\mathbf{1}_{\{\hat{N}\geq i\}} - C\,\mathbf{1}_{\{N\geq i\}}| \\
&\quad = |\hat{Q}\hat{C} - QC| + \sum_{i=1}^{\infty} (|\hat{C} - C|\,\mathbf{1}_{\{i\leq \hat{N}\wedge N\}} + |\hat{C}|\,\mathbf{1}_{\{N<i\leq \hat{N}\}} + |C|\,\mathbf{1}_{\{\hat{N}<i\leq N\}}) \\
&\quad = |\hat{Q}\hat{C} - QC| + |\hat{C} - C|(\hat{N}\wedge N) + |\hat{C}|(\hat{N} - N)^+ + |C|(N - \hat{N})^+ \\
&\quad \leq |\hat{C}||\hat{Q} - Q| + |Q||\hat{C} - C| + |\hat{C} - C|(\hat{N}\wedge N) + |\hat{C}|(\hat{N} - N)^+ + |C|(N - \hat{N})^+ \\
&\quad \leq (2|C^{(n)}| + |Q| + N + |C|)X_n \\
&\quad \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty
\end{aligned}$$

by the Converging Together Lemma. It remains to show that $\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1 \to 0$ in mean.

By the triangle inequality, we have

$$Q_n \triangleq |\|\boldsymbol{Y}_n\|_1 - \|\boldsymbol{Y}\|_1| \leq \|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1 \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty.$$

Also, by assumption,

$$\begin{aligned}
\mathbb{E}[\|\boldsymbol{Y}_n\|_1] &= \mathbb{E}\left[|\hat{Q}\hat{C}| + \sum_{i=1}^{\infty} |\hat{C}|\,\mathbf{1}_{\{\hat{N}\geq i\}}\right] \\
&= \mathbb{E}[|\hat{Q}\hat{C}| + |\hat{C}|\hat{N}] \\
&\to \mathbb{E}[|CQ| + |C|N] \\
&= \mathbb{E}[\|\boldsymbol{Y}\|_1] \quad \text{as } n \to \infty,
\end{aligned}$$

and therefore $\mathbb{E}[Q_n] \to 0$; see, e.g. [10, Theorem 5.5.2]. Now note that since $\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1 \leq \|\boldsymbol{Y}_n\|_1 + \|\boldsymbol{Y}\|_1 \leq Q_n + 2\|\boldsymbol{Y}\|_1$, we have

$$\mathbb{E}[\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1] \leq \mathbb{E}[\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1\,\mathbf{1}_{\{Q_n\leq 1\}}] + \mathbb{E}[Q_n] + 2\mathbb{E}[\|\boldsymbol{Y}\|_1\,\mathbf{1}_{\{Q_n>1\}}],$$

where $\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1\mathbf{1}_{\{Q_n\leq 1\}}$ and $\|\boldsymbol{Y}\|_1\,\mathbf{1}_{\{Q_n>1\}}$ are uniformly integrable by [27, Theorem 13.3], and hence

$$\lim_{n\to\infty} \mathbb{E}[\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_1\,\mathbf{1}_{\{Q_n\leq 1\}}] = \lim_{n\to\infty} \mathbb{E}[\|\boldsymbol{Y}\|_1\,\mathbf{1}_{\{Q_n>1\}}] = 0. \qquad \square$$

6.3. Applications

In this last section we prove the theorems regarding our applications to the analysis of information ranking algorithms and random graphs.

Proof of Theorem 3. By applying the same steps from the proof of Theorem 2 conditionally on the sigma-algebra $\mathcal{F}_n$ (with ν_n^* the probability distribution of the root vector, which is allowed to be different), we obtain

$$\begin{aligned}
\mathbb{E}[|R^{(n,k)} - R^{(k)}| \mid \mathcal{F}_n] &\leq \sum_{j=0}^{k} \mathbb{E}[|W^{(n,j)} - W^{(j)}| \mid \mathcal{F}_n] \\
&\leq d_1(\nu_n^*, \nu^*) + K_n \sum_{j=1}^{k} \left(\sum_{t=0}^{j-1} \rho_n^t \rho^{j-1-t} d_1(\mu_n, \mu) + \rho_n^{j-1} d_1(\nu_n^*, \nu^*) \right),
\end{aligned}$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

where $K_n = \max\{\mathbb{E}[N_\varnothing^{(n)} \mid \mathcal{F}_n], \mathbb{E}[|Q|]\}$. Next, for any $\varepsilon > 0$ define the event $A_{n,\varepsilon} = \{d_1(\nu_n^*, \nu^*) + d_1(\mu_n, \mu) \le \varepsilon\}$. Recall that $\rho_n \le \rho + d_1(\mu_n, \mu)$ and note that $K_n \le \max\{\mathbb{E}[N_\varnothing] + d_1(\nu_n^*, \nu^*), \mathbb{E}[|Q|]\}$. The result for fixed k is obtained by first conditioning on the event $A_{n,\varepsilon}$, as will be done for the case of $k_n \to \infty$ below, and therefore we omit the details.

For $k_n \to \infty$ and $\rho < 1$, choose $0 < \varepsilon \le (1-\rho)/2$ and define $A_{n,\varepsilon}$ as above; set $K = \max\{\mathbb{E}[N_\varnothing] + \varepsilon, \mathbb{E}[|CQ|]\}$. Then apply Markov's inequality conditionally to obtain for any $\delta > 0$,

$$\begin{aligned}
\mathbb{P}(|R^{(n,k_n)} - R^{(k_n)}| > \delta) &\le \mathbb{P}(|R^{(n,k_n)} - R^{(k_n)}| > \delta,\ A_{n,\varepsilon}) + \mathbb{P}(A_{n,\varepsilon}^c) \\
&\le \delta^{-1}\mathbb{E}[\mathbf{1}_{\{A_{n,\varepsilon}\}}\,\mathbb{E}[|R^{(n,k_n)} - R^{(k_n)}| \mid \mathcal{F}_n]] + \mathbb{P}(A_{n,\varepsilon}^c) \\
&\le \delta^{-1}\mathbb{E}\left[\mathbf{1}_{\{A_{n,\varepsilon}\}}\left\{d_1(\nu_n^*, \nu^*) + K_n \sum_{j=1}^{\infty}(1-\varepsilon)^{j-1} d_1(\nu_n^*, \nu^*)\right\}\right] \\
&\quad + \delta^{-1}\mathbb{E}\left[\mathbf{1}_{\{A_{n,\varepsilon}\}}\, K_n \sum_{j=1}^{\infty} j(1-\varepsilon)^{j-1} d_1(\mu_n, \mu)\right] + \mathbb{P}(A_{n,\varepsilon}^c) \\
&\le \delta^{-1}\left(1 \vee \frac{K}{\varepsilon^2}\right)\mathbb{E}[d_1(\nu_n^*, \nu^*) \wedge \varepsilon + d_1(\mu_n, \mu) \wedge \varepsilon] + \mathbb{P}(A_{n,\varepsilon}^c).
\end{aligned}$$

The assumption that $d_1(\nu_n^*, \nu^*) + d_1(\mu_n, \mu) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ and the bounded convergence theorem show that $|R^{(n,k_n)} - R^{(k_n)}| \xrightarrow{\mathbb{P}} 0$. This combined with the observation that $|R^{(k_n)} - \mathcal{R}| \to 0$ a.s. completes the proof. □

We now prove the theorem for the analysis of the configuration model.

Proof of Theorem 4. Let $F_n(k) = \mathbb{P}(N \le k \mid \mathcal{F}_n)$, $G_n(k) = \mathbb{P}(N_\varnothing \le k \mid \mathcal{F}_n)$, $F(k) = \mathbb{P}(N \le k)$, and $G(k) = \mathbb{P}(N_\varnothing \le k)$. Let $\{U_{\boldsymbol{i}}\}_{\boldsymbol{i} \in U}$ be a sequence of i.i.d. uniform (0, 1) random variables, independent of $\mathcal{F}_n$. Construct the two trees using the coupled vectors $\{(N_{\boldsymbol{i}}^{(n)}, N_{\boldsymbol{i}})\}_{\boldsymbol{i} \in U}$, where

$$(N_\varnothing^{(n)}, N_\varnothing) = (G_n^{-1}(U_\varnothing), G^{-1}(U_\varnothing)), \qquad (N_{\boldsymbol{i}}^{(n)}, N_{\boldsymbol{i}}) = (F_n^{-1}(U_{\boldsymbol{i}}), F^{-1}(U_{\boldsymbol{i}})) \quad \text{for } \boldsymbol{i} \ne \varnothing,$$

and $f^{-1}(t) = \inf\{x \in \mathbb{R}\colon f(x) \ge t\}$. These are the optimal couplings for which $d_1(\nu_n^*, \nu^*)$ and $d_1(\nu_n, \nu)$ are achieved.

Next, by adapting Proposition 2 to allow ν_n^* to be different, we obtain, for $j \ge 1$,

$$\mathbb{E}[|Z^{(n,j)} - Z^{(j)}| \mid \mathcal{F}_n] \le m_n^{j-1} d_1(\nu_n^*, \nu^*) + \mathbf{1}_{\{j \ge 2\}}\, m^* \sum_{t=0}^{j-2} m_n^t m^{j-2-t} d_1(\nu_n, \nu),$$

where $m = \mathbb{E}[N]$, $m_n = \mathbb{E}[N^{(n)} \mid \mathcal{F}_n]$, $m^* = \mathbb{E}[N_\varnothing]$, and $m_n^* = \mathbb{E}[N_\varnothing^{(n)} \mid \mathcal{F}_n]$.

Note that $M^{(n,j)} = Z^{(n,j)}/(m_n^* m_n^{j-1}) - Z^{(j)}/(m^* m^{j-1})$ for $j \ge 1$ is a martingale with respect to the filtration $\mathcal{G}_j = \sigma(N_{\boldsymbol{i}}^{(n)}, N_{\boldsymbol{i}} : \boldsymbol{i} \in \mathbb{N}^s, 0 \le s \le j-1)$, conditionally on $\mathcal{F}_n$. Let $A_{n,\delta} = \{d_1(\nu_n^*, \nu^*) + j_n d_1(\nu_n, \nu) \le \delta\}$ and note that by assumption, $\mathbb{P}(A_{n,\delta}) \to 0$ for any $\delta > 0$. Now let $\varepsilon > 0$ and use Doob's maximal inequality to obtain

$$\mathbb{P}\Big(\max_{1 \le j \le j_n} |M^{(n,j)}| > \varepsilon,\, A_{n,\delta}\Big) \le \frac{1}{\varepsilon}\mathbb{E}[\mathbf{1}_{\{A_{n,\delta}\}}\,\mathbb{E}[|M^{(n,j_n)}| \mid \mathcal{F}_n]].$$

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

To bound this last expectation note that for any $j \geq 1$,

$$\begin{aligned}
&\mathbb{E}[|M^{(n,j)}| \mid \mathcal{F}_n] \\
&\quad \leq \mathbb{E}\left[\frac{1}{m^* m^{j-1}}|Z^{(n,j)} - Z^{(j)}| + \left|\frac{m_n^* m_n^{j-1}}{m^* m^{j-1}} - 1\right| \frac{Z^{(n,j)}}{m_n^* m_n^{j-1}} \;\middle|\; \mathcal{F}_n\right] \\
&\quad = \frac{1}{m^* m^{j-1}}\mathbb{E}[|Z^{(n,j)} - Z^{(j)}| \mid \mathcal{F}_n] + \left|\frac{m_n^* m_n^{j-1}}{m^* m^{j-1}} - 1\right| \\
&\quad \leq \frac{1}{m^*}\left(\frac{m_n}{m}\right)^{j-1} d_1(\nu_n^*, \nu^*) + \mathbf{1}_{\{j\geq 2\}} \frac{1}{m}\sum_{t=0}^{j-2}\left(\frac{m_n}{m}\right)^t d_1(\nu_n, \nu) + \left|\frac{m_n^* m_n^{j-1}}{m^* m^{j-1}} - 1\right| \\
&\quad \leq \frac{1}{m \wedge m^*}\left(\frac{m \vee m_n}{m}\right)^{j-1} (d_1(\nu_n^*, \nu^*) + j d_1(\nu_n, \nu)) + \left|\frac{m_n^* m_n^{j-1}}{m^* m^{j-1}} - 1\right|.
\end{aligned}$$

Moreover, by (9) it follows that for $1 \leq j \leq j_n$ and on the event $A_{n,\delta}$,

$$\left(\frac{m \vee m_n}{m}\right)^{j-1} \leq \mathrm{e}^{(j-1)(m_n/m-1)^+} \leq \mathrm{e}^{(j-1)d_1(\nu_n,\nu)/m} \leq \mathrm{e}^{\delta/m}$$

and

$$\begin{aligned}
\left|\frac{m_n^* m_n^{j-1}}{m^* m^{j-1}} - 1\right| &\leq \frac{m_n^*}{m^*}\left|\left(\frac{m_n}{m}\right)^{j-1} - 1\right| + \left|\frac{m_n^*}{m^*} - 1\right| \\
&\leq \frac{m_n^*}{m^*}(j-1)\left|\frac{m_n}{m} - 1\right| \mathrm{e}^{(j-2)^+(m_n/m-1)^+} + \left|\frac{m_n^*}{m^*} - 1\right| \\
&\leq \frac{(m^* + \delta)e^{\delta/m}}{m^* m}(j-1)d_1(\nu_n, \nu) + \frac{1}{m^*} d_1(\nu_n^*, \nu^*).
\end{aligned}$$

It follows that on the event $A_{n,\delta}$,

$$\begin{aligned}
\mathbb{E}[|M^{(n,j_n)}| \mid \mathcal{F}_n] \leq &\left(\frac{e^{\delta/m}}{m \wedge m^*} + \frac{1}{m^*}\right) d_1(\nu_n^*, \nu^*) \\
&+ \left(\frac{e^{\delta/m}}{m \wedge m^*} + \frac{(m^* + \delta)e^{\delta/m}}{m^* m}\right) j_n d_1(\nu_n, \nu).
\end{aligned}$$

Hence,

$$\mathbb{P}\left(\max_{1\leq j\leq j_n} |M^{(n,j)}| > \varepsilon, A_{n,\delta}\right) \leq \frac{K}{\varepsilon}\mathbb{E}[\mathbf{1}_{\{A_{n,\delta}\}}(d_1(\nu_n^*, \nu^*) + j_n d_1(\nu_n, \nu))]$$

for some constant $K = K(\delta) < \infty$. Since $d_1(\nu_n^*, \nu^*) + j_n d_1(\nu_n, \nu) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$, we obtain, using the bounded convergence theorem, that

$$\max_{1\leq j\leq j_n} |M^{(n,j)}| \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty. \tag{11}$$

For the second statement of the theorem note that for any $j \geq 1$,

$$\frac{|Z^{(n,j)} - Z^{(j)}|}{m^{j-1}} \leq m^*|M^{(n,j)}| + m^*\left|\frac{m_n^* m_n^{j-1}}{m^* m^{j-1}} - 1\right| X^{(n,j)},$$

where $X^{(n,j)} = Z^{(n,j)}/(m_n^* m_n^{j-1})$ is a mean 1 nonnegative martingale conditionally on $\mathcal{F}_n$.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

By (11), it only remains to show that $\max_{1\le j\le j_n} |m_n^* m_n^{j-1}/m^* m^{j-1} - 1| X^{(n,j)} \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. To this end, note that for $A_{n,\delta}$ as above and some constant $H = H(\delta) < \infty$,

$$\begin{aligned}
&\mathbb{P}\Bigg(\max_{1\le j\le j_n} \left| \frac{m_n^* m_n^{j-1}}{m^* m_n^{j-1}} - 1 \right| X^{(j)} > \varepsilon,\, A_{n,\delta} \Bigg) \\
&\quad \le \mathbb{P}\Big(H(j_n d_1(\nu_n, \nu) + d_1(\nu_n^*, \nu^*)) \max_{1\le j\le j_n} X^{(n,j)} > \varepsilon,\ A_{n,\delta} \Big) \\
&\quad \le \frac{H}{\varepsilon} \mathbb{E}[\mathbf{1}_{\{A_{n,\delta}\}} (j_n d_1(\nu_n, \nu) + d_1(\nu_n^*, \nu^*)) \mathbb{E}[X^{(n,j_n)} \mid \mathcal{F}_n]] \\
&\quad = \frac{H}{\varepsilon} \mathbb{E}[\mathbf{1}_{\{A_{n,\delta}\}} (j_n d_1(\nu_n, \nu) + d_1(\nu_n^*, \nu^*))] \\
&\quad \to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

where in the second step we used Doob's maximal inequality conditionally on $\mathcal{F}_n$. This completes the proof. □

The last proof verifies the conditions of Theorem 4 for the size-biased empirical distribution.

Proof of Lemma 2. To analyze $d_1(\nu_n^*, \nu^*)$ define $F(k) = \sum_{i=0}^k f(i)$ and let $F_n(k)$ denote the empirical distribution function of F. Then

$$d_1(\nu_n^*, \nu^*) = \sum_{k=0}^\infty |F_n(k) - F(k)| = \sum_{k=0}^\infty \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{D_i > k\}} - \mathbb{P}(D > k) \right| = \frac{1}{n} \sum_{k=0}^\infty |S_n^*(k)|,$$

where $S_n^*(k) = Y_{k,1} + \cdots + Y_{k,n}$ and $Y_{k,i} = \mathbf{1}_{\{D_i > k\}} - \mathbb{P}(D > k)$. Hence, we need to show that

$$\frac{1}{n^{1-\delta'}} \sum_{k=0}^\infty |S_n^*(k)| \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty$$

for any $0 < \delta' < \frac{1}{2}$. This follows from noting that, for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\Bigg(\sum_{k=0}^\infty |S_n^*(k)| > \varepsilon n^{1-\delta'} \Bigg) &\le \frac{1}{\varepsilon n^{1-\delta'}} \sum_{k=0}^\infty \mathbb{E}[|S_n^*(k)|] \\
&\le \frac{1}{\varepsilon n^{1-\delta'}} \sum_{k=0}^\infty (\mathbb{E}[(S_n^*(k))^2])^{1/2} \\
&= \frac{1}{\varepsilon n^{1-\delta'}} \sum_{k=0}^\infty (n \operatorname{var}(Y_{k,1}))^{1/2} \\
&\le \frac{1}{\varepsilon n^{1/2-\delta'}} \sum_{k=0}^\infty (\mathbb{P}(D > k))^{1/2} \\
&\le \frac{1}{\varepsilon n^{1/2-\delta'}} \left(1 + \sum_{k=1}^\infty \left(\frac{\mathbb{E}[D^{2+\epsilon}]}{k^{2+\epsilon}} \right)^{1/2} \right) \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

where in the second inequality we used the monotonicity of the L_p-norm and in the last one Markov's inequality.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

Similarly, if we let

$$G_n(k) = \sum_{i=1}^{n} \frac{D_i}{L_n} \mathbf{1}_{\{D_i \le k+1\}} \quad \text{and} \quad G(k) = \frac{\mathbb{E}[D\,\mathbf{1}_{\{D\le k+1\}}]}{\mathbb{E}[D]} \quad \text{for } k = 0, 1, 2, \ldots,$$

then

$$\begin{aligned}
d_1(\nu_n, \nu) &= \sum_{k=0}^{\infty} |G_n(k) - G(k)| \\
&= \sum_{k=0}^{\infty} \left| \sum_{i=1}^{n} \frac{D_i}{L_n} \mathbf{1}_{\{D_i > k+1\}} - \frac{\mathbb{E}[D\,\mathbf{1}_{\{D>k+1\}}]}{\mathbb{E}[D]} \right| \\
&\le \sum_{k=0}^{\infty} \left| \sum_{i=1}^{n} D_i\, \mathbf{1}_{\{D_i > k+1\}} \left(\frac{1}{L_n} - \frac{1}{n\mathbb{E}[D]} \right) \right| \\
&\quad + \frac{1}{\mathbb{E}[D]} \sum_{k=0}^{\infty} \left| \frac{1}{n} \sum_{i=1}^{n} D_i\, \mathbf{1}_{\{D_i > k+1\}} - \mathbb{E}[D\,\mathbf{1}_{\{D>k+1\}}] \right| \\
&\le \frac{|n\mathbb{E}[D] - L_n|}{L_n} \frac{1}{n\mathbb{E}[D]} \sum_{k=0}^{\infty} \sum_{i=1}^{n} D_i\, \mathbf{1}_{\{D_i > k+1\}} \\
&\quad + \frac{1}{n\mathbb{E}[D]} \sum_{k=0}^{\infty} \left| \sum_{i=1}^{n} (D_i\, \mathbf{1}_{\{D_i > k+1\}} - \mathbb{E}[D\,\mathbf{1}_{\{D>k+1\}}]) \right| \\
&= \frac{|n\mathbb{E}[D] - L_n|}{n} \frac{n}{L_n} \frac{1}{n\mathbb{E}[D]} \sum_{i=1}^{n} D_i (D_i - 1)^+ + \frac{1}{n\mathbb{E}[D]} \sum_{k=0}^{\infty} |S_n(k)|,
\end{aligned}$$

where $S_n(k) = X_{k,1} + \cdots + X_{k,n}$ and $X_{k,i} = D_i \mathbf{1}_{\{D_i > k+1\}} - \mathbb{E}[D\,\mathbf{1}_{\{D>k+1\}}]$. By the weak law of large numbers

$$\frac{n}{L_n} \frac{1}{n\mathbb{E}[D]} \sum_{i=1}^{n} D_i (D_i - 1)^+ \xrightarrow{\mathbb{P}} \frac{\mathbb{E}[D(D-1)^+]}{(\mathbb{E}[D])^2} \quad \text{as } n \to \infty,$$

therefore it suffices to show that

$$\frac{|n\mathbb{E}[D] - L_n|}{n^{1-\delta}} \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{1}{n^{1-\delta}} \sum_{k=0}^{\infty} |S_n(k)| \xrightarrow{\mathbb{P}} 0 \quad \text{for } 0 < \delta < \min\left\{ \frac{1}{2}, \frac{\epsilon}{2+\epsilon} \right\}.$$

Since $\mathbb{E}[D^{2+\epsilon}] < \infty$, the Marcinkiewicz–Zygmund strong law gives

$$\frac{(n\mathbb{E}[D] - L_n)}{n^{1/p}} = \frac{1}{n^{1/p}} \sum_{i=1}^{n} (D_i - \mathbb{E}[D]) \to 0 \quad \text{a.s. as } n \to \infty$$

for any $1 \le p < 2$, in particular, for $1/p = 1 - \delta$. For the limit involving $S_n(k)$ let

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

$a_n = \lfloor n^{1/(2+\epsilon)} \rfloor$ and follow the same steps used to analyze $S_n^*(k)$ to obtain

$$\begin{aligned}
&\mathbb{P}\Bigg(\sum_{k=0}^{\infty} |S_n(k)| > \varepsilon n^{1-\delta}\Bigg) \\
&\quad \le \frac{1}{\varepsilon n^{1-\delta}} \sum_{k=0}^{a_n-1} (\mathbb{E}[(S_n(k))^2])^{1/2} + \frac{1}{\varepsilon n^{1-\delta}} \sum_{k=a_n}^{\infty} \mathbb{E}[|S_n(k)|] \\
&\quad \le \frac{1}{\varepsilon n^{1/2-\delta}} \sum_{k=0}^{a_n-1} (\mathrm{var}(X_{k,1}))^{1/2} + \frac{n^\delta}{\varepsilon} \sum_{k=a_n}^{\infty} \mathbb{E}[|X_{k,1}|] \\
&\quad \le \frac{1}{\varepsilon n^{1/2-\delta}} \sum_{k=0}^{a_n-1} (\mathbb{E}[D^2 \mathbf{1}_{\{D>k+1\}}])^{1/2} + \frac{n^\delta}{\varepsilon} \sum_{k=a_n}^{\infty} 2\mathbb{E}[D \mathbf{1}_{\{D>k+1\}}].
\end{aligned}$$

Using the inequality

$$\mathbb{E}[D^{2+\epsilon}] \ge \mathbb{E}[D^{2+\epsilon} \mathbf{1}_{\{D>r\}}] \ge r^{2+\epsilon-t} \mathbb{E}[D^t \mathbf{1}_{\{D>r\}}] \quad \text{for any } r \ge 1, \tag{12}$$

and for any $t \in [0, 2+\epsilon]$, we obtain

$$\begin{aligned}
\mathbb{P}\Bigg(\sum_{k=0}^{\infty} |S_n(k)| > \varepsilon n^{1-\delta}\Bigg) &= O\Bigg(\frac{1}{n^{1/2-\delta}} \sum_{k=0}^{a_n-1} \frac{1}{(k+1)^{\epsilon/2}} + n^\delta \sum_{k=a_n}^{\infty} \frac{1}{(k+1)^{1+\epsilon}}\Bigg) \\
&= O\Bigg(\frac{a_n^{1-\epsilon/2} \mathbf{1}_{\{\epsilon\neq 2\}} + (\log a_n) \mathbf{1}_{\{\epsilon=2\}}}{n^{1/2-\delta}} + n^\delta a_n^{-\epsilon}\Bigg) \\
&= O(n^{\delta-\epsilon/(2+\epsilon)} + n^{\delta-1/2} (\log n) \mathbf{1}_{\{\epsilon=2\}}) \quad \text{as } n \to \infty,
\end{aligned}$$

which converges to 0 for $\delta < \min\{\frac{1}{2}, \epsilon/(2+\epsilon)\}$. □

Acknowledgement

We would like to thank an anonymous referee for his/her comments, which were very helpful in improving the presentation and scope of the results.

References

[1] Alsmeyer, G. and Iksanov, A. (2009). A log-type moment result for perpetuities and its application to martingales in supercritical branching random walks. *Electron. J. Prob.* **14,** 289–313.

[2] Alsmeyer, G. and Meiners, M. (2012). Fixed points of inhomogeneous smoothing transforms. *J. Differ. Equ. Appl.* **18,** 1287–1304.

[3] Alsmeyer, G. and Meiners, M. (2013). Fixed points of the smoothing transform: two-sided solutions. *Prob. Theory Relat. Fields* **155,** 165–199.

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

[5] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9,** 1196–1217.

[6] Biggins, J. D. (1977). Martingale convergence in the branching random walk. *J. Appl. Prob.* **14,** 25–37.

[7] Bollobás, B. (2001). *Random Graphs*, 2nd edn. Cambridge University Press.

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

[9] Chen, N., Litvak, N. and Olvera-Cravioto, M. (2014). Ranking algorithms on directed configuration networks. Preprint. Available at http://arxiv.org/abs/1409.7443.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press

[10] Durrett, R. (2010). *Probability: Theory and Examples*, 4th edn. Cambridge University Press.
[11] Fill, J. A. and Janson, S. (2001). Approximating the limiting Quicksort distribution. *Random Structures Algorithms* **19,** 376–406.
[12] Iksanov, A. M. (2004). Elementary fixed points of the BRW smoothing transforms with infinite number of summands. *Stoch. Process. Appl.* **114,** 27–50.
[13] Iksanov, A. and Meiners, M. (2015). Fixed points of multivariate smoothing transforms with scalar weights. *ALEA Latin Amer. J. Prob. Math. Statist.* **12,** 69–114.
[14] Jelenković, P. R. and Olvera-Cravioto, M. (2010). Information ranking and power laws on trees. *Adv. Appl. Prob.* **42,** 1057–1093.
[15] Jelenković, P. R. and Olvera-Cravioto, M. (2012). Implicit renewal theorem for trees with general weights. *Stoch. Process. Appl.* **122,** 3209–3238.
[16] Liu, Q. (1998). Fixed points of a generalized smoothing transformation and applications to the branching random walk. *Adv. Appl. Prob.* **30,** 85–112.
[17] Liu, Q. (2000). On generalized multiplicative cascades. *Stoch. Process. Appl.* **86,** 263–286.
[18] Lyons, R. (1997). A simple path to Biggins' martingale convergence for branching random walk. In *Classical and Modern Branching Processes*, Springer, New York, pp. 217–221.
[19] Neininger, R. (2001). On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms* **19,** 498–524.
[20] Olvera-Cravioto, M. (2012). Tail behavior of solutions of linear recursions on trees. *Stoch. Process. Appl.* **122,** 1777–1807.
[21] Rösler, U. (1991). A limit theorem for "Quicksort". *RAIRO Inf. Théor. Appl.* **25,** 85–100.
[22] Rösler, U. and Rüschendorf, L. (2001). The contraction method for recursive algorithms. *Algorithmica* **29,** 3–33.
[23] van der Hofstad, R. (2014). Random graphs and complex networks. Preprint. Available at http://www.win.tue.nl/∼rhofstad/NotesRGCN.html.
[24] van der Hofstad, R., Hooghiemstra, G. and Van Mieghem, P. (2005). Distances in random graphs with finite variance degrees. *Random Structures Algorithms* **27,** 76–123.
[25] Villani, C. (2009). *Optimal Transport: Old and New*. Springer, Berlin.
[26] Volkovich, Y. and Litvak, N. (2010). Asymptotic analysis for personalized web search. *Adv. Appl. Prob.* **42,** 577–604.
[27] Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.

https://doi.org/10.1017/apr.2016.12 Published online by Cambridge University Press