Stochastic recursions on directed random graphs

Nicolas Fraiman   Tzu-Chi Lin   Mariana Olvera-Cravioto

October 19, 2020

Abstract

For a directed graph $G(V_n, E_n)$ on the vertices $V_n = \{1, 2, \dots, n\}$, we study the distribution of a Markov chain $\{\mathbf{R}^{(k)} : k \geq 0\}$ on $\mathbb{R}^n$ such that the ith component of $\mathbf{R}^{(k)}$, denoted $R_i^{(k)}$, corresponds to the value of the process on vertex i at time k. We focus on processes $\{\mathbf{R}^{(k)} : k \geq 0\}$ where the value of $R_i^{(k+1)}$ depends only on the values $\{R_j^{(k)} : j \to i\}$ of its inbound neighbors, and possibly on vertex attributes. We then show that, provided $G(V_n, E_n)$ converges in the local weak sense to a marked Galton-Watson process, the dynamics of the process for a uniformly chosen vertex in V_n can be coupled, for any fixed k, to a process $\{\mathcal{R}_\emptyset^{(r)} : 0 \leq r \leq k\}$ constructed on the limiting marked Galton-Watson process. Moreover, we derive sufficient conditions under which $\mathcal{R}_\emptyset^{(k)}$ converges, as $k \to \infty$, to a random variable $\mathcal{R}^*$ that can be characterized in terms of the so-called *special* endogenous solution to a branching distributional fixed-point equation. Our framework can be applied also to processes $\{\mathbf{R}^{(k)} : k \geq 0\}$ whose only source of randomness comes from the realization of the graph $G(V_n, E_n)$.

Keywords: Markov chains, stochastic recursions, interacting particle systems, distributional fixed-point equations, weighted branching processes, directed graphs.
2010 MSC: Primary: 60G10, 05C80, 60J05. Secondary: 60J85.

1 Introduction

The main motivation for this work is to provide a mathematical framework that can be used to establish a rigorous connection between the stationary distribution of Markov chains on $\mathbb{R}^n$, whose dynamics are governed by the neighbor-to-neighbor interactions on a directed graph, and the solutions to branching distributional fixed-point equations. To illustrate the type of connection we seek, consider the Markov chain $\{W_k : k \geq 0\}$ corresponding to the waiting time of the kth customer in a single-server queue with i.i.d. interarrival times and i.i.d. processing times. It is well known that, $\{W_k : k \geq 0\}$ satisfies the recursion:

$$W_{k+1} = (W_k + \chi_k - \tau_{k+1})^+, \qquad k \geq 0,$$

where χ_k is the processing time of the kth customer and τ_k is the interarrival time between the $(k-1)$th and kth customers. We also know that provided that $E[\chi_1 - \tau_1] < 0$, there exists a random variable W_∞ such that $W_k \Rightarrow W_\infty$ as $k \to \infty$, where $\Rightarrow$ denotes weak convergence. It follows that W_∞ can be characterized as the unique solution to Lindley's equation:

$$W \stackrel{\mathcal{D}}{=} (W + \chi - \tau)^+,$$

where W is independent of (χ, τ) and $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution.

In general, if $\{X_k : k \geq 0\}$ is a Feller chain on $\mathbb{R}$ whose transitions are determined via

$$X_{k+1} = \Phi(X_k, \xi_k), \qquad k \geq 0,$$

for some deterministic and continuous map Φ, and $\{\xi_k : k \geq 0\}$ a sequence of i.i.d. random variables, then, assuming $X_k \Rightarrow X_\infty$ as $k \to \infty$ for some random variable X_∞, we can expect that X_∞ will be a solution to

$$X \stackrel{\mathcal{D}}{=} \Phi\left(X, \xi\right),$$

where X is independent of ξ.

Branching distributional equations take the general form

$$X \stackrel{\mathcal{D}}{=} \Psi\left(N, Q, \{X_i, C_i : 1 \leq i \leq N\}\right),$$

where the $\{X_i\}$ are i.i.d. copies of X, independent of $(N, Q, C_1, C_2, \dots)$. These recursions often arise when analyzing structures on trees [3], divide-and-conquer algorithms [31, 28], queueing networks with synchronization requirements [45, 52], and as heuristics for the stationary behavior of random processes on sparse, locally tree-like graphs. It is this last source of problems that this paper focuses on, by providing conditions on the types of recursions and the properties of the underlying graphs, that will provably lead to a branching distributional fixed-point equation.

The set of techniques developed in this paper rely heavily on the analysis of the PageRank algorithm [59, 40, 17, 47, 35, 51], which in the context of this work corresponds to a linear recursion with no randomness other than the one used to generate the graph where it is defined. The PageRank recursion has now been extensively studied, both from the point of view of its convergence to a branching distributional fixed-point equation, and also from the point of view of the tail behavior of its solutions [6, 7, 42, 43, 50]. Interestingly, much of what is known for PageRank can be extended to include additional random noises that act on either the vertices or the edges of the graph, and the mode of convergence can be extended, for unbounded recursions, to include the convergence of certain moments. When we add the noises, our framework connects to the study of discrete-time interactive particle systems like those studied in [46], specialized to recursions that will lead to a characterization of their stationary distribution in terms of a branching distributional fixed-point equation. Specifically, the type of discrete-time interacting particle systems that our framework covers are defined on directed graphs with asymptotically no self-loops.

The remainder of the paper is organized as follows. In Section 1.1 we give a precise description of the type of recursions that we study, including some well-known examples that fit into our framework. In Section 2 we specify the type of directed random graphs for which our results will hold and state our main theorem. In Section 3 we explain how to construct solutions to branching distributional fixed-point equations, and in Section 4 we prove all our theorems. Finally, for completeness, we give in the Appendix a brief description of the random graph models that we use throughout the paper.

1.1 Structure of the recursions

Consider a directed graph $G(V_n, E_n)$ on the set of vertices $V_n = \{1, 2, \dots, n\}$ and having directed edges on the set E_n. We will assume later that $G(V_n, E_n)$ is constructed according to specific random graph models, but for now we do not impose any additional conditions. The graph does not need to be simple, nor is it required to be strongly connected.

Once the graph edges are realized, each vertex $i \in V_n$ is assigned its in-degree, D_i^-, its out-degree, D_i^+, and its attribute $\mathbf{a}_i \in \mathcal{S}'$, where $\mathcal{S}'$ is a separable metric space. The vertex attributes $\mathbf{a}_i$ may contain parameters needed to construct the graph $G(V_n, E_n)$, as well as additional parameters specific to the recursion being studied. Define

$$\mathbf{X}_i = (D_i^-, D_i^+, \mathbf{a}_i), \qquad i \in V_n.$$

To define our Markov chain $\{\mathbf{R}^{(k)} : k \geq 0\}$, start with an initial (deterministic) vector $\mathbf{r}^{(0)} = (r_0, \ldots, r_0) \in \mathbb{R}^n$ and recursively define $\mathbf{R}^{(0)} = \mathbf{r}^{(0)}$ and

$$R_i^{(k+1)} = \Phi\left(\mathbf{X}_i, \zeta_i^{(k)}, \left\{g(R_j^{(k)}, \mathbf{X}_j), \xi_{j,i}^{(k)} : j \to i\right\}\right), \qquad i \in V_n, \quad k \geq 0,$$

where the sequences $\{\zeta_i^{(k)} : i \geq 1,\, k \geq 0\}$ and $\{\xi_{i,j}^{(k)} : i, j \geq 1,\, k \geq 0\}$ consist each of i.i.d. random variables, independent of each other and of any other random variables in the graph and g is a deterministic function. We will refer to the random variables $\boldsymbol{\zeta}^{(k)} = \{\zeta_i^{(k)} : i \in V_n\}$ and the vectors $\boldsymbol{\xi}^{(k)} = \{\xi_{i,j}^{(k)} : i, j \in V_n\}$ as the "noise" at time k, with $\boldsymbol{\zeta}^{(k)}$ representing noise on the vertices and $\boldsymbol{\xi}^{(k)}$ noise on the edges.

Remark 1.1. It is important to point out that the type of recursions we study exclude cases where $R_i^{(k+1)}$ depends on $R_i^{(k)}$, since $G(V_n, E_n)$ will not in general have self-loops. In other words, the evolution of each vertex is determined only by that of its inbound neighbors, their vertex attributes, its own vertex attribute, and potentially, noises on vertices and/or edges, but not by its own history. It is this characteristic that allows the limit as $k \to \infty$ to be fully characterized in terms of a branching distributional fixed-point equation, which we do not expect to be true for more general recursions.

Examples 1.2. The following examples illustrate the type of recursions that our framework is intended to cover. The first two examples include noises at the vertices, while the second two have no noises of either kind.

1. *Generalized DeGroot model:* This model is used in social sciences to study opinion dynamics. Here, $R_i^{(k)}$ denotes the opinion of the ith individual in the population at time k,

$$R_i^{(k+1)} = cf\left(q_i, \zeta_i^{(k)}\right) + \frac{1-c}{D_i^-} \sum_{j \to i} R_j^{(k)}, \tag{1.1}$$

where $c \in [0, 1]$ is a damping factor and q_i is a vertex parameter. When $c = 0$, equation (1.1) is known as the DeGroot model [26]. When $f(q_i, \zeta_i^{(k)}) = q_i$, it is known as the Friedkin-Johnsen model [33, 34, 32, 53], and q_i represents the internal opinion or stubbornness of agent i. Another extension of the model [60] sets $f(q_i, \zeta_i^{(k)}) = \zeta_i^{(k)}$ in order to model miscommunication between the agents. This choice of $f(q_i, \zeta_i^{(k)}) = \zeta_i^{(k)}$ has also been used in [1, 24] to model production networks, in which case $R_i^{(k)}$ represents the logarithm of the output of firm i, $\zeta_i^{(k)}$ models a shock to firm i, and c is the level of interconnection in the economy.

2. *Noisy majority voter model:* This is an interacting particle system, with agents on a directed graph, and taking values in the set $\{0,1\}$. The recursion describes the "vote" of agent i at time k, where each agent either adopts the vote indicated by the majority of its inbound neighbors with some probability p, or chooses the opposite vote with probability $1-p$. Its dynamics are given by

$$R_i^{(k+1)} = \left(\zeta_i^{(k)} + 1 \left(\sum_{j \to i} R_j^{(k)} \geq \frac{D_i^-}{2} \right) \right) \mod 2, \tag{1.2}$$

where $\zeta_i^{(k)} \sim \text{Ber}(\epsilon)$. In [3, 4] the recursion is described for directed 3-regular trees, and it argued that it defines a contraction under a Wasserstein metric for specific values of ϵ. This model, defined on the square lattice, has also been studied via Monte Carlo simulation in [25]. Note that removing the noise $\zeta_i^{(k)}$ from (1.2) is equivalent to modeling the Glauber dynamics at zero temperature [16]. Recursion (1.2) has also been applied to the study of opinion dynamics in [56] and trader dynamics on financial networks in [57].

3. *Google's PageRank algorithm:* This recursion computes the rank of vertices on a directed graph of size n by assigning to each vertex i a rank $r_i = R_i/n$, where

$$R_i = (1-c)nq_i + \sum_{j \to i} \frac{c}{D_j^+} R_j,$$

where R_i is the scale-free rank of vertex i, q_i represents a personalization value, with $\mathbf{q} = (q_1, \ldots, q_n)$ a probability vector, and $c \in (0,1)$ is a damping factor. The PageRank algorithm is one of the most popular centrality measures on networks, and is used in many different areas of science and engineering, including word sense disambiguation [2], spam detection [38], citation ranking [19], visual search [44] and many others [37]. It was originally created by Brin and Page [13] to rank the webpages in the WWW. Its convergence on random graphs to the special endogenous solution of a smoothing transform has been established in [17, 47, 51], and this characterization has been used to prove the so-called "power-law hypothesis" in [59, 41, 50, 35].

4. *A model for financial cascades:* This recursion was proposed in [27] for studying banking networks

$$R_i = q_i + \sum_{j \to i} \frac{(R_j - v_j)^+ \wedge b_j}{D_j^+}, \tag{1.3}$$

where R_i is the total wealth held by bank i, and (q_i, b_i, v_i) represent its total external assets, its inter-bank loans, and its external liability, respectively. The work in [27] studies only an Erdős-Rényi graph, but can also fits our more general framework. Interestingly, (1.3) defines a contraction only if the limiting out-degree satisfies $P(D^+ = 0) > 0$.

We will now separate the different levels of randomness involved in the construction of the process $\{\mathbf{R}^{(k)} : k \geq 0\}$. The first level of randomness is due to the attributes $\{\mathbf{a}_i : i \in V_n\}$, which may influence the construction of the graph itself, i.e., the presence/absence of edges. We will identify this level of randomness through the filtration:

$$\mathscr{F}_n = \sigma(\mathbf{a}_i : 1 \leq i \leq n),$$

which we use to define the conditional probability $\mathbb{P}_n(\cdot) = E[1(\cdot)|\mathscr{F}_n]$ and its corresponding conditional expectation $\mathbb{E}_n[\cdot] = E[\cdot|\mathscr{F}_n]$.

In order to distinguish the randomness produced by the noise sequences from that of the graph construction, we will also define the filtration:

$$\mathscr{G}_n = \sigma(\mathscr{F}_n, G(V_n, E_n)),$$

which contains all the random variables involved in the construction of the graph $G(V_n, E_n)$, and leaves as the only remaining source of randomness the noise sequences. We will further use $\mathscr{G}_n$ to define the conditional probability and expectation $\mathbf{P}_n(\cdot) = E[1(\cdot)|\mathscr{G}_n]$ and $\mathbf{E}_n[\cdot] = E[\cdot|\mathscr{G}_n]$. Note that the $\{\mathbf{X}_i : i \in V_n\}$ are measurable with respect to $\mathscr{G}_n$, but not necessarily $\mathscr{F}_n$.

The main assumption that we will make on the map Φ throughout the paper, is that it be Lipschitz continuous in its recursive arguments, i.e., the $\{R_j^{(k)}\}$, and continuous in its vertex attributes, since the graphs themselves will be changing when we take the large graph limit. Intuitively, this assumption ensures that we can control consecutive iterations of the recursion through a linear map on $\mathbb{R}^n$. Since the vertex attributes $\{\mathbf{a}_i\}$ are assumed to take values on a metric space $\mathcal{S}'$, we can choose a convenient metric ρ' for our continuity assumptions below. Throughout the rest of the paper, for $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{N} \times \mathbb{N} \times \mathcal{S}'$, where $\mathbf{x} = (d^-, d^+, \mathbf{a})$ and $\tilde{\mathbf{x}} = (\tilde{d}^-, \tilde{d}^+, \tilde{\mathbf{a}})$, we define the metric:

$$\rho(\mathbf{x}, \tilde{\mathbf{x}}) = |d^- - \tilde{d}^-| + |d^+ - \tilde{d}^+| + \rho'(\mathbf{a}, \tilde{\mathbf{a}}), \tag{1.4}$$

and use $\mathcal{S} = \mathbb{N} \times \mathbb{N} \times \mathcal{S}'$ to denote the underlying metric space.

The main assumptions on the map Φ are given below.

Assumption [R]. There exist continuous functions $\sigma_-, \sigma_+, \beta : \mathcal{S} \to \mathbb{R}$ and $p \in [1, \infty)$ such that:

1) For any $\mathbf{v}, \tilde{\mathbf{v}} \in \mathbb{R}^n$,

$$\left(\mathbf{E}_n\left[\left|\Phi\left(\mathbf{X}_i, \zeta_i^{(0)}, \{v_j, \xi_{j,i}^{(0)} : j \to i\}\right) - \Phi\left(\mathbf{X}_i, \zeta_i^{(0)}, \{\tilde{v}_j, \xi_{j,i}^{(0)} : j \to i\}\right)\right|^p\right]\right)^{1/p} \leq \sum_{j \to i} \sigma_-(\mathbf{X}_i)\, |v_j - \tilde{v}_j|\,,$$

2) For any $\mathbf{x} \in \mathcal{S}$ and $r, \tilde{r} \in \mathbb{R}$,

$$|g(r, \mathbf{x}) - g(\tilde{r}, \mathbf{x})| \leq \sigma_+(\mathbf{x})|r - \tilde{r}|,$$

3) For any $\mathbf{v} \in \mathbb{R}^n$,

$$\left(\mathbf{E}_n\left[\left|\Phi\left(\mathbf{X}_i, \zeta_i^{(0)}, \{v_j, \xi_{j,i}^{(0)} : j \to i\}\right)\right|^p\right]\right)^{1/p} \leq \sum_{j \to i} \sigma_-(\mathbf{X}_i)|v_j| + \beta(\mathbf{X}_i) \quad \text{and}$$

4) One of the following holds:

   i. The matrix $\mathbf{C}$ whose (i,j)th component is $C_{i,j} = \sigma_-(\mathbf{X}_i)\sigma_+(\mathbf{X}_j)1(j \to i)$ satisfies $\|\mathbf{C}\|_p \leq K < \infty$ for any directed graph $G(V_n, E_n)$, or

ii. There exists $K < \infty$ such that if $\|\mathbf{r}^{(0)}\|_\infty \leq K$ then $\|\mathbf{R}^{(1)}\|_\infty \leq K$ for any directed graph $G(V_n, E_n)$.

In addition, suppose that the functions $g(r, \mathbf{x})$ and $\varphi(\mathbf{x}; \mathbf{v})$ are continuous in $\mathbf{x}$ P-a.s. with respect to ρ for any fixed $r \in \mathbb{R}$ and $\mathbf{v} = \{v_j : j \geq 1\} \subseteq \mathbb{R}$, where

$$\varphi(\mathbf{x}; \mathbf{v}) = \Phi\left(\mathbf{x}, \zeta, \{v_j, \xi_j : j \in B(\mathbf{x})\}\right)$$

for $\mathbf{x} = (d^-, d^+, \mathbf{a}) \in \mathcal{S}$, $B(\mathbf{x}) = \{1, 2, \ldots, d^-\}$, and $\zeta, \{\xi_j : j \geq 1\} \subseteq \mathbb{R}$ a fixed realization of the noises.

Examples 1.3. All the models in Example 1.2 satisfy Assumption [R] for different choices of p, g, σ_-, σ_+ and β, as described below:

1. *Generalized DeGroot model:* We can take any $p \geq 1$ and $g(r, \mathbf{x}) = r$, $\sigma_-(\mathbf{X}_i) = (1-c)/D_i^-$, $\sigma_+(\mathbf{X}_j) = 1$, and $\beta(\mathbf{X}_i) = c\left(\mathbb{E}_n[|f(q_i, \zeta_i)|^p]\right)^{1/p}$. In addition, if $|f(q, \zeta)| \leq K$ for some $K < \infty$, then, $\|\mathbf{R}^{(1)}\|_\infty \leq K$, and therefore, Assumption [R](ii) is satisfied.

2. *Noisy majority voter model:* We can take any $p \geq 1$ and $g(r, \mathbf{x}) = r$. To identify σ_-, σ_+ and β, note that for any vector $\mathbf{r}, \tilde{\mathbf{r}} \in \mathbb{R}^n$,

$$\left(\mathbf{E}_n\left[\left|\Phi\left(\mathbf{X}_i, \zeta_i^{(0)}, \{r_j, \xi_{j,i}^{(0)} : j \to i\}\right) - \Phi\left(\mathbf{X}_i, \zeta_i^{(0)}, \{\tilde{r}_j, \xi_{j,i}^{(0)} : j \to i\}\right)\right|^p\right]\right)^{1/p}$$
$$= \left|1\left(\frac{1}{D_i^-}\sum_{j\to i} r_j \geq \frac{1}{2}\right) - 1\left(\frac{1}{D_i^-}\sum_{j\to i}\tilde{r}_j \geq \frac{1}{2}\right)\right| \leq \frac{2}{D_i^-}\sum_{j\to i}|r_j - \tilde{r}_j|,$$

so we can take $\sigma_-(\mathbf{X}_i) = 2/D_i^-$ and $\sigma_+(\mathbf{X}_j) = 1$. In addition, by choosing $p = 1$ and ignoring the mod 2 in the recursion we obtain $\beta(\mathbf{X}_i) = P(\zeta = 1)^{1/p}$. Since $\mathbf{R}_i^{(k)} \in \{0, 1\}$ for all $i \in V_n$, then Assumption [R](ii) is satisfied.

3. *Google's PageRank algorithm:* We can take $p = 1$, $g(r, \mathbf{X}_j) = cr/D_j^+$, $\sigma_-(\mathbf{X}_i) = 1$, $\sigma_+(\mathbf{X}_j) = c/D_j^+$, and $\beta(\mathbf{X}_i) = (1-c)nq_i$. The corresponding matrix $\mathbf{C}$ satisfies $\|\mathbf{C}\|_1 \leq c < 1$.

4. *A model for financial cascades:* We can take $p = 1$ and $g(r, \mathbf{X}_j) = ((r - v_j)^+ \wedge b_j)/D_j^+$. Using the inequality $|(x \vee a) \wedge b - (y \vee a) \wedge b| \leq |x - y|$, gives that we can set $\sigma_-(\mathbf{X}_i) = 1, \sigma_+(\mathbf{X}_j) = 1/D_j^+$, and $\beta(\mathbf{X}_i) = |q_i|$. The corresponding matrix $\mathbf{C}$ satisfies $\|\mathbf{C}\|_1 = 1$.

1.2 Convergence on the graph

Our main goal in this paper is to analyze the behavior of the recursion Φ on the graph $G(V_n, E_n)$, which we propose to do by characterizing the distribution of a uniformly chosen at random component of the vector $\mathbf{R}^{(k)}$. In most of the examples that we have in mind, we can establish the existence of a stationary distribution for the Markov chain $\{\mathbf{R}^{(k)} : k \geq 0\}$ taking values in $\mathbb{R}^n$ by showing that the map Φ defines a contraction under a suitable Wasserstein metric. However, since it is possible to establish the existence of a stationary distribution in other ways, our main theorem (Theorem 2.2) will focus on characterizing the distribution of a uniformly chosen component of $\{\mathbf{R}^{(r)} : 0 \leq r \leq k\}$ for any fixed k, without specifically imposing conditions to ensure its convergence as $k \to \infty$.

That said, we include for completeness a result establishing the convergence of $\{\mathbf{R}^{(k)} : k \geq 0\}$ whenever the map Φ defines a contraction under any of the l_p norms in $\mathbb{R}^n$.

Let $\lambda_{k,n}(\cdot) = \mathbf{P}_n\left(\mathbf{R}^{(k)} \in \cdot\right)$ denote the probability measure of the random vector $\mathbf{R}^{(k)} \in \mathbb{R}^n$ given $\mathscr{G}_n$. All the proofs are given in Section 4.

Theorem 1.4. *Let $G(V_n, E_n)$ be any directed graph and suppose Assumption [R] holds. Then, provided $\|\mathbf{C}\|_p < 1$ in Assumption [R], there exists a random vector $\mathbf{R} \in \mathbb{R}^n$ distributed according to the probability measure $\lambda_n(\cdot) = \mathbf{P}_n(\mathbf{R} \in \cdot)$ such that*

$$\mathbf{E}_n\left[\left\|\mathbf{R}^{(k)} - \mathbf{R}\right\|_p^p\right] \to 0, \qquad k \to \infty$$

a.s. with respect to $\mathbf{P}_n$. Moreover, for any $k \geq 1$,

$$\left(\mathbf{E}_n\left[\left\|\mathbf{R}^{(k)} - \mathbf{R}\right\|_p^p\right]\right)^{1/p} \leq \left(\mathbf{E}_n\left[\left\|\mathbf{r}^{(0)} - \mathbf{R}^{(1)}\right\|_p^p\right]\right)^{1/p} \frac{\|\mathbf{C}\|_p^k}{1 - \|\mathbf{C}\|_p}.$$

2 Characterizing the typical behavior

As mentioned earlier, our main goal is to characterize the distribution of a uniformly chosen component of the vector $\mathbf{R}^{(k)}$, since it represents the typical behavior of a vertex in the graph $G(V_n, E_n)$ under the iterations of the map Φ. Throughout the paper we denote this random variable by $R_I^{(k)}$, where I is a uniformly chosen index in V_n. The main idea behind our characterization is that, provided the graph $G(V_n, E_n)$ is locally tree-like, $R_I^{(k)}$ will satisfy a recursion on a tree which will converge, as $n, k \to \infty$, to an endogenous solution to a branching distributional fixed-point equation. It follows that the assumptions for our main theorem will need to include conditions on the graph $G(V_n, E_n)$ ensuring its local weak convergence.

Throughout the rest of the paper we use $\mu_{k,n}(\cdot) = \mathbb{P}_n\left(R_I^{(k)} \in \cdot\right)$ to denote the distribution of $R_I^{(k)}$. Note that $\mu_{k,n}$ is a random measure since it is defined conditionally given $\mathscr{F}_n = \sigma(\mathbf{a}_i : 1 \leq i \leq n)$, which does not include the realization of the graph $G(V_n, E_n)$. Moreover, whenever the sequence of attributes $\{\mathbf{a}_i : 1 \leq i \leq n\}$ is exchangeable, the vertices in $G(V_n, E_n)$ are too, in which case $\mu_{k,n}$ represents the common marginal distribution of any component of the vector $\mathbf{R}^{(k)}$.

We now go on to impose conditions on the types of random graphs where our characterization can be shown to hold. The two models we consider are the directed configuration model and the inhomogeneous random digraph, both of which are known to converge, in the local weak convergence sense, to marked Galton-Watson processes. We refer the reader to the Appendix for a description of the two models, and proceed instead to directly state the assumptions.

Recall that the vertex attributes $\{\mathbf{a}_i : 1 \leq i \leq n\}$ take values on the separable metric space $\mathcal{S}'$, and that we have defined the space $\mathcal{S} = \mathbb{N} \times \mathbb{N} \times \mathcal{S}'$ and equipped it with the metric ρ defined via (1.4). In general, for a separable metric space $\mathcal{S}''$ equipped with the metric ρ'', we define the Wasserstein metric W_p, $p \in [1, \infty)$, via:

$$W_p(\mu, \nu) = \inf \left\{ \left(\mathbb{E}_n \left[\rho''(\hat{\mathbf{X}}, \boldsymbol{X})^p \right] \right)^{1/p} : \text{law}(\hat{\mathbf{X}}) = \mu, \text{ law}(\boldsymbol{X}) = \nu \right\}.$$

The following set of assumptions ensures the local weak convergence of the graph $G(V_n, E_n)$ to a marked Galton-Watson process, and the convergence of finitely many iterations of the map Φ. However, they do not on their own establish the convergence of infinitely many iterations of the map Φ on the limiting tree, nor the characterization we seek in terms of a distributional fixed-point equation. We point out that since we are interested in being able to include scale-free graphs in our framework, i.e., graphs whose degree distributions are regularly varying with finite means but not necessarily finite higher-order moments, we assume only convergence of the vertex attributes in W_1. Interestingly, under our assumptions, the offspring distribution of the limiting Galton-Watson process may have infinite mean.

Assumption [G]. The graph $G(V_n, E_n)$ is constructed according to either a directed configuration model (DCM) or an inhomogeneous random digraph (IRD). Parts (a) and (b) in each case refer to the construction of the graph itself, while part (c) refers to the map Φ. For the function φ in Assumption [R], $\mathbf{x} \in \mathcal{S}$, and $\mathbf{v} = \{v_j : j \geq 1\} \subseteq \mathbb{R}$, we define the random functions:

$$\omega_0(\mathbf{x}, \mathbf{v}) = \sup_{\tilde{\mathbf{x}}:\rho(\mathbf{x},\tilde{\mathbf{x}})<1} \varphi(\tilde{\mathbf{x}}; \mathbf{v}) \qquad \text{and} \qquad \omega(\mathbf{x}, \mathbf{v}) = \sup_{\tilde{\mathbf{x}}:\rho(\mathbf{x},\tilde{\mathbf{x}})<1} g(\varphi(\tilde{\mathbf{x}}; \mathbf{v}), \tilde{\mathbf{x}}).$$

- **Directed configuration model:** Let $\mathbf{a}_i = (D_i^-, D_i^+, \mathbf{b}_i)$ for $1 \leq i \leq n$ denote the extended degree sequence used to construct the graph, and for $B \subseteq \mathcal{S}'$ let

$$\upsilon_n(B) = \frac{1}{n} \sum_{i=1}^n 1((D_i^-, D_i^+, \mathbf{b}_i) \in B).$$

  a. Suppose there exists a probability measure υ on $\mathcal{S}'$ such that

$$W_1(\upsilon_n, \upsilon) \xrightarrow{P} 0 \qquad n \to \infty.$$

  b. $E[\mathscr{D}^- + \mathscr{D}^+ + \rho'(\boldsymbol{A}, \mathbf{a}_0)] < \infty$ and $E[\mathscr{D}^-] = E[\mathscr{D}^+]$ for $\boldsymbol{A} = (\mathscr{D}^-, \mathscr{D}^+, \boldsymbol{B})$ distributed according to υ and some non-random $\mathbf{a}_0 \in \mathcal{S}'$.

  c. The following limits hold as $n \to \infty$ for $\sigma_-, \sigma_+, \beta, p$ in Assumption [R]:

$$\sum_{i=1}^n \frac{D_i^+}{L_n} |g(r_0, \mathbf{X}_i)|^p \xrightarrow{P} \frac{1}{E[\mathscr{D}^+]} E[\mathscr{D}^+ |g(r_0, \boldsymbol{X}_0)|^p] < \infty,$$

$$\sum_{i=1}^n \frac{D_i^+}{L_n} \left(D_i^- \sigma_+(\mathbf{X}_i) \sigma_-(\mathbf{X}_i) \right)^p \xrightarrow{P} \frac{1}{E[\mathscr{D}^+]} E[\mathscr{D}^+ \left(\mathscr{D}^- \sigma_+(\boldsymbol{X}_0) \sigma_-(\boldsymbol{X}_0) \right)^p] =: c^p < \infty,$$

$$\sum_{i=1}^n \frac{D_i^+}{L_n} \left(\sigma_+(\mathbf{X}_i) \beta(\mathbf{X}_i) \right)^p \xrightarrow{P} \frac{1}{E[\mathscr{D}^+]} E[\mathscr{D}^+ \left(\sigma_+(\boldsymbol{X}_0) \beta(\boldsymbol{X}_0) \right)^p] < \infty,$$

$$\frac{1}{n} \sum_{i=1}^n (D_i^- \sigma_-(\mathbf{X}_i))^p \xrightarrow{P} E\left[(\mathscr{D}^- \sigma_-(\boldsymbol{X}_0))^p \right] < \infty,$$

$$\frac{1}{n}\sum_{i=1}^n \beta(\mathbf{X}_i)^p \xrightarrow{P} E[\beta(\boldsymbol{X}_0)^p] < \infty,$$

$$E\left[\omega_0(\boldsymbol{X}_0, \boldsymbol{V})^p + \mathscr{D}^+\omega(\boldsymbol{X}_0, \boldsymbol{V})^p\right] < \infty,$$

for $\boldsymbol{X}_0 = (\mathscr{D}^-, \mathscr{D}^+, \boldsymbol{A})$ distributed according to υ and $\boldsymbol{V} = \{V_j : j \geq 1\}$ a sequence of i.i.d. random variables satisfying $E[|V_1|^p] < \infty$, independent of $\boldsymbol{X}_0$.

- **Inhomogeneous random digraph:** Let $\mathbf{a}_i = (W_i^-, W_i^+, \mathbf{b}_i)$ for $1 \leq i \leq n$ denote the extended type sequence used to construct the graph, and for $B \subseteq \mathcal{S}'$ let

$$\upsilon_n(B) = \frac{1}{n}\sum_{i=1}^n 1((W_i^-, W_i^+, \mathbf{b}_i) \in B).$$

a. Suppose there exists a probability measure υ on $\mathcal{S}'$ such that

$$W_1(\upsilon_n, \upsilon) \xrightarrow{P} 0 \qquad n \to \infty.$$

b. $E[W^- + W^+ + \rho'(\boldsymbol{A}, \mathbf{a}_0)] < \infty$ for $\boldsymbol{A} = (W^-, W^+, \boldsymbol{B})$ distributed according to υ and some non-random $\mathbf{a}_0 \in \mathcal{S}'$, and

$$\frac{1}{n}\sum_{i=1}^n \sum_{1\leq j\leq n, j\neq i} \left| p_{ij}^{(n)} - (r_{ij}^{(n)} \wedge 1)\right| \xrightarrow{P} 0, \qquad n \to \infty,$$

where $r_{ij}^{(n)} = W_i^+ W_j^-/(\theta n)$ and $\theta = E[W^- + W^+]$.

c. The following limits hold as $n \to \infty$ for $\sigma_-, \sigma_+, \beta, p$ in Assumption [R]:

$$\sum_{i=1}^n \frac{W_i^+}{L_n^+}\mathbb{E}_n\left[|g(r_0, \mathbf{X}_i)|^p | \mathbf{a}_i\right] \xrightarrow{P} \frac{1}{E[W^+]}E[W^+ |g(r_0, \boldsymbol{X}_0')|^p] < \infty,$$

$$\sum_{i=1}^n \frac{W_i^+}{L_n^+}\mathbb{E}_n\left[\left(D_i^- \sigma_+(\mathbf{X}_i)\sigma_-(\mathbf{X}_i)\right)^p | \mathbf{a}_i\right] \xrightarrow{P} \frac{1}{E[W^+]}E[W^+\left(D^-\sigma_+(\boldsymbol{X}_0')\sigma_-(\boldsymbol{X}_0')\right)^p] =: c^p < \infty,$$

$$\sum_{i=1}^n \frac{W_i^+}{L_n^+}\mathbb{E}_n\left[(\sigma_+(\mathbf{X}_i)\beta(\mathbf{X}_i))^p | \mathbf{a}_i\right] \xrightarrow{P} \frac{1}{E[W^+]}E[W^+\left(\sigma_+(\boldsymbol{X}_0')\beta(\boldsymbol{X}_0')\right)^p] < \infty,$$

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}_n\left[(D_i^- \sigma_-(\mathbf{X}_i))^p \middle| \mathbf{a}_i\right] \xrightarrow{P} E[(D^-\sigma_-(\boldsymbol{X}_0))^p] < \infty,$$

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}_n\left[\beta_-(\mathbf{X}_i)^p | \mathbf{a}_i\right] \xrightarrow{P} E[\beta(\boldsymbol{X}_0)^p] < \infty,$$

$$E\left[\omega_0(\boldsymbol{X}_0, \boldsymbol{V})^p + W^+\omega(\boldsymbol{X}_0', \boldsymbol{V})^p\right] < \infty,$$

for $\boldsymbol{X}_0 = (D^-, D^+, \boldsymbol{A})$, $\boldsymbol{X}_0' = (D^-, D^+ + 1, \boldsymbol{A})$, D^- and D^+ are conditionally independent, given $(W^-, W^+, \boldsymbol{A})$, Poisson random variables with means $E[W^+]W^-/\theta$ and $E[W^-]W^+/\theta$, respectively, $(W^-, W^+, \boldsymbol{A})$ is distributed according to υ, and $\boldsymbol{V} = \{V_j : j \geq 1\}$ is a sequence of i.i.d. random variables satisfying $E[|V_1|^p] < \infty$, independent of $\boldsymbol{X}_0$.

Remark 2.1. The vertex attributes for the two random graph models are slightly different. For the DCM they are of the form $\mathbf{a}_i = (D_i^-, D_i^+, \mathbf{b}_i)$, where the D_i^- corresponds to the in-degree of vertex i, D_i^+ corresponds to its out-degree, and $\mathbf{b}_i$ may contain additional parameters needed by the map Φ. In particular, we can take $\mathcal{S}' = \mathbb{N} \times \mathbb{N} \times \mathcal{S}''$ for some other separable metric space $\mathcal{S}''$. On the other hand, for the IRD the vertex attributes are of the form $\mathbf{a}_i = (W_i^-, W_i^+, \mathbf{b}_i)$, where W_i^- and W_i^+ are strictly positive real numbers used to control the in-degree and out-degree of vertex i, respectively. In this case, we can take $\mathcal{S}' = \mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{S}''$ for some other separable metric space $\mathcal{S}''$. Note that for the DCM our definition of $\mathbf{X}_i = (D_i^-, D_i^+, \mathbf{a}_i)$ is redundant, but it is convenient to keep it this way to make the notation for the two models consistent.

Assumptions [R] and [G] allow us to establish the main convergence result in our paper. For real-valued distributions, we use d_p to denote the Wasserstein metric of order $p \in [1, \infty)$, given by

$$d_p(\mu, \nu) = \inf \left\{ (\mathbb{E}_n [|X - Y|^p])^{1/p} : \text{law}(X) = \mu, \, \text{law}(Y) = \nu \right\},$$

which admits the explicit representation:

$$d_p(\mu, \nu)^p = \int_0^1 |F^{-1}(u) - G^{-1}(u)|^p du,$$

for $F(x) = \mu((-\infty, x])$, $G(x) = \nu((-\infty, x])$, and $f^{-1}(u) = \inf\{x \in \mathbb{R} : f(x) \geq u\}$ the generalized inverse of function f.

Throughout the rest of the paper, we will use the notation $\boldsymbol{X}_0 = (D^-, D^+, \boldsymbol{A})$ to be a vector having the distribution indicated by Assumption [G]. Because of its role in the construction of the coupled marked Galton-Watson process, we denote its first component $D^- = \mathcal{N}_0$. In addition, we will use the notation $\boldsymbol{X}$ to denote the size-biased version of $\boldsymbol{X}_0$, which in the two models looks slightly different:

- For the DCM:
$$P(\boldsymbol{X} \in B) = \frac{1}{E[\mathscr{D}^+]} E \left[\mathscr{D}^+ 1((\mathscr{D}^-, \mathscr{D}^+, \boldsymbol{A}) \in B) \right],$$

- For the IRD:
$$P(\boldsymbol{X} \in B) = \frac{1}{E[W^+]} E \left[W^+ 1((D^-, D^+ + 1, \boldsymbol{A}) \in B) \right].$$

The first component of $\boldsymbol{X}$ will be denoted $\mathcal{N}$.

We are now ready to state our main theorem, which states that there exists a coupling between the trajectory of the process $\{R_I^{(r)} : 0 \leq r \leq k\}$ along a uniformly chosen vertex $I \in V_n$, and the trajectory of the root node in an equivalent process constructed on the limiting tree.

Theorem 2.2. *Suppose the map Φ satisfies Assumption [R] and the directed graph $G(V_n, E_n)$ satisfies Assumption [G]. Let $\mu_{k,n}(\cdot) = \mathbb{P}_n \left(R_I^{(k)} \in \cdot \right)$, where I is uniformly distributed in $\{1, 2, \ldots, n\}$. Then, for any fixed $k \geq 0$ there exists a sequence of random variables $\{\mathcal{R}_\emptyset^{(0)}, \mathcal{R}_\emptyset^{(1)}, \ldots, \mathcal{R}_\emptyset^{(k)}\}$ constructed on the same probability space as $\{R_I^{(0)}, R_I^{(1)}, \ldots, R_I^{(k)}\}$, such that*

$$\max_{0 \leq r \leq k} \mathbb{E}_n \left[\left| R_I^{(r)} - \mathcal{R}_\emptyset^{(r)} \right|^p \right] \xrightarrow{P} 0, \qquad n \to \infty.$$

In particular, this implies that if $\nu_k(\cdot) = P\left(\mathcal{R}_\emptyset^{(k)} \in \cdot\right)$*, then, for any fixed* $k \geq 1$*,*

$$d_p(\mu_{k,n}, \nu_k) \xrightarrow{P} 0, \qquad n \to \infty.$$

Moreover, if $c \in (0,1)$ *in Assumption [G](c), then there exists a probability measure* ν *such that*

$$d_p(\nu_k, \nu) \to 0 \quad a.s., \qquad k \to \infty,$$

where ν *is the probability measure of a random variable* $\mathcal{R}^*$ *that satisfies:*

$$\mathcal{R}^* = \Phi\left(\boldsymbol{X}_0, \zeta, \{\mathcal{V}_j, \xi_j : 1 \leq j \leq \mathcal{N}_0\}\right),$$

with the $\{\mathcal{V}_j\}$ *i.i.d. copies of* $\mathcal{V}$*, independent of* $\boldsymbol{X}_0$ *and of* $(\zeta, \{\xi_j : j \geq 1\})$*, and* $\mathcal{V}$ *the special endogenous solution to the distributional fixed-point equation:*

$$\mathcal{V} \stackrel{\mathcal{D}}{=} \Psi\left(\boldsymbol{\Upsilon}, \{\mathcal{V}_j : 1 \leq j \leq \mathcal{N}\}\right) := g\left(\Phi\left(\boldsymbol{X}, \zeta, \{\mathcal{V}_j, \xi_j : 1 \leq j \leq \mathcal{N}\}\right), \boldsymbol{X}\right),$$

where $\boldsymbol{\Upsilon} = (\boldsymbol{X}, \zeta, \{\xi_j : j \geq 1\})$*, and the* $\{\mathcal{V}_j\}$ *are i.i.d. copies of* $\mathcal{V}$*, independent of* $\boldsymbol{\Upsilon}$*.*

Remark 2.3. Note that the map that defines a distributional fixed point equation is the composition

$$\Psi\left(\mathbf{y}, \{v_j : 1 \leq j \leq d^-\}\right) = g\left(\Phi\left(\mathbf{x}, \zeta, \{v_j, \xi_j : 1 \leq j \leq d^-\}\right), \mathbf{x}\right),$$

for $\mathbf{x} = (d^-, d^+, \mathbf{a})$, $\mathbf{y} = (\mathbf{x}, \zeta, \{\xi_j : j \geq 1\})$, and $\mathbf{v} = \{v_j : j \geq 1\}$, not the map Φ itself. The noises ζ and $\{\xi_j : j \geq 1\}$ are always assumed to be independent of everything else, and are essentially "absorbed" into the marks of the Galton-Watson process where the solution $\mathcal{V}$ to the branching distributional fixed-point equation is constructed, represented by $\mathbf{y}$ above (see Section 3 for more details).

As an immediate corollary we obtain the convergence of R_I provided Φ also defines a contraction on $\mathbb{R}^n$.

Corollary 2.4. *Suppose that in addition to Assumptions [R] and [G], we have* $\limsup_{n\to\infty} \|\mathbf{C}\|_p \leq c' < 1$ *P-a.s. Let* $\mu_n(\cdot) = \mathbf{P}_n(R_I \in \cdot)$*. Then,*

$$d_p(\mu_n, \nu) \xrightarrow{P} 0, \qquad n \to \infty,$$

for ν *the limiting measure in Theorem 2.2.*

3 Solutions to branching recursions

In order to explain what the special endogenous solution to a branching distributional equation is, we will first need to construct a marked Galton-Watson process. To do this, we will start by assigning to each node in the tree a label of the form $\mathbf{i} = (i_1, i_2, ..., i_k) \in \mathcal{U}$, where $\mathcal{U} = \bigcup_{k=0}^\infty \mathbb{N}_+^k$ with the convention that $\mathbb{N}_+^0 = \{\emptyset\}$ denotes its root. To simplify the notation, we also write $\mathbf{i} = i$ instead of $\mathbf{i} = (i)$ for labels of length one. The index concatenation operation is denoted by $(\mathbf{i}, j) = (i_1, i_2, ..., i_k, j)$.

Next, let $\{(\mathcal{N}_{\mathbf{i}}, \mathbf{\Upsilon}_{\mathbf{i}}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ be a sequence of i.i.d. copies of some generic branching vector $(\mathcal{N}, \mathbf{\Upsilon})$, where $\mathcal{N}_{\mathbf{i}} \in \mathbb{N}$ represents the number of offspring of node $\mathbf{i}$ and $\mathbf{\Upsilon}_{\mathbf{i}}$ denotes its mark. Let $(\mathcal{N}_\emptyset, \mathbf{\Upsilon}_\emptyset) \stackrel{\mathcal{D}}{=} (\mathcal{N}_0, \mathbf{\Upsilon}_0)$, be a vector independent of $\{(\mathcal{N}_{\mathbf{i}}, \mathbf{\Upsilon}_{\mathbf{i}}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$, whose distribution is allowed to be different from that of the i.i.d. sequence. In the context of our paper, $\mathcal{N}_{\mathbf{i}}$ corresponds to the first component of the vector $\mathbf{\Upsilon}_{\mathbf{i}}$, but this does not always have to be the case.

For the recursions studied in this paper, we have

$$\mathbf{\Upsilon}_{\mathbf{i}} = (\boldsymbol{X}_{\mathbf{i}}, \zeta_{\mathbf{i}}, \{\xi_{(\mathbf{i},j)} : j \geq 1\}),$$

where $\boldsymbol{X}_{\mathbf{i}} \stackrel{\mathcal{D}}{=} \boldsymbol{X}_0$, the $\{\boldsymbol{X}_{\mathbf{i}} : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ are i.i.d. copies of $\boldsymbol{X}$, is independent of $\boldsymbol{X}_\emptyset$, and the noise sequence $\{(\zeta_{\mathbf{i}}, \{\xi_{(\mathbf{i},j)}\}_{j\geq 1}) : \mathbf{i} \in \mathcal{U}\}$ is i.i.d. and independent of $\{\boldsymbol{X}_{\mathbf{i}} : \mathbf{i} \in \mathcal{U}\}$. Moreover, since the number of offspring in the tree corresponds to the in-degree of vertices in $G(V_n, E_n)$, we have that $\mathcal{N}_{\mathbf{i}}$ is the first component of vector $\mathbf{\Upsilon}_{\mathbf{i}}$.

To construct the tree, use the sequence $\{\mathcal{N}_{\mathbf{i}} : \mathbf{i} \in \mathcal{U}\}$ to define the set of nodes in the kth generation according to:

$$\mathcal{A}_0 = \{\emptyset\} \qquad \text{and} \qquad \mathcal{A}_k = \{(\mathbf{i}, j) \in \mathcal{U} : \mathbf{i} \in \mathcal{A}_{k-1}, 1 \leq j \leq \mathcal{N}_{\mathbf{i}}\}, \quad k \geq 1.$$

Next, for $k \geq 1$, use this marked Galton-Watson tree to construct the random variables:

$$\begin{aligned} \mathcal{V}_{\mathbf{i}}^{(1)} = g(r_0, \boldsymbol{X}_{\mathbf{i}}) \quad \mathbf{i} \in \mathcal{A}_k, \quad \mathcal{V}_{\mathbf{i}}^{(r)} &= \Psi\left(\mathbf{\Upsilon}_{\mathbf{i}}, \left\{\mathcal{V}_{(\mathbf{i},j)}^{(r-1)} : 1 \leq j \leq \mathcal{N}_{\mathbf{i}}\right\}\right) \quad \mathbf{i} \in \mathcal{A}_{k-r},\ 2 \leq r \leq k, \\ \mathcal{R}_\emptyset^{(k)} &= \Phi\left(\boldsymbol{X}_\emptyset, \zeta_\emptyset, \left\{\mathcal{V}_j^{(k)}, \xi_j : 1 \leq j \leq \mathcal{N}_\emptyset\right\}\right), \end{aligned} \tag{3.1}$$

for some constant $r_0 \in \mathbb{R}$.

To see that $\mathcal{R}_\emptyset^{(k)}$ is finite and obtain a bound for its pth moment, note that under Assumption [R] we have for any $k \geq 1$,

$$\begin{aligned} \left(E\left[\left|\mathcal{R}_\emptyset^{(k)}\right|^p\right]\right)^{1/p} &\leq \left(E\left[\left(\sigma_-(\boldsymbol{X}_\emptyset)\sum_{j=1}^{\mathcal{N}_\emptyset}\left|\mathcal{V}_j^{(k)}\right| + \beta(\boldsymbol{X}_\emptyset)\right)^p\right]\right)^{1/p} \\ &\leq \left(E\left[\left(\sigma_-(\boldsymbol{X}_\emptyset)\sum_{j=1}^{\mathcal{N}_\emptyset}\left(E\left[\left|\mathcal{V}_j^{(k)}\right|^p\right]\right)^{1/p} + \beta(\boldsymbol{X}_\emptyset)\right)^p\right]\right)^{1/p} \\ &= \left(E\left[\left(\sigma_-(\boldsymbol{X}_0)\mathcal{N}_0\left(E\left[\left|\mathcal{V}_1^{(k)}\right|^p\right]\right)^{1/p} + \beta(\boldsymbol{X}_0)\right)^p\right]\right)^{1/p} \\ &\leq \left(E\left[\left|\mathcal{V}_1^{(k)}\right|^p\right]\right)^{1/p}\left(E\left[(\mathcal{N}_0\sigma_-(\boldsymbol{X}_0))^p\right]\right)^{1/p} + \left(E\left[\beta(\boldsymbol{X}_0)^p\right]\right)^{1/p}, \end{aligned}$$

where in the second and third inequalities we used Minkowski's inequality, the first time conditionally on $\boldsymbol{X}_\emptyset$.

A slight modification of the same steps yields

$$\left(E\left[\left|\mathcal{V}_\emptyset^{(k-1)}\right|^p\right]\right)^{1/p} \leq \left(E\left[\left|\mathcal{V}_1^{(k)}\right|^p\right]\right)^{1/p}\left(E\left[(\mathcal{N}\sigma_-(\boldsymbol{X})\sigma_+(\boldsymbol{X}))^p\right]\right)^{1/p} + \left(E\left[(\sigma_+(\boldsymbol{X})\beta(\boldsymbol{X}))^p\right]\right)^{1/p}$$

$$= c\left(E\left[\left|\mathcal{V}_1^{(k)}\right|^p\right]\right)^{1/p} + b,$$

where $b = (E[(\sigma_+(\boldsymbol{X})\beta(\boldsymbol{X}))^p])^{1/p}$. Now let $a_k = \left(E\left[\left|\mathcal{V}_\emptyset^{(k-1)}\right|^p\right]\right)^{1/p}$ and rewrite the inequality we just derived as:

$$a_k \leq b + ca_{k-1} \leq b\sum_{s=0}^{k-2} c^s + c^k a_1,$$

where $a_1^p = E[|g(r_0, \mathbf{X})|^p] < \infty$ by assumption.

Therefore, for any $k \geq 1$,

$$\left(E\left[\left|\mathcal{R}_\emptyset^{(k)}\right|^p\right]\right)^{1/p} \leq 1(k \geq 2)b\sum_{s=0}^{k-2} c^s + c^k a_1,$$

which is finite for any fixed k, and finite as $k \to \infty$ provided $c < 1$.

Theorem 2.2 shows that if $c < 1$, $\nu_k(\cdot) = P\left(\mathcal{V}_1^{(k)} \in \cdot\right)$ converges in d_p, which ensures the existence of a probability measure ν such that $d_p(\nu_k, \nu) \to 0$ as $k \to \infty$. Moreover, as the proof of the theorem will show, this measure ν is an *endogenous solution* to the branching distributional fixed-point equation

$$\mathcal{V} \stackrel{\mathcal{D}}{=} \Psi\left(\boldsymbol{\Upsilon}, \{\mathcal{V}_j : 1 \leq j \leq \mathcal{N}\}\right),$$

where the $\{\mathcal{V}_j : j \geq 1\}$ are i.i.d. copies of $\mathcal{V}$, independent of $\boldsymbol{\Upsilon}$, i.e., $\nu(\cdot) = P(\mathcal{V} \in \cdot)$. We refer the reader to [3] and [48] for a thorough discussion of the notion of *endogeny* and its characterization.

We also point out that, in general, the existence of a solution $\mathcal{V}$ to the branching distributional fixed-point equation does not require the map Ψ to define a contraction (i.e., $c \in (0, 1)$), but the contraction approach is the easiest to state with the level of generality we aim for, since other approaches (e.g., stochastic monotonicity) may require more specific knowledge of the map Ψ.

The reader may also find interesting that branching distributional fixed-point equations such as the one above do not in general have unique solutions. We refer the reader to the work in [5, 7, 9] for a detailed characterization of the multiple solutions for the linear and maximum recursions

$$\mathcal{R} \stackrel{\mathcal{D}}{=} \mathcal{Q} + \sum_{i=1}^{\mathcal{N}} \mathcal{C}_i \mathcal{R}_i \qquad \text{and} \qquad \mathcal{R} \stackrel{\mathcal{D}}{=} \mathcal{Q} \vee \bigvee_{i=1}^{\mathcal{N}} \mathcal{C}_i \mathcal{R}_i.$$

However, in the context of the present work, the solution $\mathcal{V}$ appearing in the characterization of $\mathcal{R}^*$ is, by construction, the so-called special endogenous solution.

4 Proofs

This section contains the proofs of Theorem 1.4, Theorem 2.2 and Corollary 2.4. Since the proof of Theorem 2.2 consists of several steps, to ease the reading we will start with the proofs of Theorem 2.2 and Corollary 2.4, which are more straightforward.

Proof of Theorem 1.4. Let $\{\xi_{j,i}^{(k)} : i, j \geq 0,\, k \in \mathbb{Z}\}$ and $\{\zeta_i^{(k)} : i \in V_n,\, k \in \mathbb{Z}\}$ are each sequences of i.i.d. noises independent of each other and of $\mathscr{G}_n$. Start by constructing an array of random vectors $\{\tilde{\mathbf{R}}^{(k,m)} : k \geq 0,\, m \geq 0\}$ according to the recursion $\tilde{\mathbf{R}}^{(0,m)} = \mathbf{r}^{(0)}$ for all $m \geq 0$ and

$$\tilde{R}_i^{(k+1,m)} = \Phi\left(\mathbf{X}_i, \zeta_i^{(k-m)}, \left\{g(\tilde{R}_j^{(k,m)}, \mathbf{X}_j), \xi_{j,i}^{(k-m)} : j \to i\right\}\right), \qquad i \in V_n, \quad k \geq 1.$$

Note that $\mathbf{R}^{(k)} = \tilde{\mathbf{R}}^{(k,0)}$ for all $k \geq 0$, and observe that we are shifting the sequence of noises as m changes. Moreover, since the noise sequences consist of sets of i.i.d. random variables, $\tilde{\mathbf{R}}^{(k,m)}$ and $\mathbf{R}^{(k)}$ have the same distribution for each $k \geq 0$.

Next, note that for any $m \geq 1$,

$$\begin{aligned}\left\|\mathbf{R}^{(k)} - \tilde{\mathbf{R}}^{(k+m,m)}\right\|_p &\leq \left\|\tilde{\mathbf{R}}^{(k,0)} - \tilde{\mathbf{R}}^{(k+1,1)}\right\|_p + \left\|\tilde{\mathbf{R}}^{(k+1,1)} - \tilde{\mathbf{R}}^{(k+m,m)}\right\|_p \\ &\leq \sum_{i=0}^{m-1} \left\|\tilde{\mathbf{R}}^{(k+i,i)} - \tilde{\mathbf{R}}^{(k+i+1,i+1)}\right\|_p,\end{aligned}$$

which in turn implies, by Minkowski's inequality, that

$$\left(\mathbf{E}_n\left[\left\|\mathbf{R}^{(k)} - \tilde{\mathbf{R}}^{(k+m,m)}\right\|_p^p\right]\right)^{1/p} \leq \sum_{i=0}^{m-1}\left(\mathbf{E}_n\left[\left\|\tilde{\mathbf{R}}^{(k+i,i)} - \tilde{\mathbf{R}}^{(k+i+1,i+1)}\right\|_p^p\right]\right)^{1/p}.$$

Moreover, for any $k \geq 1$, Assumption [R] gives

$$\begin{aligned}\mathbf{E}_n\left[\left\|\tilde{\mathbf{R}}^{(k+i,i)} - \tilde{\mathbf{R}}^{(k+i+1,i+1)}\right\|_p^p\right] &\leq \mathbf{E}_n\left[E\left[\left\|\tilde{\mathbf{R}}^{(k+i,i)} - \tilde{\mathbf{R}}^{(k+i+1,i+1)}\right\|_p^p \middle| \mathscr{G}_n, \tilde{\mathbf{R}}^{(k+i-1,i)}, \tilde{\mathbf{R}}^{(k+i,i+1)}\right]\right] \\ &\leq \mathbf{E}_n\left[\sum_{l=1}^{n}\left(\sum_{j\to l}\sigma_-(\mathbf{X}_l)\sigma_+(\mathbf{X}_j)\left|R_j^{(k+i-1,i)} - R_j^{(k+i,i+1)}\right|\right)^p\right] \\ &= \mathbf{E}_n\left[\sum_{l=1}^{n}\left(\sum_{j=1}^{n} C_{l,j}\left|R_j^{(k+i-1,i)} - R_j^{(k+i,i+1)}\right|\right)^p\right] \\ &\leq \mathbf{E}_n\left[\|\mathbf{C}\|_p^p\left\|\tilde{\mathbf{R}}^{(k+i-1,i)} - \tilde{\mathbf{R}}^{(k+i,i+1)}\right\|_p^p\right] \\ &\leq \left(\|\mathbf{C}\|_p^p\right)^{k+i}\mathbf{E}_n\left[\left\|\tilde{\mathbf{R}}^{(0,i)} - \tilde{\mathbf{R}}^{(1,i+1)}\right\|_p^p\right] \\ &= \|\mathbf{C}\|_p^{p(k+i)}\mathbf{E}_n\left[\left\|\mathbf{r}^{(0)} - \mathbf{R}^{(1)}\right\|_p^p\right].\end{aligned}$$

Hence,

$$\begin{aligned}\left(\mathbf{E}_n\left[\left\|\mathbf{R}^{(k)} - \tilde{\mathbf{R}}^{(k+m,m)}\right\|_p^p\right]\right)^{1/p} &\leq \sum_{i=0}^{m-1}\|\mathbf{C}\|_p^{k+i}\left(\mathbf{E}_n\left[\left\|\mathbf{r}^{(0)} - \mathbf{R}^{(1)}\right\|_p^p\right]\right)^{1/p} \\ &= \left(\mathbf{E}_n\left[\left\|\mathbf{r}^{(0)} - \mathbf{R}^{(1)}\right\|_p^p\right]\right)^{1/p}\|\mathbf{C}\|_p^k\sum_{i=0}^{m-1}\|\mathbf{C}\|_p^i\end{aligned}$$

$$\leq \left(\mathbf{E}_n \left[\left\| \mathbf{r}^{(0)} - \mathbf{R}^{(1)} \right\|_p^p \right] \right)^{1/p} \frac{\|\mathbf{C}\|_p^k}{1 - \|\mathbf{C}\|_p}.$$

It follows that the sequence $\{\lambda_{k,n} : k \geq 0\}$ is Cauchy under the Wasserstein metric:

$$W_p(\nu, \mu) := \inf \left\{ \left(\mathbf{E}_n \left[\|\mathbf{X} - \mathbf{Y}\|_p^p \right] \right)^{1/p} : \text{law}(\mathbf{X}) = \mu, \text{law}(\mathbf{Y}) = \nu \right\}.$$

We then conclude that there exists a measure λ_n on $\mathbb{R}^n$ such that

$$\lim_{k \to \infty} W_p(\lambda_{k,n}, \lambda_n) = 0 \qquad \mathbf{P}_n\text{-a.s.}$$

Since optimal couplings are guaranteed to exist, there exists $\mathbf{R}$ distributed according to λ_n for which the statement of the theorem holds. This completes the proof. □

We now give the proof of Corollary 2.4.

Proof of Corollary 2.4. Note that

$$d_p(\mu_n, \mu_{n,k})^p \leq \mathbf{E}_n \left[\left| R_I^{(k)} - R_I \right|^p \right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}_n \left[\left| R_i^{(k)} - R_i \right|^p \right] = \frac{1}{n} \mathbf{E}_n \left[\left\| \mathbf{R}^{(k)} - \mathbf{R} \right\|_p^p \right].$$

Now use the triangle inequality, Theorem 1.4 and Theorem 2.2 to obtain that for any $k \geq 1$,

$$\begin{aligned}
\lim_{n \to \infty} d_p(\mu_n, \nu) &\leq \limsup_{n \to \infty} \left(d_p(\mu_n, \mu_{k,n}) + d_p(\mu_{k,n}, \nu_k) + d_p(\nu_k, \nu) \right) \\
&\leq \limsup_{n \to \infty} \left(n^{-1/p} \left(\mathbf{E}_n \left[\left\| \mathbf{r}^{(0)} - \mathbf{R}^{(1)} \right\|_p^p \right] \right)^{1/p} \frac{\|\mathbf{C}\|_p^k}{1 - \|\mathbf{C}\|_p} + d_p(\mu_{k,n}, \nu_k) + d_p(\nu_k, \nu) \right).
\end{aligned}$$

To complete the proof, we need to show that

$$\limsup_{n \to \infty} n^{-1/p} \left(\mathbf{E}_n \left[\left\| \mathbf{r}^{(0)} - \mathbf{R}^{(1)} \right\|_p^p \right] \right)^{1/p} < \infty.$$

Note that by Minkowski inequality, we have

$$\begin{aligned}
n^{-1/p} \left(\mathbf{E}_n \left[\left\| \mathbf{r}^{(0)} - \mathbf{R}^{(1)} \right\|_p^p \right] \right)^{1/p} &\leq n^{-1/p} \mathbf{E}_n \left[\left(\|\mathbf{r}^{(0)}\|_p + \|\mathbf{R}^{(1)}\|_p \right)^p \right]^{1/p} \\
&\leq n^{-1/p} \left(\|\mathbf{r}^{(0)}\|_p + \mathbf{E}_n \left[\|\mathbf{R}^{(1)}\|_p^p \right]^{1/p} \right)
\end{aligned}$$

Then by Assumption [R](3) and Assumption [G](c), $\|\mathbf{r}^{(0)}\|_p = n^{1/p} |r_0| < \infty$ and

$$\begin{aligned}
n^{-1} \mathbf{E}_n \left[\|\mathbf{R}^{(1)}\|_p^p \right] &= \frac{1}{n} \sum_{i=1}^n \mathbf{E}_n \left[|R_i^{(1)}|^p \right] \\
&\leq \frac{1}{n} \sum_{i=1}^n \left(\sum_{j \to i} \sigma_-(\mathbf{X}_i) |r_{0,i}| + \beta(\mathbf{X}_i) \right)^p
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\left(D_i^-\sigma_-(\mathbf{X}_i)|r_0| + \beta(\mathbf{X}_i)\right)^p \\
&\leq \left(|r_0| \left(\frac{1}{n}\sum_{i=1}^{n}(D_i^-\sigma_-(\mathbf{X}_i))^p \right)^{1/p} + \left(\frac{1}{n}\sum_{i=1}^{n}\beta(\mathbf{X}_i)^p \right)^{1/p} \right)^p \\
&\xrightarrow{P} \left(|r_0| \left(E\left[\mathcal{N}\sigma_-(\boldsymbol{X}_0))^p \right] \right)^{1/p} + \left(E\left[\beta(\boldsymbol{X}_0)^p \right] \right)^{1/p} \right)^{1/p} < \infty.
\end{aligned}
$$

□

4.1 Proof of Theorem 2.2

The remainder of the paper contains the proof of Theorem 2.2, which we will separate into three main steps. The first one uses a previously established coupling theorem [51] for the exploration of the inbound neighborhood of a uniformly chosen vertex in $G(V_n, E_n)$ with a marked Galton-Watson process. The offspring distribution and that of their marks are still dependent on the graph $G(V_n, E_n)$, so the second step consists in removing this dependence. The last step establishes the convergence of infinitely many iterations of the recursion on the limiting marked Galton-Watson process.

We start by identifying the distribution of the number of offspring and the marks in the coupling theorem, ignoring the noises, which we do separately for the two models we consider for $G(V_n, E_n)$. In the coupling, the randomly chosen vertex I in $G(V_n, E_n)$ will be identified with the root of its coupled tree, denoted $\emptyset$; its number of offspring corresponds to the in-degree of vertex I, and on the tree will be denoted $\hat{N}_\emptyset$, and its mark will be the vector $\mathbf{X}_I$, which on the tree will be denoted $\hat{\mathbf{X}}_\emptyset$. Note that $\hat{\mathbf{X}}_\emptyset$ contains $\hat{N}_\emptyset$ as its first component. After the first step, any other vertex encountered in the exploration of the in-component of vertex I will be coupled with an i.i.d. copy of the vector $(\hat{N}, \hat{\mathbf{X}})$, which will receive a tree label of the form $\mathbf{i} = (i_1, \dots, i_k)$ according to the order in which they appear in the exploration. It is important to point out that due to the size-bias induced by exploring the inbound neighbors of a vertex, the distribution of $(\hat{N}, \hat{\mathbf{X}})$ will in general be different from that of $(\hat{N}_\emptyset, \hat{\mathbf{X}}_\emptyset)$, which is unaffected by the size-bias. To include the noises into the coupling we use the observation that they are by assumption independent of the graph $G(V_n, E_n)$, and therefore, we can simply attach to each of the marks in the tree the noises that correspond to their coupled vertex in $G(V_n, E_n)$ at the time we encounter it (see the proof of Theorem 4.1). Since the noises are i.i.d. from one iteration to the next, when coupling $R_I^{(k)}$ for a fixed $k \geq 1$ we do not need to know in which iteration we encountered the node, and we can simply refer to $\hat{\mathbf{Y}}_\emptyset = \left(\hat{\mathbf{X}}_\emptyset, \zeta_\emptyset, \{\xi_j : j \geq 1\}\right)$ as the full mark of the root $\emptyset$, and $\hat{\mathbf{Y}}_\mathbf{i} = \left(\hat{\mathbf{X}}_\mathbf{i}, \zeta_\mathbf{i}, \{\xi_{(\mathbf{i},j)} : j \geq 1\}\right)$ as the full mark of node $\mathbf{i} \neq \emptyset$, with the sequence $\{(\zeta_\mathbf{i}, \{\xi_{(\mathbf{i},j)} : j \geq 1\}) : \mathbf{i} \in \mathcal{U}\}$ i.i.d. and independent of $\{\hat{\mathbf{X}}_\mathbf{i} : \mathbf{i} \in \mathcal{U}\}$. The distributions of $\hat{\mathbf{X}}_\emptyset$ and $\hat{\mathbf{X}}$ for each of the two models are given below:

- **Directed configuration model:**

$$
\mathbb{P}_n\left(\hat{\mathbf{X}}_\emptyset \in B\right) = \frac{1}{n}\sum_{i=1}^{n} 1\left((D_i^-, D_i^+, \mathbf{a}_i) \in B\right)
$$

$$\mathbb{P}_n\left(\hat{\mathbf{X}} \in B\right) = \sum_{i=1}^{n} \frac{D_i^+}{L_n} \cdot 1\left(\left(D_i^-, D_i^+, \mathbf{a}_i\right) \in B\right),$$

for any measurable set $B \subseteq \mathcal{S}'$.

- **Inhomogeneous random digraph:** Let a_n, b_n be sequences satisfying $a_n \wedge b_n \xrightarrow{P} \infty$ and $a_n b_n/n \xrightarrow{P} 0$ as $n \to \infty$, and define $\bar{W}_i^- = W_i^- \wedge a_n$, $\bar{W}_i^+ = W_i^+ \wedge b_n$; set $\Lambda_n^- = \sum_{i=1}^n \bar{W}_i^-$ and $\Lambda_n^+ = \sum_{i=1}^n \bar{W}_i^+$.

$$\mathbb{P}_n\left(\hat{\mathbf{X}}_\emptyset \in B\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}_n\left(\left(D_i^-, D_i^+, \mathbf{a}_i\right) \in B\right)$$
$$\mathbb{P}_n\left(\hat{\mathbf{X}} \in B\right) = \sum_{i=1}^{n} \frac{\bar{W}_i^+}{\Lambda_n^+} \cdot \mathbb{P}_n\left(\left(D_i^-, D_i^+ + 1, \mathbf{a}_i\right) \in B\right),$$

for any measurable set $B \subseteq \mathcal{S}'$. Moreover, conditionally on $\mathscr{F}_n$, D_i^- and D_i^+ are independent Poisson random variables with means $\Lambda_n^+ \bar{W}_i^-/(\theta n)$ and $\Lambda_n^- \bar{W}_i^+/(\theta n)$, respectively.

The marked Galton-Watson process, as defined in Section 3, is now constructed for any fixed depth using the sequence of independent vectors $\{(\hat{N}_\mathbf{i}, \hat{\mathbf{Y}}_\mathbf{i}) : \mathbf{i} \in \mathcal{U}\}$, with $\{(\hat{N}_\mathbf{i}, \hat{\mathbf{Y}}_\mathbf{i}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ i.i.d. copies of $(\hat{N}, \hat{\mathbf{Y}})$. Specifically, its generations are defined recursively via:

$$\hat{A}_0 = \{\emptyset\} \qquad \text{and} \qquad \hat{A}_k = \{(\mathbf{i}, j) \in \mathcal{U} : \mathbf{i} \in \hat{A}_{k-1}, 1 \leq j \leq \hat{N}_\mathbf{i}\}, \quad k \geq 1.$$

Next, for any fixed $k \geq 1$, use this marked tree to construct the random variables:

$$\hat{V}_\mathbf{i}^{(1)} = g(r_0, \hat{\mathbf{X}}_\mathbf{i}) \quad \mathbf{i} \in \hat{A}_k,$$
$$\hat{V}_\mathbf{i}^{(r+1)} = \Psi\left(\hat{\mathbf{Y}}_\mathbf{i}, \left\{\hat{V}_{(\mathbf{i},j)}^{(r)} : 1 \leq j \leq \hat{N}_\mathbf{i}\right\}\right) \quad \mathbf{i} \in \hat{A}_{k-r},\ 1 \leq r < k, \tag{4.1}$$
$$\text{and} \quad \hat{R}_\emptyset^{(k)} = \Phi\left(\hat{\mathbf{X}}_\emptyset, \zeta_\emptyset, \left\{\hat{V}_j^{(k)}, \xi_j : 1 \leq j \leq \hat{N}_\emptyset\right\}\right), \quad k \geq 1, \tag{4.2}$$

for some constant $r_0 \in \mathbb{R}$.

The coupling theorem given below is a rewording of Theorem 6.3 in [51], where the PageRank recursion was analyzed. The result states that for any vertex i in $G(V_n, E_n)$ there exists a coupling between the breadth-first exploration of its in-component and the construction of a marked Galton-Watson process having the distribution described above. In the statement of the result, τ_i denotes the generation in the coupled tree where the first miscoupling occurs. A miscoupling occurs when the exploration of the graph encounters a vertex that had already been discovered, a multiple edge between two vertices (in the DCM), or a disagreement between the attribute of a vertex and the attribute of its tree counterpart (in the IRD). Specifically, if $\tau_i > k$, then the in-component of depth k of vertex i and the first k generations of its coupled tree are identical, up to the vertex attributes/marks. Note that having identical attributes implies $\mathbf{X}_s = \hat{\mathbf{X}}_\mathbf{i}$ if $\mathbf{i}$ is the tree label that corresponds to vertex s. However, the exploration of the two models differs in the amount of information we gain from successfully coupling a vertex. In the DCM, successfully coupling the exploration of vertex s tells us only that $\mathbf{X}_s = \hat{\mathbf{X}}_\mathbf{i}$, while in the IRD we also gain the identities of the D_s^- inbound neighbors of vertex s, say $\{t_1, \ldots, t_{D_s^-}\}$, and hence the node attributes $\hat{\mathbf{A}}_{(\mathbf{i},j)} := \left(\bar{W}_{t_j}^-, \bar{W}_{t_j}^+, \mathbf{b}_{t_j}\right)$ of the $\hat{N}_\mathbf{i} = D_s^-$ offspring of $\mathbf{i}$.

Theorem 4.1. *Suppose $G(V_n, E_n)$ satisfies Assumption [G]. Then, there exists a coupling*

$$(R_I^{(0)}, R_I^{(1)}, \dots, R_I^{(k)}, \hat{R}_\emptyset^{(0)}, \hat{R}_\emptyset^{(1)}, \dots, \hat{R}_\emptyset^{(k)})$$

such that for any fixed $k \geq 0$,

$$\mathbb{P}_n \left(\bigcup_{r=0}^{k} \left\{ R_I^{(r)} \neq \hat{R}_\emptyset^{(r)} \right\} \right) \leq \mathbb{P}_n(\tau_I \leq k) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_n \left(\tau_i \leq k \right) \xrightarrow{P} 0, \qquad n \to \infty.$$

Proof. The proof would be immediate from the statement of Theorem 6.3 in [51] if it was not for the presence of noises. To make the appropriate correspondence, suppose we are exploring the in-component of vertex I in $G(V_n, E_n)$ and we want to construct $(R_I^{(k)}, \hat{R}_\emptyset^{(k)})$. Then, if vertex j is at an inbound distance s from vertex I, it will have a tree label of the form $\mathbf{i} = (i_1, \dots, i_s)$ and a mark of the form $\hat{\mathbf{Y}}_\mathbf{i} = \mathbf{Y}_j^{(k-s)}$. To construct a different pair $(R_I^{(r)}, \hat{R}_\emptyset^{(r)})$ for $0 \leq r < k$, take the coupled explorations of the inbound neighborhood of vertex I and the tree of depth k that we used to construct $(R_I^{(k)}, \hat{R}_\emptyset^{(k)})$, but restrict them to have depth r only. To the vertices/nodes at distance r from $I/\emptyset$, assign an initial value r_0, and compute $\hat{R}_I^{(r)}$ according to (4.2) using for a node $\mathbf{i}$ at distance s from $\emptyset$ a label of the form $\hat{\mathbf{Y}}_\mathbf{i} = \mathbf{Y}_j^{(r-s)}$.

Note that for a fixed depth r, $1 \leq r \leq k$, the independence of the noises from the graph $G(V_n, E_n)$ makes all the nodes in the coupled tree, other than the root, have the same distribution. $\square$

Recall that $\mu_{k,n}(\cdot) = \mathbb{P}_n \left(R_I^{(k)} \in \cdot \right)$ and define $\nu_{k,n} = \mathbb{P}_n \left(\hat{R}_\emptyset^{(k)} \in \cdot \right)$. We will use Theorem 4.1 to show that $d_p \left(\mu_{k,n}, \nu_{k,n} \right) \xrightarrow{P} 0$ as $n \to \infty$. Before we do so, it is convenient to show that $d_p \left(\nu_{k,n}, \nu_k \right) \xrightarrow{P} 0$ as $n \to \infty$, where $\nu_k(\cdot) = P \left(\mathcal{R}_\emptyset^{(k)} \in \cdot \right)$, and $\mathcal{R}_\emptyset^{(k)}$ is the value of the recursion computed at the root of the limiting tree after k steps. Note that both $\nu_{k,n}$ and ν_k are constructed on marked Galton-Watson processes, however, the distribution of the marks $\{\hat{\mathbf{X}}_\mathbf{i} : \mathbf{i} \in \mathcal{U}\}$ depends on $\mathscr{G}_n$, while the marks $\{\boldsymbol{X}_\mathbf{i} : \mathbf{i} \in \mathcal{U}\}$ do not.

We will construct couplings of $(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0)$ and $(\hat{\mathbf{X}}, \boldsymbol{X})$ that are based on couplings of the vertex attributes $(\hat{\mathbf{A}}, \boldsymbol{A})$ and their size-biased versions $(\hat{\mathbf{A}}_b, \boldsymbol{A}_b)$:

- **Directed configuration model:**

$$\mathbb{P}_n \left(\hat{\mathbf{A}} \in B \right) = \frac{1}{n} \sum_{i=1}^{n} 1 \left((D_i^-, D_i^+, \mathbf{a}_i) \in B \right);$$

$$\mathbb{P}_n \left(\hat{\mathbf{A}}_b \in B \right) = \sum_{i=1}^{n} \frac{D_i^+}{L_n} \cdot 1 \left((D_i^-, D_i^+, \mathbf{a}_i) \in B \right);$$

$$P(\boldsymbol{A} \in B) = P((\mathscr{D}^-, \mathscr{D}^+, \boldsymbol{B}) \in B);$$

$$P(\boldsymbol{A}_b \in B) = \frac{E[1((\mathscr{D}^-, \mathscr{D}^+, \boldsymbol{B}) \in B)\mathscr{D}^+]}{E[\mathscr{D}^+]};$$

for any measurable set $B \subseteq \mathcal{S}'$.

- **Inhomogeneous random digraph:**

$$\mathbb{P}_n\left(\hat{\mathbf{A}} \in B\right) = \frac{1}{n}\sum_{i=1}^{n} 1\left((\bar{W}_i^-, \bar{W}_i^+, \mathbf{b}_i) \in B\right)$$

$$\mathbb{P}_n\left(\hat{\mathbf{A}}_b \in B\right) = \sum_{i=1}^{n} \frac{\bar{W}_i^+}{\Lambda_n^+} 1\left(\left(\bar{W}_i^-, \bar{W}_i^+ + 1, \mathbf{b}_i\right) \in B\right),$$

$$P(\boldsymbol{A} \in B) = P((W^-, W^+, \boldsymbol{B}) \in B);$$

$$P(\boldsymbol{A}_b \in B) = \frac{E[1((W^-, W^+, \boldsymbol{B}) \in B)W^+]}{E[W^+]};$$

for any measurable set $B \subseteq \mathcal{S}'$.

The convergence of the vertex attributes is established in the following lemma. Note that the convergence in W_1 is only guaranteed for $\hat{\mathbf{A}}$, but not for $\hat{\mathbf{A}}_b$, since the latter is affected by the size-bias.

Lemma 4.2. *Under Assumption [G], there exist couplings $(\hat{\mathbf{A}}, \boldsymbol{A})$ and $(\hat{\mathbf{A}}_b, \boldsymbol{A}_b)$ constructed given $\mathscr{F}_n$ such that*

$$\mathbb{E}_n\left[\rho'(\hat{\mathbf{A}}, \boldsymbol{A})\right] \xrightarrow{P} 0, \qquad \rho'(\hat{\mathbf{A}}, \boldsymbol{A}) \xrightarrow{P} 0 \qquad \textit{and} \qquad \rho'(\hat{\mathbf{A}}_b, \boldsymbol{A}_b) \xrightarrow{P} 0$$

as $n \to \infty$.

Proof. Assumption [G] gives for the DCM that $W_1(\upsilon_n, \upsilon) \xrightarrow{P} 0$ as $n \to \infty$, which gives by definition the existence of $(\hat{\mathbf{A}}, \boldsymbol{A})$ such that

$$\mathbb{E}_n\left[\rho'(\hat{\mathbf{A}}, \boldsymbol{A})\right] = W_1(\upsilon_n, \upsilon) \xrightarrow{P} 0 \qquad \text{and} \qquad \rho'(\hat{\mathbf{A}}, \boldsymbol{A}) \xrightarrow{P} 0$$

as $n \to \infty$. For the IRD, we have that $\hat{\mathbf{A}} = (\hat{W}^- \wedge a_n, \hat{W}^+ \wedge b_n, \hat{\mathbf{B}})$, where $\hat{\mathbf{W}} = (\hat{W}^-, \hat{W}^+, \hat{\mathbf{B}})$ has distribution υ_n and $a_n \wedge b_n \xrightarrow{P} \infty$ as $n \to \infty$. Assumption [G] gives $\mathbb{E}_n\left[\rho'(\hat{\mathbf{W}}, \boldsymbol{A})\right] \xrightarrow{P} 0$, and

$$\mathbb{E}_n\left[\rho'(\hat{\mathbf{A}}, \hat{\mathbf{W}})\right] \leq \mathbb{E}_n\left[(\hat{W}^- - a_n)^+ + (\hat{W}^+ - b_n)^+\right] \xrightarrow{P} 0$$

as $n \to \infty$. This establishes $\mathbb{E}_n\left[\rho'(\hat{\mathbf{A}}, \boldsymbol{A})\right] + \rho'(\hat{\mathbf{A}}, \boldsymbol{A}) \xrightarrow{P} 0$ as $n \to \infty$ for the IRD.

It remains to construct a coupling for the size-biased attributes. To start, let $\hat{\mathbf{A}} = (\hat{A}_1, \hat{A}_2, \hat{\mathbf{A}}_3)$, $\boldsymbol{A} = (A_1, A_2, \boldsymbol{A}_3)$ and define the conditional probabilities

$$\pi_n(B; x) = \mathbb{P}_n\left(\hat{\mathbf{A}} \in B \middle| \hat{A}_2 = x\right) \qquad \text{and} \qquad \pi(B; x) = P\left(\boldsymbol{A} \in B \middle| A_2 = x\right).$$

Note that $x \in \mathbb{N}_+$ if $G(V_n, E_n)$ is a DCM and $x \in \mathbb{R}_+$ if it is an IRD. Let $f : \mathcal{S}' \to \mathbb{R}$ be a continuous function such that $|f(\mathbf{x})| \leq C(1 + |\mathbf{x} - \mathbf{x}_0|)$ for some fixed element $\mathbf{x}_0 \in \mathcal{S}'$. Then, if $\hat{H}(x)$ is distributed according to $\pi_n(\cdot; x)$, the properties of W_1 (see Definition 6.7 and Theorem 6.8 in [58]), give,

$$\mathbb{E}_n\left[|f(\hat{H}(x))|\right] \mathbb{P}_n(\hat{A}_2 = x) = \mathbb{E}_n\left[|f(\hat{\mathbf{A}})| \middle| \hat{A}_2 = x\right] \mathbb{P}_n(\hat{A}_2 = x) = \mathbb{E}_n\left[|f(\hat{\mathbf{A}})| 1(\hat{A}_2 = x)\right]$$

$$\xrightarrow{P} E\left[|f(\boldsymbol{A})|1(A_2 \in x)\right] = E\left[|f(H(x))|\right] P(A_2 \in dx),$$

as $n \to \infty$, where $H(x)$ is distributed according to $\pi(\cdot; x)$. This implies that $\hat{H}(x)$ converges to $H(x)$ in W_1, and therefore, there exists a coupling $(\hat{H}(x), H(x))$ such that for each fixed x,

$$\mathbb{E}_n\left[\rho'(\hat{H}(x), H(x))\right] \xrightarrow{P} 0, \qquad n \to \infty.$$

Now construct $(\hat{A}_{b,2}, A_{b,2}) = (F_n^{-1}(U), F^{-1}(U))$ where $U \sim \text{Uniform}[0,1]$, independent of $\mathscr{F}_n$,

$$F_n(x) := \mathbb{P}_n(\hat{A}_{b,2} \le x) = \begin{cases} \sum_{i=1}^n 1(D_i^+ \le x) D_i^- / L_n, & \text{if } G(V_n, E_n) \text{ is an DCM}, \\ \sum_{i=1}^n 1(\bar{W}_i^+ \le x) \bar{W}_i^- / \Lambda_n^+, & \text{if } G(V_n, E_n) \text{ is an IRD}, \end{cases}$$

and

$$F(x) := P(A_{b,2} \le x) = \begin{cases} E[1(\mathscr{D}^+ \le x)\mathscr{D}^+]/E[\mathscr{D}^+], & \text{if } G(V_n, E_n) \text{ is an DCM}, \\ E[1(W^+ \le x) W^+]/E[W^+], & \text{if } G(V_n, E_n) \text{ is an IRD}. \end{cases}$$

Finally, set

$$(\hat{\mathbf{A}}_b, \boldsymbol{A}_b) = (\hat{H}(\hat{A}_{b,2}), H(A_{b,2})),$$

and note that

$$\rho'(\hat{\mathbf{A}}_b, \boldsymbol{A}_b) \le \rho'(\hat{H}(\hat{A}_{b,2}), H(\hat{A}_{b,2})) + \rho'(H(\hat{A}_{b,2}), H(A_{b,2})),$$

where $\rho'(H(\hat{A}_{b,2}), H(A_{b,2})) \xrightarrow{P} 0$ as $n \to \infty$ whenever $|\hat{A}_{b,2} - A_{b,2}| \xrightarrow{P} 0$ as $n \to \infty$ (which is the case for our construction). To see that $\rho'(\hat{H}(\hat{A}_{b,2}), H(\hat{A}_{b,2})) \xrightarrow{P} 0$ as $n \to \infty$, fix $M > 0$, let x_i be either D_i^+ in the DCM or $\bar{W}_i^+$ in the IRD, and set $\bar{x}(n) = n^{-1} \sum_{i=1}^n x_i$. Then,

$$\begin{aligned} \mathbb{E}_n\left[\rho'(\hat{H}(\hat{A}_{b,2}), H(A_{b,2})) 1(\hat{A}_{b,2} \le M)\right] &= \sum_{i=1}^n \mathbb{E}_n\left[\rho'(\hat{H}(x_i), H(x_i))\right] 1(x_i \le M) \mathbb{P}_n(\hat{A}_{b,2} = x_i) \\ &\le \sum_{i=1}^n \mathbb{E}_n\left[\rho'(\hat{H}(x_i), H(x_i))\right] \frac{M}{n\bar{x}(n)} \\ &\le \frac{M}{\bar{x}(n)} \mathbb{E}_n\left[\rho'(\hat{\mathbf{A}}, \boldsymbol{A})\right] \xrightarrow{P} 0, \end{aligned}$$

as $n \to \infty$, which implies that $\rho'(\hat{H}(\hat{A}_{b,2}), H(A_{b,2})) 1(\hat{A}_{b,2} \le M) \xrightarrow{P} 0$ as $n \to \infty$. Also,

$$P\left(\rho'(\hat{\mathbf{A}}_b, \boldsymbol{A}_b) 1(\hat{A}_{b,2} > M) > \epsilon\right) \le P\left(\hat{A}_{b,2} > M\right) \to P(A_{b.2} > M) \qquad n \to \infty.$$

Taking $M \to \infty$ gives that $\rho'(\hat{H}(\hat{A}_{b,2}), H(\hat{A}_{b,2})) \xrightarrow{P} 0$ as $n \to \infty$. This completes the proof. □

We now use Lemma 4.2 to construct couplings for $(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0)$ and $(\hat{\mathbf{X}}, \boldsymbol{X})$.

Lemma 4.3. *Suppose Assumption [G] holds and let $(\hat{\mathbf{A}}, \boldsymbol{A})$ and $(\hat{\mathbf{A}}_b, \boldsymbol{A}_b)$ be the couplings in Lemma 4.2. Then, there exist couplings for $(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0)$ and $(\hat{\mathbf{X}}, \boldsymbol{X})$ constructed on the same probability space as $(\hat{\mathbf{A}}, \boldsymbol{A})$ and $(\hat{\mathbf{A}}_b, \boldsymbol{A}_b)$, such that*

$$\mathbb{E}_n\left[\rho(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0)\right] \xrightarrow{P} 0, \qquad \rho(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0) \xrightarrow{P} 0 \qquad \text{and} \qquad \rho(\hat{\mathbf{X}}, \boldsymbol{X}) \xrightarrow{P} 0$$

as $n \to \infty$.

Proof. For the DCM simply set $(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0) = (\hat{\mathbf{A}}, \boldsymbol{A})$ and $(\hat{\mathbf{X}}, \boldsymbol{X}) = (\hat{\mathbf{A}}_b, \boldsymbol{A}_b)$.

For the IRD suppose

$$(\hat{\mathbf{A}}, \boldsymbol{A}) = \left(\hat{W}^-, \hat{W}^+, \hat{\mathbf{B}}, W^-, W^+, \boldsymbol{B}\right),$$

and construct

$$\begin{aligned}(\hat{\mathbf{S}}, \boldsymbol{S}) &= \left(\Lambda_n^+(\hat{W}^- \wedge a_n)/(\theta n),\ \Lambda_n^-(\hat{W}^+ \wedge b_n)/(\theta n),\ \hat{\mathbf{B}},\ E[W^+]W^-/\theta,\ E[W^-]W^+/\theta,\ \boldsymbol{B}\right)\\ &=: \left(\hat{S}_1, \hat{S}_2, \hat{\mathbf{S}}_3, S_1, S_2, \boldsymbol{S}_3\right).\end{aligned}$$

Note that our assumptions imply that $\mathbb{E}_n\left[\rho'(\hat{\mathbf{S}}, \boldsymbol{S})\right] \xrightarrow{P} 0$ as $n \to \infty$. Now let $U, U' \sim \text{Uniform}[0,1]$ be i.i.d. and independent of $(\hat{\mathbf{S}}, \boldsymbol{S})$, and take

$$(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0) = \left(G^{-1}(U;\, \hat{S}_1), G^{-1}(U';\, \hat{S}_2), \hat{\mathbf{A}}, G^{-1}(U;\, S_1), G^{-1}(U';\, S_2), \boldsymbol{A}\right),$$

where $G^{-1}(u;\lambda) = \sum_{m=0}^{\infty} m1(G(m;\lambda) \le u < G(m+1;\lambda))$ is the generalized inverse of the Poisson distribution function with mean λ. The continuity in λ of $G(\cdot;\lambda)$ and the properties of the Poisson distribution give $\mathbb{E}_n\left[\rho(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_0)\right] \xrightarrow{P} 0$ as $n \to \infty$.

We now need to construct an explicit coupling for $\hat{\mathbf{X}}$ and $\boldsymbol{X}$, which correspond to the size-biased versions of $\hat{\mathbf{X}}_\emptyset$ and $\boldsymbol{X}_0$, respectively. To do this let

$$\pi(\cdot;\mathbf{s}) = P\left((D^-, D^+ + 1, \mathbf{s}) \in \cdot\right), \qquad \mathbf{s} \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{S}'',$$

where (D^-, D^+) are independent Poisson random variables with means s_1 and s_2, respectively, for $\mathbf{s} = (s_1, s_2, \mathbf{s}_3)$.

To construct $(\hat{\mathbf{X}}, \boldsymbol{X})$ let $(\hat{\mathbf{A}}_b, \boldsymbol{A}_b)$ be the size-biased coupled vertex attribute, and let $f_n : \mathcal{S}' \to \mathcal{S}'$ be the function that maps $\hat{\mathbf{A}}$ to $\hat{\mathbf{S}}$, and $f : \mathcal{S}' \to \mathcal{S}'$ be the function that maps $\boldsymbol{A}$ to $\boldsymbol{S}$, as defined earlier in the proof. Set $(\hat{\mathbf{S}}_b, \boldsymbol{S}_b) = (f_n(\hat{\mathbf{A}}_b), f(\boldsymbol{A}_b))$. Now set

$$(\hat{\mathbf{X}}, \boldsymbol{X}) = \left(G^{-1}(U;\, \hat{S}_{b,1}), G^{-1}(U';\, \hat{S}_{b,2}) + 1, \hat{\mathbf{A}}_b, G^{-1}(U;\, S_{b,1}), G^{-1}(U';\, S_{b,2}) + 1, \boldsymbol{A}_b\right),$$

for some $U, U' \sim \text{Uniform}[0,1]$ i.i.d. and independent of $(\hat{\mathbf{S}}_b, \boldsymbol{S}_b)$. Note that

$$\rho(\hat{\mathbf{X}}, \boldsymbol{X}) \le \left|G^{-1}(U, \hat{S}_{b,1}) - G^{-1}(U, S_{b,1})\right| + \left|G^{-1}(U', \hat{S}_{b,2}) - G^{-1}(U', S_{b,2})\right| + \rho(\hat{\mathbf{A}}_b, \boldsymbol{A}_b),$$

where the continuity in λ of $G(\cdot;\lambda)$ ensures that each of the first two terms converges in probability to zero whenever $\rho(\hat{\mathbf{S}}_b, \boldsymbol{S}_b) \xrightarrow{P} 0$ as $n \to \infty$. The latter convergence is in turn implied by Assumption [G] and $\rho(\hat{\mathbf{A}}_b, \boldsymbol{A}_b) \xrightarrow{P} 0$ as $n \to \infty$. □

It will also be convenient to have the following technical results relating convergence of the means to convergence in d_p.

Lemma 4.4. *Suppose $\mathbf{Z}^{(n)} \Rightarrow \boldsymbol{Z}$ as $n \to \infty$ for some $\{\mathbf{Z}^{(n)}, \boldsymbol{Z}\} \subseteq \mathcal{S}$, and suppose f is a nonnegative continuous function satisfying*

$$\mathbb{E}_n\left[f(\mathbf{Z}^{(n)})\right] \xrightarrow{P} E\left[f(\boldsymbol{Z})\right] < \infty, \qquad n \to \infty.$$

Then, for any event $\mathcal{E}_n$ constructed on the same probability space as $\mathbf{Z}^{(n)}$ and such that $\mathbb{P}_n(\mathcal{E}_n) \xrightarrow{P} 0$ as $n \to \infty$, we have

$$\mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(\mathcal{E}_n)\right] \xrightarrow{P} 0, \qquad n \to \infty.$$

Proof. To start, fix $M > 0$ and note that

$$\begin{aligned}\mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(\mathcal{E}_n)\right] &\leq \mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(\mathcal{E}_n)1(\mathbf{Z}^{(n)}) \leq M)\right] + \mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(f(\mathbf{Z}^{(n)}) > M))\right] \\ &\leq M\mathbb{P}_n(\mathcal{E}_n) + \mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(f(\mathbf{Z}^{(n)}) > M))\right].\end{aligned}$$

Since $\mathbb{P}_n(\mathcal{E}_n) \xrightarrow{P} 0$ as $n \to \infty$, we need to show that $\mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(f(\mathbf{Z}^{(n)} > M))\right]$ can be made arbitrarily small. To this end, note that

$$\mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(f(\mathbf{Z}^{(n)}) > M))\right] = \mathbb{E}_n\left[f(\mathbf{Z}^{(n)})\right] - \mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(f(\mathbf{Z}^{(n)} \leq M))\right],$$

where by Lemma A.2 in [12] and the observation that $f(x)1(f(x) \leq M)$ is bounded and continuous a.e., we obtain that

$$\mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(f(\mathbf{Z}^{(n)}) \leq M))\right] \xrightarrow{P} E\left[f(\boldsymbol{Z})1(f(\boldsymbol{Z}) \leq M))\right], \qquad n \to \infty$$

Therefore, we have that provided M is a point of continuity,

$$\mathbb{E}_n\left[f(\mathbf{Z}^{(n)})1(f(\mathbf{Z}^{(n)} > M))\right] \xrightarrow{P} E\left[f(\mathbf{Z})\right] - E\left[f(\mathbf{Z})1(f(\mathbf{Z}) \leq M))\right] = E\left[f(\boldsymbol{Z})1(f(\boldsymbol{Z}) > M)\right]$$

as $n \to \infty$. Since M is arbitrary, take $M \to \infty$ to complete the proof. □

This last technical lemma shows the convergence in L_p for suboptimal couplings.

Lemma 4.5. *Suppose $Z^{(n)} \xrightarrow{P} Z$ as $n \to \infty$ for some $Z^{(n)}, Z \in \mathbb{R}$, and suppose that for some $p \in [1, \infty)$*

$$\mathbb{E}_n\left[|Z^{(n)}|^p\right] \xrightarrow{P} E[|Z|^p] < \infty \qquad n \to \infty,$$

then

$$\mathbb{E}_n\left[|Z^{(n)} - Z|^p\right] \xrightarrow{P} 0 \qquad n \to \infty.$$

Proof. Note that if we let $\eta_n(\cdot) = \mathbb{P}_n(Z^{(n)} \in \cdot)$ and $\eta(\cdot) = P(Z \in \cdot)$, then the assumptions of the lemma imply that $Z^{(n)} \xrightarrow{d_p} Z$ as $n \to \infty$ (see Definition 6.7 in [58]), and therefore, there exists an optimal coupling $(Z_*^{(n)}, Z_*)$ for which the convergence in L_p holds. What we need to show is that the L_p convergence holds even for suboptimal couplings provided $Z^{(n)} \xrightarrow{P} Z$ as $n \to \infty$.

To see this is the case, fix $M > 0$ to be a continuity point of η, and use Minkowski's inequality to obtain

$$\left(\mathbb{E}_n\left[\left|Z^{(n)} - Z\right|^p\right]\right)^{1/p} \leq \left(\mathbb{E}_n\left[\left|Z^{(n)} - Z\right|^p 1(|Z^{(n)}| \leq M)\right]\right)^{1/p} + \left(\mathbb{E}_n\left[\left|Z^{(n)}\right|^p 1\left(|Z^{(n)}| > M\right)\right]\right)^{1/p} + \left(\mathbb{E}_n\left[|Z|^p 1\left(|Z^{(n)}| > M\right)\right]\right)^{1/p}.$$

Now use the observation that $E\left[|Z|^p\right] < \infty$ and the dominated convergence theorem with dominating random variable $(|Z| + M)^p$ to obtain

$$E\left[\left|Z^{(n)} - Z\right|^p 1(|Z^{(n)}| \leq M)\right] \to 0 \qquad n \to \infty,$$

and also,

$$E\left[|Z|^p 1(|Z^{(n)}| > M)\right] \to E\left[|Z|^p 1(|Z| > M)\right] \qquad n \to \infty,$$

which implies that

$$\left(\mathbb{E}_n\left[\left|Z^{(n)} - Z\right|^p 1(|Z^{(n)}| \leq M)\right]\right)^{1/p} + \left(\mathbb{E}_n\left[|Z|^p 1\left(|Z^{(n)}| > M\right)\right]\right)^{1/p} \xrightarrow{P} \left(E\left[|Z|^p 1(|Z| > M)\right]\right)^{1/p}$$

as $n \to \infty$. Now use the optimal coupling to obtain

$$\begin{aligned}\left(\mathbb{E}_n\left[\left|Z^{(n)}\right|^p 1\left(|Z^{(n)}| > M\right)\right]\right)^{1/p} &= \left(\mathbb{E}_n\left[\left|Z_*^{(n)}\right|^p 1\left(|Z_*^{(n)}| > M\right)\right]\right)^{1/p} \\ &\leq \left(\mathbb{E}_n\left[\left|Z^{(n)} - Z_*\right|^p\right]\right)^{1/p} + \left(\mathbb{E}_n\left[|Z_*|^p 1\left(|Z_*^{(n)}| > M\right)\right]\right)^{1/p} \\ &\xrightarrow{P} \left(E\left[|Z|^p 1(|Z| > M)\right]\right)^{1/p},\end{aligned}$$

as $n \to \infty$. Now take the limit as $M \to \infty$ to complete the proof. □

We will now show that each of the $\hat{R}_\emptyset^{(r)}$, $0 \leq r \leq k$ in Theorem 4.1 can be coupled with another random variable $\mathcal{R}_\emptyset^{(r)}$ in such a way that $\mathbb{E}_n\left[\left|\hat{R}_\emptyset^{(r)} - \mathcal{R}_\emptyset^{(r)}\right|^p\right] \xrightarrow{P} 0$ as $n \to \infty$.

Theorem 4.6. *Let* $\nu_{k,n}(\cdot) = \mathbb{P}_n\left(\hat{R}_\emptyset^{(k)} \in \cdot\right)$ *and* $\nu_k(\cdot) = P\left(\mathcal{R}_\emptyset^{(k)} \in \cdot\right)$. *Then, under Assumptions [R] and [G] we have for any fixed* $k \geq 1$, *and* $\hat{R}_\emptyset^{(k)}$ *is constructed according to* (4.2), *there exists a version of* $\mathcal{R}_\emptyset^{(k)}$ *on the same probability space such that*

$$\mathbb{E}_n\left[\left|\hat{R}_\emptyset^{(k)} - \mathcal{R}_\emptyset^{(k)}\right|^p\right] \to 0, \qquad n \to \infty.$$

In particular, this implies that

$$d_p(\nu_{k,n}, \nu_k) \xrightarrow{P} 0, \qquad n \to \infty.$$

Proof. To start, note that whether $\mathbf{i}$ is a label on a tree depends only on the offspring numbers $\{\hat{N}_\mathbf{i}\}$ and $\{\mathcal{N}_\mathbf{i}\}$. Disagreements between $\hat{N}_\mathbf{i}$ and $\mathcal{N}_\mathbf{i}$ will lead to labels on one tree not being part of the other tree. To explicitly construct the coupled $\mathcal{R}_\emptyset^{(k)}$ the idea is to couple the marks $(\mathbf{Y}_\mathbf{i}, \boldsymbol{\Upsilon}_\mathbf{i})$ optimally for each $\mathbf{i} \in \mathcal{U}$. For this, use Lemma 4.3 to construct for each label $\mathbf{i} \in \mathcal{U}$, $\boldsymbol{\Upsilon}_\mathbf{i} = (\boldsymbol{X}_\mathbf{i}, \zeta_\mathbf{i}, \{\xi_{(\mathbf{i},j)} : j \geq 1\})$, where $\hat{\mathbf{Y}}_\mathbf{i} = (\hat{\mathbf{X}}_\mathbf{i}, \zeta_\mathbf{i}, \{\xi_{(\mathbf{i},j)} : j \geq 1\})$ and $\rho(\hat{\mathbf{X}}_\mathbf{i}, \boldsymbol{X}_\mathbf{i}) \xrightarrow{P} 0$ as $n \to \infty$. Recall that under Assumption [G] we cannot guarantee that $\mathbb{E}_n \left[\rho(\hat{\mathbf{X}}_\mathbf{i}, \boldsymbol{X}_\mathbf{i})\right] \xrightarrow{P} 0$ as $n \to \infty$, except for the root, i.e., $\mathbf{i} = \emptyset$.

Now that we have a sequence $\{\boldsymbol{\Upsilon}_\mathbf{i} : \mathbf{i} \in \mathcal{U}\}$ coupled to the sequence $\{\hat{\mathbf{Y}}_\mathbf{i} : \mathbf{i} \in \mathcal{U}\}$ we can use it to construct each of the $\mathcal{R}_\mathbf{i}^{(k-r)}$ and $\mathcal{V}_\mathbf{i}^{(k-r+1)} = g(\mathcal{R}_\mathbf{i}^{(k-r)}, \boldsymbol{X}_\mathbf{i})$ for $\mathbf{i} \in \mathcal{A}_r$. We will show that

$$\mathbb{E}_n \left[\left| \hat{V}_\mathbf{i}^{(k-r+1)} - \mathcal{V}_\mathbf{i}^{(k-r+1)} \right|^p \right] \xrightarrow{P} 0 \qquad n \to \infty,$$

for any $\mathbf{i} \in \hat{\mathcal{T}}_k \cap \mathcal{T}_k$, where $\hat{\mathcal{T}}_k$ is the tree of depth k constructed using $\{\hat{\mathbf{Y}}_\mathbf{i} : \mathbf{i}\}$ and $\mathcal{T}_k$ is the one constructed using $\{\boldsymbol{\Upsilon}_\mathbf{i} : \mathbf{i} \in \mathcal{U}\}$. Note that if $\eta_{s,n}(\cdot) = \mathbb{P}_n \left(\hat{V}_1^{(s)} \in \cdot \right)$ and $\eta_s(\cdot) = P \left(\mathcal{V}_1^{(s)} \in \cdot \right)$, then the above limits imply that $d_p(\eta_{n,s}, \eta_s) \xrightarrow{P} 0$ as $n \to \infty$ for $1 \leq s \leq k$. We will then use these results to conclude that

$$\mathbb{E}_n \left[\left| \hat{R}_\emptyset^{(k)} - \mathcal{R}_\emptyset^{(k)} \right|^p \right] \xrightarrow{P} 0 \qquad n \to \infty.$$

We will proceed inductively, starting with nodes $\mathbf{i} \in \hat{A}_k \cap \mathcal{A}_k$ and moving up towards the root. To this end, note that $\hat{R}_\mathbf{i}^{(0)} = r_0 = \mathcal{R}_\mathbf{j}^{(0)}$ for any $\mathbf{i} \in \mathbb{N}^k$. Moreover, by the continuity of g we have $\hat{V}_\mathbf{i}^{(1)} = g(r_0, \hat{\mathbf{X}}_\mathbf{i}) \xrightarrow{P} g(r_0, \boldsymbol{X}_\mathbf{i}) = \mathcal{V}_\mathbf{i}^{(1)}$ as $n \to \infty$ for any $\mathbf{i} \in \mathbb{N}^k$. Moreover, by Assumption [G] we have that $\mathbb{E}_n \left[|\hat{V}_\mathbf{i}^{(1)}|^p \right] \xrightarrow{P} E \left[|\mathcal{V}_\mathbf{i}^{(1)}|^p \right]$ as $n \to \infty$. Now use Lemma 4.5 to obtain that

$$\mathbb{E}_n \left[\left| \hat{V}_\mathbf{i}^{(1)} - \mathcal{V}_\mathbf{i}^{(1)} \right|^p \right] \xrightarrow{P} 0, \qquad n \to \infty.$$

Next, let $\mathbf{i} \in \hat{A}_r \cap \mathcal{A}_r$ for $0 < r < k$ and construct the random variables

$$\begin{aligned}
\hat{R}_\mathbf{i}^{(k-r)} &= \Phi \left(\hat{\mathbf{X}}_\mathbf{i}, \zeta_\mathbf{i}, \{ \hat{V}_{(\mathbf{i},j)}^{(k-r)}, \xi_{(\mathbf{i},j)} : 1 \leq j \leq \hat{N}_\mathbf{i} \} \right), \\
\tilde{R}_\mathbf{i}^{(k-r)} &= \Phi \left(\hat{\mathbf{X}}_\mathbf{i}, \zeta_\mathbf{i}, \{ \mathcal{V}_{(\mathbf{i},j)}^{(k-r)}, \xi_{(\mathbf{i},j)} : 1 \leq j \leq \hat{N}_\mathbf{i} \} \right), \quad \text{and} \\
\mathcal{R}_\mathbf{i}^{(k-r)} &= \Phi \left(\boldsymbol{X}_\mathbf{i}, \zeta_\mathbf{i}, \{ \mathcal{V}_{(\mathbf{i},j)}^{(k-r)}, \xi_{(\mathbf{i},j)} : 1 \leq j \leq \mathcal{N}_\mathbf{i} \} \right).
\end{aligned}$$

Note that for constructing $\tilde{R}_\mathbf{i}^{(k-r)}$ we may run into a situation where $\hat{N}_\mathbf{i} > \mathcal{N}_\mathbf{i}$, in which case we can simply sample $\mathcal{V}_{(\mathbf{i},j)}^{(k-r)}$ for $j > \mathcal{N}_\mathbf{i}$ according to $\eta_{k-r}(\cdot) = P(\mathcal{V}_1^{(k-r)} \in \cdot)$ independently of everything else. Set $\hat{V}_\mathbf{i}^{(k-r+1)} = g(\hat{R}_\mathbf{i}^{(k-r)}, \hat{\mathbf{X}}_\mathbf{i})$, $\tilde{V}_\mathbf{i}^{(k-r+1)} = g(\tilde{R}_\mathbf{i}^{(k-r)}, \hat{\mathbf{X}}_\mathbf{i})$ and $\mathcal{V}_\mathbf{i}^{(k-r+1)} = g(\mathcal{R}_\mathbf{i}^{(k-r)}, \boldsymbol{X}_\mathbf{i})$. Now use Minkowski's inequality and Assumption [R] to obtain that

$$\begin{aligned}
&\left(\mathbb{E}_n \left[\left| \hat{V}_\mathbf{i}^{(k-r+1)} - \mathcal{V}_\mathbf{i}^{(k-r+1)} \right|^p \right] \right)^{1/p} \\
&\leq \left(\mathbb{E}_n \left[\left| \hat{V}_\mathbf{i}^{(k-r+1)} - \tilde{V}_\mathbf{i}^{(k-r+1)} \right|^p \right] \right)^{1/p} + \left(\mathbb{E}_n \left[\left| \tilde{V}_\mathbf{i}^{(k-r+1)} - \mathcal{V}_\mathbf{i}^{(k-r+1)} \right|^p \right] \right)^{1/p} \\
&\leq \left(\mathbb{E}_n \left[\left(\sigma_+(\hat{\mathbf{X}}_\mathbf{i}) \sum_{j=1}^{\hat{N}_\mathbf{i}} \sigma_-(\hat{\mathbf{X}}_\mathbf{i}) \left| \hat{V}_{(\mathbf{i},j)}^{(k-r)} - \mathcal{V}_{(\mathbf{i},j)}^{(k-r)} \right| \right)^p \right] \right)^{1/p} \qquad (4.3)
\end{aligned}$$

$$+\left(\mathbb{E}_n\left[\left|\tilde{V}_{\mathbf{i}}^{(k-r+1)}-\mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p\right]\right)^{1/p}. \tag{4.4}$$

To analyze (4.3) condition on $\hat{\mathbf{X}}_{\mathbf{i}}$ and use Minkowski's inequality to obtain that

$$\begin{aligned}
&\mathbb{E}_n\left[\left(\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\sum_{j=1}^{\hat{N}_{\mathbf{i}}}\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})\left|\hat{V}_{(\mathbf{i},j)}^{(k-r)}-\mathcal{V}_{(\mathbf{i},j)}^{(k-r)}\right|\right)^p\right] \\
&\leq \mathbb{E}_n\left[\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})^p\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})^p\left(\sum_{j=1}^{\hat{N}_{\mathbf{i}}}\left(\mathbb{E}_n\left[\left|\hat{V}_{(\mathbf{i},j)}^{(k-r)}-\mathcal{V}_{(\mathbf{i},j)}^{(k-r)}\right|^p\middle|\mathbf{X}_{\mathbf{i}}\right]\right)^{1/p}\right)^p\right] \\
&\leq \mathbb{E}_n\Bigg[\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})^p\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})^p\Bigg((\hat{N}_{\mathbf{i}}\wedge\mathcal{N}_{\mathbf{i}})\left(\mathbb{E}_n\left[\left|\hat{V}_{(\mathbf{i},1)}^{(k-r)}-\mathcal{V}_{(\mathbf{i},1)}^{(k-r)}\right|^p\right]\right)^{1/p} \\
&\qquad +(\hat{N}_{\mathbf{i}}-\mathcal{N}_{\mathbf{i}})^+\left(\left(\mathbb{E}_n\left[|\hat{V}_{(\mathbf{i},1)}^{(k-r)}|^p\right]\right)^{1/p}+\left(E\left[|\mathcal{V}_{(\mathbf{i},1)}^{(k-r)}|^p\right]\right)^{1/p}\right)\Bigg)^p\Bigg] \\
&\leq \Bigg(\left(\mathbb{E}_n\left[\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})^p\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})^p\hat{N}_{\mathbf{i}}^p\right]\right)^{1/p}\left(\mathbb{E}_n\left[|\hat{V}_{(\mathbf{i},1)}^{(k-r)}-\mathcal{V}_{(\mathbf{i},1)}^{(k-r)}|^p\right]\right)^{1/p} \\
&\qquad +\left(\mathbb{E}_n\left[\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})^p\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})^p((\hat{N}_{\mathbf{i}}-\mathcal{N}_{\mathbf{i}})^+)^p\right]\right)^{1/p}\left(\left(\mathbb{E}_n\left[|\hat{V}_{(\mathbf{i},1)}^{(k-r)}|^p\right]\right)^{1/p}+\left(E\left[|\mathcal{V}_{(\mathbf{i},1)}^{(k-r)}|^p\right]\right)^{1/p}\right)\Bigg)^p,
\end{aligned}$$

where by Assumption [G] (c) we have $\mathbb{E}_n\left[\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})^p\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})^p\hat{N}_{\mathbf{i}}^p\right]\xrightarrow{P}E\left[\sigma_+(\boldsymbol{X})^p\sigma_-(\boldsymbol{X})^p\mathcal{N}^p\right]<\infty$, and by induction we have $\mathbb{E}_n\left[\left|\hat{V}_{(\mathbf{i},1)}^{(k-r)}-\mathcal{V}_{(\mathbf{i},1)}^{(k-r)}\right|^p\right]\xrightarrow{P}0$ as $n\to\infty$. The induction hypothesis also gives $\mathbb{E}_n\left[|\hat{V}_{(\mathbf{i},1)}^{(k-r)}|^p\right]\xrightarrow{P}E\left[|\mathcal{V}_{(\mathbf{i},1)}^{(k-r)}|^p\right]<\infty$ as $n\to\infty$, and by Lemma 4.4 with $f(\mathbf{x})=(\sigma_+(\mathbf{x})\sigma_-(\mathbf{x})d^-)^p$, we have

$$\mathbb{E}_n\left[\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})^p\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})^p((\hat{N}_{\mathbf{i}}-\mathcal{N}_{\mathbf{i}})^+)^p\right]\leq\mathbb{E}_n\left[f(\hat{\mathbf{X}}_{\mathbf{i}})1(\hat{N}_{\mathbf{i}}-\mathcal{N}_{\mathbf{i}}>1)\right]\xrightarrow{P}0,\qquad n\to\infty.$$

This establishes that (4.3) converges to zero in probability.

To analyze (4.4) note that the continuity of Φ and g implies that for any $\mathbf{i}\in\hat{A}_r\cap\mathcal{A}_r$,

$$\tilde{R}_{\mathbf{i}}^{(k-r)}\xrightarrow{P}\mathcal{R}_{\mathbf{i}}^{(k-r)}\qquad\text{and}\qquad\tilde{V}_{\mathbf{i}}^{(k-r+1)}\xrightarrow{P}\mathcal{V}_{\mathbf{i}}^{(k-r+1)},\qquad n\to\infty.$$

Now use Minkowski's inequality to obtain that

$$\begin{aligned}
\left(\mathbb{E}_n\left[\left|\tilde{V}_{\mathbf{i}}^{(k-r+1)}-\mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p\right]\right)^{1/p}&\leq\left(\mathbb{E}_n\left[\left|\tilde{V}_{\mathbf{i}}^{(k-r+1)}-\mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p1(\rho(\hat{\mathbf{X}}_{\mathbf{i}},\boldsymbol{X}_{\mathbf{i}})<1)\right]\right)^{1/p} \\
&\quad+\left(\mathbb{E}_n\left[\left|\tilde{V}_{\mathbf{i}}^{(k-r+1)}\right|^p1(\rho(\hat{\mathbf{X}}_{\mathbf{i}},\boldsymbol{X}_{\mathbf{i}})\geq1)\right]\right)^{1/p} \\
&\quad+\left(\mathbb{E}_n\left[\left|\mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p1(\rho(\hat{\mathbf{X}}_{\mathbf{i}},\boldsymbol{X}_{\mathbf{i}})\geq1)\right]\right)^{1/p}.
\end{aligned}$$

To show that each of the three expressions above converge to zero in probability, note that

$$\left|\tilde{V}_{\mathbf{i}}^{(k-r+1)}-\mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p1(\rho(\hat{\mathbf{X}}_{\mathbf{i}},\boldsymbol{X}_{\mathbf{i}})<1)\leq\left(|\mathcal{V}_{\mathbf{i}}^{(k-r+1)}|+\omega(\boldsymbol{X}_{\mathbf{i}},\boldsymbol{V}_{\mathbf{i}})\right)^p,$$

where $\boldsymbol{V}_{\mathbf{i}} = \{\mathcal{V}_{(\mathbf{i},j)}^{(k-r)} : j \geq 1\}$ and $E\left[|\mathcal{V}_{(\mathbf{i},1)}^{(k-r)}|^p + |\mathcal{V}_{\mathbf{i}}^{(k-r+1)}|^p\right] < \infty$, and therefore, by dominated convergence,

$$\mathbb{E}_n\left[\left|\tilde{V}_{\mathbf{i}}^{(k-r+1)} - \mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) < 1)\right] \xrightarrow{P} 0 \qquad n \to \infty.$$

Dominated convergence also gives

$$\mathbb{E}_n\left[\left|\mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right] \xrightarrow{P} 0 \qquad n \to \infty.$$

Finally, for the middle expectation use Assumption [R] and Minkowski's inequality twice (first conditionally on $\hat{\mathbf{X}}_{\mathbf{i}}$) to obtain

$$\begin{aligned}
&\left(\mathbb{E}_n\left[\left|\tilde{V}_{\mathbf{i}}^{(k-r+1)}\right|^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right]\right)^{1/p} \\
&\leq \left(\mathbb{E}_n\left[\left(\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\left(\sum_{j=1}^{\hat{N}_{\mathbf{i}}} \sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})|\mathcal{V}_{(\mathbf{i},j)}^{(k-r)}| + \beta(\hat{\mathbf{X}}_{\mathbf{i}})\right)\right)^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right]\right)^{1/p} \\
&\leq \left(\mathbb{E}_n\left[\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})^p\left(\sum_{j=1}^{\hat{N}_{\mathbf{i}}} \sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})\left(E\left[\left|\mathcal{V}_{(\mathbf{i},j)}^{(k-r)}\right|^p\right]\right)^{1/p} + \beta(\hat{\mathbf{X}}_{\mathbf{i}})\right)^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right]\right)^{1/p} \\
&= \left(\mathbb{E}_n\left[\left(\hat{N}_{\mathbf{i}}\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})\left(E\left[\left|\mathcal{V}_1^{(k-r)}\right|^p\right]\right)^{1/p} + \sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\beta(\hat{\mathbf{X}}_{\mathbf{i}})\right)^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right]\right)^{1/p} \\
&\leq \left(E\left[\left|\mathcal{V}_1^{(k-r)}\right|^p\right]\right)^{1/p}\left(\mathbb{E}_n\left[\left(\hat{N}_{\mathbf{i}}\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})\right)^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right]\right)^{1/p} \\
&\quad + \left(\mathbb{E}_n\left[\left(\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\beta(\hat{\mathbf{X}}_{\mathbf{i}})\right)^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right]\right)^{1/p}.
\end{aligned}$$

Now use Lemma 4.4 and Assumption [G] (c) to obtain that

$$\mathbb{E}_n\left[\left(\hat{N}_{\mathbf{i}}\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\sigma_-(\hat{\mathbf{X}}_{\mathbf{i}})\right)^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right] + \mathbb{E}_n\left[\left(\sigma_+(\hat{\mathbf{X}}_{\mathbf{i}})\beta(\hat{\mathbf{X}}_{\mathbf{i}})\right)^p 1(\rho(\hat{\mathbf{X}}_{\mathbf{i}}, \boldsymbol{X}_{\mathbf{i}}) \geq 1)\right] \xrightarrow{P} 0$$

as $n \to \infty$.

We have thus shown that

$$\mathbb{E}_n\left[\left|\hat{V}_{\mathbf{i}}^{(k-r+1)} - \mathcal{V}_{\mathbf{i}}^{(k-r+1)}\right|^p\right] \xrightarrow{P} 0, \qquad n \to \infty,$$

which also implies $d_p(\eta_{k-r+1,n}, \eta_{k-r+1}) \xrightarrow{P} 0$ as $n \to \infty$. To complete the proof, note that the same arguments used above establish that $\hat{R}_\emptyset^{(k)} \xrightarrow{P} \mathcal{R}_\emptyset^{(k)}$ as $n \to \infty$. Now use the same steps used to derive (4.3) and (4.4) to obtain that

$$\begin{aligned}
&\left(\mathbb{E}_n\left[\left|\hat{R}_\emptyset^{(k)} - \mathcal{R}_\emptyset^{(k)}\right|^p\right]\right)^{1/p} \\
&\leq \left(\mathbb{E}_n\left[\left(\sum_{j=1}^{\hat{N}_\emptyset} \sigma_-(\hat{\mathbf{X}}_\emptyset)\left|\hat{V}_j^{(k)} - \mathcal{V}_j^{(k)}\right|\right)^p\right]\right)^{1/p} + \left(\mathbb{E}_n\left[\left|\tilde{R}_\emptyset^{(k)} - \mathcal{R}_\emptyset^{(k)}\right|^p\right]\right)^{1/p}
\end{aligned}$$

$$\leq \left(\mathbb{E}_n\left[|\hat{V}_1^{(k)} - \mathcal{V}_1^{(k)}|^p\right]\mathbb{E}_n\left[\sigma_-(\mathbf{X}_\emptyset)^p \hat{N}_\emptyset^p\right]\right)^{1/p}$$
$$+ \left(\mathbb{E}_n\left[\sigma_-(\hat{\mathbf{X}}_\emptyset)^p((\hat{N}_\emptyset - \mathcal{N}_\emptyset)^+)^p\right]\right)^{1/p}\left(\left(\mathbb{E}_n\left[|\hat{V}_1^{(k)}|^p\right]\right)^{1/p} + \left(E\left[|\mathcal{V}_1^{(k)}|^p\right]\right)^{1/p}\right)$$
$$+ \left(\mathbb{E}_n\left[\left|\tilde{R}_\emptyset^{(k)} - \mathcal{R}_\emptyset^{(k)}\right|^p 1(\rho(\mathbf{X}_\emptyset, \boldsymbol{X}_\emptyset) < 1)\right]\right)^{1/p}$$
$$+ \left(E\left[\left|\mathcal{V}_1^{(k)}\right|^p\right]\right)^{1/p}\left(\mathbb{E}_n\left[\left(\hat{N}_\emptyset\sigma_-(\hat{\mathbf{X}}_\emptyset)\right)^p 1(\rho(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_\emptyset) \geq 1)\right]\right)^{1/p}$$
$$+ \left(\mathbb{E}_n\left[\beta(\hat{\mathbf{X}}_\emptyset)^p 1(\rho(\hat{\mathbf{X}}_\emptyset, \boldsymbol{X}_\emptyset) \geq 1)\right]\right)^{1/p} + \left(\mathbb{E}_n\left[\left|\mathcal{R}_\emptyset^{(k)}\right|^p 1(\rho(\mathbf{X}_\emptyset, \boldsymbol{X}_\emptyset) \geq 1)\right]\right)^{1/p},$$

which converge in probability to zero as $n \to \infty$ as before. This last step also implies that $d_p(\nu_{k,n}, \nu_k) \xrightarrow{P} 0$ as $n \to \infty$. □

We are now ready to show that the coupling $(R_I^{(0)}, R_I^{(1)}, \ldots, R_I^{(k)}, \hat{R}_\emptyset^{(0)}, \hat{R}_\emptyset^{(1)}, \ldots, \hat{R}_\emptyset^{(k)})$ provided by Theorem 4.1 satisfies $\mathbb{E}_n\left[\left|R_I^{(r)} - \hat{R}_\emptyset^{(r)}\right|^p\right] \xrightarrow{P} 0$ as $n \to \infty$ for each $0 \leq r \leq k$. Note that since each of the $(R_I^{(r)}, \hat{R}_\emptyset^{(r)})$ are constructed according to (4.2) by simply shifting the noises being used, it suffices to prove the result for $r = k$.

Theorem 4.7. *Suppose the map Φ satisfies Assumption [R] and $G(V_n, E_n)$ satisfies Assumption [G]. Then, for any fixed $k \geq 1$, there exists a random variable $\hat{R}_\emptyset^{(k)}$ constructed according to* (4.2)*, such that*

$$\mathbb{E}_n\left[\left|R_I^{(k)} - \hat{R}_\emptyset^{(k)}\right|^p\right] \xrightarrow{P} 0, \qquad n \to \infty.$$

Proof. We start by constructing $R_I^{(k)}$ and $\hat{R}_\emptyset^{(k)}$ according to the coupling described in Theorem 4.1. Note that if Assumption [R](4)(ii) holds, then $\sup_{k\geq 0}\left\|\mathbf{R}^{(k)}\right\|_\infty \leq K < \infty$ for any directed graph, which implies that $\sup_{k\geq 0}\left|\hat{R}_\emptyset^{(k)}\right| \leq K$ P-a.s. Hence, if we let $\mathcal{I}_I^{(s)} = 1(R_I^{(s)} \neq \hat{R}_\emptyset^{(s)})$ and use Theorem 4.1, then

$$\mathbb{E}_n\left[\left|R_I^{(k)} - \hat{R}_\emptyset^{(k)}\right|^p\right] = \mathbb{E}_n\left[\left|R_I^{(k)} - \hat{R}_\emptyset^{(k)}\right|^p \mathcal{I}_I^{(k)}\right] \leq (2K)^p \mathbb{P}_n\left(\mathcal{I}_I^{(k)}\right) \xrightarrow{P} 0$$

as $n \to \infty$.

Let us then assume from this point onwards that the recursion is not bounded and instead satisfies $\|\mathbf{C}\|_p \leq K < \infty$ for any directed graph, where $\mathbf{C} \in \mathbb{R}^{n\times n}$ is the matrix whose (i,j)th component is $C_{i,j} = \sigma_-(\mathbf{X}_i)\sigma_+(\mathbf{X}_j)1(j \to i)$. Next, note that

$$\left(\mathbb{E}_n\left[\left|R_I^{(k)} - \hat{R}_\emptyset^{(k)}\right|^p\right]\right)^{1/p} = \left(\mathbb{E}_n\left[\left|R_I^{(k)} - \hat{R}_\emptyset^{(k)}\right|^p \mathcal{I}_I^{(k)}\right]\right)^{1/p}$$
$$\leq \left(\mathbb{E}_n\left[\left|R_I^{(k)}\right|^p \mathcal{I}_I^{(k)}\right]\right)^{1/p} + \left(\mathbb{E}_n\left[\left|\hat{R}_\emptyset^{(k)}\right|^p \mathcal{I}_I^{(k)}\right]\right)^{1/p}.$$

To analyze the first expectation note that we can condition on the exploration of the graph and the coupled tree up to the completion of generation k, to obtain that

$$\mathbb{E}_n\left[\left|R_I^{(k)}\right|^p \mathcal{I}_I^{(k)}\right] = \mathbb{E}_n\left[\mathbf{E}_n\left[\left|R_I^{(k)}\right|^p\right]\mathcal{I}_I^{(k)}\right],$$

where the only source of randomness inside the expectation $\mathbf{E}_n\left[\left|R_I^{(k)}\right|^p\right]$ is due to the noises. Now note that Assumption [R] gives that

$$\begin{aligned}\left(\mathbf{E}_n\left[\left|R_I^{(k)}\right|^p\right]\right)^{1/p} &\leq \left(\mathbf{E}_n\left[\mathbf{E}_n\left[\left|R_I^{(k)}\right|^p \middle| \mathbf{R}^{(k-1)}\right]\right]\right)^{1/p} \\ &\leq \left(\mathbf{E}_n\left[\left(\sum_{j\to I}\sigma_-(\mathbf{X}_I)\sigma_+(\mathbf{X}_j)|R_j^{(k-1)}| + \beta(\mathbf{X}_I)\right)^p\right]\right)^{1/p} \\ &\leq \sum_{j\to I}\sigma_-(\mathbf{X}_I)\sigma_+(\mathbf{X}_j)\left(\mathbf{E}_n\left[\left|R_j^{(k-1)}\right|^p\right]\right)^{1/p} + \beta(\mathbf{X}_I).\end{aligned}$$

Next, construct for each inbound neighbor of vertex I its coupled pair according to Theorem 4.1. More precisely, for each $j \to I$, let $\hat{R}_{\emptyset(j)}^{(k-1)}$ denote its tree version of the recursion. The index notation $\emptyset(j)$ means that the root of the tree where $\hat{R}_{\emptyset(j)}^{(k-1)}$ is constructed corresponds to the jth inbound neighbor of I, and that the tree construction disregards the exploration of all other inbound neighbors of I as well as I itself. We point out that on the event that $R_I^{(k)} \neq \hat{R}_\emptyset^{(k)}$ it is possible for the $\{\hat{R}_{\emptyset(j)}^{(k-1)} : j \to I\}$ to be dependent due to miscouplings, although it is also possible for $\hat{R}_{\emptyset(j)}^{(k-1)}$ to equal one of the $\hat{R}_s^{(k-1)}$ constructed on a subtree of node $\emptyset$. Now note that

$$\left(\mathbf{E}_n\left[\left|R_j^{(k-1)}\right|^p\right]\right)^{1/p} \leq \left(\mathbf{E}_n\left[\left|R_j^{(k-1)} - \hat{R}_{\emptyset(j)}^{(k-1)}\right|^p\right]\right)^{1/p} + \left(\mathbf{E}_n\left[\left|\hat{R}_{\emptyset(j)}^{(k-1)}\right|^p\right]\right)^{1/p}$$

It follows from Minkowski's inequality that

$$\begin{aligned}\left(\mathbb{E}_n\left[\left|R_I^{(k)}\right|^p \mathcal{I}_I^{(k)}\right]\right)^{1/p} &\leq \left(\mathbb{E}_n\left[\mathcal{I}_I^{(k)}\left(\sum_{j\to I}\sigma_-(\mathbf{X}_I)\sigma_+(\mathbf{X}_j)\left(\mathbf{E}_n\left[\left|R_j^{(k-1)}\right|^p\right]\right)^{1/p} + \beta(\mathbf{X}_I)\right)^p\right]\right)^{1/p} \\ &\leq \left(\mathbb{E}_n\left[\left(\sum_{j\to I}\mathcal{I}_I^{(k)}\sigma_-(\mathbf{X}_I)\sigma_+(\mathbf{X}_j)\left(\mathbf{E}_n\left[\left|R_j^{(k-1)} - \hat{R}_{\emptyset(j)}^{(k-1)}\right|^p\right]\right)^{1/p}\right)^p\right]\right)^{1/p} \\ &\quad + \left(\mathbb{E}_n\left[\mathcal{I}_I^{(k)}\left(\sum_{j\to I}\sigma_-(\mathbf{X}_I)\sigma_+(\mathbf{X}_j)\left(\mathbf{E}_n\left[\left|\hat{R}_{\emptyset(j)}^{(k-1)}\right|^p\right]\right)^{1/p} + \beta(\mathbf{X}_I)\right)^p\right]\right)^{1/p}.\end{aligned}$$

Next, fix $\epsilon > 0$ and choose $M > 0$ such that $\max_{0\leq s\leq k} E\left[\left((|\mathcal{R}_\emptyset^{(s)}| - M)^+\right)^p\right] < \epsilon^p$. Note that by Minkowski's inequality again we obtain

$$\begin{aligned}\left(\mathbb{E}_n\left[\left|R_I^{(k)}\right|^p \mathcal{I}_I^{(k)}\right]\right)^{1/p} &\leq \left(\mathbb{E}_n\left[\left(\sum_{j=1}^n C_{I,j}\left(\mathbf{E}_n\left[\left|R_j^{(k-1)} - \hat{R}_{\emptyset(j)}^{(k-1)}\right|^p\right]\right)^{1/p}\right)^p\right]\right)^{1/p} \\ &\quad + \left(\mathbb{E}_n\left[\left(\sum_{j=1}^n C_{I,j}\left(\mathbf{E}_n\left[\left((|\hat{R}_{\emptyset(j)}^{(k-1)}| - M)^+\right)^p\right]\right)^{1/p}\right)^p\right]\right)^{1/p}\end{aligned}$$

$$+ \left(\mathbb{E}_n \left[\mathcal{I}_I^{(k)} \left(\sum_{j \to I} \sigma_-(\mathbf{X}_I) \sigma_+(\mathbf{X}_j) M + \beta(\mathbf{X}_I) \right)^p \right] \right)^{1/p}$$

$$=: \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[\left((\mathbf{C}\boldsymbol{\Delta}^{(k-1)})_i \right)^p \right] \right)^{1/p} + \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[((\mathbf{C}\boldsymbol{\Gamma}^M)_i)^p \right] \right)^{1/p}$$

$$+ \left(\mathbb{E}_n \left[\mathcal{I}_I^{(k)} \left(M \sigma_-(\mathbf{X}_I) \sum_{j \to I} \sigma_+(\mathbf{X}_j) + \beta(\mathbf{X}_I) \right)^p \right] \right)^{1/p},$$

where $\boldsymbol{\Delta}^{(k-1)}, \boldsymbol{\Gamma}^M \in \mathbb{R}^n$ are the vectors whose ith components are $\Delta_i^{(k-1)} = \left(\mathbf{E}_n \left[\left| R_i^{(k-1)} - \hat{R}_{\emptyset(i)}^{(k-1)} \right|^p \right] \right)^{1/p}$ and $\Gamma_i^M = \left(\mathbf{E}_n \left[\left((|\hat{R}_{\emptyset(i)}^{(k-1)}| - M)^+ \right)^p \right] \right)^{1/p}$. To further bound the first two expectations, note that

$$\left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[\left((\mathbf{C}\boldsymbol{\Delta}^{(k-1)})_i \right)^p \right] \right)^{1/p} + \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[((\mathbf{C}\boldsymbol{\Gamma}^M)_i)^p \right] \right)^{1/p}$$
$$= \left(\mathbb{E}_n \left[\|\mathbf{C}\boldsymbol{\Delta}^{(k-1)}\|_p^p \right] \right)^{1/p} + \left(\mathbb{E}_n \left[\|\mathbf{C}\boldsymbol{\Gamma}^M\|_p^p \right] \right)^{1/p}$$
$$\leq \left(\mathbb{E}_n \left[\|\mathbf{C}\|_p^p \cdot \|\boldsymbol{\Delta}^{(k-1)}\|_p^p \right] \right)^{1/p} + \left(\mathbb{E}_n \left[\|\mathbf{C}\|_p^p \cdot \|\boldsymbol{\Gamma}^M\|_p^p \right] \right)^{1/p}$$
$$\leq K \left(\mathbb{E}_n \left[\|\boldsymbol{\Delta}^{(k-1)}\|_p^p \right] \right)^{1/p} + K \left(\mathbb{E}_n \left[\|\boldsymbol{\Gamma}^M\|_p^p \right] \right)^{1/p}$$
$$= K \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[(\Delta_i^{(k-1)})^p \right] \right)^{1/p} + K \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[(\Gamma_i^M)^p \right] \right)^{1/p}$$
$$= K \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[\left| R_i^{(k-1)} - \hat{R}_{\emptyset(i)}^{(k-1)} \right|^p \right] \right)^{1/p} + K \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n \left[\left((|\hat{R}_{\emptyset(i)}^{(k-1)}| - M)^+ \right)^p \right] \right)^{1/p}$$
$$= K \left(\mathbb{E}_n \left[\left| R_I^{(k-1)} - \hat{R}_{\emptyset}^{(k-1)} \right|^p \right] \right)^{1/p} + K \left(\mathbb{E}_n \left[\left((|\hat{R}_{\emptyset}^{(k-1)}| - M)^+ \right)^p \right] \right)^{1/p}.$$

Furthermore, Minkowski's inequality gives

$$\left(\mathbb{E}_n \left[\left| \hat{R}_{\emptyset}^{(k)} \right|^p \mathcal{I}_I^{(k)} \right] \right)^{1/p} \leq \left(\mathbb{E}_n \left[\left((|\hat{R}_{\emptyset}^{(k)}| - M)^+ \right)^p \right] \right)^{1/p} + M \left(\mathbb{E}_n \left[\mathcal{I}_I^{(k)} \right] \right)^{1/p}.$$

So we have shown so far that

$$\left(\mathbb{E}_n \left[\left| R_I^{(k)} - \hat{R}_{\emptyset}^{(k)} \right|^p \right] \right)^{1/p} \leq K \left(\mathbb{E}_n \left[\left| R_I^{(k-1)} - \hat{R}_{\emptyset}^{(k-1)} \right|^p \right] \right)^{1/p} + K \left(\mathbb{E}_n \left[\left((|\hat{R}_{\emptyset}^{(k-1)}| - M)^+ \right)^p \right] \right)^{1/p}$$
$$+ \left(\mathbb{E}_n \left[\mathcal{I}_I^{(k)} \left(M \sigma_-(\mathbf{X}_I) \sum_{j \to I} \sigma_+(\mathbf{X}_j) + \beta(\mathbf{X}_I) \right)^p \right] \right)^{1/p}$$
$$+ \left(\mathbb{E}_n \left[\left((|\hat{R}_{\emptyset}^{(k)}| - M)^+ \right)^p \right] \right)^{1/p} + M \left(\mathbb{E}_n \left[\mathcal{I}_I^{(k)} \right] \right)^{1/p}.$$

Hence, if we let $a_{s,n} = \left(\mathbb{E}_n\left[\left|R_I^{(s)} - \hat{R}_\emptyset^{(s)}\right|^p\right]\right)^{1/p}$, $b_{s,n} = \left(\mathbb{E}_n\left[\left((|\hat{R}_\emptyset^{(s)}| - M)^+\right)^p\right]\right)^{1/p}$, and

$$c_{s,n} = \left(\mathbb{E}_n\left[\mathcal{I}_I^{(s)}\left(M\sigma_-(\mathbf{X}_I)\sum_{j\to I}\sigma_+(\mathbf{X}_j) + \beta\mathbf{X}_I\right)^p\right]\right)^{1/p} + M\left(\mathbb{E}_n\left[\mathcal{I}_I^{(s)}\right]\right)^{1/p},$$

we have shown that

$$\begin{aligned}
a_{k,n} &\le Ka_{k-1,n} + Kb_{k-1,n} + b_{k,n} + c_{k,n} \\
&\le K(Ka_{k-2,n} + Kb_{k-2,n} + b_{k-1,n} + c_{k-1,n}) + Kb_{k-1,n} + b_{k,n} + c_{k,n} \\
&\le K^k a_{0,n} + \sum_{s=1}^{k} K^s b_{k-s,n} + \sum_{s=0}^{k-1} K^s(b_{k-s,n} + c_{k-s,n}) \\
&\le K^k b_{0,n} + \sum_{s=1}^{k} K^{k-s}(2b_{s,n} + c_{s,n}),
\end{aligned}$$

where in the last equality we used the observation that $R_I^{(0)} = r_0 = \hat{R}_\emptyset^{(0)}$.

Note that by Theorem 4.6, we have that for any $0 \le s \le k$,

$$b_{s,n} \xrightarrow{P} \left(E\left[\left((|\mathcal{R}_\emptyset^{(s)}| - M)^+\right)^p\right]\right)^{1/p} < \epsilon$$

as $n \to \infty$. We will show that $c_{s,n} \xrightarrow{P} 0$ as $n \to \infty$ for any $1 \le s \le k$, from where we will conclude that

$$\limsup_{n\to\infty}\left(\mathbb{E}_n\left[\left|R_I^{(s)} - \hat{R}_\emptyset^{(s)}\right|^p\right]\right)^{1/p} \le \sum_{s=0}^{k} K^s 2\epsilon,$$

and taking $\epsilon \downarrow 0$ will yield the statement of the theorem.

To see that $c_{s,n} \xrightarrow{P} 0$ as $n \to \infty$, note that $\mathbb{E}_n\left[\mathcal{I}_I^{(s)}\right] \xrightarrow{P} 0$ by Theorem 4.1. Now fix $\mathbf{x}_0 \in \mathcal{S}$, $H > 0$ and define $\mathcal{J}_i = 1(\rho(\mathbf{X}_i, \mathbf{x}_0) > H)$. Then,

$$\begin{aligned}
&\left(\mathbb{E}_n\left[\mathcal{I}_I^{(s)}\left(M\sigma_-(\mathbf{X}_I)\sum_{j\to I}\sigma_+(\mathbf{X}_j) + \beta(\mathbf{X}_I)\right)^p\right]\right)^{1/p} \\
&\le \left(\mathbb{E}_n\left[\mathcal{I}_I^{(s)}\left(M\sigma_-(\mathbf{X}_I)\sum_{j\to I}\sigma_+(\mathbf{X}_j)\mathcal{J}_j\right)^p\right]\right)^{1/p} \\
&\quad + \left(\mathbb{E}_n\left[\mathcal{I}_I^{(s)}\left(M\sigma_-(\mathbf{X}_I)\sum_{j\to I}\sigma_+(\mathbf{X}_j)(1-\mathcal{J}_j) + \beta(\mathbf{X}_I)\right)^p\right]\right)^{1/p} \\
&\le M\left(\mathbb{E}_n\left[\left(\sum_{j=1}^{n} C_{I,j}\mathcal{J}_j\right)^p\right]\right)^{1/p}
\end{aligned}$$

$$+ M \sup_{\mathbf{x}:\rho(\mathbf{x},\mathbf{x}_0)\leq H} \sigma_+(\mathbf{x}) \left(\mathbb{E}_n \left[\mathcal{I}_I^{(s)} \left(\sigma_-(\mathbf{X}_I) D_I^- + \beta(\mathbf{X}_I) \right)^p \right] \right)^{1/p}.$$

Moreover, for $\boldsymbol{\mathcal{J}} = (\mathcal{J}_1, \ldots, \mathcal{J}_n)'$,

$$\mathbb{E}_n \left[\left(\sum_{j=1}^n C_{I,j} \mathcal{J}_j \right)^p \right] = \frac{1}{n} \mathbb{E}_n \left[\|\mathbf{C}\boldsymbol{\mathcal{J}}\|_p^p \right] \leq \frac{1}{n} \mathbb{E}_n \left[\|\mathbf{C}\|_p^p \cdot \|\boldsymbol{\mathcal{J}}\|_p^p \right] \leq \frac{K^p}{n} \mathbb{E}_n \left[\|\boldsymbol{\mathcal{J}}\|_p^p \right] = K^p \mathbb{E}_n[\mathcal{J}_I].$$

Also, the proof of Lemma 4.3 can be adjusted (by setting $a_n = b_n = \infty$ for the IRD) to show that there exists a coupling $(\mathbf{X}_I, \boldsymbol{X}_0)$ such that $\rho(\mathbf{X}_I, \boldsymbol{X}_0) \xrightarrow{P} 0$ as $n \to \infty$, and Assumption [G](c) implies that $\mathbb{E}_n \left[(\sigma_-(\mathbf{X}_I) D_I^-)^p \right] \xrightarrow{P} E\left[(\sigma_-(\boldsymbol{X}_0)\mathcal{N}_0)^p \right]$ and $\mathbb{E}_n \left[\beta(\mathbf{X}_I)^p \right] \to E\left[\beta(\boldsymbol{X}_0)^p \right]$. Therefore, by Lemma 4.4,

$$\mathbb{E}_n \left[\mathcal{I}_I^{(s)} \left(\sigma_-(\mathbf{X}_I) D_I^- + \beta(\mathbf{X}_I) \right)^p \right] \xrightarrow{P} 0,$$

as $n \to \infty$. Hence, we have shown that

$$c_{n,s} \leq MK\mathbb{E}_n \left[\mathcal{J}_I \right] + o_P(1) \xrightarrow{P} MKP \left(\rho(\boldsymbol{X}_0, \mathbf{x}_0) > H \right),$$

where $o_P(1)$ denotes a term that converges to zero in probability as $n \to \infty$. Now take $H \to \infty$ to complete the proof. □

To complete the proof of Theorem 2.2, it remains to show that if c in Assumption [G] (c) satisfies $c \in (0, 1)$, then there exists a probability measure ν such that $\nu_k \xrightarrow{d_p} \nu$ as $k \to \infty$.

Theorem 4.8. *Let $\nu_k(\cdot) = P\left(\mathcal{R}_\emptyset^{(k)} \in \cdot \right)$. Then, if $c^p = E\left[(\mathcal{N}\sigma_+(\boldsymbol{X})\sigma_-(\boldsymbol{X}))^p \right] \in (0, 1)$, then, there exists a probability measure ν on $\mathbb{R}$ such that*

$$d_p(\nu_k, \nu) \to 0, \qquad k \to \infty.$$

Moreover, ν is the probability measure of a random variable $\mathcal{R}^*$ that satisfies:

$$\mathcal{R}^* = \Phi \left(\boldsymbol{X}_0, \zeta, \{ \mathcal{V}_j, \xi_j : 1 \leq j \leq \mathcal{N}_0 \} \right),$$

with the $\{\mathcal{V}_j\}$ i.i.d. copies of $\mathcal{V}$, independent of $\boldsymbol{X}_0$ and of $(\zeta, \{\xi_j : j \geq 1\})$, and $\mathcal{V}$ the special endogenous solution to the distributional fixed-point equation:

$$\mathcal{V} \stackrel{\mathcal{D}}{=} \Psi \left(\boldsymbol{\Upsilon}, \{ \mathcal{V}_j : 1 \leq j \leq \mathcal{N} \} \right) = g \left(\Phi \left(\boldsymbol{X}, \zeta, \{ \mathcal{V}_j, \xi_j : 1 \leq j \leq \mathcal{N} \} \right), \boldsymbol{X} \right),$$

where $\boldsymbol{\Upsilon} = (\boldsymbol{X}, \zeta, \{\xi_j : j \geq 1\})$, and the $\{\mathcal{V}_j\}$ i.i.d. copies of $\mathcal{V}$, independent of $\boldsymbol{\Upsilon}$.

Proof. Start by defining for $\mathbf{i} \in \mathcal{A}_k$, $\mathcal{V}_\mathbf{i}^{(1)} = g(r_0, \boldsymbol{X}_\mathbf{i})$, and for $\mathbf{i} \in \mathcal{A}_r$, $1 \leq r \leq k$,

$$\mathcal{V}_\mathbf{i}^{(r)} = g(\mathcal{R}_\mathbf{i}^{(r-1)}, \boldsymbol{X}_\mathbf{i}) \quad \text{and} \quad \mathcal{R}_\mathbf{i}^{(r)} = \Phi \left(\boldsymbol{X}_\mathbf{i}, \zeta_\mathbf{i}, \{ \mathcal{V}_{(\mathbf{i},j)}^{(r)}, \xi_{(\mathbf{i},j)} : 1 \leq j \leq \mathcal{N}_\mathbf{i} \} \right).$$

Let $\eta_k(\cdot) = P(\mathcal{V}_1^{(k)} \in \cdot)$, for $k \geq 1$. Note that it suffices to show that there exists a probability measure η such that $d_p(\eta_k, \eta) \to 0$ as $k \to \infty$, since then we can take ν to be the probability measure of

$$\mathcal{R}^* := \Phi\left(\boldsymbol{X}_0, \zeta, \{\mathcal{V}_j, \xi_j : 1 \leq j \leq \mathcal{N}_0\}\right),$$

where the $\{\mathcal{V}_j : j \geq 1\}$ are i.i.d. with common distribution η, independent of $\boldsymbol{X}_0$ and of $(\zeta, \{\xi_j : j \geq 1\})$. To this end, we will show that the map

$$\Psi\left(\boldsymbol{X}_{\mathbf{i}}, \zeta_{\mathbf{i}}, \{\mathcal{V}_{(\mathbf{i},j)}^{(r)}, \xi_{(\mathbf{i},j)} : 1 \leq j \leq \mathcal{N}_{\mathbf{i}}\}\right) = g\left(\Phi\left(\boldsymbol{X}_{\mathbf{i}}, \zeta_{\mathbf{i}}, \{\mathcal{V}_{(\mathbf{i},j)}^{(r)}, \xi_{(\mathbf{i},j)} : 1 \leq j \leq \mathcal{N}_{\mathbf{i}}\}\right), \boldsymbol{X}_{\mathbf{i}}\right)$$

defines a strict contraction under d_p. To start, note that for any $m \geq 1$,

$$d_p(\eta_k, \eta_{k+m}) \leq d_p(\eta_k, \eta_{k+1}) + d_p(\eta_{k+1}, \eta_{k+m}) \leq \sum_{i=0}^{m-1} d_p(\eta_{k+i}, \eta_{k+i+1}).$$

Next, construct $\mathcal{V}_1^{(k+i)}$ according to (3.1), and use the same sequence of $\{\boldsymbol{\Upsilon}_{\mathbf{i}} : \mathbf{i} \in \mathcal{U}\}$ to construct

$$\mathcal{W}_{\mathbf{i}}^{(1)} = g(r_0, \boldsymbol{X}_{\mathbf{i}}), \qquad \mathbf{i} \in \mathcal{A}_{k+i+1},$$

and

$$\mathcal{W}_{\mathbf{i}}^{(r)} = \Psi\left(\boldsymbol{\Upsilon}_{\mathbf{i}}, \left\{\mathcal{W}_{(\mathbf{i},j)}^{(r-1)} : 1 \leq j \leq \mathcal{N}_{\mathbf{i}}\right\}\right) \quad \mathbf{i} \in \mathcal{A}_{k+i+1-r}, \quad 2 \leq r \leq k+i+1-r.$$

Note that $\mathcal{V}_1^{(k+i)}$ has distribution η_{k+i} and $\mathcal{W}_1^{(k+i+1)}$ has distribution η_{k+i+1}. Now note that by Assumption [R] followed by Minkowski's inequality (conditioned on $\boldsymbol{X}_1$), we have

$$\begin{aligned}
&d_p(\eta_{k+i}, \eta_{k+i+1}) \\
&\leq \left(E\left[\left|\mathcal{V}_1^{(k+i)} - \mathcal{W}_1^{(k+i+1)}\right|\right]\right)^{1/p} \\
&= \left(E\left[\left|g\left(\Phi\left(\boldsymbol{X}_1, \zeta_1, \left\{\mathcal{V}_{(1,j)}^{(k+i-1)}, \xi_{(1,j)} : 1 \leq j \leq \mathcal{N}_1\right\}\right), \boldsymbol{X}_1\right)\right.\right.\right. \\
&\qquad \left.\left.\left. - g\left(\Phi\left(\boldsymbol{X}_1, \zeta_1, \left\{\mathcal{W}_{(1,j)}^{(k+i)}, \xi_{(1,j)} : 1 \leq j \leq \mathcal{N}_1\right\}\right), \boldsymbol{X}_1\right)\right|\right]\right)^{1/p} \\
&\leq \left(E\left[\sigma_+(\boldsymbol{X}_1)^p E\left[\left|\Phi\left(\boldsymbol{X}_1, \zeta_1, \left\{\mathcal{V}_{(1,j)}^{(k+i-1)}, \xi_{(1,j)} : 1 \leq j \leq \mathcal{N}_1\right\}\right)\right.\right.\right.\right. \\
&\qquad \left.\left.\left.\left. - \Phi\left(\boldsymbol{X}_1, \zeta_1, \left\{\mathcal{W}_{(1,j)}^{(k+i)}, \xi_{(1,j)} : 1 \leq j \leq \mathcal{N}_1\right\}\right)\right|^p \middle| \boldsymbol{X}_1\right]\right]\right)^{1/p} \\
&\leq \left(E\left[\sigma_+(\boldsymbol{X}_1)^p \left(\sum_{j=1}^{\mathcal{N}_1} \sigma_-(\boldsymbol{X}_1)\left|\mathcal{V}_{(1,j)}^{(k+i-1)} - \mathcal{W}_{(1,j)}^{(k+i)}\right|\right)^p\right]\right)^{1/p} \\
&\leq \left(E\left[\sigma_+(\boldsymbol{X}_1)^p \left(\sum_{j=1}^{\mathcal{N}_1} \sigma_-(\boldsymbol{X}_1)\left(E\left[\left|\mathcal{V}_{(1,j)}^{(k+i-1)} - \mathcal{W}_{(1,j)}^{(k+i)}\right|^p\right]\right)^{1/p}\right)^p\right]\right)^{1/p} \\
&= \left(E\left[\left|\mathcal{V}_{(1,1)}^{(k+i-1)} - \mathcal{W}_{(1,1)}^{(k+i)}\right|^p\right]\right)^{1/p} \left(E\left[(\mathcal{N}_1 \sigma_+(\boldsymbol{X}_1)\sigma_-(\boldsymbol{X}_1))^p\right]\right)^{1/p} \\
&= c\left(E\left[\left|\mathcal{V}_{(1,1)}^{(k+i-1)} - \mathcal{W}_{(1,1)}^{(k+i)}\right|^p\right]\right)^{1/p}
\end{aligned}$$

$$
\begin{aligned}
&\le c^{k+i-1} \left(E\left[\left| \mathcal{V}_{\mathbf{i}}^{(1)} - \mathcal{W}_{\mathbf{i}}^{(2)} \right|^p \right] \right)^{1/p} \qquad (\text{for } \mathbf{i} \in \mathcal{A}_{k+i}) \\
&\le c^{k+i-1} \left(\left(E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + \left(E\left[|\mathcal{V}_1^{(2)}|^p \right] \right)^{1/p} \right) \\
&\le c^{k+i-1} \left(\left(E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + \left(E\left[\left(\sigma_+(\boldsymbol{X}_1) \left(\sum_{j=1}^{\mathcal{N}_1} \sigma_-(\boldsymbol{X}_1) |\mathcal{V}_j^{(1)}| + \beta(\boldsymbol{X}_1) \right) \right)^p \right] \right)^{1/p} \right) \\
&\le c^{k+i-1} \left(\left(E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + \left(E\left[\left(\sigma_+(\boldsymbol{X}_1) \left(\mathcal{N}_1 \sigma_-(\boldsymbol{X}_1) \left(E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + \beta(\boldsymbol{X}_1) \right) \right)^p \right] \right)^{1/p} \right) \\
&\le c^{k+i-1} \left(\left(E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + c \left(E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + \left(E\left[(\sigma_+(\boldsymbol{X}_1)\beta(\boldsymbol{X}_1))^p \right] \right)^{1/p} \right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
d_p(\eta_k, \eta_{k+m}) &\le \sum_{i=0}^{m-1} c^{k+i-1} \left(\left((1+c) E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + \left(E\left[(\sigma_+(\boldsymbol{X}_1)\beta(\boldsymbol{X}_1))^p \right] \right)^{1/p} \right) \\
&\le \left(((1+c) E\left[|g(r_0, \boldsymbol{X})|^p \right])^{1/p} + \left(E\left[(\sigma_+(\boldsymbol{X})\beta(\boldsymbol{X}))^p \right] \right)^{1/p} \right) \frac{c^{k-1}}{1-c}
\end{aligned}
$$

which converges to zero as $k \to \infty$ since both expectations are finite by Assumption [G](c). Since the Wasserstein space with metric d_p is complete, there exists a probability measure η on $\mathbb{R}$ such that $d_p(\eta_k, \eta) \to 0$ as $k \to \infty$.

To see that the measure η we constructed as the limit of $\{\eta_k : k \ge 0\}$ is endogenous (see Definition 7 in [3]), suppose that we construct two random variables $\mathcal{V}_1^{(k)}$ and $\mathcal{W}_1^{(k)}$, both constructed recursively using the map Ψ, one where we use the i.i.d. sequence $\{\mathcal{V}_{\mathbf{i}}^{(1)} : \mathbf{i} \in \mathcal{A}_k\}$ as a starting point, and the other where we use the i.i.d. sequence $\{\mathcal{W}_{\mathbf{i}}^{(1)} : \mathbf{i} \in \mathcal{A}_k\}$ as a starting point, with both sequences distributed according to η, and independent of each other. For all other nodes $\mathbf{i} \in \bigcup_{r=0}^{k-1} \mathcal{A}_r$ we use the same sequence of vectors $\{\boldsymbol{\Upsilon}_{\mathbf{i}}\}$ for the two constructions. Now note that the same arguments used in this proof show that

$$
\begin{aligned}
\left(E\left[\left| \mathcal{V}_1^{(k)} - \mathcal{W}_1^{(k)} \right|^p \right] \right)^{1/p} &\le c^{k-1} \left(\left(E\left[|\mathcal{V}_1^{(1)}|^p \right] \right)^{1/p} + \left(E\left[|\mathcal{W}_1^{(1)}|^p \right] \right)^{1/p} \right) \\
&= 2 \left(\int_{-\infty}^{\infty} |x|^p \eta(dx) \right)^{1/p} c^{k-1}.
\end{aligned}
$$

Now take the limit as $k \to \infty$ to prove that the bivariate uniqueness property holds (see Theorem 11 in [3]), and therefore η is endogenous. □

Appendix

This appendix contains a brief description of the two directed random graph models used in the paper. Both of them have the property that their local neighborhoods converge in the local weak sense to a marked Galton-Watson process, which is the key to being able to define branching distributional fixed-point equations. Both of these models are suitable for modeling scale-free real-world networks with arbitrarily dependent in-degrees and out-degrees. Since real-world networks tend to be highly inhomogeneous, with many of them having degree distributions that follow power-laws, our assumptions ensure only that the in-degree and out-degree of a randomly chosen vertex have finite mean, allowing infinite higher order moments.

4.2 Directed configuration model

One model that produces graphs from any prescribed (graphical) degree sequence is the configuration or pairing model [10, 55], which assigns to each vertex in the graph a number of half-edges equal to its target degree and then randomly pairs half-edges to connect vertices.

We assume that each vertex i in the graph has a degree vector $\mathbf{D}_i = (D_i^-, D_i^+) \in \mathbb{N} \times \mathbb{N}$, where D_i^- and D_i^+ are the in-degree and out-degree of vertex i, respectively. In order for us to be able to draw the graph, we assume that the degree sequence $\{\mathbf{D}_i : 1 \leq i \leq n\}$ satisfies

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

Note that in order for the sum of the in-degrees to be equal to that of the out-degrees, it may be necessary to consider a double sequence $\{\mathbf{D}_i^{(n)} : i \geq 1, n \geq 1\}$ rather than a unique sequence; i.e., it may be convenient to allow $\mathbf{D}_i^{(n)} \neq \mathbf{D}_i^{(m)}$ for $n \neq m$.

Formally, the DCM can be defined as follows.

Definition 4.9. Let $\{\mathbf{D}_i : 1 \leq i \leq n\}$ be a degree sequence and let $V_n = \{1, 2, \ldots, n\}$ denote the nodes in the graph. To each node i assign D_i^- inbound half-edges and D_i^+ outbound half-edges. Enumerate all L_n inbound half-edges, respectively outbound half-edges, with the numbers $\{1, 2, \ldots, L_n\}$, and let $\mathbf{x}_n = (x_1, x_2, \ldots, x_{L_n})$ be a random permutation of these L_n numbers, chosen uniformly at random from the possible $L_n!$ permutations. The DCM with degree sequence $\{\mathbf{D}_i : 1 \leq i \leq n\}$ is the directed graph $G(V_n, E_n)$ obtained by pairing the x_ith outbound half-edge with the ith inbound half-edge.

We point out that instead of generating the permutation $\mathbf{x}_n$ of the outbound half-edges up front, one can construct the graph one vertex at a time, by pairing each of the inbound half-edges with an outbound half-edge, randomly chosen with equal probability from the set of unpaired outbound half-edges.

We emphasize that the DCM is in general a multi-graph, that is, it can have self-loops and multiple edges in the same direction. However, provided the pairing process does not create self-loops or multiple edges, the resulting graph is uniformly chosen among all graphs having the prescribed degree sequence. If one chooses this degree sequence according to a power-law, one immediately obtains a scale-free graph. It was shown in [18] 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. 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 any of our theorems, and therefore we do not require the DCM to result in a simple graph.

4.3 Inhomogeneous random digraphs

Alternatively, one could think of obtaining inhomogeneous degrees as a consequence of how likely different nodes are to have an edge between them. In the spirit of the classical Erdős-Rényi graph [30, 36, 8, 39, 11, 29], we assume that whether there is an edge between vertices i and j is determined by a coin-flip, independently of all other edges. Several models capable of producing graphs with inhomogeneous degrees while preserving the independence among edges have been suggested in the literature, including: the Chung-Lu model [20, 21, 22, 23], the Norros-Reittu model (or Poissonian random graph) [49, 55, 54], and the generalized random graph [55, 14, 54], to name a few. In all of these models, the inhomogeneity of the degrees is created by allowing the success probability of each coin-flip to depend on the "attributes" of the two vertices being connected; the scale-free property can then be obtained by choosing the attributes according to a power-law.

In order to obtain inhomogeneous degree distributions, to each vertex $i \in V_n$ we assign a *type* $\mathbf{W}_i = (W_i^-, W_i^+) \in \mathbb{R}_+ \times \mathbb{R}_+$. The W_i^- and W_i^+ will be used to determine how likely vertex i is to have inbound/outbound neighbors. As for the DCM, it may be necessary to consider a double sequence $\{\mathbf{W}_i^{(n)} : i \geq 1, n \geq 1\}$ rather than a unique sequence. With some abuse of notation, we will use $\mathscr{F}_n = \sigma(\mathbf{W}_i : 1 \leq i \leq n)$ to denote the sigma algebra generated by the type sequence, and define $\mathbb{P}_n(\cdot) = P(\cdot|\mathscr{F}_n)$ to be the conditional probability given the type sequence. In the main body of the paper we used $\mathscr{F}_n$ to denote the filtration of the vertex attributes $\{\mathbf{a}_i : 1 \leq i \leq n\}$, which include the types being used here.

We now define our family of random digraphs using the conditional probability, given the type sequence, that edge $(i, j) \in E_n$,

$$p_{ij}^{(n)} \triangleq \mathbb{P}_n\left((i,j) \in E_n\right) = 1 \wedge \frac{W_i^+ W_j^-}{\theta n}(1 + \varphi_n(\mathbf{W}_i, \mathbf{W}_j)), \qquad 1 \leq i \neq j \leq n, \tag{4.5}$$

where $-1 < \varphi_n(\mathbf{W}_i, \mathbf{W}_j) = \varphi(n, \mathbf{W}_i, \mathbf{W}_j, \mathscr{W}_n)$ a.s. is a function that may depend on the entire sequence $\mathscr{W}_n := \{\mathbf{W}_i : 1 \leq i \leq n\}$, on the types of the vertices (i, j), or exclusively on n, and $0 < \theta < \infty$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}(W_i^- + W_i^+) \xrightarrow{P} \theta, \qquad n \to \infty.$$

In the context of [12, 15], definition (4.5) corresponds to the so-called rank-1 kernel, i.e., $\kappa(\mathbf{W}_i, \mathbf{W}_j) = \kappa_+(\mathbf{W}_i)\kappa_-(\mathbf{W}_j)$, with $\kappa_+(\mathbf{W}) = W^+/\sqrt{\theta}$ and $\kappa_-(\mathbf{W}) = W^-/\sqrt{\theta}$.

References

[1] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Networks, shocks, and systemic risk. Technical report, National Bureau of Economic Research, 2015.

[2] E. Agirre, O. L. de Lacalle, and A. Soroa. Random walks for knowledge-based word sense disambiguation. *Computational Linguistics*, 40(1):57–84, 2014.

[3] D. J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equation. *The Annals of Applied Probability*, 15(2):1047–1110, 2005.

[4] G. Alsmeyer. Random recursive equations and their distributional fixed points. Unpublished manuscript, available at `https://www.uni-muenster.de/Stochastik/lehre/SS11/StochRekGleichungen/book.pdf`, 2012.

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

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

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

[8] T. L. Austin, R. E. Fagen, W. F. Penney, and J. Riordan. The number of components in random linear graphs. *The Annals of Mathematical Statistics*, 30:747–754, 1959.

[9] J .D. Biggins. Lindley-type equations in the branching random walk. *Stochastic Processes and their Applications*, 75(1):105–133, 1998.

[10] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, pages 311–316, 1980.

[11] B. Bollobás. *Random graphs.* Cambridge University Press, 2001.

[12] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31:3–122, 2007.

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

[14] T. Britton, M. Deijfen, and A. Martin-Läf. Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics*, 124(6):1377–1397, 2006.

[15] J. Cao and M. Olvera-Cravioto. Connectivity of a general class of inhomogeneous random digraphs. *Random Structures & Algorithms*, 56(3):722–774, 2020.

[16] C. Castellano and R. Pastor-Satorras. Zero temperature glauber dynamics on complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(05):P05001, 2006.

[17] N. Chen, N. Livtak, and M. Olvera-Cravioto. Generalized PageRank on directed configuration networks. *Random Structures & Algorithms*, 51(2):237–274, 2017.

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

[19] 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.

[20] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6:125–145, 2002.

[21] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. In *Proceedings of National Academy of Sciences*, volume 99, pages 15879–15882, 2002a.

[22] F. Chung and L. Lu. The volume of the giant component of a random graph with given expected degrees. *SIAM Journal on Discrete Mathematics*, 20(2):395–411, 2006a.

[23] F. Chung and L. Lu. *Complex graphs and networks*, volume 107. CBMS Regional Conference Series in Mathematics, 2006b.

[24] L. Damonte, G. Como, and F. Fagnani. Systemic risk and network intervention. *arXiv preprint arXiv:1912.08631*, 2019.

[25] M. J. de Oliveira. Isotropic majority-vote model on a square lattice. *Journal of Statistical Physics*, 66(1-2):273–281, 1992.

[26] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[27] A. Deo and S. Juneja. Limiting distributional fixed points in systemic risk graph models. In *2019 Winter Simulation Conference*, pages 878–889. IEEE, 2019.

[28] L. Devroye. On the probabilistic worst-case time of FIND. *Algorithmica*, 31:291–303, 2001.

[29] R. Durrett. *Random graph dynamics, Cambridge Series in Statistics and Probabilistic Mathematics*. Cambridge University Press, 2007.

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

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

[32] P. Frasca, C. Ravazzi, R. Tempo, and H. Ishii. Gossips and prejudices: Ergodic randomized dynamics in social networks. *IFAC Proceedings Volumes*, 46(27):212–219, 2013.

[33] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[34] N. E. Friedkin and E. C. Johnson. Social influence networks and opinion change. *Advances in Group Processes*, 16(1):1–29, 1999.

[35] A. Garavaglia, R. van der Hofstad, and N. Litvak. Local weak convergence for pagerank. *The Annals of Applied Probability*, 30(1):40–79, 2020.

[36] E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30:1141–1144, 1959.

[37] D. F. Gleich. Pagerank beyond the web. *SIAM Review*, 57(3):321–363, 2015.

[38] Z. Gyongyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with trustrank. In *Proceedings of the 30th international conference on very large data bases (VLDB)*, 2004.

[39] S. Janson, T. Luczak, and A. Rucinski. *Random graphs*. Wiley-Interscience, 2000.

[40] P. R. Jelenković and M. Olvera-Cravioto. Information ranking and power laws on trees. *Advances in Applied Probability*, 42(4):1057–1093, 2010.

[41] P. R. Jelenković and M. Olvera-Cravioto. Information ranking and power laws on trees. *Advances in Applied Probability*, 42(4):1057–1093, 2010.

[42] P. R. Jelenković and M. Olvera-Cravioto. Implicit renewal theorem and power tails on trees. *Advances in Applied Probability*, 44:528–561, 2012.

[43] P. R. Jelenković and M. Olvera-Cravioto. Implicit renewal theorem for trees with general weights. *Stochastic Processes and their Applications*, 122:3209–3238, 2012.

[44] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1877–1890, 2008.

[45] F.I. Karpelevich, M.Ya. Kelbert, and Yu.M. Suhov. Higher-order Lindley equations. *Stochastic Process. Appl.*, 53:65–96, 1994.

[46] D. Lacker, K. Ramanan, and R. Wu. Local weak convergence and propagation of ergodicity for sparse networks of interacting processes, 2020.

[47] J. Lee and M. Olvera-Cravioto. PageRank on inhomogeneous random digraphs. *Stochastic Processes and their Applications*, 130(4):1–57, 2020.

[48] T. Mach, A. Sturm, and J. M. Swart. A new characterization of endogeny. *Mathematical Physics, Analysis and Geometry*, 21(4):30, 2018.

[49] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75, 2006.

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

[51] M. Olvera-Cravioto. PageRank's behavior under degree correlations. *To appear in Annals of Applied Probability*, pages 1–43, 2019. arXiv:1909.09744.

[52] M. Olvera-Cravioto and O. Ruiz-Lacedelli. Stationary waiting time in parallel queues with synchronization. *Mathematics of Operations Research*, 2020.

[53] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii. Ergodic randomized algorithms and dynamics over networks. *IEEE Transactions on Control of Network Systems*, 2(1):78–87, 2014.

[54] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs. *Journal of Statistical Physics*, 133:169–202, 2008.

[55] R. van der Hofstad. *Random graphs and complex networks.* Cambridge University Press, 2016.

[56] A. L. M. Vilela and A. J. F. de Souza. Majority-vote model with a bimodal distribution of noises in small-world networks. *Physica A: Statistical Mechanics and its Applications*, 488:216–223, 2017.

[57] A. L. M. Vilela, C. Wang, K. P. Nelson, and H. E. Stanley. Majority-vote model for financial markets. *Physica A: Statistical Mechanics and its Applications*, 515:762–770, 2019.

[58] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[59] Y. Volkovich and N. Litvak. Asymptotic analysis for personalized web search. *Advances in Applied Probability*, 42(2):577–604, 2010.

[60] M. Yang, X. Qu, Z. Cao, and X. Yang. Innovation governs everything eventually: Extensions of the degroot model. *Acta Mathematicae Applicatae Sinica, English Series*, 33(1):35–42, 2017.