LOCAL WEAK LIMITS FOR COLLAPSED BRANCHING PROCESSES WITH RANDOM OUT-DEGREES

SAYAN BANERJEE, PRABHANKA DEKA, MARIANA OLVERA-CRAVIOTO

Abstract. We obtain local weak limits in probability for Collapsed Branching Processes (CBP), which are directed random networks obtained by collapsing random-sized families of individuals in a general continuous-time branching process. The local weak limit of a given CBP, as the network grows, is shown to be a related continuous-time branching process stopped at an independent exponential time. This is done through an explicit coupling of the in-components of vertices with the limiting object. We also show that the in-components of a finite collection of uniformly chosen vertices locally weakly converge (in probability) to i.i.d. copies of the above limit, reminiscent of propagation of chaos in interacting particle systems. We obtain as special cases novel descriptions of the local weak limits of directed preferential and uniform attachment models. We also outline some applications of our results for analyzing the limiting in-degree and PageRank distributions.

1. Introduction

We analyze in this paper an evolving directed random graph model that is obtained by collapsing a continuous-time branching process driven by a general Markovian pure birth process. Our model corresponds to a graph process where incoming individuals (nodes) arrive in families, or groups, each having a random number of individuals. Upon arrival, each member of the family chooses one other existing node with probability proportional to a function f (called the attachment function) of its degree (one plus the number of inbound edges), and connects to it using a directed outbound edge. The members of each family arrive sequentially, so that if the first family has D_1^+ members, then the $(D_1^+ + 1)$th individual belongs to the second family. For a graph with n groups, the process continues according to this rule until all the $S_n = D_1^+ + \cdots + D_n^+$ individuals have been connected to the graph, at which point we proceed to merge all the individuals in each family into a single "group vertex". The result is a directed multigraph $G(V_n, E_n)$, having vertices $V_n = \{1, 2, \ldots, n\}$ and edges in the set E_n, that models the connections among families. Our choice of notation D_i^+ for the size of the ith family comes from the observation that D_i^+ becomes the out-degree (number of outbound edges) of vertex i in V_n.

When the attachment function is linear (or constant), $G(V_n, E_n)$ constructed in this way corresponds to a directed preferential attachment graph with random out-degrees and random additive fitness (respectively, a directed uniform attachment graph with random out-degrees). In this case, vertices in the graph can be thought of as individuals, who upon arrival choose a random number of vertices to connect to, with edges always pointing from younger vertices to older ones.

For the case where D_i^+ is a fixed constant d for all i, the above model was introduced in [10] under the name of Collapsed Branching Process (CBP), and we will use this nomenclature for our more general model. The limiting degree distribution was investigated in [10]. The analysis of the model with random out-degrees and, in particular, the description

---

2010 *Mathematics Subject Classification.* 05C80, 60J80, 41A60, 60B10.

Key words and phrases. Local weak limit, collapsed branching processes, continuous-time branching processes, random out-degrees, couplings, directed preferential attachment, in-degree, PageRank, power laws.

of local weak limits (asymptotics of neighborhoods of typical vertices in large networks) were left as open problems.

The main focus of this paper is to describe the local weak limit for the collapsed branching process (CBP) graph for a general form of the attachment function. Local weak convergence is informally described as the phenomenon where finite neighborhoods of a uniformly chosen vertex in a growing sequence of finite graphs converges weakly to the neighborhood of the root in some limiting graph (finite or infinite). This concept was introduced in [1, 4] and has turned out to be an indispensable tool in understanding the local geometry of large graphs through, for example, the degree and PageRank distributions (see Section 4 for the definition and properties of this centrality measure popularized by Google), the size of giant components and the behavior of random walks on them. Moreover, local weak convergence has been used to show that a variety of random graphs are *locally tree-like*, that is, non-tree graph sequences have local limits which are trees. This phenomenon also applies to our model, as will be seen below. See [22] for a detailed treatment of local weak convergence and its applications.

For random graphs, most of the existing results on local weak convergence concern *static* random graphs like Erdős-Renyi graphs, inhomogeneous random graphs and the configuration model [22, Chapters 2, 3, 4 and references therein]. Here, the randomness comes from the degree distribution and edge connection probabilities but there is no temporal evolution, making the roles of vertices *exchangeable*. Consequently, the local weak limits for most of those graphs correspond to Galton-Watson trees where, in particular, every vertex in the same generation has the same progeny distribution. However, there has been very little work on local weak convergence of dynamic random graphs (graphs evolving over time). The main obstacle is that the time evolution assigns ages to the vertices and the local geometry around a vertex depends crucially on its age. The local weak limit of the CBP in the tree case was obtained in [21], building on the work of [15, 12], where the limit was shown to be distributed as the same CBP but stopped at an independent exponential time. The directed preferential attachment model DPA(d, β) with fixed out-degrees d and fixed additive fitness β (CBP with all out-degrees d and attachment function $f(k) = k + \beta/d$) was analyzed in [5] using an encoding of the graph in terms of a Pólya urn type scheme (see [5, Theorem 2.1]) to describe the local limit as the so-called *Pólya-point graph* [5, Section 2.3.2]. The DPA$(1, \beta)$ with random β was studied in [14]. Local limits for dynamic random trees where the attachment probabilities are non-local functionals of the vertex (like its PageRank) were recently obtained in [2]. The local weak limit of preferential attachment type models with random i.i.d. out-degrees and fixed additive fitness was derived in [9] using a generalization of the Pólya-point graph. We remark here that the Pólya representation of the pre-limit graph process, which forms the starting point of the results in [5, 14, 9], is intrinsic to a linear attachment function and does not extend to more general attachment schemes.

The main contribution of the current article is the construction of an explicit coupling between the exploration of the in-component of a uniformly chosen vertex in the CBP graph (with general f satisfying mild assumptions) and its local weak limit, described by a marked continuous-time branching process (CTBP) stopped at an independent exponential time. The CTBP appearing in the limit has a simple description: it is a Crump-Mode-Jagers (CMJ) branching process (also appearing in [15, 12, 21]) where each vertex reproduces according to a point process that equals a random number (distributed as the out-degree) of independent superimposed copies of a Markovian pure birth process ξ_f (see (1)). The rate of the exponential time equals the *Malthusian rate* of the CMJ process driven by ξ_f (see Assumption 1(i) below). Our coupling is well-defined for the simultaneous exploration of the in-components of *all the vertices* in the graph. In particular, the in-components of any finite collection of uniformly chosen vertices are successfully coupled with high probability with i.i.d. copies of the local weak limit. This is reminiscent of the phenomenon of *propagation of chaos* in interacting particle systems [7]. This type of coupling, which could include in applications the addition of vertex attributes that may depend on the out-degrees, has been established for static graphs in [19], where it is referred to as a *strong coupling.* Our results for the strong coupling of the CBP graph are summarized in Theorem 1 and yield local weak convergence in probability as a corollary (Corollary 1.1). Note that our results establish the local weak convergence of (the in-components in) the DPA(d, β) and the CBP tree, obtained respectively in [5] and [21], as special cases. The local weak convergence implies the joint distributional convergence of the empirical in-degree and PageRank distributions (Corollary 1.2). Asymptotics of the in-degree distribution for regularly varying out-degrees are quantified for the preferential and uniform attachment models in Proposition 1.

We note that, stemming from our interest in the asymptotic behavior of the degree and PageRank distributions, our couplings and local limits apply to the in-components only, rather than to the joint in- and out-components described in [5, 9]. However, our description of the limit is somewhat more intuitive and analytically tractable for refined large deviations analysis, as one can apply a host of tools from the well-studied theory of continuous-time Markov chains and CMJ processes (see the discussion after Corollary 1.2). This was already exhibited in [3] where the Pólya-point graph was redescribed as a randomly stopped CMJ process (our local limit for fixed out-degree and linear f) to compute the tail exponent of the limiting PageRank distribution in the DPA(d, β) model. Moreover, we require the out-degree distribution to only have finite first moment compared to higher moments required by [9]. We believe that a generalization of our techniques will lead to the joint local limit of the in- and out-components, and the limiting infinite tree can be described in terms of a forest of independent CMJ processes (run until different random times) emanating from vertices of the limiting out-component. We leave this analysis for a later work.

The rest of the paper is organized as follows. Section 2 provides a detailed construction of the CBP; Section 3 contains our main theorem establishing the strong coupling, and Section 4 describes some applications of the local limit to the analysis of the PageRank and in-degree distributions. Finally, Section 5 contains all the proofs, with the description of the coupling in subsections 5.1 and 5.2.

2. The collapsed branching process

To construct a collapsed branching process (CBP) with random out-degrees we start by defining a continuous time branching process (CTBP) $\boldsymbol{\xi}$ driven by a Markovian pure birth process $\{\xi_f(t) : t \geq 0\}$ satisfying $\xi_f(0) = 0$ and having birth rates

$$P(\xi_f(t + dt) = k + 1 | \xi_f(t) = k) = f(k + 1)dt + o(dt),$$

where $f : \mathbb{N} \to \mathbb{R}_+$ is a nonnegative function. Number each of the nodes in this CTBP according to the order of their arrival, with the root being labeled node 1. Let σ_i denote the time of arrival of node i in the CTBP and let $\mathcal{T}(t)$ denote the discrete skeleton of the graph determined by $\boldsymbol{\xi}$ at time t, with directed edges pointing from an offspring to its parent.

Independently of the CTBP $\boldsymbol{\xi}$, we will also construct a sequence of i.i.d. random variables $\{D_i^+ : i \geq 1\}$ taking values on $\mathbb{N} := \{1, 2, \ldots\}$ and having distribution function $H(x) = P(D_1^+ \leq x)$ with $\mu := E[D_1^+] < \infty$. We will use this sequence $\mathbf{D}_n = \{D_i^+ : 1 \leq i \leq n\}$ in combination with $\boldsymbol{\xi}$ to construct a vertex-weighted directed graph $G(V_n, E_n)$ with vertex set $V_n := \{1, 2, \ldots, n\}$. Let $S_k = D_1^+ + \cdots + D_k^+$, $S_0 = 0$.

To start, define the sets

$$V(i) = \{S_{i-1} + 1, S_{i-1} + 2, \ldots, S_i\}, \qquad i \geq 1.$$

The directed graph $G(V_n, E_n)$ is obtained by collapsing all the nodes in $V(i)$ into the vertex i, matching the outbound edges of its D_i^+ nodes with the merged vertices their parents in $\mathcal{T}(\sigma_{S_n})$ belong to. Note that $G(V_n, E_n)$ is a multigraph, since it may contain self-loops and multiple edges in the same direction between the same two vertices. Figure 1 illustrates the collapsing procedure and the resulting multigraph.

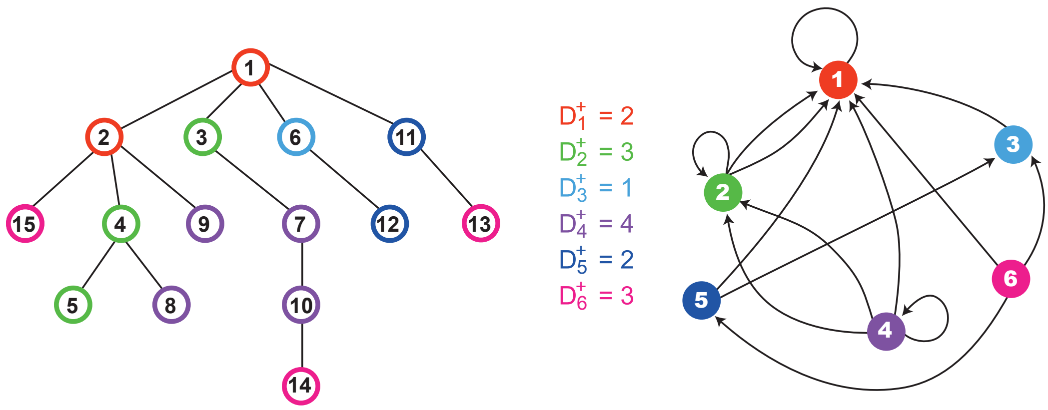

FIGURE 1. Collapsed branching process. On the left the tree $\mathcal{T}(\sigma_{S_6})$, on the right the corresponding graph $G(V_6, E_6)$.

Note that if f is linear, i.e., $f(k) = ck + \beta$ for some constants c, β satisfying $c + \beta > 0$, then $G(V_n, E_n)$ can be seen as an evolving random directed rooted graph sequence $\{G_\ell\}_{\ell \geq 1}$ where G_1 contains one vertex having D_1^+ outbound edges pointing towards itself, and for $\ell \geq 2$, G_ℓ is constructed from $G_{\ell-1}$ by adding one vertex, labeled ℓ, having D_ℓ^+ outbound edges that are connected, one at a time, with the kth edge, $1 \leq k \leq D_\ell^+$, choosing to connect to vertex i with probability:

$$P\left(k^{th} \text{ outbound edge of } \ell \text{ attaches to } i \,\middle|\, G_{\ell-1}, D_\ell^+\right)$$
$$= \begin{cases} \frac{cD_i(\ell-1,k-1)+\beta D_i^+}{\sum_{j=1}^{\ell}(cD_j(\ell-1,k-1)+\beta D_j^+)}, & 1 \leq i \leq \ell - 1, \\ \\ \frac{cD_\ell(\ell-1,k-1)+\beta(k-1)}{\sum_{j=1}^{\ell}(cD_j(\ell-1,k-1)+\beta D_j^+)}, & i = \ell, \end{cases}$$

where $D_i(\ell - 1, k - 1)$ is the total degree (in-degree plus out-degree) of vertex i after $k - 1$ edges of vertex ℓ are attached. The case $c = 1$ corresponds to a directed preferential attachment graph, while the case $c = 0$ corresponds to the directed uniform attachment graph, both with random out-degrees distributed according to H and *random additive fitness* (coming from the βD_i^+ term).

Remark 1. *To avoid confusion, we will always refer to the vertices in* $G(V_n, E_n)$ *as "vertex/vertices", while we will refer to individuals in the CTBP* $\boldsymbol{\xi}$, *or its discrete skeleton* $\{\mathcal{T}(t) : t \geq 0\}$, *as "node/nodes".*

3. Coupling with a marked continuous time branching process

The existence of a local weak limit for $G(V_n, E_n)$ requires that we impose some conditions on the function f that drives the CTBP $\boldsymbol{\xi}$.

To start, define

$$\hat{\rho}(\theta) := E\left[\int_0^\infty e^{-\theta s}\xi_f(ds)\right] = E\left[\sum_{n=1}^\infty e^{-\theta\tau_n}\right] = E\left[\sum_{n=1}^\infty e^{-\theta\sum_{i=1}^n \chi_i/f(i)}\right]$$
$$= \sum_{n=1}^\infty \prod_{i=1}^n E\left[e^{-\theta\chi_i/f(i)}\right] = \sum_{n=1}^\infty \prod_{i=1}^n \frac{1}{\theta/f(i)+1} = \sum_{n=1}^\infty \prod_{i=1}^n \frac{f(i)}{\theta+f(i)},$$

where $\{\chi_i : i \geq 1\}$ is a sequence of i.i.d. exponential random variables with rate one, and τ_n is the time of the nth birth of $\{\xi_f(t) : t \geq 0\}$.

Assumption 1. *Suppose the out-degrees $\{D_i^+ : i \geq 1\}$ are i.i.d., with $\mu = E[D_1^+] < \infty$. In addition, suppose the following hold:*

(i) There exists $\lambda > 0$ such that $\hat{\rho}(\lambda) = 1$.
(ii) $f(k) \leq C_f k$, $k \geq 1$ for some constant $C_f < \infty$.
(iii) $f_* := \inf_{i\geq 1} f(i) > 0$.
(iv) Let $\underline{\theta} := \inf\{\theta > 0 : \hat{\rho}(\theta) < \infty\}$ and suppose that

$$\lim_{\theta\searrow\underline{\theta}} \hat{\rho}(\theta) > 1.$$

Note that $\lambda > 0$ in Assumption 1 is the Malthusian rate of $\{\xi_f(t) : t \geq 0\}$.

The local weak limit of $G(V_n, E_n)$ is given by a marked continuous time branching process, whose discrete marked skeleton (the graph obtained by removing time labels from the nodes but retaining their marks) at time t will be denoted $\mathcal{T}^c(t, \boldsymbol{\mathcal{D}})$, where $\boldsymbol{\mathcal{D}} := \{\mathcal{D}_k : k \geq 1\}$ is an i.i.d. sequence having distribution H. The marks $\boldsymbol{\mathcal{D}}$ play a role in the construction of the local weak limit, and become vertex marks in its discrete skeleton. To describe this CTBP, define for each $k \geq 1$

$$\bar{\xi}_f^{(k)} = \sum_{i=1}^{\mathcal{D}_k} \xi_f^{k,i}, \tag{1}$$

where $\mathcal{D}_k$ is the kth element of $\boldsymbol{\mathcal{D}}$ and the $\{\xi_f^{k,i} : i \geq 1, k \geq 1\}$ are i.i.d. copies of ξ_f. Each node in the tree is indexed in the order in which it arrives, and node k has as its mark $\mathcal{D}_k$. Let $\mathcal{T}^c(t, \boldsymbol{\mathcal{D}})$ denote the discrete marked skeleton at time t of a marked CTBP driven by $\{(\mathcal{D}_k, \bar{\xi}_f^{(k)}) : k \geq 1\}$, conditionally on the root being born at time $t = 0$. We will denote the corresponding unmarked discrete skeleton by $\mathcal{T}^c(t)$.

Throughout the paper we will use $\mathcal{G}_i^{(n)}$ to denote the subgraph of $G(V_n, E_n)$ rooted at vertex i that corresponds with the exploration of its in-component, where the exploration is such that we only follow the inbound edges, but also keep track of the out-degrees $\{D_j^+ : j \in \mathcal{G}_i^{(n)}\}$ as we go; however, we do not follow the outbound edges. As before, $\mathcal{G}_i^{(n)}$ denotes the unmarked graph, while $\mathcal{G}_i^{(n)}(\mathbf{D}^+)$ the marked one. With some abuse of notation, we will write $j \in \mathcal{G}_i^{(n)}$ to refer to a vertex in $\mathcal{G}_i^{(n)}$.

Definition 3.1. *We say that two multigraphs $G(V, E)$ and $G(V', E')$ are* isomorphic *if there exists a bijection $\theta : V \to V'$ such that $l(i) = l(\theta(i))$ and $e(i, j) = e(\theta(i), \theta(j))$ for all $i \in V$ and all $(i, j) \in E$, where $l(i)$ is the number of self-loops of vertex i and $e(i, j)$ is the number of edges from vertex i to vertex j. In this case, we write $G \simeq G'$.*

For nodes in trees, we use the symbol $\emptyset$ to denote the root and enumerate its vertices with labels of the form $\mathbf{i} = (i_1, \ldots, i_k) \in \mathbb{N}^k$, where $(i_1, \ldots, i_k, i_{k+1})$ is the i_{k+1}th offspring of node $(i_1, \ldots, i_k)$; nodes with labels $i \in \mathbb{N}$ are offspring of the root $\emptyset$. Define the Ulam-Harris set $\mathcal{U} = \bigcup_{k=0}^\infty \mathbb{N}^k$, with the convention that $\mathbb{N}^0 \equiv \{\emptyset\}$. With a slight abuse of notation, we will write $\mathcal{D}_{\mathbf{j}}$ to refer to the mark of a node labeled $\mathbf{j} \in \mathcal{U}$. As mentioned earlier, we will also index nodes in dynamic trees in the order in which they arrive. In this case, for $j \in \mathbb{N}$, $j \in \mathcal{T}$ will denote the jth node to arrive in the tree $\mathcal{T}$.

Now we state the main result in the paper.

Theorem 1. *Suppose Assumption 1 holds. Then, for the CBP $G(V_n, E_n)$ we have:*

i) For $n \in \mathbb{N}$, if I_n is uniformly chosen in V_n, independently of anything else, then, there exists a coupling $\mathcal{P}_n$ of $\mathcal{G}_{I_n}^{(n)}(\mathbf{D}^+)$ and $(\chi, \{\mathcal{T}^c(t, \boldsymbol{\mathcal{D}}) : t \geq 0\})$, where $\boldsymbol{\mathcal{D}} := \{\mathcal{D}_k : k \geq 1\}$ is an i.i.d. sequence having distribution H and χ is an exponential random variable with rate λ, independent of $\{\mathcal{T}^c(t, \boldsymbol{\mathcal{D}}) : t \geq 0\}$, such that the event

$$C_{I_n} = \left\{ \mathcal{G}_{I_n}^{(n)} \simeq \mathcal{T}^c(\chi), \bigcap_{\mathbf{j} \in \mathcal{T}^c(\chi)} \{D^+_{\theta(\mathbf{j})} = \mathcal{D}_{\mathbf{j}}\} \right\},$$

where $\theta : \mathcal{U} \to \mathbb{N}$ is the bijection defining $\mathcal{G}_{I_n}^{(n)} \simeq \mathcal{T}^c(\chi)$, satisfies

$$\mathcal{P}_n \left(C_{I_n} \right) \to 1, \qquad as \qquad n \to \infty.$$

ii) Fix $m \in \mathbb{N}$. For $n \in \mathbb{N}$ and $\{I_{n,j} : 1 \leq j \leq m\}$ i.i.d. and uniformly chosen in V_n, independently of anything else, there exists a coupling $\mathcal{P}_{n,m}$ of $\left(\mathcal{G}_{I_{n,j}}^{(n)}(\mathbf{D}^+) \right)_{1 \leq j \leq m}$ and i.i.d. copies of $(\chi, \{\mathcal{T}^c(t, \boldsymbol{\mathcal{D}}) : t \geq 0\})$, denoted $(\chi_j, \{\mathcal{T}_j^c(t, \boldsymbol{\mathcal{D}}) : t \geq 0\})_{1 \leq j \leq m}$, such that the events $C_{I_{n,j}}$ defined as in part (i) satisfy

$$\mathcal{P}_{n,m} \left(\bigcap_{j=1}^{m} C_{I_{n,j}} \right) \to 1, \qquad as \qquad n \to \infty.$$

Theorem 1 implies, in particular, the local weak convergence in probability of $G(V_n, E_n)$, recorded in the following corollary.

Corollary 1.1. *Suppose Assumption 1 holds. For any fixed finite tree T and any deterministic sequence $\{d_{\mathbf{j}} : \mathbf{j} \in \mathcal{U}\} \subseteq \mathbb{N}$,*

$$\frac{1}{n} \sum_{i=1}^{n} 1 \left(\mathcal{G}_i^{(n)} \simeq T, \bigcap_{\mathbf{j} \in T} \{D^+_{\theta_i(\mathbf{j})} = d_{\mathbf{j}}\} \right) \xrightarrow{P} P \left(\mathcal{T}^c(\chi) \simeq T, \bigcap_{\mathbf{j} \in T} \{\mathcal{D}_{\mathbf{j}} = d_{\mathbf{j}}\} \right),$$

where $\theta_i : \mathcal{U} \to \mathbb{N}$ is the bijection defining $\mathcal{G}_i^{(n)} \simeq T$ and $\mathcal{T}^c(\chi, \boldsymbol{\mathcal{D}})$ is the marked discrete skeleton of the (randomly stopped) CTBP from the theorem.

In the sequel, we will suppress the dependence of the marked discrete skeleton $\mathcal{T}^c(\cdot, \boldsymbol{\mathcal{D}})$ on the marks $\boldsymbol{\mathcal{D}}$, and simply write $\mathcal{T}^c(\cdot)$, when there is no risk of ambiguity.

4. Applications of the local limit

In this section we give some basic applications of the local limit $\mathcal{T}^c(\chi)$ for understanding asymptotic properties of the distributions of the in-degree and the PageRank of a typical vertex in the original CBP. We will assume throughout that Assumption 1 holds.

PageRank, introduced by Brin and Page [20], is a celebrated centrality measure on networks whose goal is to rank vertices in a graph according to their 'popularity'. Specifically, the PageRank score of vertex v corresponds to the long-run proportion of time that a certain random walk spends on vertex v, hence the 'popularity' interpretation. The PageRank of a vertex is known to be heavily influenced by its in-degree and by the PageRank scores of its close inbound neighbors [18]. Formally, PageRank is defined as follows.

Let $G = G(V, E)$ be a directed network with vertices V and edge set E. For each vertex $v \in V$, let d_v^- and d_v^+ denote its in-degree and out-degree, respectively. Writing $|V|$ for the number of vertices in the graph and the vertices as $\{1, \ldots, |V|\}$, let A denote the adjacency matrix of G defined as the $|V|\times|V|$ matrix whose (i,j)th element is the number of directed edges from vertex i to vertex j. Let Δ be the diagonal matrix whose ith element is $1/d_i^+$ if $d_i^+>0$ and 0 if $d_i^+=0$. Let P be the matrix product ΔA with the zero rows replaced with the probability vector $\mathbf{q}=|V|^{-1}\mathbf{1}$. Note that P is a stochastic matrix (all rows sum to one). The PageRank vector $\boldsymbol{\pi}=(\pi_1,\dots,\pi_{|V|})$, with *damping factor* $c\in(0,1)$, is defined as the solution to the following system of equations:

$$\boldsymbol{\pi}=\boldsymbol{\pi}(cP)+(1-c)\mathbf{q}.$$

As the matrix cP is substochastic (its rows sum to c), the system of equations is guaranteed to have a unique solution given by:

$$\boldsymbol{\pi}=(1-c)\mathbf{q}(I-cP)^{-1}=(1-c)\mathbf{q}\sum_{k=0}^{\infty}(cP)^k.$$

We will consider the *scale-free* PageRank $\boldsymbol{R}:=|V|\boldsymbol{\pi}$. For $n\in\mathbb{N}$ and $1\le i\le n$, let $D_i^-(n),R_i(n)$ denote the in-degree and (scale-free) PageRank of the ith vertex in the CBP $G(V_n,E_n)$. Let $\mathcal{N}_\emptyset,\mathcal{R}_\emptyset$ be the in-degree and PageRank of the root in the coupled marked tree $\mathcal{T}^c(\chi)$ from Theorem 1. Then the following holds as a corollary of Theorem 1.

Corollary 1.2. *For any continuity point r of the distribution function of $\mathcal{R}_\emptyset$ and $k\in\mathbb{N}\cup\{0\}$, we have*

$$\frac{1}{n}\sum_{i=1}^{n}1(D_i^-(n)\ge k,R_i(n)>r)\xrightarrow{P}P(\mathcal{N}_\emptyset\ge k,\mathcal{R}_\emptyset>r)$$

as $n\to\infty$.

The above corollary follows from Corollary 1.1 exactly as in the proof of [3, Theorem 4.6] (see also [11, Theorem 2.1]) and its proof is omitted. Corollary 1.2 not only identifies the limiting joint in-degree and PageRank distribution, but the limiting objects are especially tractable from the point of view of moment computations and large deviations analysis, as they are defined as explicit functionals of the continuous time Markov chain $\{\mathcal{T}^c(t):t\ge0\}$ and an independent exponential random variable χ. Specifically, $\mathcal{N}_\emptyset\overset{d}{=}\bar{\xi}_f^{(k)}(\chi)=\sum_{i=1}^{\mathcal{D}_k}\xi_f^{k,i}(\chi)$, and $\mathcal{X}_\emptyset:=\mathcal{R}_\emptyset/\mathcal{D}_\emptyset$ satisfies a renewal-type equation which we now describe. For $t\ge0$, let $\mathcal{N}_\emptyset(t)$ and $\mathcal{R}_\emptyset(t)$ denote the in-degree and PageRank of the root $\emptyset$ in $\mathcal{T}^c(t)$. Let $\{\sigma_i^\emptyset\}_{i\ge1}$ denote the birth times of the children of the root in $\{\mathcal{T}^c(t):t\ge0\}$. For $t\ge0$ and $i\ge1$, let $\mathcal{D}_i$ and $\mathcal{R}_i(t)$ denote the mark (out-degree) and PageRank, respectively, of the ith child of $\emptyset$ in $\mathcal{T}^c(t+\sigma_i^\emptyset)$. Write $\mathcal{X}_\emptyset(t):=\mathcal{R}_\emptyset(t)/\mathcal{D}_\emptyset$ and $\mathcal{X}_i(t):=\mathcal{R}_i(t)/\mathcal{D}_i$. Then we have the following identity:

$$\mathcal{X}_\emptyset\overset{d}{=}\mathcal{X}_\emptyset(\chi)=\frac{c}{\mathcal{D}_\emptyset}\sum_{i=1}^{\mathcal{N}_\emptyset(\chi)}\mathcal{X}_i(\chi-\sigma_i^\emptyset)+\frac{1-c}{\mathcal{D}_\emptyset}.$$

The above identity has a nice recursive structure because, conditionally on $\{\sigma_i^\emptyset\}_{i\ge1}$, $\{\mathcal{X}_i(\cdot):i\ge1\}$ are i.i.d. having the same distribution as $\mathcal{X}_\emptyset(\cdot)$. This identity enables the detailed analysis of the distribution of $\mathcal{X}_\emptyset$ through renewal theoretic techniques [17, 16], generator-based methods [3], and the extensive theory of Crump-Mode-Jagers branching processes [12, 15]. This approach was already used in [3] for the DPA(d,β) model.

A future companion paper to this work will present the large deviations behavior of the distributions of the in-degree and the PageRank of a typical vertex in the general CBP. As that work will show, this behavior is heavily determined by the attachment function f and spans the entire range of distributions, from exponential tails to regularly varying ones. However, for illustration purposes, we include here some observations about the limiting in-degree distribution in the preferential and uniform attachment cases.

Proposition 1. *Let $\mathcal{N}_\emptyset$ denote the in-degree of the root of $\mathcal{T}^c(\chi)$. Let $h(d) := H(d) - H(d-1) = P(\mathcal{D} = d)$, $d \in \mathbb{N}$, be the pmf of the node marks. Then,*

(1) **Preferential attachment:** *If $f(k) = k+\beta$, with $\beta > -1$, then for any $x \in \mathbb{N}\cup\{0\}$,*

$$P(\mathcal{N}_\emptyset = x) = \sum_{d=1}^{\infty} h(d)(2+\beta)\frac{\Gamma(2+\beta+d(\beta+1))\,\Gamma(x+d(\beta+1))}{\Gamma(d(\beta+1))\,\Gamma(x+d(\beta+1)+3+\beta)},$$

where $\Gamma(\cdot)$ is the Gamma function. In particular, if $h(d) = d^{-\gamma}L(d)$ for some finite $\gamma \geq 2$ and slowly varying function $L(\cdot)$ (with $E[\mathcal{D}] < \infty$), then

$$P(\mathcal{N}_\emptyset = x) = (1+o(1))K_\gamma P(XY > x)$$

as $x \to \infty$, where $P(X > x) = (1+\beta+x)^{-3-\beta}$ for $x > -\beta$ is a Type II Pareto random variable, independent of Y, and $P(Y = d) = d^{-\gamma-1}l(d)L(d)/K_\gamma$ for $d \in \mathbb{N}$, where $K_\gamma = \sum_{d=1}^{\infty} d^{-\gamma-1}l(d)L(d)$ and

$$l(d) := \frac{(2+\beta)\Gamma(2+\beta+d(\beta+1))}{d^{2+\beta}\Gamma(d(\beta+1))} \to (2+\beta)(\beta+1)^{2+\beta}, \qquad d \to \infty.$$

Moreover, $x \mapsto P(XY > x)$ is regularly varying with tail index $\min\{3+\beta, \gamma\}$, and

$$K_\gamma P(XY > x) = (1+o(1))\begin{cases} K_\gamma E[Y^{3+\beta}]x^{-3-\beta}, & \text{if } \gamma > 3+\beta \\ E[(X^+)^\gamma]l(\infty)L(x)x^{-\gamma}, & \text{if } 2 \leq \gamma < 3+\beta, \end{cases}$$

as $x \to \infty$.

(2) **Uniform attachment:** *If $f(k) \equiv \beta$ for some $\beta > 0$, then for any $x \in \mathbb{N}\cup\{0\}$,*

$$P(\mathcal{N}_\emptyset = x) = \sum_{d=1}^{\infty} h(d)\frac{1}{d+1}\left(1+\frac{1}{d}\right)^{-x}.$$

In particular, if $h(d) = d^{-\gamma}L(d)$ for some finite $\gamma \geq 2$ and slowly varying function $L(\cdot)$ (with $E[\mathcal{D}] < \infty$), then

$$P(\mathcal{N}_\emptyset = x) = (1+o(1))E[W^\gamma]x^{-\gamma}L(x)$$

as $x \to \infty$, where W is an exponential random variable with mean one.

The above result shows in particular that, for the preferential attachment case with regularly varying out-degree distribution with exponent γ, the in-degree is regularly varying with the same exponent as the out-degree if $\gamma \leq 3+\beta$. Otherwise, the tail exponent matches the degree exponent in the tree case (our model with all out-degrees equal to one). Comparing this result to the degree distribution in the preferential attachment model with random out-degrees but deterministic additive fitness studied in [8] (see [8, Proposition 1.4]), we observe that if we choose in our model the fitness parameter to be $\beta = \delta/\mu$ (to make it on average δ after the collapsing procedure) and set the additive fitness parameter of the model in [8] to be δ, then we obtain the same transition in the tail exponent for both models. The model in [8] can also be exactly obtained via a closely related CBP where the attachment function f_v of an incoming vertex v depends on its out-degree D^+ as $f_v(k) = k + \beta/D^+$, $k \in \mathbb{N}$. Since the construction of the CBP can be done conditionally on the out-degree sequence, one can obtain local limits for this variant using a similar construction to the one used here.

For the uniform attachment case, we observe the (somewhat surprising) phenomenon that, although the in-degree distribution has exponential tails for deterministic out-degrees, making the out-degree distribution regularly varying also makes the in-degree distribution regularly varying with the same exponent.

5. Proofs

The rest of the paper contains the proofs of Theorems 1 and Corollaries 1.1 and 1.

5.1. **Coupling Construction for $m = 1$.** Let $\{\mathcal{T}(t) : t \geq 0\}$ denote the discrete skeleton of the CTBP $\boldsymbol{\xi}$ described earlier. Recall that the graph $G(V_n, E_n)$ is obtained by simply collapsing the nodes in the sets $\{V(i) : 1 \leq i \leq n\}$. Clearly, it suffices to run $\boldsymbol{\xi}$ until the time σ_{S_n}.

Fix $i \in V_n$ and define

$$t_{n,i} := \frac{1}{\lambda} \log(n/i).$$

We will now explain how to construct a coupling between $\mathcal{G}_i^{(n)}$ and a tree $\mathcal{T}_i^c(t_{n,i})$ that evolves according to the law of $\{\mathcal{T}^c(t) : t \geq 0\}$. Recall that in order to simplify the notation, we have omitted the dependence on the marks $\mathbf{D}^+$ and $\boldsymbol{\mathcal{D}}$, and simply write $\mathcal{G}_i^{(n)}$ and $\mathcal{T}_i^c(t_{n,i})$ for the marked graphs.

This coupling will only be successful with high probability for large values of i, so although it will be well-defined for any $i \geq 1$, it will most likely fail to satisfy $\mathcal{G}_i^{(n)} \simeq \mathcal{T}_i^c(t_{n,i})$ if i is not sufficiently large.

We start by sampling the out-degrees $\{D_i^+ : 1 \leq i \leq n\}$ and the tree $\mathcal{T}(\sigma_{S_n})$. Note that at this point the entire graph $G(V_n, E_n)$ has been sampled, so steps 1-4 in the construction below are deterministic. Specifically, we will copy (re-use) some of the vertices in $\mathcal{G}_i^{(n)}$ and their birth times to construct $\mathcal{T}_i^c(t_{n,i})$, but potentially ignore others. To start, for $j \geq 1$, let $\kappa_i(j)$ denote the label in $G(V_n, E_n)$ of the jth oldest vertex in $\mathcal{G}_i^{(n)}$, with $\kappa_i(1) = i$. Nodes in $\{\mathcal{T}_i^c(t) : t \geq 0\}$ will be of two kinds, those that we will copy from $\mathcal{G}_i^{(n)}$ and those that will be generated independently. The construction below collects the vertices that are copied from $\mathcal{G}_i^{(n)}$ onto nodes in $\{\mathcal{T}_i^c(t) : t \geq 0\}$ in the sets J_i and J_i^*. It will also create additional 'dummy nodes' in $\{\mathcal{T}_i^c(t) : t \geq 0\}$ when certain types of miscouplings occur, and those will be collected in the set J_i^d but will receive no labels.

We will use $\tau_{i,j}$ to denote the birth time in $\mathcal{T}_i^c(\cdot)$ of the node corresponding to the jth vertex to be added to $J_i \cup J_i^*$. Vertices in $\mathcal{G}_i^{(n)}$ will be explored in the order of their ages. When the jth explored vertex has multiple nodes with outgoing edges to vertices in J_i, a dummy vertex is created for every such node except the oldest of these nodes. For each such node, say ω', the associated birth time in $\mathcal{T}_i^c$ is denoted by $\tau_{i,\omega'}$. The purpose of these dummy nodes is to ensure that the law of $\mathcal{T}_i^c(\cdot)$ agrees with that of $\mathcal{T}^c(\cdot)$ (even when the coupling between $\mathcal{T}_i^c(\cdot)$ and $\mathcal{G}_i^{(n)}$ is broken). The time in $\{\mathcal{T}_i^c(t) : t \geq 0\}$ will be tracked by the internal clock s_i^*. In the following construction, the last explored vertex in $J_i \cup J_i^*$ before the current vertex will be denoted by κ^*, which we will call the exploration number.

1. Fix $i \in V_n$ and initialize the internal clock $s_i^* = 0$ and the exploration number $\kappa^* = \kappa_i(1) = i$. The root node of $\{\mathcal{T}_i^c(t) : t \geq 0\}$ is born at time $\tau_{i,1} = 0$, i.e., $|\mathcal{T}_i^c(0)| = 1$, and is assigned as its mark $\mathcal{D}_{i,1} = D_i^+$.
2. If none of the nodes in $V(i)$ create a self-loop, initialize the sets $J_i = \{i\}$ and $J_i^* = \varnothing$, and move on to the next step. Else, initialize the sets $J_i^* = \{i\}$ and $J_i = \varnothing$, and go to step 4.
3. For $j = 2, \ldots, |\mathcal{G}_i^{(n)}|$ do the following:
   a. Determine $\kappa_i(j)$ on the exploration of $\mathcal{G}_i^{(n)}$.
   b. If there is a node in $V(\kappa_i(j))$ that attaches to a vertex in J_i, go to step 3(c). Otherwise, update $j = j + 1$ and go back to step 3.
   c. Set $\tau_{i,j} = s_i^* + \sigma_\omega - \sigma_{S_{\kappa^*}}$, where ω is the oldest node in $V(\kappa_i(j))$ connecting to a vertex in the set J_i. Let $\kappa_i(p)$ be the ancestor of ω in the set J_i. Update $s_i^* = \tau_{i,j}$ and add a node labelled j to $\{\mathcal{T}_i^c(t) : t \geq 0\}$ born at time $\tau_{i,j}$, connected to p, and having mark $\mathcal{D}_{i,j} = D^+_{\kappa_i(j)}$.
   d. If $V(\kappa_i(j))$ creates no self loops nor multiple edges with any of the nodes corresponding to vertices in J_i, update $J_i = J_i \cup \{\kappa_i(j)\}$. Else, update $J_i^* =$

$J_i^* \cup \{\kappa_i(j)\}$. If $\kappa_i(j)$ was added to J_i^* and there are $l \geq 2$ edges from nodes in $V(\kappa_i(j))$ to vertices in J_i, for each of the $l-1$ nodes ω' in $V(\kappa_i(j))$ with $\sigma_{\omega'} > \sigma_\omega$, create a 'dummy node' and add it to J_i^d. Attach this node (without label) to $\mathcal{T}_i^c(\cdot)$, at the node corresponding to the vertex in J_i where the associated edge coming from ω' was incident, and assign to this node the birth time $\tau_{i,\omega'} = s_i^* + \sigma_{\omega'} - \sigma_\omega$.

e. Set $j = j+1$, update $\kappa^* = \kappa_i(j)$, and go back to step 3.

4. Update the internal clock to $s_i^* = s_i^* + \sigma_{S_n} - \sigma_{S_{\kappa^*}}$.
5. To each $\kappa_i(j) \in J_i^*$, attach an independent copy of $\mathcal{T}^c(s_i^* - \tau_{i,j})$ conditioned on its root having mark $\mathcal{D}_{i,j}$ (these are the vertices where miscouplings occurred). For a dummy node ω' in J_i^d, sample an independent mark $\mathcal{D}_{\omega'}$ from the out-degree distribution H and attach an independent copy of $\mathcal{T}^c((s_i^* - \tau_{i,\omega'})^+)$ conditioned on the root having mark $\mathcal{D}_{\omega'}$.

The construction returns $\mathcal{T}_i^c(s_i^*)$. If $s_i^* < t_{n,i}$, let $\mathcal{T}_i^c(s_i^*)$ continue evolving according to the law of $\{\mathcal{T}^c(t) : t \geq 0\}$ until time $t_{n,i}$.

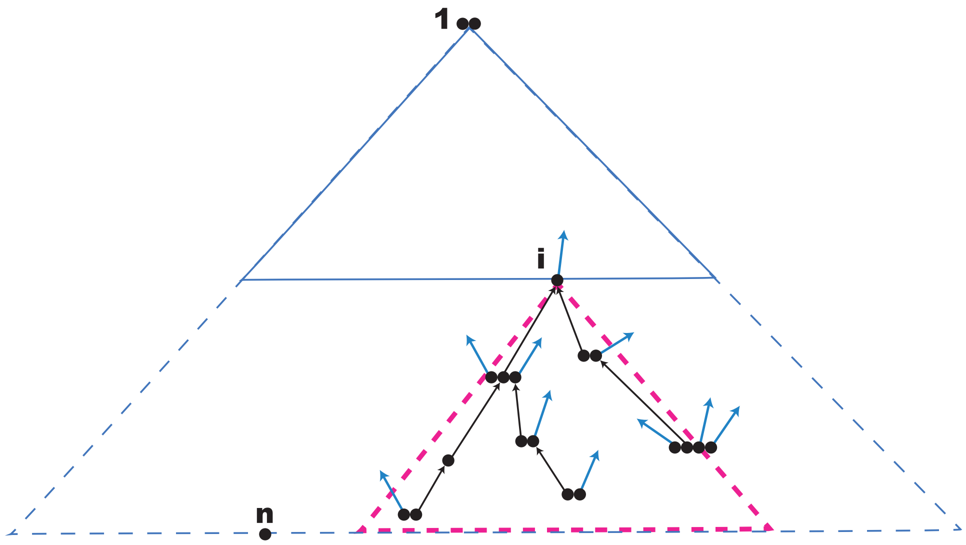

FIGURE 2. The lifted tree $\mathcal{T}(\sigma_{S_n})$ and the in-component of vertex i, $\mathcal{G}_i^{(n)}$. Nodes in each family $V(j)$, $j \geq 1$ are depicted as if they were all born at the same time and are labeled according to the vertices of $G(V_n, E_n)$ they give rise to. In this figure, $V(1) = \{1, 2\}$, $V(i) = \{S_i\}$ and $V(n) = \{S_n\}$. This depiction of $\mathcal{G}_i^{(n)}$ shows a successful coupling with its local limit.

Remark 2. *Note that in steps 3(c) and 3(d) we allow nodes in $V(\kappa_i(j))$ to have outgoing edges to vertices in J_i^* since these edges do not appear in the coupled tree $\mathcal{T}_i^c(t_{n,i})$ (descendants of nodes corresponding to vertices in J_i^* will be generated independently in step 5). The tree structure is always preserved; the unique edge connecting node j in $\mathcal{T}_i^c(t_{n,i})$ to its parent node (corresponding to a vertex in J_i) is copied from $\mathcal{G}_i^{(n)}$ in step 3(c).*

Remark 3. *Note that the coupling between $\mathcal{G}_i^{(n)}$ and $\mathcal{T}_i^c(t_{n,i})$ can break for one of the following two reasons:*

a. $|J_i^*| \geq 0$, *which means miscouplings in steps 2 or 3(d) occurred, or,*

b. $|J_i^*| = 0$ *but* $|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{G}_i^{(n)}|$, *which would happen if* $\{\mathcal{T}_i^c(t) : t \geq 0\}$ *has births in between times* s_i^* *and* $t_{n,i}$.

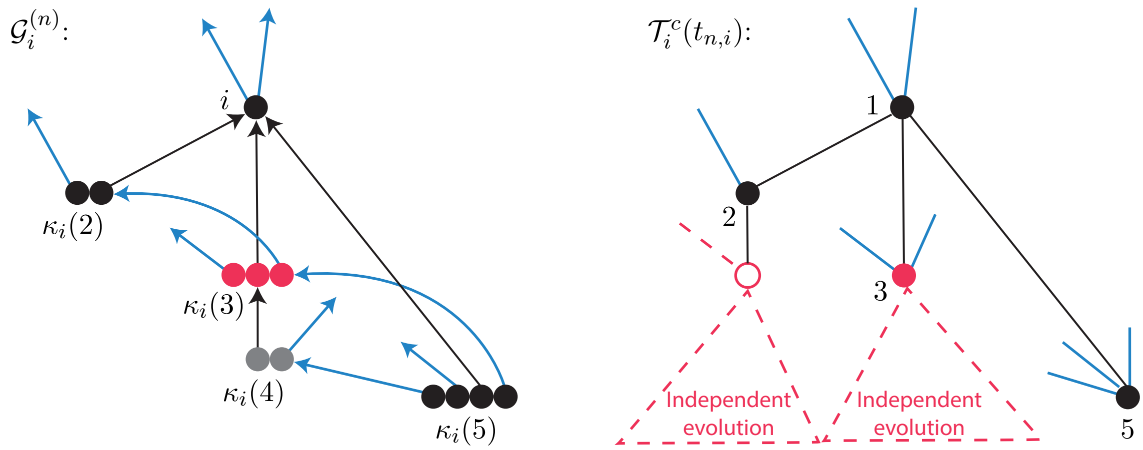

FIGURE 3. Coupling of $\mathcal{G}_i^{(n)}$ and $\mathcal{T}_i^c(t_{n,i})$. On the left we have a depiction of $\mathcal{G}_i^{(n)}$, and on the right, its coupled tree $\mathcal{T}_i^c(t_{n,i})$. Vertices $i, \kappa_i(2)$, and $\kappa_i(5)$ are in J_i, while vertex $\kappa_i(3)$ is in J_i^*. Vertex $\kappa_i(4)$ is not in $J_i \cup J_i^*$ since it was skipped in step 3(b). The miscoupling caused by vertex $\kappa_i(3)$ generated one 'dummy node' in $\mathcal{T}_i^c(t_{n,i})$ with an independently generated mark, the one depicted as an offspring of node 2. Vertex $\kappa_i(5)$ did not cause a miscoupling since it did not create a self-loop nor did it attach to more than one vertex in J_i.

5.2. **Coupling Construction for $m \geq 2$.** For some finite $m \geq 2$ and $\underline{\mathbf{i}} = \{i_1, \dots, i_m\} \subseteq [n]$, we similarly construct a coupling between $\{\mathcal{G}_{i_1}^{(n)}, \dots, \mathcal{G}_{i_m}^{(n)}\}$ and independent trees $\{\mathcal{T}_{i_1;\underline{\mathbf{i}}}^c(t_{n,i_1}), \dots, \mathcal{T}_{i_m;\underline{\mathbf{i}}}^c(t_{n,i_m})\}$, all of which evolve according to the law of $\{\mathcal{T}^c(t) : t \geq 0\}$.

As for the $m = 1$ case, sample the out-degrees $\{D_i^+ : 1 \leq i \leq n\}$ and the tree $\mathcal{T}(\sigma_{S_n})$. Without loss of generality, assume $i_1 \leq i_2 \leq \ldots \leq i_m$. Set $\{\mathcal{T}_{i_1;\underline{\mathbf{i}}}^c(t) : t \geq 0\} = \{\mathcal{T}_{i_1}^c(t) : t \geq 0\}$ following the steps for the $m = 1$ case. Further, define the sets $J_{i_1;\underline{\mathbf{i}}} = J_{i_1}$, $J_{i_1;\underline{\mathbf{i}}}^* = J_{i_1}^*$, $S_{\underline{\mathbf{i}}} = \{\kappa_{i_1}(j) : 1 \leq j \leq |G_{i_1}^{(n)}|\}$, and the internal clock $s_{i_1;\underline{\mathbf{i}}}^* = s_{i_1}^*$.
To construct $\mathcal{T}_{i_2;\underline{\mathbf{i}}}^c(t_{n,i_2})$, we again use the same construction, with the only difference being that the set $J_{i_2}^*$ is replaced by a larger set $J_{i_2;\underline{\mathbf{i}}}^*$: a vertex is put in $J_{i_2;\underline{\mathbf{i}}}^*$ if it creates a loop or multiple edges with vertices in (the current) J_{i_2} or any vertex in $S_{\underline{\mathbf{i}}}$. Vertices in $J_{i_2;\underline{\mathbf{i}}}^*$ and the dummy nodes in $J_{i_2;\underline{\mathbf{i}}}^d$ undergo independent subsequent evolution as described in Step 5 of the coupling construction. We also require that for vertices in $J_{i_2}^* \subseteq J_{i_2;\underline{\mathbf{i}}}^*$, we use the same independent copies in the construction of Step 5. Return the internal clock time $s_{i_2;\underline{\mathbf{i}}}$ and the tree $\mathcal{T}_{i_2;\underline{\mathbf{i}}}^c(s_{i_2;\underline{\mathbf{i}}})$, and update the set $S_{\underline{\mathbf{i}}} = S_{\underline{\mathbf{i}}} \cup \{\kappa_{i_2}(j) : 1 \leq j \leq |G_{i_2}^{(n)}|\}$.
Iterate the above process successively for $i_3, \dots, i_m$ to construct m trees

$$\{\mathcal{T}_{i_1;\underline{\mathbf{i}}}^c(s_{i_1;\underline{\mathbf{i}}}),\ \dots, \mathcal{T}_{i_m;\underline{\mathbf{i}}}^c(s_{i_m;\ \underline{\mathbf{i}}})\},$$

which are all independent of each other and their evolution has the same law as $\{\mathcal{T}^c(t) : t \geq 0\}$.

Remark 4. *Note that the coupling between* $(\mathcal{G}_{i_1}^{(n)}, \dots, \mathcal{G}_{i_m}^{(n)})$ *and* $(\mathcal{T}_{i_1;\underline{\mathbf{i}}}^c(t_{n,i_1}), \dots, \mathcal{T}_{i_m;\underline{\mathbf{i}}}^c(t_{n,i_m}))$ *can break for one of the following two reasons:*

a. $|J_{i_\ell;\underline{\mathbf{i}}}^*| > 0$ *for some* $1 \leq \ell \leq m$, *which means miscouplings in steps 2 or 3(d) occurred, or,* *b.* $|J^*_{i_\ell;\underline{\mathbf{i}}}| = 0$ *for all* $1 \leq \ell \leq m$ *but* $|\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t_{n,i_\ell})| \neq |\mathcal{G}^{(n)}_{i_\ell}|$ *for some* ℓ*, which would happen if* $\{\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t) : t \geq 0\}$ *has births in between times* $s^*_{i_\ell;\underline{\mathbf{i}}}$ *and* $t_{n,i}$*.*

5.3. **Proof of Theorem 1.** The proof of Theorem 1 is split into several lemmas. Although the description of the coupling is given by first sampling $\{D^+_i : 1 \leq i \leq n\}$ and then constructing the tree $\{\mathcal{T}(t) : t \geq 0\}$ until time σ_{S_n}, note that the two can be done simultaneously. This will be the approach followed through most of the proofs in this section, under the assumption that D^+_i is sampled from distribution H at time $\sigma_{S_{i-1}}$.

To ease the reading of this section we have compiled the notation that is used repeatedly:

- $\mathscr{G}_t$: denotes the filtration generated by the construction of $\{\mathcal{T}(t) : t \geq 0\}$ along with the sequence $\{D^+_i : i \geq 1\}$ up to time t.
- σ_k: denotes the birth time of the kth node in $\{\mathcal{T}(t) : t \geq 0\}$.
- $S_k = D^+_1 + \cdots + D^+_k$.
- $V(i) = \{S_{i-1}+1, S_{i-1}+2, \ldots, S_i\}$.
- $\{\mathcal{T}^c(t) : t \geq 0\}$: generic version of the discrete skeleton of a CTBP driven by $\{(\mathcal{D}_k, \bar{\xi}^{(k)}_f) : k \geq 1\}$.
- $\Lambda(t) = |\mathcal{T}^c(t)| + \sum_{j=1}^{|\mathcal{T}^c(t)|} \mathcal{D}_j$: number of nodes plus the sum of the marks in $\mathcal{T}^c(t)$.
- $\mathcal{G}^{(n)}_i$: subgraph of $G(V_n, E_n)$ obtained from exploring the in-component of vertex i, with the out-degrees of its vertices as marks.
- $\kappa_i(j)$ is the label in $G(V_n, E_n)$ of the jth oldest vertex in $\mathcal{G}^{(n)}_i$.
- $\{\mathcal{T}^c_i(t) : t \geq 0\}$: discrete skeleton of the tree rooted at vertex $i \in G(V_n, E_n)$ at time t, as constructed in the coupling from Section 5.1.
- $\{\mathcal{D}_{i,j} : j \geq 1\}$: marks of the nodes in $\{\mathcal{T}^c_i(t) : t \geq 0\}$.
- $D^-_v(t)$ is the in-degree of node v in $\mathcal{T}(t)$.
- $\mathscr{F}^{(i)}_t$ sigma-algebra generated by $\{\mathcal{T}^c_i(s), \{\mathcal{D}_{i,j} : j \in \mathcal{T}^c_i(s)\} : 0 \leq s \leq t\}$.
- $\Lambda_i(t) = |\mathcal{T}^c_i(t)| + \sum_{j=1}^{|\mathcal{T}^c_i(t)|} \mathcal{D}_{i,j}$: number of nodes plus the sum of the marks in $\mathcal{T}^c_i(t)$.
- s^*_i: time at which the construction of $\{\mathcal{T}^c_i(t) : t \geq 0\}$ ends in Section 5.1.
- J_i: set of vertices of $\mathcal{G}^{(n)}_i$ that were successfully coupled to nodes in $\mathcal{T}^c_i(s^*_i)$.
- J^*_i: set of vertices of $\mathcal{G}^{(n)}_i$ that caused miscouplings with $\mathcal{T}^c_i(s^*_i)$.
- $s_{i_\ell,\underline{\mathbf{i}}}, J_{i_\ell;\underline{\mathbf{i}}}; 1 \leq \ell \leq m$: corresponding objects in the construction in Section 5.2.

Lemma 2. *For any* $t > 0$*, we have*

$$E\left[|\mathcal{T}^c(t)| + \sum_{j \in \mathcal{T}^c(t)} \mathcal{D}_j\right] \leq 1 + \mu e^{C_f(\mu+1)t}.$$

Proof. Let $\{\xi_l(t) : t \geq 0\}$ be a Markovian pure birth process satisfying $\xi_l(0) = 0$ and having birth rates

$$P(\xi_l(t+dt) = k+1 | \xi_l(t) = k) = C_f(k+1)dt + o(dt), \ k \geq 0.$$

Let $\{\mathcal{D}_i : i \geq 1\}$ be an i.i.d. sequence distributed according to H, and define

$$\bar{\xi}^{(k)}_l = \sum_{i=1}^{\mathcal{D}_k} \xi^{k,i}_l,$$

where the $\{\xi^{k,i}_l : i \geq 1, k \geq 1\}$ are i.i.d. copies of ξ_l. Let $\hat{\mathcal{T}}^c(t)$ be the discrete skeleton of a marked CTBP driven by $\{\bar{\xi}^{(k)}_l : k \geq 1\}$ at time $t \geq 0$ conditionally on the root being born at time $t = 0$.

Then, by Assumption 1 we have that

$$|\mathcal{T}^c(t)| + \sum_{j \in \mathcal{T}^c(t)} \mathcal{D}_j \leq_{\text{s.t.}} |\hat{\mathcal{T}}^c(t)| + \sum_{j \in \hat{\mathcal{T}}^c(t)} \mathcal{D}_j.$$

Let $X(t) := |\hat{\mathcal{T}}^c(t)| + \sum_{j\in\hat{\mathcal{T}}^c(t)} \mathcal{D}_j, t \geq 0$. Note that, if $N_v(t)$ denotes the number of offspring of node v in $\hat{\mathcal{T}}^c(t)$ and $h(m) = P(\mathcal{D}_1 = m)$, then for $t \geq 0$ and small $\Delta > 0$,

$$\begin{aligned} P(X(t+\Delta) = X(t)+m+1|X(t)) &= C_f\left(\sum_{v\in\hat{\mathcal{T}}^c(t)} (N_v(t)+\mathcal{D}_v)\right) h(m)\Delta + h(m)o(\Delta) \\ &= C_f(X(t)-1)h(m)\Delta + h(m)o(\Delta). \end{aligned}$$

Hence, recalling $\mu = E[\mathcal{D}_1]$,

$$\begin{aligned} E[X(t+\Delta)] - E[X(t)] &= C_f\Delta\, E[X(t)-1]\sum_{m=1}^{\infty}(m+1)h(m) + o(\Delta) \\ &= C_f\Delta\, E[X(t)-1](\mu+1) + o(\Delta). \end{aligned}$$

From this, writing $M(t) := E[X(t)]$, we obtain the differential equation

$$M'(t) = C_f(\mu+1)(M(t)-1),\ t \geq 0.$$

Solving this equation gives

$$M(t) = 1 + (M(0)-1)e^{C_f(\mu+1)t},\ t \geq 0.$$

The lemma follows upon noting $M(0) = E[X(0)] = 1+\mu$. □

Lemma 3. *We have that*

$$\sup_{j,k\geq m}\left|\sigma_k - \sigma_j - \frac{1}{\lambda}\log(k/j)\right| \to 0, \qquad P\text{-a.s. as } m\to\infty.$$

Proof. Note that under Assumption 1, using [15, Theorem 6.3], there exists an almost surely positive random variable Θ such that:

$$me^{-\lambda\sigma_m} \to \Theta, \qquad P\text{-a.s. as } m\to\infty.$$

Equivalently, we have that:

$$-\sigma_m + \frac{1}{\lambda}\log m \to \frac{1}{\lambda}\log\Theta, \qquad P\text{-a.s. as } m\to\infty.$$

From here it follows that

$$\sup_{j,k\geq m}\left|\sigma_k - \sigma_j - \frac{1}{\lambda}\log(k/j)\right| \leq 2\sup_{k\geq m}\left|-\sigma_k + \frac{1}{\lambda}\log k - \frac{1}{\lambda}\log\Theta\right| \to 0 \qquad P\text{-a.s.}$$

as $m\to\infty$. □

Lemma 4. *We have*

i)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} P\left(|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{T}_i^c(\sigma_{S_n} - \sigma_{S_i})|\right) = 0,$$

ii)

$$\lim_{n\to\infty}\frac{1}{n^m}\sum_{\underline{\mathbf{i}}\subseteq[n]:|\underline{\mathbf{i}}|=m}\sum_{\ell=1}^{m} P\left(\left\{|\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t_{n,i_\ell})| \neq |\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(\sigma_{S_n} - \sigma_{S_\ell})|\right\}\right) = 0.$$

Proof. Fix $\epsilon > 0$ and note that

$$\begin{aligned} &P\left(|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{T}_i^c(\sigma_{S_n} - \sigma_{S_i})|\right) \\ &\leq P\left(|\mathcal{T}_i^c(t_{n,i}-\epsilon)| < |\mathcal{T}_i^c(t_{n,i}+\epsilon)|\right) + P\left(|\sigma_{S_n} - \sigma_{S_i} - t_{n,i}| > \epsilon\right). \end{aligned} \tag{2}$$

To bound the first probability let $\mathscr{F}_t^{(i)} = \sigma(\mathcal{T}_i^c(s), \{\mathcal{D}_{i,j} : j \in \mathcal{T}_i^c(s)\} : 0 \leq s \leq t)$ and $\Lambda_i(t) = |\mathcal{T}_i^c(t)| + \sum_{j=1}^{|\mathcal{T}_i^c(t)|}\mathcal{D}_{i,j}$. Next, note that conditionally on $\mathscr{F}^{(i)}_{t_{n,i}-\epsilon}$, the next birth in

$\{\mathcal{T}_i^c(t) : t \geq 0\}$ will happen in an exponential time that has a rate that, by Assumption 1, is bounded from above by $C_f \Lambda_i(t_{n,i} - \epsilon)$. Therefore,

$$\begin{aligned}
&P\left(|\mathcal{T}_i^c(t_{n,i} - \epsilon)| < |\mathcal{T}_i^c(t_{n,i} + \epsilon)|\right) \\
&\leq E\left[P\left(|\mathcal{T}_i^c(t_{n,i} - \epsilon)| < |\mathcal{T}_i^c(t_{n,i} + \epsilon)| \,\middle|\, \mathscr{F}^{(i)}_{t_{n,i}-\epsilon}\right)\right] \\
&\leq E\left[P\left(\mathrm{Exp}(C_f \Lambda_i(t_{n,i} - \epsilon)) \leq 2\epsilon\right) \middle| \Lambda_i(t_{n,i} - \epsilon)\right)\right] \\
&\leq E\left[1 - e^{-C_f 2\epsilon \Lambda_i(t_{n,i})}\right]. \qquad (3)
\end{aligned}$$

Now let $I_n = \lceil nU \rceil$, where U is a uniform $[0,1]$ independent of everything else, and note that by Lemma 3 and the strong law of large numbers,

$$\begin{aligned}
\lim_{n\to\infty} \left|\sigma_{S_n} - \sigma_{S_{I_n}} - t_{n,I_n}\right| &\leq \lim_{n\to\infty} \left|\sigma_{S_n} - \sigma_{S_{I_n}} - \frac{1}{\lambda}\log(S_n/S_{I_n})\right| \\
&\quad + \lim_{n\to\infty} \frac{1}{\lambda}\left|\log\left(S_n I_n/(n S_{I_n})\right)\right| = 0. \qquad P\text{-a.s.}
\end{aligned}$$

Finally, letting $\chi = -(1/\lambda)\log U$, note that $t_{n,I_n} \leq \chi$, and use (2) to obtain that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} P\left(|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{T}_i^c(\sigma_{S_n} - \sigma_{S_i})|\right) \\
&\leq E\left[1 - e^{-C_f 2\epsilon \Lambda_{I_n}(t_{n,I_n})}\right] + P\left(\left|\sigma_{S_n} - \sigma_{S_{I_n}} - t_{n,I_n}\right| > \epsilon\right) \\
&\leq E\left[1 - e^{-C_f 2\epsilon \Lambda(\chi)}\right] + P\left(\left|\sigma_{S_n} - \sigma_{S_{I_n}} - t_{n,I_n}\right| > \epsilon\right) \\
&\to E\left[1 - e^{-C_f 2\epsilon \Lambda(\chi)}\right],
\end{aligned}$$

as $n \to \infty$. Now take $\epsilon \downarrow 0$ to complete the proof.
Part (ii) follows exactly as (i) upon using the union bound and the fact that, for any $t > 0$, $\underline{\mathbf{i}} \subseteq [n]$ with $|\underline{\mathbf{i}}| = m$ and $\ell \in \{1, \ldots, m\}$, $\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t)$ has the same law as $\mathcal{T}^c_{i_\ell}(t)$. □

Recall that s_i^* denotes the internal clock in the coupling construction in Section 5.1.

Lemma 5. *For any $i \in V_n$ set $a_i = i^{1/2-\delta}$ for some $0 < \delta < 1/2$, and define $\vartheta = C_f(\mu+1)/\lambda$. Let $\Lambda_i(t) = |\mathcal{T}_i^c(t)| + \sum_{j=1}^{|\mathcal{T}_i^c(t)|} \mathcal{D}_{i,j}$ and define the event*

$$E_{n,i} = \left\{\Lambda_i(t_{n,i} \vee s_i^*) \leq a_i\right\}.$$

Then, for any constant $c_\vartheta \in (\vartheta/(\vartheta + 1/2 - \delta), 1)$,

$$P(E^c_{n,i}) \leq (\mu+1) n^{-c_\vartheta(1/2+\vartheta-\delta)+\vartheta} + 1(i < n^{c_\vartheta}) + P\left(|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{T}_i^c(\sigma_{S_n} - \sigma_{S_i})|\right).$$

Proof. Note that since $s_i^* \leq \sigma_{S_n} - \sigma_{S_i}$, then for any $\epsilon > 0$ we have

$$\begin{aligned}
P(E^c_{n,i}) &\leq P\left(|\mathcal{T}_i^c(t_{n,i})| + \sum_{j=1}^{|\mathcal{T}_i^c(t_{n,i})|} \mathcal{D}_{i,j} > a_i\right) + P\left(|\mathcal{T}_i^c(t_{n,i})| < |\mathcal{T}_i^c(\sigma_{S_n} - \sigma_{S_i})|\right) \\
&\leq \frac{E[\Lambda_i(t_{n,i})]}{a_i} + P\left(|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{T}_i^c(\sigma_{S_n} - \sigma_{S_i})|\right).
\end{aligned}$$

Now use Lemma 2 to obtain that for $\vartheta = C_f(\mu+1)/\lambda$, and any $1 > c_\vartheta > \vartheta/(1/2+\vartheta-\delta)$,

$$\begin{aligned}\frac{E[\Lambda_i(t_{n,i})]}{a_i} &\leq \frac{1+\mu e^{C_f(\mu+1)t_{n,i}}}{a_i} 1(i \geq n^{c_\vartheta}) + 1(i < n^{c_\vartheta}) \\ &\leq \frac{1+\mu(n/i)^\vartheta}{i^{1/2-\delta}} 1(i \geq n^{c_\vartheta}) + 1(i < n^{c_\vartheta}) \\ &\leq (\mu+1)\frac{n^\vartheta}{i^{1/2+\vartheta-\delta}} 1(i \geq n^{c_\vartheta}) + 1(i < n^{c_\vartheta}) \\ &\leq (\mu+1) n^{-c_\vartheta(1/2+\vartheta-\delta)+\vartheta} + 1(i < n^{c_\vartheta}).\end{aligned}$$

□

Remark 5. *The above lemma readily extends to* $E_{n,i_\ell;\underline{\mathbf{i}}} = \{\Lambda_{i_\ell;\underline{\mathbf{i}}}(t_{n,i} \vee s^*_{i_\ell;\underline{\mathbf{i}}}) \leq a_{i_\ell}\}$*, where* $\Lambda_{i_\ell;\underline{\mathbf{i}}} = |\mathcal{T}_{i_\ell;\underline{\mathbf{i}}}(t)| + \sum_{j=1}^{|\mathcal{T}_{i_\ell;\underline{\mathbf{i}}}(t)|} \mathcal{D}_{(i_\ell;\underline{\mathbf{i}}),j}$ *when using the construction for* $m \geq 2$ *in Section 5.2. We obtain the same bound (replacing* i *by* i_ℓ*) for* $P(E^c_{n,i_\ell;\underline{\mathbf{i}}})$.

Lemma 6. *i) For any* $i \in V_n$ *set* $a_i = i^{1/2-\delta}$ *for some* $0 < \delta < 1/2$*. Define the event* $E_{n,i}$ *as in Lemma 5. Then,*

$$P\left(E_{n,i} \cap \{|J^*_i| \geq 1\}\right) \leq 1(i=1) + \frac{2C_f \mu i^{-2\delta}}{f_*}.$$

ii) For any $\underline{\mathbf{i}} = (i_1, \ldots, i_m) \subseteq [n]$*, define the events* $E_{n,i_\ell;\underline{\mathbf{i}}}$ *as in Remark 5 with* $a_{i_\ell} = i_\ell^{1/2-\delta}$ *for some* $0 < \delta < 1/2$*. Then*

$$P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}} \cap \bigcup_{\ell=1}^m \{|J^*_{i_\ell;\underline{\mathbf{i}}}| \geq 1\}\right) \leq 1(\min_{1\leq \ell\leq m} i_\ell = 1) + \left(1 \wedge \frac{2C_f m^2 \mu (\max_{1\leq\ell\leq m} i_\ell)^{1-2\delta}}{f_*(\min_{1\leq\ell\leq m} i_\ell)}\right).$$

Proof. i) Define the events

$$H_{i,j} := \{\kappa_i(j) \text{ is the oldest vertex in } J^*_i\},$$

with $H_{i,j} = \emptyset$ if $j > |\mathcal{G}_i^{(n)}|$. Note that

$$P\left(E_{n,i} \cap \{|J^*_i| \geq 1\}\right) = P\left(E_{n,i} \cap \bigcup_{j=1}^{a_i} H_{i,j}\right) = \sum_{j=1}^{a_i} P\left(E_{n,i} \cap H_{i,j}\right).$$

To analyze the last probabilities focus on the construction of the tree $\{\mathcal{T}^c_i(t) : t \geq 0\}$, and note that for $H_{i,j}$ to happen it must be that one of the $D^+_{\kappa_i(j)} = \mathcal{D}_{i,j}$ nodes in $V(\kappa_i(j))$ will attach to one of the younger nodes in $V(\kappa_i(j))$ or to one of the nodes in $\bigcup_{r=1}^{j-1} V(\kappa_i(r))$, and up until this point there have been no miscouplings. For time $t \geq 0$, denote by $s^*_i(t)$ the internal time in the tree process $\mathcal{T}^c_i(\cdot)$ accrued after all vertices in $\mathcal{G}_i^{(n)}$ with at least one node born by time t have been explored. Define the event

$$E_i(t) = \left\{ |\mathcal{T}^c_i(s^*_i(t))| + \sum_{j=1}^{|\mathcal{T}^c_i(s^*_i(t))|} \mathcal{D}_{i,j} \leq a_i \right\},$$

which satisfies $E_{n,i} \subseteq E_i(t)$ for all $0 \leq t \leq t_{n,i} \vee s^*_i$. Next, let $S_{\kappa_i(j)-1} \leq \omega \leq S_{\kappa_i(j)}$ be the oldest node in $V(\kappa_i(j))$ that is born to $\bigcup_{r=1}^{j-1} V(\kappa_i(r))$, and note that conditionally on $\mathscr{G}_{\sigma_\omega}$ the event $H_{i,j}$ will happen if any of the nodes $v \in V(\kappa_i(j)), v > \omega$, is such that v is born to one of the nodes in $U_v := \bigcup_{r=1}^{j-1} V(\kappa_i(r)) \cup \{S_{\kappa_i(j)-1}, \ldots, v-1\}$. Let $\mathcal{H}_{i,j,v}$ be the event that node $v > \omega$ is the first node such that v is born to a node in U_v. Then, recalling that $D_u^-(t)$ is the in-degree of node u in $\mathcal{T}(t)$, using Assumption 1,

$$\begin{aligned}
1(E_i(\sigma_\omega))P(H_{i,j}|\mathscr{G}_{\sigma_\omega}) &\le \sum_{v=\omega+1}^{S_{\kappa_i(j)}} 1(E_i(\sigma_\omega))P(\mathcal{H}_{i,j,v}|\mathscr{G}_{\sigma_\omega}) \\
&= \sum_{v=\omega+1}^{S_{\kappa_i(j)}} E\left[1(E_i(\sigma_{v-1}))P(\mathcal{H}_{i,j,v}|\mathscr{G}_{\sigma_{v-1}})\middle|\mathscr{G}_{\sigma_\omega}\right] \\
&= \sum_{v=\omega+1}^{S_{\kappa_i(j)}} E\left[1(E_i(\sigma_{v-1}))\cdot\frac{\sum_{u\in U_v} f(D_u^-(\sigma_{v-1})+1)}{\sum_{u\in\mathcal{T}(\sigma_{v-1})} f(D_u^-(\sigma_{v-1})+1)}\middle|\mathscr{G}_{\sigma_\omega}\right] \\
&\le \sum_{v=\omega+1}^{S_{\kappa_i(j)}} E\left[1(E_i(\sigma_{v-1}))\cdot\frac{\sum_{u\in U_v} C_f(D_u^-(\sigma_{v-1})+1)}{\sum_{u\in\mathcal{T}(\sigma_{v-1})} f_*(v-1)}\middle|\mathscr{G}_{\sigma_\omega}\right] \\
&\le \frac{C_f}{f_*\omega}\sum_{v=\omega+1}^{S_{\kappa_i(j)}} E\left[1(E_i(\sigma_{v-1}))\left(|\mathcal{T}_i^c(\sigma_{v-1})| + \sum_{k=1}^{|\mathcal{T}_i^c(\sigma_{v-1})|}\mathcal{D}_{i,k}\right)\middle|\mathscr{G}_{\sigma_\omega}\right] \\
&\le \frac{C_f a_i \mathcal{D}_{i,j}}{f_* S_{\kappa_i(j)-1}}.
\end{aligned}$$

It follows that

$$P(E_{n,i}\cap H_{i,j}) \le \frac{C_f a_i}{f_*}E\left[\frac{\mathcal{D}_{i,j}}{S_{\kappa_i(j)-1}}\right],$$

and since $S_i \ge i$ for all $i \ge 1$,

$$\begin{aligned}
P(E_{n,i}\cap\{|J_i^*|\ge 1\}) &\le 1\wedge\left(\frac{C_f a_i}{f_*}\sum_{j=1}^{a_i} E\left[\frac{\mathcal{D}_{i,j}}{S_{\kappa_i(j)-1}}\right]\right) \le 1\wedge\left(\frac{C_f\mu a_i^2}{f_*(i-1)}\right) \\
&\le 1(i=1) + \frac{2C_f\mu i^{-2\delta}}{f_*}.
\end{aligned}$$

To prove (ii), for any $\underline{\mathbf{i}} = (i_1,\dots,i_m) \subseteq [n]$, define events

$$H_{(i_\ell;\underline{\mathbf{i}}),j} := \{\kappa_{i_\ell}(j) \text{ is the oldest vertex in } J^*_{i_\ell;\underline{\mathbf{i}}}\},$$

with $H_{(i_\ell;\underline{\mathbf{i}}),j} = \emptyset$ if $j > |G_{i_\ell}^{(n)}|$. Then we can write

$$\begin{aligned}
&P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}}\cap\bigcup_{\ell=1}^m\{|J^*_{i_\ell;\underline{\mathbf{i}}}|\ge 1\}\right) \\
&= P\left(\bigcap_{u=1}^m E_{n,i_u;\underline{\mathbf{i}}}\cap\left(\bigcup_{\ell=1}^m\left(\bigcap_{k=1}^{\ell-1}\{|J^*_{i_k;\underline{\mathbf{i}}}|=0\}\right)\cap\left(\bigcup_{j=1}^{a_{i_\ell}} H_{(i_\ell;\underline{\mathbf{i}}),j}\right)\right)\right) \\
&\le \sum_{\ell=1}^m\sum_{j=1}^{a_{i_\ell}} P\left(\bigcap_{u=1}^m E_{n,i_u;\underline{\mathbf{i}}}\cap\left(\bigcap_{k=1}^{\ell-1}\{|J^*_{i_k;\underline{\mathbf{i}}}|=0\}\right)\cap H_{(i_\ell;\underline{\mathbf{i}}),j}\right),
\end{aligned}$$

where we use the convention that $\cap_{k=1}^0$ is the null set. Note that for $H_{(i_\ell;\underline{\mathbf{i}}),j}$ to happen, one of the $D^+_{\kappa_{i_\ell}(j)} = \mathcal{D}_{i_\ell,j}$ nodes in $V(\kappa_{i_\ell}(j))$ will attach to one of the younger nodes in $V(\kappa_{i_\ell}(j))$ or to one of the nodes in $\bigcup_{k=1}^{\ell-1}\bigcup_{r=1}^{a_{i_k}} V(\kappa_{i_k}(r))\cup\bigcup_{r=1}^{j-1} V(\kappa_{i_\ell}(r))$. Then, using the same arguments as in proof of part (i), we get that

$$P\left(\bigcap_{u=1}^{m} E_{n,i_u;\underline{\mathbf{i}}} \cap \left(\bigcap_{k=1}^{\ell-1}\{|J^*_{i_k;\underline{\mathbf{i}}}|=0\}\right)\cap H_{(i_\ell;\underline{\mathbf{i}}),j}\right) \le \frac{C_f(\sum_{k=1}^{\ell} a_{i_k})}{f_*} E\left[\frac{\mathcal{D}_{i_\ell,j}}{S_{\kappa_\ell(j)-1}}\right].$$

Therefore,

$$\begin{aligned} P\left(\bigcap_{\ell=1}^{m} E_{n,i_\ell;\underline{\mathbf{i}}} \cap \bigcup_{\ell=1}^{m}\{|J^*_{i_\ell;\underline{\mathbf{i}}}|\ge 1\}\right) &\le 1\wedge\left(\sum_{l=1}^{m}\frac{C_f(\sum_{k=1}^{\ell} a_{i_k})}{f_*}\sum_{j=1}^{a_{i_\ell}} E\left[\frac{\mathcal{D}_{i_\ell,j}}{S_{\kappa_{i_\ell}(j)-1}}\right]\right) \\ &\le 1\wedge\left(\frac{C_f m^2\mu(\max_{1\le\ell\le m} a_{i_\ell})^2}{f_*((\min_{1\le\ell\le m} i_\ell)-1)}\right) \\ &\le 1\left(\min_{1\le\ell\le m} i_\ell = 1\right) + \left(1\wedge\frac{2C_f m^2\mu(\max_{1\le\ell\le m} i_\ell)^{1-2\delta}}{f_*(\min_{1\le\ell\le m} i_\ell)}\right). \end{aligned}$$

□

Remark 6. *Let $U_1,\dots,U_m$ be independent random variables, uniformly distributed in $[0,1]$, and set $I_{n,\ell}=\lceil nU_\ell\rceil$. Then it is routine to check that*

$$1\wedge\frac{2C_f m^2\mu(\max_{1\le\ell\le m} I_{n,\ell})^{1-2\delta}}{f_*(\min_{1\le\ell\le m} I_{n,\ell})}\xrightarrow{P} 0$$

as $n\to\infty$. This fact, in conjunction with part (ii) of the above lemma, will be used in the proof of Theorem 1(ii).

Lemma 7. *Let $E_{n,i}$ and $E_{n,i_\ell;\underline{\mathbf{i}}}$ be the events defined in Lemma 5 and Remark 5. Then,*

i)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} P\left(E_{n,i}\cap\{|J^*_i|=0\}\cap\left\{|\mathcal{T}^c_i(t_{n,i})|\ne|\mathcal{G}^{(n)}_i|\right\}\right)=0.$$

ii)

$$\lim_{n\to\infty}\frac{1}{n^m}\sum_{\underline{\mathbf{i}}\subseteq[n]:|\underline{\mathbf{i}}|=m} P\left(\bigcap_{\ell=1}^{m} E_{n,i_\ell;\underline{\mathbf{i}}}\cap\bigcap_{\ell=1}^{m}\{|J^*_{i_\ell;\underline{\mathbf{i}}}|=0\}\cap\bigcup_{\ell=1}^{m}\{|\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t_{n,i_\ell})|\ne|\mathcal{G}^{(n)}_{i_\ell}|\}\right)=0.$$

Proof. Note that on the event $\{|J^*_i|=0\}$ we have that

$$\sigma_{S_n}-\sigma_{S_i}-\sum_{j=2}^{|J_i|}(\sigma_{S_{\kappa_i(j)}}-\sigma_{S_{\kappa_i(j)-1}})\le s^*_i\le\sigma_{S_n}-\sigma_{S_i}.$$

Moreover, since on the event $\{|J^*_i|=0\}$ we have that $\mathcal{T}^c_i(s^*_i)\simeq\mathcal{G}^{(n)}_i$, in order for $|\mathcal{T}^c_i(t_{n,i})|\ne|\mathcal{G}^{(n)}_i|$ to happen it must be that either $s^*_i<t_{n,i}$ and $\{\mathcal{T}^c_i(t):t\ge 0\}$ had births in $(s^*_i,t_{n,i})$, or $s^*_i>t_{n,i}$ and $\{\mathcal{T}^c_i(t):t\ge 0\}$ had births in $(t_{n,i},s^*_i)$.

Fix $\epsilon>0$, and note that

$$\begin{aligned} &P\left(E_{n,i}\cap\{|J^*_i|=0\}\cap\left\{|\mathcal{T}^c_i(t_{n,i})|\ne|\mathcal{G}^{(n)}_i|\right\}\right) \\ &=P\left(E_{n,i}\cap\{|J^*_i|=0\}\cap\{|\mathcal{T}^c_i(s^*_i)|<|\mathcal{T}^c_i(t_{n,i})|\}\right) \\ &\quad+P\left(E_{n,i}\cap\{|J^*_i|=0\}\cap\{|\mathcal{T}^c_i(t_{n,i})|<|\mathcal{T}^c_i(s^*_i)|\}\right) \\ &\le P\left(E_{n,i}\cap\{|J^*_i|=0\}\cap\{|\mathcal{T}^c_i(t_{n,i}-\epsilon)|<|\mathcal{T}^c_i(t_{n,i})|\}\right) \qquad (4)\\ &\quad+P\left(E_{n,i}\cap\{|J^*_i|=0\}\cap\{s^*_i<t_{n,i}-\epsilon\}\right) \qquad (5)\\ &\quad+P\left(|\mathcal{T}^c_i(t_{n,i})|<|\mathcal{T}^c_i(\sigma_{S_n}-\sigma_{S_i})|\right). \end{aligned}$$

Using the same steps leading to (3), we obtain that (4) is bounded by

$$P\left(|\mathcal{T}_i^c(t_{n,i}-\epsilon)| < |\mathcal{T}_i^c(t_{n,i})|\right) \leq E\left[1 - e^{-C_f\epsilon\Lambda_i(t_{n,i})}\right].$$

To analyze the probability in (5), note that by Assumption 1 we have that on the event $\{|J_i^*| = 0, s_i^* < t_{n,i}\}$,

$$\sum_{j=2}^{|J_i|}(\sigma_{S_{\kappa_i(j)}} - \sigma_{S_{\kappa_i(j)-1}}) \leq_{\text{s.t.}} \text{Erlang}\left(\sum_{j=1}^{|\mathcal{T}_i^c(t_{n,i})|}\mathcal{D}_{i,j},\, S_i\right) =: \mathcal{E}_i.$$

It follows that

$$\begin{aligned}
&P\left(E_{n,i}\cap\{|J_i^*|=0\}\cap\{s_i^* < t_{n,i}-\epsilon\}\right)\\
&\leq P\left(E_{n,i}\cap\{|J_i^*|=0\}\cap\{t_{n,i}-\epsilon > s_i^* \geq \sigma_{S_n}-\sigma_{S_i}-\epsilon/2\}\right)\\
&\quad + P\left(E_{n,i}\cap\{|J_i^*|=0\}\cap\left\{s_i^* < t_{n,i},\, \sum_{j=2}^{|J_i|}(\sigma_{S_{\kappa_i(j)}} - \sigma_{S_{\kappa_i(j)-1}}) > \epsilon/2\right\}\right)\\
&\leq P\left(t_{n,i}-\sigma_{S_n}+\sigma_{S_i} > \epsilon/2\right) + P\left(E_{n,i}\cap\{\mathcal{E}_i > \epsilon/2\}\right).
\end{aligned}$$

By Markov's inequality we have that

$$P\left(E_{n,i}\cap\{\mathcal{E}_i > \epsilon/2\}\right) \leq P(\text{Erlang}(a_i, S_i) > \epsilon/2) \leq 1\wedge E\left[\frac{2a_i}{\epsilon S_i}\right] \leq 1\wedge\frac{2a_i}{\epsilon i}.$$

Finally, let $I_n = \lceil nU\rceil$, where U is a uniform $[0,1]$ independent of everything else. Then, we conclude that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n P\left(E_{n,i}\cap\{|J_i^*|=0\}\cap\left\{|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{G}_i^{(n)}|\right\}\right)\\
&\leq E\left[1-e^{-C_f\epsilon\Lambda_{I_n}(t_{n,I_n})}\right] + E\left[\frac{2a_{I_n}}{\epsilon I_n}\wedge 1\right] + P\left(t_{n,I_n}-\sigma_{S_n}+\sigma_{S_{I_n}} > \epsilon/2\right)\\
&\quad + \frac{1}{n}\sum_{i=1}^n P\left(|\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{T}_i^c(\sigma_{S_n}-\sigma_{S_i})|\right)\\
&\to E\left[1-e^{-C_f\epsilon\Lambda(\chi)}\right], \qquad n\to\infty,
\end{aligned}$$

by Lemma 4(i) and the same arguments used in its proof. Taking $\epsilon\downarrow 0$ completes the proof of (i).

To prove (ii), for any $\underline{\mathbf{i}} = (i_1,\ldots,i_m)\subseteq[n]$, note that

$$\begin{aligned}
&P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}}\cap\bigcap_{\ell=1}^m\{|J_{i_\ell;\underline{\mathbf{i}}}^*|=0\}\cap\bigcup_{\ell=1}^m\{|\mathcal{T}_{i_\ell;\underline{\mathbf{i}}}^c(t_{n,i_\ell})| \neq |\mathcal{G}_{i_\ell}^{(n)}|\}\right)\\
&\leq\sum_{\ell=1}^m P\left(\bigcap_{j=1}^{\ell} E_{n,i_j;\underline{\mathbf{i}}}\cap\bigcap_{j=1}^{\ell}\{|J_{i_j;\underline{\mathbf{i}}}^*|=0\}\cap\{|\mathcal{T}_{i_\ell;\underline{\mathbf{i}}}^c(t_{n,i_\ell})| \neq |\mathcal{G}_{i_\ell}^{(n)}|\}\right).
\end{aligned}$$

Then note that (ii) follows from the same arguments as (i) upon observing that, on the event $\bigcap_{j=1}^{\ell} E_{n,i_j;\underline{\mathbf{i}}}\cap\bigcap_{j=1}^{\ell}\{|J_{i_j;\underline{\mathbf{i}}}^*|=0\}$, $s_{i_\ell;\underline{\mathbf{i}}}^* - \sigma_{S_n} - \sigma_{S_{i_\ell}}$ is stochastically dominated by an Erlang$(\ell a_{i_\ell}, S_{i_\ell})$ random variable.

□

We are now ready to prove Theorem 1.

Proof of Theorem 1(i). Start by conditioning on $I_n = i$. Next, let $a_i = i^{1/2-\delta}$ for some $0 < \delta < 1/2$ and $E_{n,i}$ be the event defined in Lemma 5. To start, define the event

$$F_{n,i} = \left\{ \mathcal{G}_i^{(n)} \simeq \mathcal{T}_i^c(t_{n,i}), \bigcap_{\mathbf{j} \in \mathcal{T}_i^c(t_{n,i})} \{D^+_{\theta_i(\mathbf{j})} = \mathcal{D}_{\mathbf{j}}\} \right\},$$

where θ_i is the bijection defining $\mathcal{G}_i^{(n)} \simeq \mathcal{T}_i^c(t_{n,i})$. Let U be uniformly distributed in $[0,1]$, independent of everything else, and set $I_n = \lceil nU \rceil$. We will start by showing that $P(F_{n,I_n}) \to 1$ as $n \to \infty$. Observe that, by virtue of the coupling construction in Section 5.1, on the event $\{|J_i^*| = 0\} \cap \left\{ |\mathcal{T}_i^c(t_{n,i})| = |\mathcal{G}_i^{(n)}| \right\}$, one can obtain a bijection θ_i by requiring $\theta_i(\mathbf{j}) := \kappa_i(j)$, where $\kappa_i(j)$ is the enumeration of the vertex in $\mathcal{G}_i^{(n)}$ corresponding to $\mathbf{j} \in \mathcal{T}_i^c(t_{n,i})$ in the exploration of $\mathcal{G}_i^{(n)}$. From this and Remark 3, note that

$$\begin{aligned} P(F_{n,i}) &\geq P\left(\{|J_i^*| = 0\} \cap \left\{ |\mathcal{T}_i^c(t_{n,i})| = |\mathcal{G}_i^{(n)}| \right\} \cap E_{n,i} \right) \\ &\geq 1 - P\left(E_{n,i}^c\right) - P\left(E_{n,i} \cap \{|J_i^*| \geq 1\}\right) \\ &\quad - P\left(E_{n,i} \cap \{|J_i^*| = 0\} \cap \left\{ |\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{G}_i^{(n)}| \right\} \right). \end{aligned}$$

The limit $P(F_{n,I_n}) \to 1$ as $n \to \infty$ will follow once we show that

$$\begin{aligned} \Delta_n := \frac{1}{n} \sum_{i=1}^n & \left(P(E_{n,i}^c) + P(E_{n,i} \cap \{|J_i^*| \geq 1\}) \right. \\ & \left. + P(E_{n,i} \cap \{|J_i^*| = 0\} \cap \left\{ |\mathcal{T}_i^c(t_{n,i})| \neq |\mathcal{G}_i^{(n)}| \right\} \right) \to 0 \end{aligned}$$

as $n \to \infty$. Note that by part (i) of Lemmas 4, 5, 6, 7 we have that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Delta_n \leq \lim_{n\to\infty} \left\{ E\left[1(I_n < n^{c_\vartheta}) + 1(I_n = 1) + \frac{2C_f \mu I_n^{-2\delta}}{f_*} \right] \right\} \leq \lim_{n\to\infty} P(U < n^{c_\vartheta - 1}) = 0.$$

This in turn implies that

$$P(F_{n,I_n}) \to 1, \qquad n \to \infty. \tag{6}$$

Finally, note that, recalling $\chi = -(1/\lambda)\log U$ and setting and $\theta(\mathbf{j}) := \theta_{I_n}(\mathbf{j})$,

$$\begin{aligned} &\left| P\left(\mathcal{G}_{I_n}^{(n)} \simeq \mathcal{T}_{I_n}^c(\chi), \bigcap_{\mathbf{j} \in \mathcal{T}_{I_n}^c(\chi)} \{D^+_{\theta(\mathbf{j})} = \mathcal{D}_{\mathbf{j}}\} \right) - 1 \right| \\ &\leq |P(F_{n,I_n}) - 1| + P\left(\mathcal{T}_{I_n}^c(\chi) \not\simeq \mathcal{T}_{I_n}^c(t_{n,I_n}) \right). \end{aligned}$$

To see that $P\left(\mathcal{T}_{I_n}^c(\chi) \not\simeq \mathcal{T}_{I_n}^c(t_{n,I_n}) \right) \to 0$ as $n \to \infty$, note that for $\epsilon_n = n^{-1/2}$,

$$\begin{aligned} &P\left(\mathcal{T}_{I_n}^c(\chi) \not\simeq \mathcal{T}_{I_n}^c(t_{n,I_n}) \right) \\ &= P\left(|\mathcal{T}_{I_n}^c(\chi)| > |\mathcal{T}_{I_n}^c(t_{n,I_n})| \right) \\ &\leq P(\chi - t_{n,I_n} > \epsilon_n) + P\left(|\mathcal{T}_{I_n}^c(\chi)| > |\mathcal{T}_{I_n}^c(\chi - \epsilon_n)| \} \right) \\ &\leq P\left(\frac{\lceil nU \rceil}{nU} > e^{\lambda \epsilon_n} \right) + P\left(\mathrm{Exp}\left(C_f \left(|\mathcal{T}_{I_n}^c(\chi - \epsilon_n)| + \sum_{j \in \mathcal{T}_{I_n}^c(\chi - \epsilon_n)} \mathcal{D}_j \right) \right) \leq \epsilon_n \right) \\ &\leq \frac{1}{n(e^{\lambda \epsilon_n} - 1)} + E\left[1 - e^{-C_f \epsilon_n \left(|\mathcal{T}_{I_n}^c(\chi)| + \sum_{j \in \mathcal{T}_{I_n}^c(\chi)} \mathcal{D}_j \right)} \right] \to 0, \end{aligned} \tag{7}$$

as $n \to \infty$. Noting that $\mathcal{T}_{I_n}^c(\chi)$ has the same law as $\mathcal{T}^c(\chi)$ (with $\mathcal{T}^c(\cdot)$ independent of χ) completes the proof of Theorem 1(i). □

We now give the proof of part (ii) of Theorem 1.

Proof of Theorem 1(ii). Recall the events $F_{n,i}$ defined in the proof of part (i) of Theorem 1. Let $U_1, \ldots, U_m$ be independent random variables, uniformly distributed in $[0,1]$, and set $I_{n,\ell} = \lceil nU_\ell \rceil$. We first show that $P(\bigcap_{\ell=1}^m F_{n,I_{n,\ell}}) \to 1$ as $n \to \infty$. Note that for any $\underline{\mathbf{i}} = (i_1, \ldots, i_m) \subseteq [n]$,

$$\begin{aligned}
P\left(\bigcap_{\ell=1}^m F_{n,i_\ell}\right) &\geq P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}} \cap \bigcap_{\ell=1}^m \{|J^*_{i_\ell;\underline{\mathbf{i}}}| = 0\} \cap \bigcap_{\ell=1}^m \left\{|\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t_{n,i_\ell})| = |\mathcal{G}^{(n)}_{i_\ell}|\right\}\right) \\
&\geq 1 - \sum_{\ell=1}^m P\left(E^c_{n,i_\ell;\underline{\mathbf{i}}}\right) - P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}} \cap \bigcup_{\ell=1}^m \{|J^*_{i_\ell;\underline{\mathbf{i}}}| \geq 1\}\right) \\
&\quad - P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}} \cap \bigcap_{\ell=1}^m \{|J^*_{i_\ell;\underline{\mathbf{i}}}| = 0\} \cap \bigcup_{\ell=1}^m \left\{|\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t_{n,i_\ell})| \neq |\mathcal{G}^{(n)}_{i_\ell}|\right\}\right).
\end{aligned}$$

So it suffices to prove

$$\begin{aligned}
&\frac{1}{n^m} \sum_{\underline{\mathbf{i}} \subseteq [n]: |\underline{\mathbf{i}}| = m} \left[\sum_{\ell=1}^m P(E^c_{n,i_\ell;\underline{\mathbf{i}}}) + P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}} \cap \bigcup_{\ell=1}^m \{|J^*_{i_\ell;\underline{\mathbf{i}}}| \geq 1\}\right)\right. \\
&\left. + P\left(\bigcap_{\ell=1}^m E_{n,i_\ell;\underline{\mathbf{i}}} \cap \bigcap_{\ell=1}^m \{|J^*_{i_\ell;\underline{\mathbf{i}}}| = 0\} \cap \bigcup_{\ell=1}^m \left\{|\mathcal{T}^c_{i_\ell;\underline{\mathbf{i}}}(t_{n,i_\ell})| \neq |\mathcal{G}^{(n)}_{i_\ell}|\right\}\right)\right] \to 0
\end{aligned}$$

as $n \to \infty$. But this follows directly once we similarly apply part (ii) of Lemmas 4, 6, 7 and Lemma 5, Remark 5, Remark 6. Finally, we have

$$\begin{aligned}
\left|\mathcal{P}_{n,m}\left(\bigcap_{\ell=1}^m C_{I_{n,\ell}}\right) - 1\right| &\leq \left|P\left(\bigcap_{\ell=1}^m F_{n,I_{n,\ell}}\right) - 1\right| + \sum_{\ell=1}^m P(\mathcal{T}^c_{I_{n,\ell}}(\chi_\ell) \not\simeq \mathcal{T}^c_{I_{n,\ell}}(t_{n,I_{n,\ell}})) \\
&\leq \left|P\left(\bigcap_{\ell=1}^m F_{n,I_{n,\ell}}\right) - 1\right| + mP(\mathcal{T}^c_{I_n}(\chi) \not\simeq \mathcal{T}^c_{I_n}(t_{n,I_n})).
\end{aligned}$$

We already showed that both terms converge to 0 as $n \to \infty$, which completes the proof. □

Proof of Corollary 1.1. Fix a finite tree T and a deterministic sequence of marks $\{d_{\mathbf{j}} : \mathbf{j} \in \mathcal{U}\}$. For $n \in \mathbb{N}$ and $\{I_{n,k} : k = 1, 2\}$ i.i.d. sampled uniformly from V_n, independently of anything else, recall the coupling $\mathcal{P}_{n,2}$ of $\left(\mathcal{G}^{(n)}_{I_{n,1}}, \mathcal{G}^{(n)}_{I_{n,2}}\right)$ with $(\mathcal{T}^c_1(\chi_1), \mathcal{T}^c_2(\chi_2))$, where $\mathcal{T}^c_1(\chi_1)$ and $\mathcal{T}^c_2(\chi_2)$ are i.i.d. copies of $\mathcal{T}^c(\chi)$. Denote the corresponding expectation by $\mathbb{E}_{n,2}$.

To simplify the notation, for $i \in V_n$ and $k = 1, 2$, define the events

$$F_i = \left\{\mathcal{G}^{(n)}_i \simeq T, \bigcap_{\mathbf{j} \in T} \{D^+_{\theta_i(\mathbf{j})} = d_{\mathbf{j}}\}\right\} \text{ and } \hat{F}_k = \left\{\mathcal{T}^c_k(\chi_k) \simeq T, \bigcap_{\mathbf{j} \in T} \{\mathcal{D}^{(k)}_{\mathbf{j}} = d_{\mathbf{j}}\}\right\},$$

where θ_i is the bijection that defines the isomorphism $\mathcal{G}^{(n)}_i \simeq T$, and $\mathcal{D}^{(k)}_{\mathbf{j}}$ denotes the mark of the node indexed $\mathbf{j}$ in $\mathcal{T}^c_k(\chi_k)$. Note that

$$\begin{aligned}
&\mathbb{E}_{n,2}\left(\frac{1}{n}\sum_{i=1}^n 1(F_i) - \mathcal{P}_{n,2}(\hat{F}_1)\right)^2 \\
&= \mathbb{E}_{n,2}\left(\frac{1}{n^2}\sum_{i,j=1}^n 1(F_i)1(F_j)\right) - 2\mathcal{P}_{n,2}(F_{I_{n,1}})\mathcal{P}_{n,2}(\hat{F}_1) + \left(\mathcal{P}_{n,2}(\hat{F}_1)\right)^2. \qquad (8)
\end{aligned}$$

By Theorem 1(i), with C_{I_n} defined in its statement,

$$|\mathcal{P}_{n,2}(F_{I_{n,1}}) - \mathcal{P}_{n,2}(\hat{F}_1)| \leq \mathcal{P}_n(C_{I_n}^c) \to 0$$

as $n \to \infty$. Moreover, by Theorem 1(ii) (applied for $m = 2$), with $C_{I_{n,1}}, C_{I_{n,2}}$ defined in its statement,

$$\left|\mathbb{E}_{n,2}\left[\frac{1}{n^2}\sum_{i,j=1}^{n} 1(F_i)1(F_j)\right] - \left(\mathcal{P}_{n,2}(\hat{F}_1)\right)^2\right| = \left|\mathbb{E}_{n,2}\left[\frac{1}{n^2}\sum_{i,j=1}^{n} 1(F_i)1(F_j)\right] - \mathcal{P}_{n,2}(\hat{F}_1 \cap \hat{F}_2)\right|$$

$$= \left|\mathcal{P}_{n,2}(F_{I_{n,1}} \cap F_{I_{n,2}}) - \mathcal{P}_{n,2}(\hat{F}_1 \cap \hat{F}_2)\right| \leq \mathcal{P}_{n,2}(C_{I_{n,1}}^c \cup C_{I_{n,2}}^c) \to 0$$

as $n \to \infty$. Using the above two observations in (8), we conclude

$$\mathbb{E}_{n,2}\left[\left(\frac{1}{n}\sum_{i=1}^{n} 1(F_i) - P\left(\mathcal{T}^c(\chi) \simeq T, \bigcap_{\mathbf{j}\in T}\{\mathcal{D}_{\mathbf{j}} = d_{\mathbf{j}}\}\right)\right)^2\right] \to 0$$

as $n \to \infty$. The result follows from this. □

Proof of Proposition 1. (1) *Preferential attachment:* The expression for $P(\mathcal{N}_\emptyset = x)$ follows from [10, Corollary 1.4]. To get the exponent for a regularly varying out-degree distribution, observe that, by Stirling's formula,

$$\frac{\Gamma(x + d(\beta+1))}{\Gamma(x + d(\beta+1) + 3 + \beta)} = (x + d(\beta+1))^{-3-\beta}(1 + O((x + d(\beta+1))^{-1})),$$

$$l(d) := \frac{(2+\beta)\Gamma(2+\beta+d(\beta+1))}{d^{2+\beta}\Gamma(d(\beta+1))} = (2+\beta)(\beta+1)^{2+\beta}(1 + O(d^{-1})).$$

Using the expression for $P(\mathcal{N}_\emptyset = x)$ in [10, Corollary 1.4] and the first estimate above, we can write

$$P(\mathcal{N}_\emptyset = x) = (1 + O(x^{-1}))\sum_{d=1}^{\infty} d^{2+\beta-\gamma}(x + (\beta+1)d)^{-3-\beta}\tilde{L}(d),$$

where $\tilde{L}(d) := l(d)L(d) \sim L(d)$ as $d \to \infty$.

Fix $\epsilon \in (0,1)$ and define $b(x) = \lfloor x^{1+\epsilon} \rfloor$. Note that, using [6, Proposition 1.5.10],

$$\begin{aligned}
P(\mathcal{N}_\emptyset = x) &= O\left(\sum_{d=b(x)+1}^{\infty} \tilde{L}(d)d^{-1-\gamma}\right) \\
&\quad + (1 + O(x^{-1}))\sum_{d=1}^{b(x)} d^{2+\beta-\gamma}(x + (\beta+1)d)^{-3-\beta}\tilde{L}(d) \\
&= O\left(b(x)^{-\gamma}\tilde{L}(b(x))\right) + (1 + O(x^{-1}))K_\gamma P(XY > x)
\end{aligned}$$

where $P(Y = d) = d^{-\gamma-1}\tilde{L}(d)/K_\gamma$ for $d \in \mathbb{N}$ and $K_\gamma = \sum_{d=1}^{\infty} d^{-\gamma-1}\tilde{L}(d)$ and $P(X > x) = (1+\beta+x)^{-3-\beta}$ for $x > -\beta$ is a Type II Pareto random variable, independent of Y. The assertion about regular variation of $x \mapsto P(XY > x)$ follows from [13, Lemma 4.1]. By Breiman's theorem (see [13, Lemma 4.2]), $P(XY > x) \sim E[Y^{3+\beta}]P(X > x) \sim E[Y^{3+\beta}]x^{-3-\beta}$ if $\gamma > 3+\beta$ and $P(XY > x) \sim E[(X^+)^\gamma]P(Y > x) \sim E[(X^+)^\gamma]K_\gamma^{-1}\tilde{L}(x)x^{-\gamma}$ if $2 \leq \gamma < 3+\beta$, as $x \to \infty$. Since

$$b(x)^{-\gamma}\tilde{L}(b(x)) = o\left(x^{-\gamma}\right) = o\left(P(XY > x)\right)$$

as $x \to \infty$, we obtain that

$$P(\mathcal{N}_\emptyset = x) = (1 + o(1))K_\gamma P(XY > x)$$

as $x \to \infty$.

(2) *Uniform attachment:* The expression for $P(\mathcal{N}_\emptyset = x)$ follows from [10, Corollary 1.6]. To get the lower bound on the in-degree distribution for regularly varying out-degree distribution, observe that

$$P(\mathcal{N}_\emptyset = x) = \sum_{d=1}^{\infty} d^{-\gamma-1} L(d) \left(1 + \frac{1}{d}\right)^{-x-1} = E[1/\mathcal{D}] E\left[\left(1 + \frac{1}{Y'}\right)^{-x-1}\right]$$
$$= E[1/\mathcal{D}] E\left[e^{-(x+1)\log(1+1/Y')}\right] = E[1/\mathcal{D}] P\left(\frac{W}{\log(1+1/Y')} > x + 1\right)$$

where W is an exponential random variable with rate one, independent of Y', and $P(Y' = d) = d^{-\gamma-1}L(d)/E[1/\mathcal{D}]$ for $d \in \mathbb{N}$.

Now note that the random variable $V = 1/\log(1 + 1/Y')$ satisfies, as $x \to \infty$,

$$P(V > x) = P\left(Y' > 1/(e^{1/x} - 1)\right) = P(Y' > x(1 + o(1))) = (1 + o(1))(E[1/\mathcal{D}])^{-1} x^{-\gamma} L(x),$$

and is, therefore, regularly varying with tail index γ. Breiman's theorem gives now

$$E[1/\mathcal{D}]P(WV > x+1) = (1+o(1))E[W^\gamma]E[1/\mathcal{D}]P(V > x+1) = (1+o(1))E[W^\gamma]x^{-\gamma}L(x)$$

as $x \to \infty$.

□

Acknowledgements: We thank Shankar Bhamidi for helpful discussions and advice throughout the course of writing this article. Research supported in part by the NSF RTG award (DMS-2134107). Sayan Banerjee and Prabhanka Deka were supported in part by the NSF-CAREER award (DMS-2141621).

References

[1] David Aldous and J. Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

[2] Sayan Banerjee, Shankar Bhamidi, and Xiangying Huang. Co-evolving dynamic networks. *arXiv preprint arXiv:2203.11877*, 2022.

[3] Sayan Banerjee and Mariana Olvera-Cravioto. PageRank asymptotics on directed preferential attachment networks. *The Annals of Applied Probability*, 32(4):3060–3084, 2022.

[4] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. In *Selected Works of Oded Schramm*, pages 533–545. Springer, 2011.

[5] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. Asymptotic behavior and distributional limits of preferential attachment graphs. *The Annals of Probability*, 42(1):1–40, 2014.

[6] Nicholas H Bingham, Charles M Goldie, Jozef L Teugels, and JL Teugels. *Regular variation*. Number 27. Cambridge university press, 1989.

[7] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. *arXiv preprint arXiv:2203.00446*, 2022.

[8] Maria Deijfen, Henri Van Den Esker, Remco Van Der Hofstad, and Gerard Hooghiemstra. A preferential attachment model with random initial degrees. *Arkiv för matematik*, 47(1):41–72, 2009.

[9] Alessandro Garavaglia, Rajat Subhra Hazra, Remco van der Hofstad, and Rounak Ray. Universality of the local limit of preferential attachment models. *arXiv preprint arXiv:2212.05551*, 2022.

[10] Alessandro Garavaglia and Remco van der Hofstad. From trees to graphs: collapsing continuous-time branching processes. *Journal of Applied Probability*, 55(3):900–919, 2018.

[11] Alessandro Garavaglia, Remco van der Hofstad, and Nelly Litvak. Local weak convergence for PageRank. *The Annals of Applied Probability*, 30(1):40–79, 2020.

[12] Peter Jagers and Olle Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, 16(2):221–259, 1984.

[13] Hedegaard Anders Jessen and Thomas Mikosch. Regularly varying functions. *Publications de L'institut Mathematique*, 80(94):171–192, 2006.

[14] Tiffany YY Lo. Weak local limit of preferential attachment random trees with additive fitness. *arXiv preprint arXiv:2103.00900*, 2021.

[15] Olle Nerman. On the convergence of supercritical general (CMJ) branching processes. *Probability Theory and Related Fields*, 57(3):365–395, 1981.

[16] Seppo Niemi and Esa Nummelin. On non-singular renewal kernels with an application to a semigroup of transition kernels. *Stochastic Processes and their Applications*, 22(2):177–202, 1986.
[17] Esa Nummelin. Uniform and ratio limit theorems for markov renewal and semi-regenerative processes on a general state space. In *Annales de l'IHP Probabilités et statistiques*, volume 14, pages 119–143, 1978.
[18] Mariana Olvera-Cravioto. PageRank's behavior under degree correlations. *Annals of Applied Probability*, 3(1):1403–1442, 2021.
[19] Mariana Olvera-Cravioto. Strong couplings for static locally tree-like random graphs. *Journal of Applied Probability*, 4(59):1261–1285, 2022.
[20] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
[21] Anna Rudas, Bálint Tóth, and Benedek Valkó. Random trees and general branching processes. *Random Structures Algorithms*, 31(2):186–202, 2007.
[22] Remco Van Der Hofstad. *Random graphs and complex networks Vol. 2*. Cambridge Series in Statistical and Probabilistic Mathematics *(to appear)*, 2023.

Department of Statistics and Operations Research, 304 Hanes Hall, University of North Carolina, Chapel Hill, NC 27599
Email address: sayan@email.unc.edu, deka@live.unc.edu, molvera@email.unc.edu