

Degree centrality and root finding in growing random networks*

Sayan Banerjee[†] Xiangying Huang[‡]

Abstract

We consider growing random networks $\{\mathcal{G}_n\}_{n \geq 1}$ where, at each time, a new vertex attaches itself to a collection of existing vertices via a fixed number $m \geq 1$ of edges, with probability proportional to a function f (called attachment function) of their degree. It was shown in [BB21] that such network models exhibit two distinct regimes: (i) the persistent regime, corresponding to $\sum_{i=1}^{\infty} f(i)^{-2} < \infty$, where the top K maximal degree vertices fixate over time for any given K , and (ii) the non-persistent regime, with $\sum_{i=1}^{\infty} f(i)^{-2} = \infty$, where the identities of these vertices keep changing infinitely often over time. In this article, we develop root finding algorithms using the empirical degree structure and local network information based on a snapshot of such a network at some large time. In the persistent regime, the algorithm is purely based on degree centrality, that is, for a given error tolerance $\varepsilon \in (0, 1)$, there exists $K_\varepsilon \in \mathbb{N}$ such that for any $n \geq 1$, the confidence set for the root in \mathcal{G}_n , which contains the root with probability at least $1 - \varepsilon$, consists of the top K_ε maximal degree vertices. In particular, the size of the confidence set is stable in the network size. Upper and lower bounds on K_ε are explicitly characterized in terms of the error tolerance ε and the attachment function f . In the non-persistent regime, for an appropriate choice of $r_n \rightarrow \infty$ at a rate much smaller than the diameter of the network, the neighborhood of radius r_n around the maximal degree vertex is shown to contain the root with high probability, and a size estimate for this set is obtained. It is shown that, when $f(k) = k^\alpha, k \geq 1$, for any $\alpha \in (0, 1/2]$, this size grows at a smaller rate than any positive power of the network size.

Keywords: attachment functions; continuous time branching processes; network centrality measures; degree centrality; persistence; root finding algorithms.

MSC2020 subject classifications: 05C82; 60J85; 60J28.

Submitted to EJP on February 21, 2022, final version accepted on March 7, 2023.

*SB was supported in part by the NSF-CAREER award (DMS-2141621) and the NSF RTG award (DMS-2134107).

[†]Department of Statistics and Operations Research, Hanes Hall, University of North Carolina, Chapel Hill, NC 27599. E-mail: sayan@email.unc.edu

[‡]Department of Statistics and Operations Research, Hanes Hall, University of North Carolina, Chapel Hill, NC 27599. E-mail: zoehuang@unc.edu

1 Introduction

The last two decades have seen an explosion in the application of temporal networks in diverse scenarios, including modeling and analyzing connectivity in groups of individuals, spread of epidemics, opinions and rumors, just to name a few. Owing to their large size, it is often infeasible to track these networks over time; instead, one observes a snapshot of the network at some large time and asks about its nascent states. In this article, we are interested in growing random networks where an incoming vertex connects to a collection of existing vertices via a fixed number of edges, with probability proportional to a function f of their degree, called the attachment function. We ask the following: can one detect the root (first vertex in the network) to a quantifiable degree of accuracy based on observing a single snapshot of this network process? This question connects to the broad field of *network archaeology* [NK11] where one is interested in reconstructing the past of a network from its present, including connectivity structure, central nodes, etc. Although related questions have been studied in a more applied setting, a rigorous mathematical analysis has only recently been initiated (see Section 1.2). Most mathematical works focus on the uniform and linear preferential attachment networks (when f is respectively constant and linearly increasing) owing to their analytical tractability. However, from a modeling perspective, such assumptions seem rather ad hoc and ideally the attachment function should be inferred based on the network data at hand. Hence, it is important to address the above questions in a more general setting with minimal assumptions on the attachment function. In this article, we ask to what extent the information on the empirical degree structure of the network snapshot can aid in root detection. This is part of our continuing program to understand network archaeology for growing random networks driven by general attachment functions (see also [BBC22, BB21, BB22]).

1.1 Model definition

We define a slightly more general version of the class of networks studied in this paper. Fix an attachment function $f : \mathbb{N} \rightarrow (0, \infty)$ and an \mathbb{N} -valued (possibly random) attachment sequence $\{m_i : i \geq 1\}$. We will construct a sequence of random graphs $\{\mathcal{G}_n : n \geq 1\}$ as follows:

(i) Let \mathcal{G}_1 be the graph with two vertices $\{v_0, v_1\}$ and m_1 edges between them, where v_0 is also referred to as the root.

(ii) Fix $n > 1$. Given we have obtained \mathcal{G}_{n-1} , \mathcal{G}_n is constructed from \mathcal{G}_{n-1} by adding one vertex v_n and m_n directed edges, each with one end connected to v_n , and the other ends of these edges are connected sequentially to one of the existing vertices $\{v_i : 0 \leq i \leq n-1\}$ in \mathcal{G}_{n-1} according to the following rule. For $1 \leq j \leq m_n$, the j -th edge has its other end connected to v_i ($0 \leq i \leq n-1$) with probability proportional to $f(\text{degree of } v_i)$ (the degree is computed before the connection is made).

We will find it convenient to index the graph sequence by the number of attached edges (with both ends already connected to respective vertices). Let $s_n := \sum_{i=1}^n m_i$ for $n \geq 1$ and $s_0 = 0$. Mathematically, we construct the following (directed) random graph sequence $\{\mathcal{G}_k^* : k \geq s_1\}$ where \mathcal{G}_k^* has k attached edges. For any $n \geq 2$ and any $s_{n-1} < k \leq s_n$, \mathcal{G}_k^* has $n+1$ vertices. For $l \geq 0$, let $d_0(l)$ denote the degree of the root after the l -th edge is attached. For $i \geq 1$, denote by $d_i(l)$, $l > s_{i-1}$, the degree of the $(i+1)$ -th vertex (i.e. vertex v_i) after the l -th edge is attached and set $d_i(l) = 0$ for all $l \leq s_{i-1}$. For $s_{n-1} < k \leq s_n$, an edge e_k is added to the graph, with one end attached to

v_n , and for $0 \leq i \leq n - 1$,

$$P(\text{other end of } e_k \text{ attached to } v_i | \mathcal{G}_j^*, j \leq k - 1) = \frac{f(d_i(k - 1))}{\sum_{j=0}^{n-1} f(d_j(k - 1))}. \quad (1.1)$$

Thus, $\{\mathcal{G}_n : n \geq 1\}$ has the same law as $\{\mathcal{G}_{s_n}^* : n \geq 1\}$. We will denote the vertex set of \mathcal{G}_n by $V(\mathcal{G}_n)$.

Two important special cases are: (i) $f(k) = k + \beta$ with $k \geq 1$ for some $\beta \geq 0$, which is known as the *linear preferential attachment model*, and (ii) $f \equiv 1$, which is called the *uniform attachment model*.

1.2 Centrality measures, persistence and root finding algorithms

Let \mathbb{G} be a collection of (possibly labeled) graphs. We write \mathbb{G}° for the corresponding collection of unlabeled graphs and let $V(\mathbb{G}^\circ)$ denote its vertex set. A centrality measure $\Psi : V(\mathbb{G}^\circ) \rightarrow \mathbb{R}_+$ gives each vertex in the graph a ‘score’ based on some notion of ‘centrality’ in an appropriate geometry of the unlabeled graph. In our setup, $\mathbb{G} = \{\mathcal{G}_n : n \geq 1\}$ is a growing sequence of random graphs and our goal is to recover the root from an unlabeled copy of \mathcal{G}_n for some large n . The idea is to choose a centrality measure Ψ that *correlates strongly with the age* of the vertex. Then, for a given error tolerance $\varepsilon \in (0, 1)$, we can choose the $K_\varepsilon(n)$ most central vertices (as determined by Ψ) in \mathcal{G}_n for an appropriate $K_\varepsilon(n) \in \mathbb{N}$ and ensure that the root of \mathcal{G}_n is contained in this set with probability at least $1 - \varepsilon$.

There are several existing notions of centrality, some of which we now discuss.

- (i) *Degree centrality* [DM09, Gal13, BB21] quantifies the centrality of a vertex in a general network simply based on how large its degree is. For dynamic networks, the older vertices exist in the network for a longer time and one expects them to have a higher degree.
- (ii) *Jordan (or centroid) centrality* [BDL17, JL18, JL16, BB22] quantifies centrality in trees by looking at the size of the maximal subtree rooted at each vertex. The smaller this size, the more comparable the sizes are of all the subtrees rooted at this vertex, and the more central this vertex is.
- (iii) *Rumor Centrality* [SZ11, SZ12, KL16] was originally developed in connection to finding the most likely source of rumor spread or infection spread for a susceptible-infected (SI) epidemic model on a background random graph. This measure has been successfully used in devising root finding algorithms for the uniform attachment model in [BDL17].

There are several other notions of centrality [Fre77, New05, BV14], but only the ones above have so far been rigorously mathematically analyzed in connection with root detection. Recently, [BMR15, CDKM15] and [BEMR17, LP19, DR18] have respectively used degree and Jordan centrality in reconstruction of seeds (initial graph from which the network starts) from a single snapshot of the network at a large time in linear preferential attachment and uniform attachment models.

A very desirable property of centrality measures is *persistence* as defined below.

Definition 1.1. Fix $K \geq 1$ and a network centrality measure Ψ . For an evolving network sequence $\{\mathcal{G}_n : n \geq 1\}$, say that the sequence is (Ψ, K) persistent if there exists $n^* < \infty$ a.s. such that for all $n \geq n^*$ the optimal K vertices $(v_{1,\Psi}^{(n)}, v_{2,\Psi}^{(n)}, \dots, v_{K,\Psi}^{(n)})$, as measured by the centrality measure Ψ , remain the same and further the relative ordering amongst these K optimal vertices remain the same.

Persistence ensures that the identities of the most central vertices fixate over time. This implies that one can use the centrality measure Ψ to construct confidence sets

for the root, with given error tolerance ε , whose size only depend on ε and not on the network size, that is, $K_\varepsilon(n) \equiv K_\varepsilon$. This stability with respect to network size makes the confidence set robust for large networks.

Persistence of the degree centrality measure was first shown for the linear preferential attachment model in [Gal13]. The phenomenon of persistence for the general class of models described in Section 1.1 was first investigated in [BB21]. Building on the techniques of [DM09] (which investigated persistence for a related dynamic random graph model), they showed the existence of two distinct regimes: (i) the *persistent regime* when $\sum_{i=1}^{\infty} \frac{1}{f(i)^2} < \infty$, and (ii) the *non-persistent regime* when $\sum_{i=1}^{\infty} \frac{1}{f(i)^2} = \infty$. We now summarize some of the results of [BB21]. In the following, for any $v \in \mathcal{G}_n$, $\Psi(v)$ is the degree of v in \mathcal{G}_n . Moreover, m_i can be random for the results stated below.

- (i) *Persistent regime*: Assume f is non-decreasing and there exists $C_f > 0$ such that $f(i) \leq C_f i$ for all $i \geq m_1$. Also, suppose $\sum_{i=1}^{\infty} \frac{1}{f(i)^2} < \infty$ and that, almost surely,

$$\limsup_{n \rightarrow \infty} \frac{1}{\log s_n} \sum_{i=1}^{m_n-1} \frac{1}{f(i)} \leq \frac{1}{8C_f}.$$

Then, the random graph sequence $\{\mathcal{G}_n\}_{n \geq 1}$ is (Ψ, K) persistent for any fixed $K \geq 1$. In case $m_i = 1$ for all $i \geq 1$ (the tree case), the monotonicity and sub-linearity assumptions on f can be replaced by Assumption (A2) below.

- (ii) *Non-persistent regime*: If $\sum_{i=1}^{\infty} \frac{1}{f(i)^2} = \infty$ and a ‘continuity’ assumption (3.18) on the attachment function is satisfied, then, almost surely, the random graph sequence $\{\mathcal{G}_n\}_{n \geq 1}$ is not (Ψ, K) persistent for any $K \geq 1$.

The above suggests that, in the persistent regime, there is hope for devising degree centrality based root finding algorithms which produce confidence sets that are stable in the network size. In the non-persistent regime, the stability of purely degree based algorithms in the network size seems unlikely. It is then natural to ask how much of the network connectivity information is required to obtain efficient root finding algorithms.

In [BB22], it was shown that the Jordan centrality measure (defined only for the tree case $m_i \equiv 1$) exhibits persistence more generally whenever the attachment function f satisfies Assumptions (A1)–(A3) below. This observation was used to obtain explicit confidence set bounds for root finding algorithms that grow polynomially in ε^{-1} (ε being the error tolerance). However, this centrality measure is applicable only for tree networks. Moreover, it is computationally expensive as computing the Jordan centrality measure for any single vertex requires knowledge of the connectivity structure of the whole network.

This leads to some natural questions. (i) Can we devise root finding algorithms for the dynamic networks defined in Section 1.1 for $m_i \equiv m \geq 1$ in the persistent regime ($\sum_{i=1}^{\infty} \frac{1}{f(i)^2} < \infty$), purely based on the empirical degree distribution, that are stable in the network size? In this case, can we give explicit bounds for the confidence set? (ii) In the non-persistent regime, can we obtain root finding algorithms using degree centrality and minimal local connectivity information around high degree vertices? What is the size of the associated confidence set and how does it grow with network size? These questions, besides their theoretical appeal, are extremely important from an applications perspective as most real world networks are large and far from tree like. Degree based centrality measures are much easier to compute than their global analogues (like Jordan centrality) as they are purely local. Moreover, they are well-defined and applicable on non-tree networks.

1.3 Our contribution

We attempt to answer the above questions in the current article. In the process, we develop a quantitative approach to degree centrality. Theorem 2.2 obtains sharp upper and lower bounds on the size of the confidence set for the root (or budget) in terms of the inverse error tolerance for the linear preferential attachment model for general $m \geq 1$. Theorem 2.3 addresses the general persistent regime. For trees, the bounds are almost optimal. Even for $m > 1$, the bounds are polynomial in the inverse error tolerance and are explicitly obtained in terms of the attachment function. It is noted in Remark 2.4 that, in the persistent regime, not only does one obtain purely local root finding algorithms that do not require a network tree structure as in [BB22], but also the bounds obtained are nearly as good as, and in most cases tighter than, those for Jordan centrality. A surprising observation, noted in the same remark, is that, as seen explicitly through the size of the confidence set in the linear attachment case in Theorem 2.2, the efficacy of root detection using degree centrality gets better with increasing m . Thus, although the network becomes increasingly non-tree like for large m , the increased positive reinforcement of degrees allows the older vertices to gain in degree even further.

Next, we analyze the non-persistent regime. In Theorem 2.5, we show that a neighborhood of the maximal degree vertex of an appropriately chosen radius $r_n \rightarrow \infty$ contains the root with high probability. For given error tolerance ε , we also quantify the minimum network size for the above set to contain the root with probability at least $1 - \varepsilon$. We note in Remark 2.6 that for large n the radius r_n is much smaller than the diameter of the network, and for sublinear attachment functions $f(k) = k^\alpha, k \geq 1$, for any $\alpha \in (0, 1/2]$, the size of the confidence set grows at a smaller rate than any positive power of the network size. In Remark 2.7 we list some results (which essentially follow from [BB21]) on the degree distribution of fixed vertices and the maximum degree as the network grows. In particular, the degrees of fixed vertices scale similarly as the maximum degree in the persistent regime, but show a stark contrast in the non-persistent regime.

Although we use a number of ideas from [BB21] and [BB22], there are several significant new techniques and ideas developed here, some of which we now describe. We introduce a novel embedding of the network in the case $m_i \equiv m > 1$ into a collapsed continuous time branching process (see Section 3) which lets us bypass the monotonicity assumptions required on f in [BB21] in the non-tree case. We also develop a unified approach to budget lower bounds (see Section 5) and apply it in Lemmas 5.1, 7.1 and 7.3. A major technical advance in the current article is the detailed quantification of the asymptotic behavior of degrees, population sizes of associated continuous time branching processes (CTBP), the age of the maximal degree vertex and its distance from the root. To do this, we exploit properties of recursive distributional equations which characterize limiting normalized population sizes for CTBP (see Lemma 3.5) extending the program of [BB22]. Moreover, the proof of Theorem 2.5 relies on novel connections with asymptotics of the height of CTBP (maximal distance of root to any vertex), and characterizing the behavior of functionals of associated Laplace transforms near zero (see Lemmas 8.1 and 8.2).

We note here that for linear preferential attachment networks, there is an extensive body of literature (see [BRST01, PRR17, Sén21] and references therein) obtaining detailed asymptotics for the degree evolution of vertices, the limits of (appropriately normalized) degrees to explicit random variables and rates of convergence to these limits. It would be interesting to obtain bounds on the confidence sets by appealing to these results in the linear case and compare them to the ones obtained in the current article. However, our goal here is to develop general techniques that are robust enough to apply to a much wider class of attachment functions, and we stick to this principle

throughout the paper.

Section 2 states the main results and an outline of proof techniques. Section 3 describes branching process embeddings for the network and collects several tail estimates for important quantities of this branching process. Section 4 collects properties of some key martingales used to track degrees of individual vertices. Section 5 describes a universal approach for obtaining budget lower bounds by showing that, if the root has relatively small degree in the graph \mathcal{G}_K , then with probability at least $1/2$, its centrality score cannot beat those of an $\Omega(K)$ number of vertices at any subsequent time. Probabilities of such configurations in \mathcal{G}_K are then estimated in respective cases. Section 6, 7, 8 are respectively devoted to the proofs of Theorems 2.2, 2.3 and 2.5.

We note here that, for linear preferential attachment models, there is an alternate notion of progressive exploration towards finding the root in large networks where one performs local moves towards vertices of higher degree [BBC⁺12, FP17, BK10]. However, this leads to a confidence set that grows polylogarithmically in n (as opposed to our confidence set, whose size is stable in n), with the computational advantage of only having to explore a small part of the graph (compared to our method requiring the degree of each vertex in the network). We plan to investigate such exploratory root finding algorithms for more general attachment networks in a future article.

Notation: Throughout this paper, C, C', C'' will be used to denote generic positive constants whose values might change from line to line. Unless otherwise stated, all constants are implicitly assumed to depend on m (number of edges attached per new vertex). We also set the convention ‘ $\sum_{i=a}^b = 0$ ’ whenever $a > b$. For two càdlàg processes χ_1, χ_2 , we will write $\chi_1(\cdot) \stackrel{d}{\geq} \chi_2(\cdot)$ if there is a coupling $(\tilde{\chi}_1(\cdot), \tilde{\chi}_2(\cdot))$ such that $\tilde{\chi}_1(t) \geq \tilde{\chi}_2(t)$ for all $t \geq 0$.

We now define some key objects used in this article. Let $\{E_i : i \geq 1\}$ be i.i.d. exponential random variables with mean 1. For $k \geq 1$, define

$$S_k(l) := \sum_{i=1}^l \frac{E_i}{f^k(i)}, \quad l \in \mathbb{N}. \tag{1.2}$$

Define the point process ξ by

$$\xi(t) := \sum_{n=1}^{\infty} 1_{\{S_1(n) \leq t\}}, \quad t \geq 0. \tag{1.3}$$

We will write $\sigma_l := S_1(l)$, $l \geq 1$, to denote the arrival time of the l -th individual in the above point process. For $A \in \mathbb{N}_0$, denote by $\xi_A(\cdot)$ the point process obtained as in (1.3) with attachment function $f_A(\cdot) := f(A + \cdot)$ replacing $f(\cdot)$. Note that $\xi_0(\cdot) \stackrel{d}{=} \xi(\cdot)$. By convention, for any $A \in \mathbb{N}_0$, we take $\xi_A(t) = 0$ for $t < 0$.

Define the functions

$$\Phi_k(l) := \sum_{i=1}^{l-1} \frac{1}{f^k(i)}, \quad l, k \in \mathbb{N}. \tag{1.4}$$

Extend f to a function on $[1, \infty)$ by requiring $f(x) = f(\lfloor x \rfloor)$ where $x \geq 1$. Extend $\Phi_k(\cdot)$, $k = 1, 2$, to $[1, \infty)$ via linear interpolation and set $\Phi_k(z) = 0$ for $z \in [0, 1]$. Note that, using the above extensions of f, Φ_1, Φ_2 ,

$$\Phi_k(x) = \int_1^x \frac{1}{f^k(z)} dz, \quad x \geq 1, k = 1, 2.$$

Define

$$\mathcal{K}(t) := \Phi_2 \circ \Phi_1^{-1}(t), \quad t \geq 0.$$

We will use the notation $\Phi_k(\infty) := \lim_{l \rightarrow \infty} \Phi_k(l)$.

1.4 Assumptions on attachment functions

Let $\mu(t) := E(\xi(t)), t \geq 0$. We can naturally obtain a non-negative measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ from μ (the intensity measure of ξ) which we also denote by μ . We define the Laplace transform of μ by

$$\hat{\mu}(\theta) = \int_0^\infty e^{-\theta t} \mu(dt), \quad 0 \leq \theta < \infty.$$

It is known that (see [RTV07, Equation (3)])

$$\hat{\mu}(\theta) = \sum_{k=1}^{\infty} \prod_{i=1}^k \frac{f(i)}{\theta + f(i)}.$$

Throughout this article, we have the following assumptions on f :

(A1) Every attachment function satisfies $f(k) \rightarrow \infty$ as $k \rightarrow \infty$, and f can grow at most linearly, i.e., there exists $C_f > 0$ such that $f(i) \leq C_f i$ for all $i \geq m_1$.

(A2) Define $\underline{\lambda} := \inf\{\lambda > 0 : \hat{\mu}(\lambda) < \infty\}$. We assume,

$$\underline{\lambda} < \infty \quad \text{and} \quad \lim_{\lambda \downarrow \underline{\lambda}} \hat{\mu}(\lambda) > 1.$$

Assumption (A2) together with the monotonicity of $\hat{\mu}(\cdot)$ implies that there exists a unique $\lambda^* := \lambda^*(f)$ such that

$$\hat{\mu}(\lambda^*) = 1. \tag{1.5}$$

This value λ^* is often referred to as the *Malthusian rate* of growth parameter. The above assumptions are relatively standard in the continuous time branching process literature, and will be essential for continuous time embeddings of our discrete network process discussed in Section 3.

(A3) With λ^* denoting the Malthusian rate of growth parameter, $\limsup_{i \rightarrow \infty} f(i)/i < \lambda^*$.

Observe that Assumption (A1) implies that $f_* := \min_{i \geq m_1} f(i) > 0$.

Remark 1.2. Suppose f satisfies Assumptions (A1) and (A2). Then, by [BB22, Lemma 2.3], Assumption (A3) is satisfied if any of the following conditions hold: (a) there exists $C^* \geq 0$ such that $\lim_{i \rightarrow \infty} f(i)/i = C^*$; (b) $\limsup_{i \rightarrow \infty} f(i)/i < f_*$. In particular (affine) preferential attachment and sublinear preferential attachment (there exists $\alpha \in (0, 1)$ such that $f(k) \leq k^\alpha$ for all $k \geq 1$) satisfy Assumptions (A1)-(A3).

2 Main results

Our main results give bounds on the minimum number of central vertices, measured by some centrality measure, required to produce a confidence set of the root for a given error tolerance. This is quantified through the notion of *budget* as defined below.

Let \mathbb{G} be a collection of (possibly labeled) graphs, and let \mathbb{G}° denote the corresponding set of unlabeled graphs. For any $K \geq 1$ and a centrality measure Ψ , we will write $H_{K, \Psi} : \mathbb{G} \rightarrow V(\mathbb{G})$ to denote the function that maps $G \in \mathbb{G}$ to a subset of K most central vertices in G as measured by Ψ (ties being broken arbitrarily).

Definition 2.1. Fix $0 < \varepsilon < 1$ and $K \geq 1$. A mapping $H_{K, \Psi}$ is called a *budget K root finding algorithm with error tolerance ε for the collection of random graphs $\{\mathcal{G}_n : n \geq 1\}$* if

$$\liminf_{n \rightarrow \infty} P(v_0 \in H_{K, \Psi}(\mathcal{G}_n)) \geq 1 - \varepsilon. \tag{2.1}$$

We will write $K_\Psi(\varepsilon)$ to denote the smallest budget K such that $H_{K, \Psi}(\cdot)$ is a root finding algorithm with error tolerance ε for the collection of random graphs $\{\mathcal{G}_n : n \geq 1\}$. If there is no finite K satisfying (2.1), we set $K_\Psi(\varepsilon) = \infty$.

2.1 The persistent regime

In this section, we restrict attention to Ψ being the degree centrality measure, that is $\Psi(v) = \deg(v)$, $v \in V(\mathbb{G}^\circ)$. We also focus on the persistent regime, namely, attachment functions satisfying $\Phi_2(\infty) < \infty$. Our first result gives sharp upper and lower bounds on $K_\Psi(\varepsilon)$ when f is linear.

Theorem 2.2 (Linear attachment). *Suppose $f(i) = i + \beta$ for $i \geq 1$, for some $\beta \geq 0$, and $m_i \equiv m \geq 1$. There exist positive constants C_1, C'_1, C_2 , depending on m, β , such that for any $\varepsilon \in (0, 1)$,*

$$\frac{C'_1}{\varepsilon^{\frac{2m+\beta}{m(m+\beta)}}} \leq K_\Psi(\varepsilon) \leq \frac{C_1}{\varepsilon^{\frac{2m+\beta}{m(m+\beta)}}} \exp\left(\sqrt{C_2 \log \frac{1}{\varepsilon}}\right).$$

The next result considers the persistent regime in generality. Part (i) gives near optimal bounds for $K_\Psi(\varepsilon)$ in the tree case in terms of the Malthusian rate λ^* and $f_* := \min_{i \geq m_1} f(i)$. Part (ii) gives upper and lower bounds for $K_\Psi(\varepsilon)$ in the non-tree case that are polynomial in ε^{-1} .

Theorem 2.3. *Suppose the attachment function f satisfies Assumptions (A1)-(A3). Moreover, assume $\Phi_2(\infty) < \infty$.*

(i) *Suppose $m_i \equiv m = 1$. For any fixed $\delta \in (0, 1)$, there exist positive constants C_1 (not depending on δ) and C_δ (depending on δ) such that for all $\varepsilon \in (0, 1)$,*

$$\frac{C_1}{\varepsilon^{\frac{\lambda^*}{f_*}}} \leq K_\Psi(\varepsilon) \leq \frac{C_\delta}{\varepsilon^{(1-\delta)\frac{\lambda^*}{f_*}}}.$$

(ii) *Suppose $m_i \equiv m > 1$. For any $\delta \in (0, 1)$, there exist positive constants C_1 (not depending on δ) and \bar{C}_δ (depending on δ) such that for all $\varepsilon \in (0, 1)$,*

$$\frac{C_1}{\varepsilon^{\frac{f_*}{m f(m)}}} \leq K_\Psi(\varepsilon) \leq \frac{\bar{C}_\delta}{\varepsilon^{(1-\delta)\frac{f_*}{m f(m)}}}, \quad \varepsilon \in (0, 1).$$

Remark 2.4. Several remarks are in order.

- Note that if $f(k) \leq g(k)$ for all $k \geq 1$, then $\lambda^*(f) \leq \lambda^*(g)$. Using this and the known Malthusian rates for uniform attachment and linear preferential attachment, we conclude $f_* \leq \lambda^* \leq 2C_f$. From this observation, it follows that the bounds in (i) are indeed tighter than those in (ii) (applied to $m = 1$).
- In [BB22], polynomial upper and lower bounds in ε^{-1} are derived for $K_\Psi(\varepsilon)$ when Ψ is the Jordan centrality measure. However, this measure is only applicable in the tree case and is computationally expensive as the score for each vertex requires information on the size of all the subtrees rooted at that vertex. The degree centrality measure works on non-tree networks and its computation only requires the degree of each vertex. Moreover, the exponent of ε^{-1} in the upper bound on $K_\Psi(\varepsilon)$ for the Jordan centrality measure obtained in [BB22, Theorem 3.1 (a)] is $2C_f/f_*$ (taking $\beta = 0$ there), whereas the corresponding exponent in (i) above is $\lambda^*/(1-\delta)f_*$ (for arbitrarily small $\delta > 0$). As $\lambda^*/f_* \leq 2C_f/f_*$ (see (a)), we conclude that, in the tree case under the persistent regime, the obtained upper bound on the size of the confidence set for the root using degree centrality is comparable to, and in most cases tighter than, that in [BB22] using Jordan centrality.
- In the tree case, when f is linear, from Theorem 2.2 with $m = 1$ and [BB22, Corollary 3.3 (c)], the exact exponent of ε^{-1} in the order of $K_\Psi(\varepsilon)$ in both degree and Jordan centrality measures is seen to be $(2 + \beta)/(1 + \beta)$. In this sense, degree centrality does just as good as Jordan centrality on trees for linear f , although it is computationally far more efficient.

- (d) The exponent of ε^{-1} in the upper bound in Theorem 2.3 (ii) equals that in Theorem 2.2 (taking $\delta \sim 0$) in the linear case with $\beta = 0$. As Theorem 2.2 captures the exact exponent, the upper bound in Theorem 2.3 (ii), in this sense, is the best one can obtain. In this case, the lower bound exponent is clearly not tight, and it is unclear if there are examples of f for which the lower bound is tight in the non-tree case. The technical difficulty in getting sharp bounds in this set-up is that the underlying network does not admit an embedding into a ‘true’ continuous time branching process (see Section 3) which renders many of the tools we use to obtain refined asymptotics in the tree case inapplicable.
- (e) The exponent of ε^{-1} in the bounds obtained in Theorem 2.2 is seen to decrease as m grows. This surprising observation implies that, for large m , although the network becomes increasingly non-tree like, the efficacy of root detection using degree centrality increases. This can be intuitively explained by the stronger positive reinforcement of degrees as m grows: each new vertex comes in with more edges which tend to increase the degrees of high degree vertices even more.
- (f) The constants appearing in the above theorems implicitly depend on m . It will be interesting to make this dependence more explicit. It will also be interesting to extend the above results to non-constant and possibly random attachment sequences $\{m_i\}$. We will address this in subsequent work.

2.2 The non-persistent regime

Consider the non-persistent regime where $\Phi_2(\infty) = \infty$ and suppose $f(k) \rightarrow \infty$ as $k \rightarrow \infty$. In this regime, under some mild assumptions on f (see 1, 2 in Theorem 2.5), the identity of the maximal degree vertex changes infinitely often as the network grows [BB21]. Moreover, for any fixed collection of vertices, the probability of the maximal degree vertex lying in this collection goes to zero as the network grows (see (2.2)). Thus, there is no hope of obtaining a confidence set for the root solely based on empirical degree information that is stable in the network size. On the other hand, [BB22] showed that a stable confidence set can be obtained for tree networks, but this requires connectivity information for the whole network via Jordan centrality and hence is computationally expensive on large networks. In this section, we exhibit how the connectivity information in a neighborhood around the maximal degree vertex of radius much smaller than the network diameter can be used to furnish confidence sets of size much smaller than the network size.

In the tree case, [BB21, Theorem 4.12] sheds light on the asymptotics of the age of the maximal degree vertex as a function of network size. More precisely, define $\mathcal{I}_n^* := \inf\{1 \leq i \leq n : d_i(n) \geq d_j(n) \forall j \leq n\}$ for $n \geq 1$. Then, [BB21, Theorem 4.12] shows that

$$\frac{\log \mathcal{I}_n^*}{\mathcal{K}(\frac{1}{\lambda^*} \log n)} \xrightarrow{P} \frac{(\lambda^*)^2}{2}, \text{ as } n \rightarrow \infty. \quad (2.2)$$

Thus, if a comparison is obtained between the age of the maximal degree vertex and its distance from the root, that is, $\rho_n := \text{dist}(v_{\mathcal{I}_n^*}, v_1)$, then the above result can be used to obtain the asymptotics of ρ_n . Then, one can expect to ‘find the root’ inside a ball around the maximal degree vertex of radius comparable to ρ_n . This is the idea behind the following theorem. In Theorem 2.5, r_n can be thought of as ρ_n and b_n represents the approximate number of vertices in the ball of radius r_n around the maximal degree vertex.

For a sequence $\{r_n\}_{n \geq 1} \subset \mathbb{N}$, we define $B_n(v, r_n)$ to be the set of all vertices in \mathcal{G}_n which are at graph distance at most r_n from v . Let $|B_n(v, r_n)|$ denote the number of vertices in $B_n(v, r_n)$.

Theorem 2.5. Suppose $m_i \equiv m = 1$. Suppose $f(i) \leq C_0 i^\alpha$ for $i \geq 1$, for some $C_0 > 0$ and $\alpha \in (0, 1/2]$. Moreover, assume that $f(k) \rightarrow \infty$ as $k \rightarrow \infty$. Also assume that $\mathcal{K}(\cdot) := \Phi_2 \circ \Phi_1^{-1}(\cdot)$ satisfies the following ‘continuity’ assumptions:

1. $\lim_{\delta \downarrow 0} \limsup_{t \rightarrow \infty} \frac{\mathcal{K}((1+\delta)t)}{\mathcal{K}(t)} = 1$.
2. There exists positive constants t', D such that $\mathcal{K}(3t) \leq D\mathcal{K}(t)$ for all $t \geq t'$.

Define

$$r_n := c_1 \lambda^* \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right),$$

$$b_n := \exp\left(\frac{8}{1-\alpha} r_n \log\left(\frac{\log n}{2\lambda^* r_n}\right)\right),$$

where c_1 is a positive constant (independent of n) defined in Lemma 3.12. There exist $C, C' > 0$ (independent of n) such that for any $\varepsilon \in (0, 1)$,

$$P(v_0 \in B_n(v_{\max}(n), r_n), |B_n(v_{\max}(n), r_n)| \leq b_n) \geq 1 - \varepsilon \text{ for all } n \geq \exp\left(\lambda^* \mathcal{K}^{-1}\left(\frac{\log(C/\varepsilon)}{C'}\right)\right).$$

In words, the above theorem says that, with high probability, the root lies within a neighborhood of radius r_n around the maximal degree vertex and the size of this (confidence) set is at most b_n . For arbitrary $\varepsilon > 0$, it also gives a lower bound on the network size for the root to lie in this set with probability at least $1 - \varepsilon$. This lower bound is especially useful in the non-persistent regime as the size of the confidence set grows with network size and this makes the result applicable for fixed large n .

Remark 2.6. We make the following comments.

(a) To compute the above confidence set for the root, one only needs to know the maximal degree vertex (which can be obtained from the empirical degree distribution) and the network in a neighborhood of radius r_n around this vertex. Under the hypotheses of Theorem 2.5, $r_n / \log n \rightarrow 0$ as $n \rightarrow \infty$, by (3.23). As the diameter of \mathcal{G}_n (same order as the height) is typically $O(\log n)$ [HJ16, Theorem 13.2], the radius of this neighborhood is much smaller than the network diameter.

(b) For any $\alpha \in (0, 1/2)$, taking $f(k) = k^\alpha$ with $k \geq 1$, it is easy to verify that $\mathcal{K}(t) = O(t^{(1-2\alpha)/(1-\alpha)})$. Thus, $r_n = O((\log n)^{(1-2\alpha)/(1-\alpha)})$. Moreover, b_n , which can be seen as the size of the confidence set which contains the root with high probability, is of order

$$\exp\left(O\left((\log n)^{(1-2\alpha)/(1-\alpha)} \log \log n\right)\right),$$

which grows smaller than any positive power of n .

For $\alpha = 1/2$, $\mathcal{K}(t) = O(\log t)$ and thus $r_n = O(\log \log n)$, $b_n = \exp(O((\log n) \log \log n))$. Further, as α approaches 0, r_n gets closer to the diameter $\log n$ whereas, for any $\alpha > 1/2$, $r_n = O(1)$ (note that r_n is well-defined for any $\alpha \in [0, 1]$). Thus, Theorem 2.5 ‘interpolates’ between the persistent regime and the uniform attachment model (which can be seen as the ‘worst case’ for root detection using empirical degree distribution and local information).

(c) A natural question that arises is whether an appropriately constructed ‘purely local’ centrality measure Ψ (the Ψ -value of each vertex requires knowledge of the network only in an $O(1)$ radius) can be shown to be persistent in the (degree) non-persistent regime. This, in turn, would lead to a stable budget $K_\Psi(\varepsilon)$ that does not depend on network size.

Preliminary calculations show that natural guesses like the ‘ j -hop degree centrality measure’, defined as the sum of degrees of all vertices in a neighborhood of radius j around a vertex, will not work for any fixed $j \geq 1$. To see this in the setup of (b) above

for $j = 1$, one can study the evolution of the sum of degrees in a 1-neighborhood of a vertex through the continuous time processes

$$\xi_A^{(2)}(t) = 2\xi_A(t) + \sum_{i=1}^{\xi_A(t)} \xi_{0,i}(t - \sigma_{A,i}), \quad t \geq 0, \quad A \in \mathbb{N}_0,$$

where ξ_A is defined after (1.3), $\sigma_{A,i}$ are the event times in ξ_A , and $\xi_{0,i}$ are i.i.d. copies of ξ defined in (1.3). Suppose at some time T , a new vertex v enters the network when the root has a large degree A . As $\Omega(A)$ of its neighbors have comparable $\Omega(A)$ degrees, with high probability, the sum of degrees in the 1-neighborhood around the root at time T is $\geq \eta A^2$, for some small $\eta > 0$. Then, $1 + \xi_0^{(2)}(t)$ and $\eta A^2 + \xi_A^{(2)}(t)$ respectively denote the sum of degrees in the 1-neighborhood of v and a lower bound on the corresponding quantity for the root, at time $t + T$. The techniques in [BB21] can be used to observe that persistence requires, for sufficiently large A , with high probability, $\eta A^2 + \xi_A^{(2)}(t) > 1 + \xi_0^{(2)}(t)$ for all $t \geq 0$, for sufficiently small $\eta > 0$. It can be shown (using exact computation or the martingales defined in Section 4) that, for any $A \in \mathbb{N}$, $E\left(\xi_A^{(2)}(t) - \xi_0^{(2)}(t)\right) = O(t^{(1+\alpha)/(1-\alpha)})$ whereas $\text{Var}\left(\xi_A^{(2)}(t) - \xi_0^{(2)}(t)\right) \geq Ct^{3/(1-\alpha)}$ for large t , for some $C > 0$. This indicates that, for large t , the fluctuations of $\eta A^2 + \xi_A^{(2)}(\cdot) - 1 - \xi_0^{(2)}(\cdot)$ are of a higher order than its drift, and hence the associated sum of degrees processes cross each other infinitely often, implying lack of persistence.

We will investigate this question in broader generality in a future article.

Remark 2.7 (Degree evolution of fixed vertices and maximum degree asymptotics). The approach of [BB21] and the current article can, in addition, be used to furnish, for any attachment function satisfying Assumptions (A1)-(A3), asymptotics of the degrees of the first k vertices of the network for any finite k , and the maximum degree, as the network grows. In particular, it follows from [BB21] (see Theorem 4.9 and equation (8.23) there) that in the persistent regime ($\Phi_2(\infty) < \infty$), with $m_i \equiv 1$, the degree $d_k(n)$ of the k -th vertex when the network has $n \geq k$ edges satisfies

$$d_k(n) = \Phi_1^{-1}\left(\frac{1}{\lambda^*} \log n + X_k(n)\right)$$

where $X_k(n)$ converges almost surely to some finite random variable $X_k(\infty)$ as $n \rightarrow \infty$. A similar statement holds for the maximum degree $d_{\max}(n)$ as described in [BB21, Theorem 4.9]. For linear f , the non-tree network, and asymptotics on it, can be obtained from the tree network by a collapsing procedure [GvdH18, BDOC23]. In this case, Φ_1 is logarithmic and this implies that $n^{-1/\lambda^*} d_k(n)$ and $n^{-1/\lambda^*} d_{\max}(n)$ converge to non-degenerate random variables, as was noted in detail in several previous works (eg. [PRR17, S en21]). With some more work, it should be possible to extend the above convergence to joint convergence for any finite collection of vertices, for any attachment function satisfying Assumptions (A1)-(A3). For general f , characterizing the joint distribution of $\{X_k(\infty) : k \in \mathbb{N}_0\}$, the limiting rescaled maximum degree, and obtaining results in the non-tree case, are interesting open questions.

In the tree case for the non-persistent regime ($\Phi_2(\infty) = \infty$), when $\mathcal{K}(\cdot)$ satisfies a ‘slow-variation’ type assumption, namely $\lim_{\delta \downarrow 0} \limsup_{t \rightarrow \infty} \frac{\mathcal{K}((1+\delta)t)}{\mathcal{K}(t)} = 1$, there is a stark difference between the degree growth of fixed vertices and the maximum degree as the network size increases. From [BB21, Lemma 7.9], and Lemma 3.2 and (3.5) below, it can be shown that for any fixed k ,

$$\frac{\Phi_1(d_k(n)) - \frac{1}{\lambda^*} \log n}{\sqrt{\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)}} \Rightarrow N(0, 1) \quad \text{as } n \rightarrow \infty.$$

However, from [BB21, Theorem 4.12],

$$\frac{\Phi_1(d_{\max}(n)) - \frac{1}{\lambda^*} \log n}{\mathcal{K}(\frac{1}{\lambda^*} \log n)} \xrightarrow{P} \frac{\lambda^*}{2} \text{ as } n \rightarrow \infty.$$

We hope to explore these questions in more detail in future work.

2.3 Outline of proof techniques

We end this section with an outline of techniques used to prove our main results.

1. In Section 3, we discuss embeddings of our discrete time network process into continuous time branching processes where individuals reproduce according to point processes with inter-arrival times having means determined by the attachment function. Among other things, a key technical advantage of such embeddings is that, informally speaking, they allow for a ‘time-changed’ interpretation of the process which enables us to bypass the random denominator appearing in the network attachment probabilities (1.1). In the tree case, this embedding is standard and goes by the name of Athreya–Karlin embedding. For the non-tree case, we introduce a novel embedding into a collapsed branching process (CBP) formed by collapsing batches of individuals into single vertices, appropriately preserving the connectivity structure. Such CBPs have appeared recently [GvdH18, BDOC23], but our CBP differs from them and allows for an exact embedding of the network for all attachment functions (as opposed to just the linear and uniform case in [GvdH18, BDOC23]). Moreover, this embedding lets us bypass the monotonicity assumptions on f in the non-tree case required in several prior works (eg. [BB21]). For these branching processes, we discuss refined quantitative estimates on hitting times and distributional properties of limits of renormalized population sizes (Section 3.1), asymptotics of the age of maximal degree vertices (Section 3.2) and height of the skeleton of the branching process (Section 3.3). Although we build on several prior works (discussed in respective sections), the obtained estimates are mostly new, are of independent interest, and form the backbone of our proofs.
2. In Section 4, we recall some large deviations properties of a class of martingales (defined in (4.1)) from [BB21] which have proved to be a crucial tool in analyzing the degree evolution of fixed vertices and competition between their degrees as the network grows.
3. In Section 5, a universal strategy to obtain budget lower bounds is described where the probability of the root being in the confidence set is lower bounded in terms of an event involving the root having an ‘unusually low’ degree in the network (see Lemma 5.1).
4. The proofs of Theorems 2.2 and 2.3 are completed in Sections 6 and 7. The strategy discussed above, along with the obtained branching process estimates, is used in obtaining budget lower bounds. Obtaining budget upper bounds involves controlling the probabilities of two events: A_n (see (6.1)) being the event that the root degree grows slower than a certain rate in the network size, and $E_n \cap A_n$, where E_n (see (6.2)) is the event where the degree of the root is eventually exceeded by that of the $(n + 1)$ -th connecting vertex as the network grows. The former is bounded using branching process estimates and the later is bounded using the martingale large deviation results from Section 4.
5. Finally, in Section 8, Theorem 2.5 is proved by combining the estimates on hitting times, age of the maximal degree vertex and height of the branching process with Laplace transform asymptotics of the associated reproduction point process.

3 Branching process embedding

Branching process embedding in the case $m_i \equiv 1$. A crucial technique to our proofs is an appropriate embedding of the discrete dynamics described in Section 1.1 into a continuous time branching process.

Definition 3.1. A continuous time branching process (CTBP) driven by f , written as $\{BP_f(t) : t \geq 0\}$, is defined to be a branching process started with one individual at time $t = 0$ and such that every individual in this system has an offspring distribution that is an independent copy of the point process ξ defined in (1.3).

We will denote the number of individuals (size) of the branching process $BP_f(t)$ at time t by $|BP_f(t)|$.

Let $BP_f^{(0)}(\cdot), BP_f^{(1)}(\cdot)$ be two independent branching processes driven by ξ as in Definition 3.1 started from v_0 and v_1 respectively. Define $\widetilde{BP}_f(t) := BP_f^{(0)}(t) \cup BP_f^{(1)}(t)$.

For $n \geq 2$, we label the vertices in $\widetilde{BP}_f(t)$ in the order of their arrivals into the system. For $i = 0, 1$, $BP_f^{(i)}(t)$ can be viewed as a labelled tree rooted at v_i with edges between parents and their offspring. Upon adding an edge between v_0 and v_1 , $\widetilde{BP}_f(t)$ becomes a tree, and we assign its root to be v_0 . Define the sequence of stopping times

$$T_l := \inf\{t \geq 0 : |\widetilde{BP}_f(t)| = l + 1\}, \quad l \geq 1, \quad (3.1)$$

that is, the arrival time of v_l into the system. Note that $T_1 = 0$. The following property follows from properties of Exponential distributions, and is the starting point of the Athreya–Karlin embedding [AK68].

Lemma 3.2. Fix the attachment function f . Let $\{\mathcal{G}_l : l \geq 1\}$ be the sequence of random trees constructed in Section 1.1 using f as the attachment function and with $m_i \equiv 1$, and let $\{\widetilde{BP}_f(t) : t \geq 0\}$ be the CTBP constructed as above. Then we have

$$\{\widetilde{BP}_f(T_l), l \geq 1\} \stackrel{d}{=} \{\mathcal{G}_l : l \geq 1\}.$$

Collapsed branching process embedding in the case $m_i \equiv m > 1$. The embedding into a continuous time branching process does not carry over directly in the case $m_i \equiv m > 1$. Nevertheless, we propose the following ‘continuous time interpolation’ between \mathcal{G}_l and \mathcal{G}_{l+1} using an adaptive collapsing operation on a continuous time branching process.

- Define the discrete time graph-valued process $\{CBP_f(l) : l \in \mathbb{N}\}$ as follows. $CBP_f(1)$ is the graph \mathcal{G}_1 with two vertices v_0, v_1 connected by m edges. Define $\tau_0 = \tau_1 = 0$.
- Suppose for $l \geq 1$, we have constructed the random time τ_l and the process $\{CBP_f(j) : 1 \leq j \leq l\}$. To obtain $CBP_f(l+1)$, each individual $v \in CBP_f(l)$, say with degree $d(v, l)$ in $CBP_f(l)$, independently reproduces according to the point process

$$\tilde{\xi}^{d(v,l)}(\cdot) \stackrel{d}{=} \xi_{d(v,l)-1}(t)$$

where, for any $k \in \mathbb{N}$, $\xi_{k-1}(t)$ is as defined in the line below (1.3) with the attachment function $f^{(v)}(\cdot) := f(\cdot + k - 1)$.

The newly born individuals do not reproduce until there are m newborns since τ_l . Define Θ_{l+1} to be the first time since τ_l when there are m newly born individuals $v_{l+1,1}, \dots, v_{l+1,m}$ and set $\tau_{l+1} := \tau_l + \Theta_{l+1}$. Then we collapse the vertices $v_{l+1,1}, \dots, v_{l+1,m}$ to a single vertex v_{l+1} , and define $CBP_f(l+1)$ to be the graph obtained after this collapsing operation. Note that $d(v_{l+1}, l+1) = m$. For any

$j \in \mathbb{N}_0$, define the following process that represents the ‘continuous time’ degree of v_j :

$$\xi^{m,(v_j)}(t) := \sum_{l=j}^{\infty} \tilde{\xi}^{d(v_j,l)}((t \wedge \tau_{l+1}) - \tau_l), \quad t \geq \tau_j, \tag{3.2}$$

where we take $\tilde{\xi}^{d(v_j,l)}(s) = 0$ for $s \leq 0$. Note that $d(v_j, l) = m + \xi^{m,(v_j)}(\tau_l)$ for any $j \in \mathbb{N}_0, l \geq j$.

The following lemma is easy to check using properties of Exponential distributions.

Lemma 3.3. *Fix an attachment function f and an integer $m > 1$. Let $\{\mathcal{G}_l : l \geq 1\}$ be the sequence of random trees constructed in Section 1.1 using f as the attachment function and with $m_i \equiv m$, and let $\{CBP_f(l) : l \in \mathbb{N}\}$ be the random graph process constructed as above. Then we have*

$$\{CBP_f(l), l \geq 1\} \stackrel{d}{=} \{\mathcal{G}_l : l \geq 1\}.$$

Moreover, the processes $\xi^{m,(v_j)}(\cdot + \tau_j), j \in \mathbb{N}_0$, are independent and have the same distribution as the point process defined in (1.3) with attachment function $f^{(m)}(k) := f(k + m - 1)$ for $k \in \mathbb{N}$, that is, $\xi^{m,(v_j)}(t + \tau_j) \stackrel{d}{=} \xi_{m-1}(t), t \geq 0$.

The above construction is reminiscent of the collapsed continuous-time branching processes defined in [GvdH18] (see also [BDOC23]). However, the two constructions are different as the newly born vertices in (τ_l, τ_{l+1}) are not allowed to reproduce in the above construction, while they are in [GvdH18].

Simple birth process with immigration. We describe one of the simple examples of continuous time branching processes which will be useful later, the simple birth process with immigration. Let $\{Y_{\nu,\beta}(t) : t \geq 0\}$ denote the simple birth process with birth rate $\nu > 0$ and immigration rate $\beta > 0$. Given the current population size $Y_{\nu,\beta}(t)$, the average instantaneous rate of growth is $\nu Y_{\nu,\beta}(t) + \beta$. More precisely,

$$\begin{aligned} P(Y_{\nu,\beta}(t + dt) - Y_{\nu,\beta}(t) = 1 | \mathcal{F}(t)) &= (\nu Y_{\nu,\beta}(t) + \beta)dt + o(dt), \\ P(Y_{\nu,\beta}(t + dt) - Y_{\nu,\beta}(t) \geq 2 | \mathcal{F}(t)) &= o(dt), \end{aligned}$$

where $\{\mathcal{F}(t) : t \geq 0\}$ is the natural filtration of the process.

Set $Y_{\nu,\beta}(0) = 1$ and define $M(\theta, t) = \mathbb{E}(e^{\theta Y_{\nu,\beta}(t)})$ be the associated moment generating function. Then $M(\theta, t)$ satisfies

$$\frac{\partial M}{\partial t} = \nu(e^\theta - 1) \frac{\partial M}{\partial \theta} + \beta(e^\theta - 1)M,$$

with initial condition $M(\theta, 0) = e^\theta$ (see, e.g. [All10, Section 6.4.4]). The solution to the above differential equation is

$$M(\theta, t) = \frac{e^\theta}{((1 - e^\theta)e^{\nu t} + e^\theta)^{1+\beta/\nu}}, \tag{3.3}$$

which is valid for $\theta < \log(e^{\nu t}/(e^{\nu t} - 1))$.

3.1 Asymptotics for population size and associated hitting times for CTBP

In this section we will collect some properties of the branching process introduced in Definition 3.1 that will be used later. Throughout this section, we will work under the assumptions (A1)-(A3) for f . The following theorem was recorded in [BB22] using classical results of [JN84, BG79].

Theorem 3.4 (Theorem 3.5 in [BB22]). *Suppose the attachment function f satisfies Assumptions (A1)-(A3). There exists a strictly positive random variable W_∞ such that, with λ^* being the Malthusian rate of growth parameter in (1.5),*

$$e^{-\lambda^* t} |BP_f(t)| \xrightarrow{a.s., L^2} W_\infty, \quad \text{as } t \rightarrow \infty. \tag{3.4}$$

Furthermore, W_∞ is an absolutely continuous or singular continuous random variable (with respect to the Lebesgue measure) and is supported on all of \mathbb{R}_+ .

Taking $t = T_n$, where T_n is defined in (3.1), the above theorem implies that

$$T_n - \frac{1}{\lambda^*} \log n \xrightarrow{a.s.} -\frac{1}{\lambda^*} \log W_\infty \quad \text{as } n \rightarrow \infty. \tag{3.5}$$

The following lemma gives estimates for the tail behavior of the distribution of W_∞ near zero.

Lemma 3.5. *Let W_∞ be as in Theorem 3.4. For any $a \in [0, f(1)/\lambda^*]$, there exists $C_a > 0$ such that*

$$P(W_\infty \leq x) \leq C_a x^a, \quad x > 0.$$

The proof of the above lemma relies on the crucial fact that the random variable W_∞ satisfies the following recursive distributional equation:

$$W_\infty \stackrel{d}{=} \sum_{i=1}^{\infty} e^{-\lambda^* \sigma_i} W_{\infty,i}, \tag{3.6}$$

where $\{\sigma_i\}_{i \geq 1}$ are the arrival times of the point process ξ defined in (1.3), $\{W_{\infty,i}\}_{i \geq 1}$ are i.i.d copies of W_∞ that are independent of $\{e^{-\lambda^* \sigma_i}\}_{i \geq 1}$. In particular, the moment generating function $\phi(t) = Ee^{-tW_\infty}$, $t \geq 0$, of W_∞ satisfies

$$\phi(t) = E \prod_{i \geq 1} \phi(A_i t) \tag{3.7}$$

where $A_i = e^{-\lambda^* \sigma_i}$. These recursive identities follow directly from Theorem 3.4 and the dynamics of the branching process BP_f . The proof of Lemma 3.5 will require the following lemma from [Liu01].

Lemma 3.6 ([Liu01]). *Let $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a bounded function and let A be a positive random variable such that for some $0 < p < 1$, $t_0 \geq 0$ and all $t > t_0$,*

$$\phi(t) \leq pE\phi(At).$$

If $pE(A^{-a}) < 1$ for some $0 < a < \infty$, then $\phi(t) = O(t^{-a})(t \rightarrow \infty)$ and $\int_0^\infty \phi(t)t^{a-1}dt < \infty$.

Proof of Lemma 3.5. Since f satisfies Assumption (A2) and $f(k) \rightarrow \infty$ as $k \rightarrow \infty$ by Assumption (A1), Lemma 1 of [RTV07] gives

$$E(\hat{\xi}(\lambda^*) \log^+ \hat{\xi}(\lambda^*)) < \infty,$$

where $\hat{\xi}(\lambda^*) := \int_0^\infty e^{-\lambda^* t} \xi(dt)$. Hence, by Proposition 1.1 of [Ner81], $P(W_\infty > 0) = 1$. Recalling the moment generating function ϕ of W_∞ , we obtain $\lim_{t \rightarrow \infty} \phi(t) = 0$. In particular, there exists $t'_0 > 0$ such that for all $t > t'_0$, $\phi(t) < \frac{1}{2}$. Let $\delta \in (0, 1)$ be arbitrarily fixed and define $N_\delta := \sum_{i=1}^\infty 1_{\{A_i > \delta\}}$, where we recall $A_i = e^{-\lambda^* \sigma_i}$. Note that $P(N_\delta < \infty) = 1$, since $\lim_{i \rightarrow \infty} \sigma_i = \infty$ almost surely (which follows from the fact that, by the at-most-linear growth of f required by Assumption (A1), $\sum_{i=1}^\infty \frac{1}{f(i)} = \infty$).

By the functional equation (3.7) and the fact that $\phi(A_it) \leq 1$ for all t , we see that when $t > t_0/\delta$,

$$\phi(t) \leq E[\phi(A_1t)2^{-N_\delta+1}] = p_\delta E\phi(\tilde{A}_1t)$$

where $p_\delta = E[2^{-N_\delta+1}]$ and \tilde{A}_1 is a positive random variable whose distribution is determined by

$$Eg(\tilde{A}_1) = \frac{1}{p_\delta} E[g(A_1)2^{-N_\delta+1}]$$

for any bounded and measurable function g . Note that for $EA_1^{-a} = Ee^{\lambda^* a \sigma_1} < \infty$ for any $a \in [0, f(1)/\lambda^*)$. Choose and fix any such a . We can write

$$p_\delta E\tilde{A}_1^{-a} = E[A_1^{-a}2^{-N_\delta+1}].$$

Moreover, almost surely, $N_\delta \rightarrow \infty$ as $\delta \rightarrow 0$. Hence, by the dominated convergence theorem,

$$\limsup_{\delta \rightarrow 0} E[A_1^{-a}2^{-N_\delta+1}] = 0.$$

Hence, choosing $\delta \in (0, 1)$ small enough, we have $p_\delta E\tilde{A}_1^{-a} < 1$. Applying Lemma 3.2, we conclude that for any $a \in [0, f(1)/\lambda^*)$, there exists $C'_a > 0$ such that $\phi(t) \leq C'_a t^{-a}$ for any $t > 0$. It follows that for any $x > 0$,

$$P(W_\infty \leq x) \leq \inf_{\theta > 0} e^{\theta x} \phi(\theta) \leq C'_a \inf_{\theta > 0} \exp(\theta x - a \log \theta).$$

Optimizing the above bound in θ leads to the bound claimed in the lemma. □

Recall $\{T_n\}_{n \geq 1}$ from (3.1). The following lemma gives tail estimates for the distribution of T_n .

Lemma 3.7. *Let $\{\gamma_n\}_{n \geq 1}$ be a non-decreasing sequence such that $\lim_{n \rightarrow \infty} \gamma_n = \infty$.*

(i) *Suppose in addition $\limsup_{n \rightarrow \infty} \frac{\gamma_n}{\log n} = 0$. For any $a \in [0, f(1)/\lambda^*)$, there exists $\tilde{C}_a > 0$ such that for $n \geq 3$,*

$$P(T_n \geq \frac{1}{\lambda^*} \log n + \gamma_n) \leq C e^{-a \lambda^* \gamma_n}. \tag{3.8}$$

(ii) *For any $k \in \mathbb{N}$, there exists some $C_k > 0$ so that for $n \geq 1$,*

$$P(T_n < \frac{1}{\lambda^*} \log n - \gamma_n) \leq C_k e^{-k \lambda^* \gamma_n}. \tag{3.9}$$

Proof. We begin with the proof of (3.8). To simplify notation let $\sigma'_n = \frac{1}{\lambda^*} \log n + \gamma_n$. First note that

$$\begin{aligned} P(T_n \geq \sigma'_n) &= P(\widetilde{BP}_f(\sigma'_n) \leq n + 1) = P(BP_f^{(0)}(\sigma'_n) + BP_f^{(1)}(\sigma'_n) \leq n + 1) \\ &\leq 2P(BP_f(\sigma'_n) \leq (n + 1)/2) = 2P\left(e^{-\lambda^* \sigma'_n} BP_f(\sigma'_n) \leq e^{-\lambda^* \sigma'_n} (n + 1)/2\right), \end{aligned}$$

where the second line follows from the fact that $BP_f^{(0)}(\cdot)$ and $BP_f^{(1)}(\cdot)$ are i.i.d. copies of $BP_f(\cdot)$. Let $x_n = e^{-\lambda^* \sigma'_n} (n + 1)/2$. According to Theorem 3.7 in [BB22] there exist finite positive constants C_1, C_2, δ (depending on f) such that for any $t \geq 0$,

$$P\left(\sup_{s \in [t, \infty)} |e^{-\lambda^* s} BP_f(s) - W_\infty| > e^{-\delta t}\right) \leq C_1 e^{-C_2 t}. \tag{3.10}$$

Hence,

$$P\left(e^{-\lambda^* \sigma'_n} BP_f(\sigma'_n) \leq x_n\right) \leq P(|e^{-\lambda^* \sigma'_n} BP_f(\sigma'_n) - W_\infty| > e^{-\delta \sigma'_n}) + P(W_\infty \leq x_n + e^{-\delta \sigma'_n}).$$

By Lemma 3.5 for any $a \in [0, f(1)/\lambda^*)$, there exists $C_a > 0$ such that for $n \geq 3$,

$$P(W_\infty \leq x_n + e^{-\delta \sigma'_n}) \leq C_a (x_n + e^{-\delta \sigma'_n})^a. \tag{3.11}$$

When n is sufficiently large, $e^{-\delta \sigma'_n} \ll x_n$ (this follows from the assumption that $\gamma_n = o(\log n)$). Hence, collecting (3.10) and (3.11), there exists $\tilde{C}_a > 0$ so that for $n \geq 3$,

$$P(T_n \geq \frac{1}{\lambda^*} \log n + \gamma_n) \leq \tilde{C}_a e^{-a \lambda^* \gamma_n}.$$

It remains to prove (3.9). Let $\sigma''_n = \frac{1}{\lambda^*} \log n - \gamma_n$. Note that

$$\begin{aligned} P(T_n < \sigma''_n) &= P(BP_f^{(0)}(\sigma''_n) + BP_f^{(1)}(\sigma''_n) \geq n + 1) \leq 2P(BP_f(\sigma''_n) \geq n/2) \\ &= 2P(e^{-\lambda^* \sigma''_n} BP_f(\sigma''_n) \geq e^{-\lambda^* \sigma''_n} \cdot n/2) \\ &= 2P(e^{-\lambda^* \sigma''_n} BP_f(\sigma''_n) \geq e^{\lambda^* \gamma_n} / 2) \\ &\leq 2^{k+1} e^{-k \lambda^* \gamma_n} E \left[\left(\sup_{t \geq 0} e^{-\lambda^* t} |BP_f(t)| \right)^k \right] \end{aligned}$$

for any positive integer k . Recall that $\{\sigma_i : i \geq 1\}$ denote the arrival times of the point process ξ defined in (1.3). By Theorem 4 in [MR19], if for any positive integer k , $E(\sum_{i=1}^\infty e^{-\lambda^* \sigma_i})^k < \infty$ and $E(|BP_f(t)|^k) < \infty$ for every $t \geq 0$, then

$$C_k := 2^{k+1} E \left[\left(\sup_{t \geq 0} e^{-\lambda^* t} |BP_f(t)| \right)^k \right] < \infty.$$

To complete the proof it suffices to check the two conditions hold. By Proposition 5.7 in [BB22], the random variable $Y := \sum_{i=1}^\infty e^{-\lambda^* \sigma_i}$ satisfies $E(e^{\delta Y}) < \infty$ for some $\delta > 0$. In particular, $E(\sum_{i=1}^\infty e^{-\lambda^* \sigma_i})^k < \infty$ for any $k \in \mathbb{N}$. Moreover, by Assumption (A1), $BP_f(t)$ is dominated by a Yule process with a linear attachment function. Hence, by (3.3), it is clear that $E(|BP_f(t)|^k) < \infty$ for every $t \geq 0$. \square

3.2 Asymptotics for age of maximal degree vertex in CTBP

The proof of Theorem 2.5 relies on a precise quantification of the tail probabilities of the age of the (oldest) vertex with maximal degree in a continuous time branching process. This is achieved in this subsection.

For $a \geq 0, i \in \mathbb{N}$, let $B_a^{(i)}$ be the birth time of the i -th individual born at or after time a and let $\xi^{(i)}$ denote the point process of times when this individual reproduces, measured relative to its own birth time (i.e. $\xi^{(i)}(s') = k$ if individual i has k children at time $s' + B_a^{(i)}$). For $s \geq 0$, let $D_a^{(i)}(s) := \Phi_1(\xi^{(i)}(s))$. Again, note that for any $a \geq 0$, $\{D_a^{(i)}(\cdot) : i \in \mathbb{N}\}$ are i.i.d. processes, each having the same distribution as $D_0^{(1)}(\cdot)$. For any $0 \leq a < b$, let $n[a, b]$ denote the number of individuals born to the branching process in the time interval $[a, b]$.

For any $0 \leq a < b \leq c$ with $n[a, b] > 0$, let

$$D_{a,b}^{\max}(c) := \sup\{D_a^{(i)}(c - B_a^{(i)}) : i \in \mathbb{N} \text{ such that } B_a^{(i)} \leq b\} \tag{3.12}$$

denote the maximum of the degrees of vertices born in the time interval $[a, b]$ at time c . We define $D_{a,b}^{\max}(c) = 0$ if $c < b$ or $n[a, b] = 0$. Define

$$\mathcal{I}_c^*(t) := \sup\{B_0^{(i)} : D_0^{(i)}(t - B_0^{(i)}) = D_{0,t}^{\max}(t)\}. \tag{3.13}$$

Notice that this definition is different from that of $\mathcal{I}_c^*(t)$ in [BB21] (there is an inf in place of sup in [BB21]). However, technically, they can be analyzed using similar tools. The following is the main result of this subsection.

Lemma 3.8. *Suppose $\mathcal{K}(\cdot) := \Phi_2 \circ \Phi_1^{-1}(\cdot)$ satisfies the continuity assumptions in Theorem 2.5. For any $\varepsilon > 0$,*

$$P\left(\mathcal{I}_c^*(T_n) > \left(\frac{\lambda^*}{2} + \varepsilon\right) \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right) \leq C \exp\left(-C' \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right), \quad n \geq 3,$$

for some positive constants C, C' .

The proof of Lemma 3.8 relies on quantitative versions of Lemmas 7.12-7.14 in [BB21] described in the following three lemmas.

Lemma 3.9. *Assume that the attachment function f satisfies Assumptions (A1)-(A3). For any $0 \leq a < b$, there exists $\delta_1 > 0$ such that for any $\delta \in (0, \delta_1)$, there exist $C, C' > 0$ that depend on f, δ, a, b such that*

$$P(n[at, bt] \leq e^{(\lambda^* - \delta)bt}) \leq C e^{-C't}, \quad P(n[0, bt] > e^{(\lambda^* + \delta)bt}) \leq C e^{-\delta bt}.$$

Proof. Observe that, for any $\delta > 0$,

$$P(n[at, bt] \leq e^{(\lambda^* - \delta)bt}) \leq P(n[0, at] > e^{(\lambda^* - \delta)bt}) + P(n[0, bt] \leq 2e^{(\lambda^* - \delta)bt}). \quad (3.14)$$

To estimate the first term in the bound (3.14), note that for any $\delta \in (0, \lambda^*(b-a)/(b+a))$,

$$P(n[0, at] > e^{(\lambda^* - \delta)bt}) \leq P(n[0, at] > e^{(\lambda^* + \delta)at}) \leq \sup_{s \geq 0} E\left(e^{-\lambda^* s} n[0, s]\right) e^{-\delta at}, \quad (3.15)$$

where the supremum above is finite by [Ner81, Proposition 2.2] and Assumption (A2).

To estimate the second term in the bound (3.14), recall that by Theorem 3.4, $e^{-\lambda^* bt} n[0, bt]$ converges almost surely to a random variable W_∞ as $t \rightarrow \infty$. Write

$$\begin{aligned} P(n[0, bt] \leq 2e^{(\lambda^* - \delta)bt}) &= P(e^{-\lambda^* bt} n[0, bt] \leq 2e^{-\delta bt}) \\ &\leq P\left(\left|e^{-\lambda^* bt} n[0, bt] - W_\infty\right| > e^{-\delta bt}\right) + P(W_\infty \leq 3e^{-\delta bt}). \end{aligned} \quad (3.16)$$

By Theorem 3.7 in [BB22], there exist finite positive constants C_1, C_2, δ_0 (depending on f) such that for any $t \geq 0$,

$$P\left(\sup_{s \in [t, \infty)} |e^{-\lambda^* s} n[0, s] - W_\infty| > e^{-\delta_0 t}\right) \leq C_1 e^{-C_2 t}.$$

Hence, for any $\delta \in (0, \delta_0/b)$,

$$P\left(\left|e^{-\lambda^* bt} n[0, bt] - W_\infty\right| > e^{-\delta bt}\right) \leq P\left(\sup_{s \in [t, \infty)} |e^{-\lambda^* s} n[0, s] - W_\infty| > e^{-\delta_0 t}\right) \leq C_1 e^{-C_2 t}.$$

Moreover, using Lemma 3.5 with $a = f(1)/(2\lambda^*)$, we obtain $C' > 0$ such that

$$P(W_\infty \leq 3e^{-\delta bt}) \leq C' e^{-f(1)\delta bt/(2\lambda^*)}.$$

Using the above two bounds in (3.16),

$$P(n[0, bt] \leq 2e^{(\lambda^* - \delta)bt}) \leq C_1 e^{-C_2 t} + C' e^{-f(1)\delta bt/(2\lambda^*)}. \quad (3.17)$$

The first bound in the lemma, with $\delta_1 := [\lambda^*(b-a)/(b+a)] \wedge [\delta_0/b]$, now follows upon using (3.15) and (3.17) in (3.14). The second bound follows similarly as (3.15). \square

Recall $\mathcal{K}(t) := \Phi_2 \circ \Phi_1^{-1}(t), t \geq 0$.

Lemma 3.10. Assume $\Phi_2(\infty) = \infty$ and the attachment function f satisfies Assumption (A1). Also assume that

$$\lim_{\delta \downarrow 0} \limsup_{t \rightarrow \infty} \frac{\mathcal{K}((1 + \delta)t)}{\mathcal{K}(t)} = 1. \tag{3.18}$$

Then for any $0 \leq a < b$ and any $\eta \in (0, \sqrt{2\lambda^*b}/2)$, there exists $C, C' > 0$ (depending on a, b, η) such that

$$\begin{aligned} P\left(D_{a\mathcal{K}(t), b\mathcal{K}(t)}^{\max}(t - \eta\mathcal{K}(t)) < t - b\mathcal{K}(t) + (\sqrt{2\lambda^*b} - 2\eta)\mathcal{K}(t)\right) &\leq Ce^{-C'\mathcal{K}(t)}, \\ P\left(D_{a\mathcal{K}(t), b\mathcal{K}(t)}^{\max}(t + \eta\mathcal{K}(t)) > t - a\mathcal{K}(t) + (\sqrt{2\lambda^*b} + 2\eta)\mathcal{K}(t)\right) &\leq Ce^{-C'\mathcal{K}(t)}. \end{aligned} \tag{3.19}$$

Proof. The proof follows from that of Lemma 7.13 in [BB21]. The major difference is that Lemma 3.9 above is used to produce quantitative bounds for the events in (3.19) (in Lemma 7.13 of [BB21], the probabilities were shown to converge to zero without quantitative bounds).

Recall δ_1 from Lemma 3.9 for the given a, b . Take $\varepsilon \in (0, 1)$, $\delta \in (0, \delta_1)$ small enough such that

$$\sqrt{2\lambda^*b} - \eta < \frac{\sqrt{2(\lambda^* - \delta)b}}{(1 + \varepsilon)^2}, \quad \sqrt{2\lambda^*b} + \eta > \frac{\sqrt{2(\lambda^* + \delta)b}}{(1 - \varepsilon)^2}. \tag{3.20}$$

For any $z \geq -\eta/2, t \geq t_2(z)$ for some $t_2 = t_2(z)$ depending on z , estimate (7.52) in [BB21] produces the following estimate for any $i \in \mathbb{N}$:

$$P\left(D_{a\mathcal{K}(t)}^{(i)}(t - \eta\mathcal{K}(t) - b\mathcal{K}(t)) < t - b\mathcal{K}(t) + z\mathcal{K}(t)\right) \leq 1 - \exp\left(-\frac{(1 + \varepsilon)^4(z + \eta)^2\mathcal{K}(t)}{2}\right).$$

Using this and the first bound in Lemma 3.9, we obtain

$$\begin{aligned} &P\left(D_{a\mathcal{K}(t), b\mathcal{K}(t)}^{\max}(t - \eta\mathcal{K}(t)) < t - b\mathcal{K}(t) + z\mathcal{K}(t)\right) \\ &\leq P(n[a\mathcal{K}(t), b\mathcal{K}(t)] \leq e^{(\lambda^* - \delta)b\mathcal{K}(t)}) \\ &\quad + P\left(D_{a\mathcal{K}(t)}^{(i)}(t - \eta\mathcal{K}(t) - b\mathcal{K}(t)) < t - b\mathcal{K}(t) + z\mathcal{K}(t) \text{ for all } i \leq \lfloor e^{(\lambda^* - \delta)b\mathcal{K}(t)} \rfloor + 1\right) \\ &\leq C_1e^{-C_2\mathcal{K}(t)} + \left(1 - \exp\left(-\frac{(1 + \varepsilon)^4(z + \eta)^2\mathcal{K}(t)}{2}\right)\right)^{e^{(\lambda^* - \delta)b\mathcal{K}(t)}} \\ &\leq C_1e^{-C_2\mathcal{K}(t)} + \exp\left(-\exp\left((\lambda^* - \delta)b\mathcal{K}(t) - \frac{(1 + \varepsilon)^4(z + \eta)^2\mathcal{K}(t)}{2}\right)\right). \end{aligned}$$

Note that, by (3.20) and as $\eta \in (0, \sqrt{2\lambda^*b}/2)$, $\frac{\sqrt{2(\lambda^* - \delta)b}}{(1 + \varepsilon)^2} - \eta > \sqrt{2\lambda^*b} - 2\eta > 0$. For any $-\eta/2 \leq z < \frac{\sqrt{2(\lambda^* - \delta)b}}{(1 + \varepsilon)^2} - \eta$ we have

$$-\exp\left((\lambda^* - \delta)b\mathcal{K}(t) - \frac{(1 + \varepsilon)^4(z + \eta)^2\mathcal{K}(t)}{2}\right) \leq -\exp(C_{z,\eta}\mathcal{K}(t))$$

for some positive constant $C_{z,\eta}$ depending only on z, η . Hence, taking $z = \sqrt{2\lambda^*b} - 2\eta$,

$$\begin{aligned} &P\left(D_{a\mathcal{K}(t), b\mathcal{K}(t)}^{\max}(t - \eta\mathcal{K}(t)) < t - b\mathcal{K}(t) + z\mathcal{K}(t)\right) \\ &\leq C_1e^{-C_2\mathcal{K}(t)} + \exp(-\exp(C_{z,\eta}\mathcal{K}(t))) \\ &\leq C_3e^{-C_2\mathcal{K}(t)} \end{aligned}$$

for sufficiently large t , where we have used $\mathcal{K}(t) = \Phi_2 \circ \Phi_1^{-1}(t) \rightarrow \infty$ as $t \rightarrow \infty$ (which is a consequence of $\Phi_1(\infty) = \Phi_2(\infty) = \infty$). Note that the constants C_2, C_3 can depend on b, η . This proves the first bound in (3.19).

The second bound in (3.19) follows similarly upon using the second bound in Lemma 3.9 and following the calculations (7.57)–(7.60) in [BB21]:

$$\begin{aligned} &P\left(D_{a\mathcal{K}(t), b\mathcal{K}(t)}^{\max}(t + \eta\mathcal{K}(t)) > t - a\mathcal{K}(t) + w\mathcal{K}(t)\right) \\ &\leq P(n[0, b\mathcal{K}(t)] > e^{(\lambda^* + \delta)b\mathcal{K}(t)}) + \exp\left((\lambda^* + \delta)b\mathcal{K}(t) - \frac{(w - \eta)^2(1 - \varepsilon)^4\mathcal{K}(t)}{2}\right) \\ &\leq C e^{-\delta b\mathcal{K}(t)} + \exp(-C_{w, \eta}\mathcal{K}(t)) \end{aligned}$$

where $w = \sqrt{2\lambda^*b} + 2\eta$. □

Lemma 3.11. *Assume $\Phi_2(\infty) = \infty$. Moreover, assume that there exists positive constants t', D such that $\mathcal{K}(3t) \leq D\mathcal{K}(t)$ for all $t \geq t'$. Then there exists $A_0 > 0$ such that*

$$P(D_{A_0\mathcal{K}(t), t}^{\max}(t + \mathcal{K}(t)) > t) \leq C e^{-C'\mathcal{K}(t)}$$

for some constants $C, C' > 0$.

Proof. Define the event

$$\mathcal{E}_A := \left\{ \sup_{t \geq 0} e^{-\lambda^*t} n[0, t] \leq A \right\}.$$

By Corollary 3.8 in [BB22],

$$P(\mathcal{E}_A^c) \leq C \log A/A^2.$$

By (7.63) of [BB21], we have some $j_0 \geq 2$ and $t^* > 0$ such that for all $t \geq t^*$ and $j_0 \leq j \leq 1 + (2t/\mathcal{K}(t))$,

$$P(D_{j\mathcal{K}(t), (j+1)\mathcal{K}(t)}^{\max}(t + \mathcal{K}(t)) > t, \mathcal{E}_A) \leq A \exp\left(-\frac{(j-1)^2}{8D}\mathcal{K}(t)\right).$$

By (7.64) of [BB21], we have for any $t \geq t^*$,

$$\begin{aligned} P(D_{j_0\mathcal{K}(t), t}^{\max}(t + \mathcal{K}(t)) > t, \mathcal{E}_A) &\leq \sum_{j=j_0}^{1+\lfloor t/\mathcal{K}(t) \rfloor} P(D_{j\mathcal{K}(t), (j+1)\mathcal{K}(t)}^{\max}(t + \mathcal{K}(t)) > t, \mathcal{E}_A) \\ &\leq \sum_{j=j_0}^{1+\lfloor t/\mathcal{K}(t) \rfloor} A \exp\left(-\frac{(j-1)^2}{8D}\mathcal{K}(t)\right) \leq CA \exp\left(-\frac{(j_0-1)^2}{8D}\mathcal{K}(t)\right), \end{aligned}$$

where C is a positive constant not depending on t, A . Finally,

$$\begin{aligned} P(D_{j_0\mathcal{K}(t), t}^{\max}(t + \mathcal{K}(t)) > t) &\leq P(D_{j_0\mathcal{K}(t), t}^{\max}(t + \mathcal{K}(t)) > t, \mathcal{E}_A) + P(\mathcal{E}_A^c) \\ &\leq CA \exp\left(-\frac{(j_0-1)^2}{8D}\mathcal{K}(t)\right) + C \log A/A^2. \end{aligned}$$

The lemma now follows upon taking $A = \exp\left(\frac{(j_0-1)^2}{16D}\mathcal{K}(t)\right)$. □

Proof of Lemma 3.8. Write $u^* := \lambda^*/2$. Take any $p > 0$, whose value will be appropriately chosen later. Let $t_n^+ = \frac{1}{\lambda^*} \log n + p\mathcal{K}(\frac{1}{\lambda^*} \log n)$ and $t_n^- = \frac{1}{\lambda^*} \log n - p\mathcal{K}(\frac{1}{\lambda^*} \log n)$. Note that

$$\begin{aligned} &P\left(\mathcal{I}_c^*(T_n) \leq (u^* + \varepsilon)\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right) \\ &\geq P\left(\mathcal{I}_c^*(s) \leq (u^* + \varepsilon)\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right) \text{ for all } s \in [t_n^-, t_n^+], T_n \in [t_n^-, t_n^+]\right). \end{aligned}$$

Hence,

$$\begin{aligned}
 &P\left(\mathcal{I}_c^*(T_n) > (u^* + \varepsilon)\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right) \\
 &\leq P\left(\mathcal{I}_c^*(s) > (u^* + \varepsilon)\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right) \text{ for some } s \in [t_n^-, t_n^+]\right) + P(T_n > t_n^+) + P(T_n < t_n^-).
 \end{aligned}
 \tag{3.21}$$

Consider the function

$$H(u) = -u + \sqrt{2\lambda^*u}, \quad u \geq 0.$$

This function has a unique maximum at $u = u^* := \lambda^*/2$ with maximum value $H(u^*) = \lambda^*/2 = u^*$. Recall A_0 from Lemma 3.11 and take $A_1 := \lambda^* + A_0$.

Let $\varepsilon_0 \in (0, \lambda^*/4)$ be small enough that $\max\{H(u^*) - H(u^* - \varepsilon), H(u^*) - H(u^* + \varepsilon)\} < \sqrt{2\lambda^*(u^* - \varepsilon)}$ for all $\varepsilon < \varepsilon_0$. Fix any $\varepsilon \in (0, \varepsilon_0)$. Let $\zeta := \min\{H(u^*) - H(u^* - \varepsilon), H(u^*) - H(u^* + \varepsilon)\}$. Partition the set $(u^* + \varepsilon, A_1]$ into disjoint intervals $[u_i, v_i], i \in \mathbb{J}$, of mesh $\zeta/3$. Let $p = \min\{1, \zeta/12\}$. Then for any $t > 0$,

$$\begin{aligned}
 &P(\mathcal{I}_c^*(s) \leq (u^* + \varepsilon)\mathcal{K}(t) \text{ for all } s \in [t - p\mathcal{K}(t), t + p\mathcal{K}(t)]) \\
 &\geq P\left(D_{(u^* - \varepsilon)\mathcal{K}(t), u^*\mathcal{K}(t)}^{\max}(s) \geq t + (H(u^*) - \zeta/3)\mathcal{K}(t), \right. \\
 &\quad \left. D_{u_i\mathcal{K}(t), v_i\mathcal{K}(t)}^{\max}(s) \leq t + (H(v_i) + \zeta/2)\mathcal{K}(t) \text{ for all } i \in \mathbb{J}, \right. \\
 &\quad \left. D_{A_1\mathcal{K}(t), t}^{\max}(s) \leq t, \text{ for all } s \in [t - p\mathcal{K}(t), t + p\mathcal{K}(t)]\right).
 \end{aligned}$$

Hence, taking complements and applying the union bound,

$$\begin{aligned}
 &P(\mathcal{I}_c^*(s) > (u^* + \varepsilon)\mathcal{K}(t) \text{ for some } s \in [t - p\mathcal{K}(t), t + p\mathcal{K}(t)]) \\
 &\leq P\left(D_{(u^* - \varepsilon)\mathcal{K}(t), u^*\mathcal{K}(t)}^{\max}(t - p\mathcal{K}(t)) < t + (H(u^*) - \zeta/3)\mathcal{K}(t)\right) \\
 &\quad + \sum_{i \in \mathbb{J}} P\left(D_{u_i\mathcal{K}(t), v_i\mathcal{K}(t)}^{\max}(t + p\mathcal{K}(t)) > t + (H(v_i) + \zeta/2)\mathcal{K}(t)\right) \\
 &\quad + P\left(D_{A_1\mathcal{K}(t), t}^{\max}(t + p\mathcal{K}(t)) > t\right).
 \end{aligned}$$

The first term in the product can be estimated using Lemma 3.10 by taking $\eta = \zeta/6, a = u^* - \varepsilon, b = u^*$, and the sum can be estimated using the same lemma by taking $\eta = \zeta/12, a = u_i, b = v_i$ for each $i \in \mathbb{J}$ and noting that the cardinality of \mathbb{J} does not depend on t . The last term can be estimated from Lemma 3.11 upon noting that $A_1 > A_0$. Hence, we obtain upon putting $t = \frac{1}{\lambda^*} \log n$,

$$P\left(\mathcal{I}_c^*(s) > (u^* + \varepsilon)\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right) \text{ for some } s \in [t_n^-, t_n^+]\right) \leq Ce^{-C'\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)}. \tag{3.22}$$

To estimate the last two terms in (3.21), we will use Lemma 3.7 with $\gamma_n := \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)$. Note that, as $\Phi_2(\infty) = \infty, \lim_{t \rightarrow \infty} \mathcal{K}(t) = \infty$. Moreover, as $f(n) \rightarrow \infty$ as $n \rightarrow \infty$ by Assumption (A1), for any $\delta > 0$, there exists $n_\delta \in \mathbb{N}$ such that $f(n) > \delta^{-1}$ for all $n \geq n_\delta$. Hence, for $n > n_\delta$,

$$\Phi_2(n) \leq \sum_{k=1}^{n_\delta} \frac{1}{f^2(k)} + \delta \sum_{k=n_\delta+1}^n \frac{1}{f(k)} \leq \sum_{k=1}^{n_\delta} \frac{1}{f^2(k)} + \delta\Phi_1(n).$$

In particular, $\limsup_{n \rightarrow \infty} \frac{\Phi_2(n)}{\Phi_1(n)} \leq \delta$. As $\delta > 0$ is arbitrary, $\limsup_{n \rightarrow \infty} \frac{\Phi_2(n)}{\Phi_1(n)} = 0$. Hence,

$$\lim_{t \rightarrow \infty} \frac{\mathcal{K}(t)}{t} = \lim_{t \rightarrow \infty} \frac{\Phi_2 \circ \Phi_1^{-1}(t)}{\Phi_1 \circ \Phi_1^{-1}(t)} = 0. \tag{3.23}$$

Therefore, $\gamma_n := \mathcal{K}(\frac{1}{\lambda^*} \log n)$ satisfies the hypotheses of Lemma 3.7, and thus

$$P(T_n < t_n^-) + P(T_n > t_n^+) \leq C e^{-C' \mathcal{K}(\frac{1}{\lambda^*} \log n)}. \tag{3.24}$$

Using (3.22) and (3.24) in (3.21), we have

$$P\left(\mathcal{I}_c^*(T_n) > (u^* + \varepsilon) \mathcal{K}(\frac{1}{\lambda^*} \log n)\right) \leq C e^{-C' \mathcal{K}(\frac{1}{\lambda^*} \log n)}. \quad \square$$

3.3 Height of branching processes

Recall the continuous time branching process $BP_f(t)$, viewed as a labelled tree (vertices labelled in order of arrival). Let H_t denote the height of this tree at time t , that is, the maximal distance of a vertex in $BP_f(t)$ from the root. We will need the following tail estimate for the height from [HJ16].

Lemma 3.12 (Lemma 13.15 of [HJ16]). *For every $r > 0$, there exists c_r such that, for large t ,*

$$P(H_t \geq c_r t) \leq e^{-rt}.$$

4 Martingales tracking degrees

For $A \in \mathbb{N}_0$, recall the definition of $f_A(\cdot)$ and $\xi_A(\cdot)$ from (1.3). Define the process

$$M_A(t) := \sum_{i=1}^{\xi_A(t)} \frac{1}{f_A(i)} - t, \quad t \geq 0, \tag{4.1}$$

where we recall our convention that $\sum_{i=1}^0 \frac{1}{f_A(i)} = 0$. This process turns out to be the crucial technical tool used in [BB21], as well as the current article, to track degrees of individual vertices in the random graph process. We collect some properties of this process, proved in [BB21]. For $A \in \mathbb{N}_0$, write $f_*(A) := \inf_{i \geq A+1} f(i)$.

Lemma 4.1 (Lemmas 7.1, 7.3 and 7.4 in [BB21]). *For any $A \in \mathbb{N}_0$, the process $M_A(\cdot)$ defined in (4.1) is a martingale with respect to its natural filtration.*

If $\Phi_2(\infty) < \infty$, then there exists positive constants $x_0, x'_0, x''_0, C_1, C_2$ (independent of A) such that for any $A \in \mathbb{N}_0$,

$$P\left(\inf_{s < \infty} M_A(s) \leq -x\right) \leq \begin{cases} C_1 \exp(-C_2 x^2) & \text{if } x_0 \leq x < x'_0 \sqrt{f_*(A)}, \\ C_1 \exp(-C_2 \sqrt{f_*(A)} x) & \text{if } x \geq x'_0 \sqrt{f_*(A)}. \end{cases} \tag{4.2}$$

and

$$P\left(\sup_{s < \infty} M_A(s) \geq x\right) \leq C_1 \exp(-C_2 x^2) \quad \text{for all } x \geq x''_0. \tag{4.3}$$

Moreover, if $\Phi_2(\infty) < \infty$, $M_A(t) \rightarrow M_A(\infty)$ almost surely, and the law of $M_A(\infty)$ is absolutely continuous with respect to the Lebesgue measure.

5 A universal approach for budget lower bounds

Our general approach for obtaining lower bounds on $K_\Psi(\varepsilon)$ is to connect $\liminf_{n \rightarrow \infty} P(v_0 \notin H_{\lfloor \beta K/2 \rfloor, \Psi}(\mathcal{G}_n))$, for some $\beta \in (0, 1)$ and any $K \geq 2$, to the probability of the event $\mathcal{A}_{\beta, d^*, K}$ defined below. For $K \geq 1$, $d^* \geq m$, $\beta \in (0, 1)$, define

$$\mathcal{A}_{\beta, d^*, K} = \{v_0 \text{ has degree } \leq d^* \text{ in } \mathcal{G}_K \text{ and there are at least } \lfloor \beta K \rfloor \text{ other vertices with degree } \geq d^* \text{ in } \mathcal{G}_K\}.$$

Our proofs for lower bounds on the budget will involve lower bounds on $P(\mathcal{A}_{\beta, d^*, K})$ for appropriate choices of β, d^* . This, by Lemma 5.1, will produce lower bounds on $K_{\Psi}(\varepsilon)$.

Following Definition 2.1, for $1 \leq j \leq n$ we will denote by $H_{j, \Psi}(\mathcal{G}_n)$ the set of j vertices in \mathcal{G}_n with the largest Ψ values (ties being broken arbitrarily). As Ψ is taken to be the degree centrality measure here, $H_{j, \Psi}(\mathcal{G}_n)$ is the set of j vertices with the largest degrees in \mathcal{G}_n (ties being broken arbitrarily).

Lemma 5.1. *Suppose $m_i \equiv m \geq 1$ and $\Phi_2(\infty) < \infty$. For any $d^* \geq m$, $\beta \in (0, 1]$ and $K \geq 4\beta^{-1}$,*

$$\liminf_{n \rightarrow \infty} P(v_0 \notin H_{\lfloor \beta K/2 \rfloor, \Psi}(\mathcal{G}_n)) \geq \frac{1}{2} P(\mathcal{A}_{\beta, d^*, K}).$$

Proof. We will use the construction and terminology of the collapsed branching process embedding defined in Section 3. Let \mathcal{F}_K be the natural (stopped) filtration associated with the continuous time description of the collapsed branching process up to stopping time τ_K . For $0 \leq i \leq K, t \geq 0$, write $d'_i(t) := \xi^{(m), v_i}(t + \tau_K - \tau_i)$ to denote the ‘continuous time’ degree of v_i at time $t + \tau_K$, conditioned on the filtration \mathcal{F}_K . By definition we see that $d'_i(0)$ represents the degree of v_i in $CBP_f(K)$ (which is known given \mathcal{F}_K) for each $0 \leq i \leq K$.

Let \mathcal{U}_K denote the set of indices $i \in \{1, \dots, K\}$ for which the degree of v_i in \mathcal{G}_K is at least d^* . Conditioned on the filtration \mathcal{F}_K , and on the event $\mathcal{A}_{\beta, d^*, K}$, $|\mathcal{U}_K| \geq \lfloor \beta K \rfloor$, and the degree of v_0 is at most d^* . It follows from the construction of the collapsed branching process embedding that $\{d'_i(t)\}_{0 \leq i \leq K}$ are independent. So, we can construct a collection of i.i.d. processes $\{\zeta_{d^*}^{(i)}(t) : t \geq 0\}_{i \in \mathcal{U}_K \cup \{0\}}$, each having the same law as $\zeta_{d^*}(t) := \xi_{d^*-1}(t) + d^*, t \geq 0$, such that for all $t \geq 0$, $d'_i(t) \geq \zeta_{d^*}^{(i)}(t)$ for all $i \in \mathcal{U}_K$, and $d'_0(t) \leq \zeta_{d^*}^{(0)}(t)$.

By Lemma 4.1, for $i \in \mathcal{U}_K \cup \{0\}$, the process

$$M^{(i)}(t) := \sum_{i=d^*}^{\zeta_{d^*}^{(i)}(t)-1} \frac{1}{f(i)} - t \stackrel{d}{=} \sum_{i=1}^{\xi_{d^*-1}(t)} \frac{1}{f_{d^*-1}(i)} - t, \quad t \geq 0,$$

(defined conditional on \mathcal{F}_K , on the event $\mathcal{A}_{\beta, d^*, K}$) is a martingale which converges almost surely to a random limit $M^{(i)}(\infty)$ whose law is absolutely continuous with respect to the Lebesgue measure.

As $\{M_i(\infty) : i \in \mathcal{U}_K \cup \{0\}\}$ are i.i.d. random variables with absolute continuous laws with respect to the Lebesgue measure, for any $i \in \mathcal{U}_K \cup \{0\}$ and $1 \leq j \leq |\mathcal{U}_K| + 1$, the probability of $M^{(i)}(\infty)$ being the j -th largest in $\{M^{(l)}(\infty)\}_{l \in \mathcal{U}_K \cup \{0\}}$ is exactly $1/(|\mathcal{U}_K| + 1)$. Therefore, on the event $\mathcal{A}_{\beta, d^*, K}$,

$$\begin{aligned} & \liminf_{n \rightarrow \infty} P(v_0 \notin H_{\lfloor \beta K/2 \rfloor, \Psi}(\mathcal{G}_n) | \mathcal{F}_K) \\ & \geq \liminf_{n \rightarrow \infty} P(d'_0(\tau_n - \tau_K) \text{ is not among the } \lfloor \beta K/2 \rfloor \text{ largest values in } \{d'_i(\tau_n - \tau_K)\}_{i \in \mathcal{U}_K \cup \{0\}}) \\ & \geq \liminf_{n \rightarrow \infty} P(\zeta_{d^*}^{(0)}(\tau_n - \tau_K) \text{ is not among the } \lfloor \beta K/2 \rfloor \text{ largest values in } \{\zeta_{d^*}^{(i)}(\tau_n - \tau_K)\}_{i \in \mathcal{U}_K \cup \{0\}}) \\ & = \liminf_{n \rightarrow \infty} P(M^{(0)}(\tau_n - \tau_K) \text{ is not among the } \lfloor \beta K/2 \rfloor \text{ largest values in } \{M^{(i)}(\tau_n - \tau_K)\}_{i \in \mathcal{U}_K \cup \{0\}}) \\ & \geq P(M^{(0)}(\infty) \text{ is not among the } \lfloor \beta K/2 \rfloor \text{ largest values in } \{M^{(i)}(\infty)\}_{i \in \mathcal{U}_K \cup \{0\}}) \\ & = \sum_{j=\lfloor \beta K/2 \rfloor + 1}^{|\mathcal{U}_K| + 1} \frac{1}{|\mathcal{U}_K| + 1} \geq \frac{1}{2}. \end{aligned}$$

Taking expectations,

$$\begin{aligned} & \liminf_{n \rightarrow \infty} P(v_0 \notin H_{\lfloor \beta K/2 \rfloor, \Psi}(\mathcal{G}_n)) \\ &= \liminf_{n \rightarrow \infty} E \left[P(v_0 \notin H_{\lfloor \beta K/2 \rfloor, \Psi}(\mathcal{G}_n) | \mathcal{F}_K) 1_{\mathcal{A}_{\beta, d^*, K}} \right] \geq \frac{1}{2} P(\mathcal{A}_{\beta, d^*, K}). \end{aligned} \quad \square$$

6 Linear attachment: proof of Theorem 2.2

6.1 Lower bound

We will use Lemma 5.1 with $\beta = 1$ and $d^* = m$. In this case, $\mathcal{A}_{\beta, d^*, K}$ simplifies to

$$\mathcal{A}_K = \{v_0 \text{ has degree } m \text{ in } \mathcal{G}_K\}.$$

A lower bound for $P(\mathcal{A}_K)$ is obtained in the following lemma.

Lemma 6.1. *There exists $C > 0$ depending on f such that for all $K \geq 1$, $P(\mathcal{A}_K) \geq C/K^{\frac{m(m+\beta)}{2m+\beta}}$.*

Proof. Let $i_0 \geq 1$ be such that $\frac{m+\beta}{(2m+\beta)i_0+\beta} \leq 1/2$ and write $C_0 = P(\mathcal{A}_{i_0})$. Then, for $K > i_0$, $\mathcal{A}_K = \mathcal{A}_{i_0} \cup \{\text{no edge from } v_{i_0+1}, \dots, v_K \text{ attaches to } v_0\}$. Hence,

$$\begin{aligned} P(\mathcal{A}_K) &\geq P(\mathcal{A}_{i_0}) \prod_{i=i_0}^{K-1} \prod_{im < k \leq (i+1)m} \left(1 - \frac{f(m)}{\sum_{j=0}^i f(d_j(k-1))} \right) \\ &\geq P(\mathcal{A}_{i_0}) \prod_{i=i_0}^{K-1} \left(1 - \frac{f(m)}{\sum_{j=0}^i f(d_j(s_i))} \right)^m \\ &= C_0 \prod_{i=i_0}^{K-1} \left(1 - \frac{m+\beta}{(2m+\beta)i+\beta} \right)^m \\ &= C_0 \exp \left(m \sum_{i=i_0}^{K-1} \log \left(1 - \frac{m+\beta}{(2m+\beta)i+\beta} \right) \right). \end{aligned}$$

Using the observation that $\log(1-x) \geq -x - x^2$ for $x \in (0, 1/2]$, we have

$$P(\mathcal{A}_K) \geq C \exp \left(-\frac{m(m+\beta)}{2m+\beta} \sum_{i=i_0}^{K-1} \frac{1}{i+\beta/(2m+\beta)} \right) \geq C' \exp \left(-\frac{m(m+\beta)}{2m+\beta} \log K \right),$$

for positive constants C, C' not depending on K . □

Proof of lower bound in Theorem 2.2. Lemma 5.1 and Lemma 6.1 together yield

$$P(v_0 \notin H_{\lfloor K/2 \rfloor, \Psi}(\mathcal{G}_n)) \geq C/K^{\frac{m(m+\beta)}{2m+\beta}}$$

for some $C > 0$, which then leads to the budget lower bound

$$K_{\Psi}(\varepsilon) \geq \left(\frac{C}{\varepsilon} \right)^{\frac{2m+\beta}{m(m+\beta)}}. \quad \square$$

6.2 Upper bound

To obtain the budget upper bound, we will show that, with high probability, the degree of the root grows at a certain rate (see event A_n below). On this event, the degree of the root has enough of a head start so that the degree of the $(n+1)$ -th added vertex (for large n) never crosses that of the root at any future time with high probability.

Precise quantitative bounds on the associated probabilities is key to obtaining the budget upper bound.

Throughout this section $m_i \equiv m \geq 1$. All constants that appear below can depend on m, β , but not n .

Let $\{\alpha_n\}$ be an increasing sequence that goes to infinity as n goes to infinity, which will be specified later. Recall that $s_n = \sum_{i=1}^n m_i = mn$. For $i \geq 1$ and $l > s_{i-1}$, recall that $d_i(l)$ denotes the degree of v_i in \mathcal{G}_l^* (the random graph, index by attached edges, at time l). Define the following events for $n \geq 1$,

$$A_n = \{\Phi_1(d_0(s_n)) \geq \alpha_n\} \tag{6.1}$$

and

$$E_n = \{d_n(k) \geq d_0(k) \text{ for some } k \geq s_n\}. \tag{6.2}$$

Our goal is to find some $n_0(\varepsilon) \in \mathbb{N}$ so that for all $n \geq n_0(\varepsilon)$, $P(E_n) \leq \varepsilon$. This leads to an upper bound on the budget $K_\Psi(\varepsilon) \leq n_0(\varepsilon)$. This will be done in two steps. Step 1 estimates $P(E_n \cap A_n)$ and Step 2 estimates $P(A_n^c)$.

Step 1: The first step is to obtain an upper bound on $P(E_n \cap A_n)$ following the same approach as in the proof of Theorem 4.2 in [BB21]. To be self-contained we will present the calculation. The expressions are intentionally left in terms of general f till (6.4) as they are also used in the proof of Theorem 2.3.

For $A \in \mathbb{N}_0$, recall that $\xi_A(\cdot)$ denotes the point process obtained with attachment function $f_A(\cdot) = f(A + \cdot)$. Define $\zeta_1^{(n)}(t) = d_0(n) + \xi_{d_0(s_n)-1}(t)$ and $\zeta_2^{(n)}(t) = m + \xi_{m-1}(t)$, and the corresponding martingales

$$M_1^{(n)}(t) = \sum_{k=d_0(s_n)}^{\zeta_1^{(n)}(t)-1} \frac{1}{f(k)} - t = \sum_{k=1}^{\xi_{d_0(s_n)-1}(t)} \frac{1}{f_{d_0(s_n)-1}(k)} - t,$$

$$M_2^{(n)}(t) = \sum_{k=m}^{\zeta_2^{(n)}(t)-1} \frac{1}{f(k)} - t = \sum_{k=1}^{\xi_{m-1}(t)} \frac{1}{f_{m-1}(k)} - t.$$

Recall $\Phi_1(\cdot)$ from (1.4). Then

$$\begin{aligned} &P(E_n \cap A_n) \\ &\leq P(M_2^{(n)}(t) + \Phi_1(m) \geq M_1^{(n)}(t) + \Phi_1(d_0(s_n)) \text{ for some } t \geq 0, \Phi_1(d_0(s_n)) \geq \alpha_n) \\ &\leq P\left(\inf_{s < \infty} M_{d_0(s_n)}(s) \leq -\frac{\Phi_1(d_0(s_n)) - \Phi_1(m)}{2}, \Phi_1(d_0(s_n)) \geq \alpha_n\right) \\ &\quad + P\left(\sup_{s < \infty} M_m(s) \geq \frac{\Phi_1(d_0(s_n)) - \Phi_1(m)}{2}, \Phi_1(d_0(s_n)) \geq \alpha_n\right). \end{aligned}$$

Choosing n sufficiently large such that $\frac{\alpha_n - \Phi_1(m)}{2} \geq \frac{\alpha_n}{4}$, we have

$$\begin{aligned} P(E_n \cap A_n) &\leq P\left(\inf_{s < \infty} M_{d_0(s_n)}(s) \leq -\frac{\alpha_n}{4}, \Phi_1(d_0(s_n)) \geq \alpha_n\right) \\ &\quad + P\left(\sup_{s < \infty} M_m(s) \geq \frac{\alpha_n}{4}, \Phi_1(d_0(s_n)) \geq \alpha_n\right). \end{aligned}$$

Assume n_0 is large enough so that $\frac{\alpha_n - \Phi_1(m)}{2} \geq \frac{\alpha_n}{4}$ for all $n \geq n_0$, and $\frac{\alpha_{n_0}}{4} \geq \max\{x_0, x_0''\}$, where x_0, x_0'' are the constants appearing in Lemma 4.1. It follows from the lemma that

there exist constants $C, C' > 0$ such that for all $n \geq n_0$,

$$\begin{aligned} & P\left(\inf_{s < \infty} M_{d_0(s_n)}(s) \leq -\frac{\alpha_n}{4}, \Phi_1(d_0(s_n)) \geq \alpha_n\right) \\ & \leq E\left[\left(C \exp(-C'(\alpha_n^2)) + C \exp\left(-C' \sqrt{f_*(d_0(s_n))} \alpha_n\right)\right) 1_{\{\Phi_1(d_0(s_n)) \geq \alpha_n\}}\right] \\ & \leq C \exp(-C' \alpha_n^2) + C \exp\left(-C' \sqrt{f_*(\lfloor \Phi_1^{-1}(\alpha_n) \rfloor)} \alpha_n\right) =: a_n, \end{aligned} \tag{6.3}$$

and

$$P\left(\sup_{s < \infty} M_m(s) \geq \frac{\alpha_n}{4}, \Phi_1(d_0(s_n)) \geq \alpha_n\right) \leq C \exp(-C' \alpha_n^2) =: b_n. \tag{6.4}$$

Since $f(i) = i + \beta$, $f_*(\lfloor \Phi_1^{-1}(\alpha_n) \rfloor) = \lfloor \Phi_1^{-1}(\alpha_n) \rfloor + 1 + \beta \geq C \exp(\alpha_n)$. Hence, setting $\alpha_n = M(\log n)^{1/2}$, $n \geq 1$, for a fixed large M , we obtain a constant $C_1 > 0$ such that

$$P(E_n \cap A_n) \leq a_n + b_n \leq C_1 n^{-m}, \quad n \geq 1.$$

Step 2: Next we will estimate $P(A_n^c)$.

Recall that $\{\mathcal{G}_k^* : k \geq s_1\}$ denotes the edge indexed random graph sequence. $\{\mathcal{G}_n : n \geq 1\}$ has the same law as $\{\mathcal{G}_{s_n}^* : n \geq 1\}$. Recall $f(i) = i + \beta$ for $i \geq 1$.

We will construct another sequence of graphs $\{\bar{\mathcal{G}}_k : k \geq s_1\}$ by attaching one vertex with one edge at each step using attachment function $\bar{f}(i) = i + \beta/m$. The construction starts with a fixed graph $\bar{\mathcal{G}}_{s_1}$ with vertices $\{\bar{v}_1, \dots, \bar{v}_{2m}\}$ and an edge between \bar{v}_i and \bar{v}_{i+m} for each $1 \leq i \leq m$.

Observe that $\bar{\mathcal{G}}_k$ has vertices $\{\bar{v}_i : 1 \leq i \leq k + m\}$. For $k \geq s_1$, $1 \leq i \leq k + m$, let $\bar{d}_i(k)$ denote the degree of vertex \bar{v}_i in $\bar{\mathcal{G}}_k$. Define $d_0''(k) = \sum_{i=1}^m \bar{d}_i(k)$. Note that $d_0''(s_1) = d_0(s_1) = m$. For any $k \geq s_1$,

$$P(d_0''(k+1) = d_0''(k) + 1 | \bar{d}_i(j), 1 \leq i \leq m, j \leq k) = \frac{\sum_{i=1}^m \bar{f}(\bar{d}_i(k))}{\sum_{j=1}^{k+m} \bar{f}(\bar{d}_j(k))}.$$

Lemma 6.2. *There exists a coupling of the random graph sequences $\{\mathcal{G}_k^* : k \geq s_1\}$ and $\{\bar{\mathcal{G}}_k : k \geq s_1\}$ such that $d_0''(k) \leq d_0(k)$ for all $k \geq s_1$.*

Proof. We proceed by induction. Recall that $d_0''(s_1) = d_0(s_1) = m$. Suppose for some $n \geq 1$ and some $s_n \leq k < s_{n+1}$, we have coupled the random graph sequences $\{\mathcal{G}_l^* : s_1 \leq l \leq k\}$ and $\{\bar{\mathcal{G}}_l : s_1 \leq l \leq k\}$ such that $d_0''(i) \leq d_0(i)$ for $i \leq k$. We want to extend the coupling to $s_1 \leq l \leq k + 1$ such that $d_0''(k + 1) \leq d_0(k + 1)$. For $s_n \leq k < s_{n+1}$, in the construction of the original $\{\mathcal{G}_j^*\}$,

$$P(\text{other end of } e_{k+1} \text{ attached to } v_0 | \mathcal{G}_j^*, j \leq k) = \frac{f(d_0(k))}{\sum_{j=0}^n f(d_j(k))}.$$

Note that, for $s_n \leq k < s_{n+1}$,

$$\sum_{j=1}^{k+m} \bar{f}(\bar{d}_j(k)) = 2k + \frac{(k+m)\beta}{m} \geq 2k + (n+1)\beta = \sum_{j=0}^n f(d_j(k)).$$

From this relation and the induction assumption, for $s_n \leq k < s_{n+1}$,

$$\frac{f(d_0(k))}{\sum_{j=0}^n f(d_j(k))} \geq \frac{f(d_0''(k))}{\sum_{j=1}^{k+m} \bar{f}(\bar{d}_j(k))} = \frac{\sum_{i=1}^m \bar{f}(\bar{d}_i(k))}{\sum_{j=1}^{k+m} \bar{f}(\bar{d}_j(k))},$$

that is,

$$P(\text{other end of } e_{k+1} \text{ attached to } v_0 \mid \mathcal{G}_j^*, j \leq k) \geq P(\text{other end of } e_{k+1} \text{ attached to one of } \bar{v}_1, \dots, \bar{v}_m \mid \bar{\mathcal{G}}_j, j \leq k).$$

This can be used to extend the coupling to $\{\mathcal{G}_l^* : s_1 \leq l \leq k + 1\}$ and $\{\bar{\mathcal{G}}_l : s_1 \leq l \leq k + 1\}$ such that $d_0''(i) \leq d_0(i)$ for $i \leq k + 1$. Using the Kolmogorov extension theorem, this completes the proof. \square

From Lemma 6.2 it follows that, for $n \geq 1$,

$$P(A_n^c) = P(\Phi_1(d_0(s_n)) < \alpha_n) \leq P(\Phi_1(d_0''(s_n)) < \alpha_n) \leq P(\cap_{i=1}^m \{\Phi_1(\bar{d}_i(s_n)) < \alpha_n\}),$$

where recall $\Phi_1(l) = \sum_{i=1}^{l-1} \frac{1}{f(i)}$. Note that, for any $1 \leq i \leq m, k \geq s_1$, the ‘offspring’ subtree of \bar{v}_i (subtree corresponding to the connected component of $\bar{\mathcal{G}}_k$ containing \bar{v}_i , rooted at \bar{v}_i) can be described by an embedding into the continuous time branching process (analogous to Lemma 3.2)

$$\widetilde{BP}_{\bar{f}}^{(i)}(t) := BP_{\bar{f}}^{(i)}(t) \cup BP_{\bar{f}}^{(i+m)}(t),$$

where $\{BP_{\bar{f}}^{(i)}(t) : 1 \leq i \leq 2m\}$ are independent branching processes as in Definition 3.1 constructed using the attachment function \bar{f} . Hence, for any $n \geq 1$, the degrees $\{\bar{d}_i(s_n) : 1 \leq i \leq m\}$ are independent and have the same distribution, which leads to

$$P(A_n^c) \leq P(\Phi_1(\bar{d}_1(s_n)) < \alpha_n)^m. \tag{6.5}$$

Similarly as in (1.3), we define $\bar{\xi}$ to be the point process under the attachment function \bar{f} , which is a Yule process with birth rate 1 and immigration rate β/m . For $k > s_1$, let $\bar{T}_k := \inf\{t \geq 0 : \sum_{i=1}^{2m} |BP_{\bar{f}}^{(i)}(t)| = k\}$, and set $\bar{T}_{s_1} = 0$. Let $\bar{\sigma}_n = \frac{m \log n}{2m + \beta} - \gamma_n$, where $\{\gamma_n\}_{n \geq 1}$ is an increasing sequence diverging to infinity, such that $\gamma_n / \log n \rightarrow 0$ as $n \rightarrow \infty$, that will be chosen later. There exists $c > 0$ such that, for all large l , $\Phi_1(l) = \sum_{i=1}^{l-1} \frac{1}{f(i)} = \sum_{i=1}^{l-1} \frac{1}{i + \beta} \geq \log l - \log c$, which implies $\Phi_1^{-1}(l) \leq ce^l$ for large l . Then, for sufficiently large n ,

$$\begin{aligned} P(\Phi_1(\bar{d}_1(s_n)) < \alpha_n) &= P(\Phi_1(\bar{\xi}(\bar{T}_{s_n})) < \alpha_n) \\ &\leq P(\bar{T}_{s_n} < \bar{\sigma}_n) + P(\bar{\xi}(\bar{\sigma}_n) \leq \Phi_1^{-1}(\alpha_n)) \\ &\leq P(\bar{T}_{s_n} < \bar{\sigma}_n) + P(\bar{\xi}(\bar{\sigma}_n) \leq ce^{\alpha_n}). \end{aligned}$$

It follows from (3.3) that for any $\theta > 0, n \in \mathbb{N}$,

$$\begin{aligned} P(\bar{\xi}(\bar{\sigma}_n) \leq ce^{\alpha_n}) &= P(\exp(-\theta \bar{\xi}(\bar{\sigma}_n)) \geq \exp\{-\theta ce^{\alpha_n}\}) \leq \exp\{\theta ce^{\alpha_n}\} M(-\theta, \bar{\sigma}_n) \\ &= \exp\{\theta ce^{\alpha_n}\} \cdot \frac{e^{-\theta}}{((1 - e^{-\theta})e^{\bar{\sigma}_n} + e^{-\theta})^{1 + \beta/m}}. \end{aligned}$$

Hence, choosing $\theta = e^{-\alpha_n}$, we obtain for any $n \in \mathbb{N}$,

$$\begin{aligned} P(\bar{\xi}(\bar{\sigma}_n) \leq ce^{\alpha_n}) &\leq \frac{e^c}{((1 - e^{-e^{-\alpha_n}})e^{\bar{\sigma}_n} + e^{-e^{-\alpha_n}})^{1 + \beta/m}} \\ &\leq Ce^{-(\bar{\sigma}_n - \alpha_n)(1 + \beta/m)} = Cn^{-\frac{m + \beta}{2m + \beta}} e^{(\alpha_n + \gamma_n)(1 + \beta/m)}. \end{aligned} \tag{6.6}$$

It remains to estimate $P(\bar{T}_{s_n} < \bar{\sigma}_n)$. Let $\{E_k : k \geq 2m\}$ be a set of i.i.d. exponential random variables with intensity 1. Note that, for any $l \geq m + 1$,

$$\begin{aligned} \bar{T}_l &\stackrel{d}{=} \sum_{k=m}^{l-1} \frac{E_k}{\sum_{j=1}^{k+m} \bar{f}(\bar{d}_j(k))} = \sum_{k=m}^{l-1} \frac{E_k}{\sum_{j=1}^{k+m} (\bar{d}_j(k) + \beta/m)} \\ &= \sum_{k=m}^{l-1} \frac{E_k}{2k + (k+m)\beta/m} = \sum_{k=m}^{l-1} \frac{E_k}{(2 + \beta/m)k + \beta}. \end{aligned}$$

Take $\theta_n = \log n$. Then, for sufficiently large n ,

$$\begin{aligned} P(\bar{T}_{s_n} < \bar{\sigma}_n) &= P(e^{-\theta_n \bar{T}_{s_n}} \geq e^{-\theta_n \bar{\sigma}_n}) \leq e^{\theta_n \bar{\sigma}_n} E e^{-\theta_n \bar{T}_{s_n}} \\ &= e^{\theta_n \bar{\sigma}_n} \exp\left(-\sum_{k=m}^{s_n-1} \log\left(1 + \frac{\theta_n}{(2 + \beta/m)k + \beta}\right)\right) \\ &\leq C e^{\theta_n \bar{\sigma}_n} \exp\left(-\sum_{k=[\theta_n]+1}^{s_n-1} \frac{\theta_n}{(2 + \beta/m)k + \beta}\right) \quad (\text{by Taylor expansion}) \\ &\leq C e^{\theta_n \bar{\sigma}_n} \exp\left(-\frac{\theta_n m}{2m + \beta} (\log n - \log \log n - C')\right) \\ &= C e^{-\theta_n (\gamma_n - \frac{m}{2m+\beta} \log \log n - \frac{C' m}{2m+\beta})}. \end{aligned}$$

Take $\gamma_n = \frac{2m \log \log n}{2m+\beta}$. Then, for any $M' > 0$,

$$P(\bar{T}_{s_n} < \bar{\sigma}_n) \leq n^{-M'} \tag{6.7}$$

when n is sufficiently large. Collecting the terms (6.5), (6.6) and (6.7) we have for any $M' > 0, \theta \geq 1$,

$$\begin{aligned} P(A_n^c) &\leq (P(\bar{T}_{s_n} < \bar{\sigma}_n) + P(\bar{\xi}(\bar{\sigma}_n) \leq c e^{\alpha_n}))^m \\ &\leq \left(C n^{-\frac{m+\beta}{2m+\beta}} \exp((1 + \beta/m)(\alpha_n + \gamma_n)) + n^{-M'}\right)^m \\ &\leq C' n^{-\frac{m(m+\beta)}{2m+\beta}} e^{(\alpha_n + \gamma_n)(m+\beta)} + C' n^{-M' m}. \end{aligned}$$

Take $M' = 1$. Recall from Step 1 that $\alpha_n = M(\log n)^{1/2}$ for some fixed M , and recall $\gamma_n = \frac{2m \log \log n}{2m+\beta}$. Then,

$$P(A_n^c) \leq C'' n^{-\frac{m(m+\beta)}{2m+\beta}} \exp\left(C(\log n)^{1/2}\right), \quad n \geq 1.$$

This completes Step 2.

Proof of upper bound in Theorem 2.2. Combining the error probabilities obtained in Step 1 and Step 2, we have

$$\begin{aligned} P(E_n) &\leq P(E_n \cap A_n) + P(A_n^c) \leq c_1 n^{-m} + C'' n^{-\frac{m(m+\beta)}{2m+\beta}} \exp\left(C(\log n)^{1/2}\right) \\ &\leq (c_1 + C'') n^{-\frac{m(m+\beta)}{2m+\beta}} \exp\left(C(\log n)^{1/2}\right). \end{aligned}$$

This leads to the upper bound on $K_\Psi(\cdot)$:

$$K_\Psi(\varepsilon) \leq \frac{C_1}{\varepsilon^{\frac{2m+\beta}{m(m+\beta)}}} \exp\left(\sqrt{C_2 \log \frac{1}{\varepsilon}}\right), \quad \varepsilon \in (0, 1),$$

for positive constants C_1, C_2 depending on m, β but not ε . □

7 General attachment: proof of Theorem 2.3

In this section, we prove Theorem 2.3 in the course of the following lemmas.

7.1 When $m = 1$

Lemma 7.1. *Under the assumptions in Theorem 2.3 with $m = 1$, there exists some $C_1 > 0$ so that*

$$K_\Psi(\varepsilon) \geq C_1 \varepsilon^{-\frac{\lambda^*}{f^*}}, \quad \text{for all } \varepsilon \in (0, 1).$$

Proof. Let i_0 be the minimum index such that $f(i_0) = f^*$. Note that such i_0 always exists because $f(k) \rightarrow \infty$ as $k \rightarrow \infty$ by Assumption (A1). We will apply Lemma 5.1 with $d^* = i_0$ and an appropriate choice of β to be fixed later. For $K \geq 1$, $\beta \in (0, 1]$, recall the event

$$\mathcal{A}_{\beta, i_0, K} = \{v_0 \text{ has degree } \leq i_0 \text{ in } \mathcal{G}_K \text{ and there are at least } \lfloor \beta K \rfloor \\ \text{other vertices with degree } \geq i_0 \text{ in } \mathcal{G}_K\}.$$

First we give a lower bound on $P(\mathcal{A}_{\beta, i_0, K})$ for an appropriate choice of β . Recall the embedding of our discrete tree process in the continuous time branching process described in Lemma 3.2. Recall the branching process $BP_f^{(1)}(\cdot)$ representing the descendants of v_1 . Let $\{\xi_{1,i}(\cdot) : i \in \mathbb{N}\}$ denote the point processes of reproduction of the vertices descending from v_1 . For $\delta \in (0, 1)$, let $t_\delta := \frac{1}{2\lambda^*} \log(1/\delta)$ and $p := P(\xi(t_\delta) \geq i_0)$ ($\xi(\cdot)$ is defined in (1.3)). Define the events

$$G_{1,K} := \left\{ BP_f^{(1)}\left(\frac{1}{\lambda^*} \log(\delta K)\right) \geq (\delta K)/2 \right\}, \\ G_{2,K} := \left\{ \sum_{i=1}^{\lfloor \delta K/2 \rfloor} 1_{\{\xi_{1,i}(t_\delta) \geq i_0\}} \geq p\delta K/3 \right\}, \\ G_{3,K} := \left\{ \frac{1}{\lambda^*} \log K - t_\delta < T_K < \frac{1}{\lambda^*} \log K + t_\delta \right\}.$$

Then we can observe that on $G_{1,K} \cap G_{3,K}$ at least $(\delta K)/2$ vertices have arrived before time $\frac{1}{\lambda^*} \log(\delta K)$, and hence, the point process of reproduction for each of these vertices has run for at least time $t_\delta = \frac{1}{2\lambda^*} \log(1/\delta)$ before time T_K . Thus, each of these vertices independently has degree $\geq i_0$ in $\widetilde{BP}_f(T_K)$ with probability at least p . Event $G_{2,K}$ then implies that there are at least $\lfloor \delta p K/3 \rfloor$ vertices with degree $\geq i_0$ in \mathcal{G}_K . Therefore,

$$\mathcal{A}_{\delta p/3, i_0, K} \supseteq G_{1,K} \cap G_{2,K} \cap G_{3,K} \cap \{\xi^{(0)}(T_K) \leq i_0\},$$

where $\xi^{(0)}(t)$ denotes the degree of v_0 at time t . Since $\xi^{(0)}(t)$ and $BP^{(1)}(t)$ are independent,

$$P(\mathcal{A}_{\delta p/3, i_0, K}) \geq P(\xi^{(0)}(T_K) \leq i_0 | G_{3,K}) \cdot P(G_{1,K} \cap G_{2,K} \cap G_{3,K}). \quad (7.1)$$

The first term in the lower bound above satisfies

$$P(\xi^{(0)}(T_K) \leq i_0 | G_{3,K}) \geq P(\{\xi^{(0)}(T_K) \leq i_0\} \cap G_{3,K}) \\ \geq E \left[P \left(\left\{ \xi^{(0)} \left(\frac{1}{\lambda^*} \log K + t_\delta \right) \leq i_0 \right\} \cap G_{3,K} \middle| T_K \right) \right] \\ \geq E \left[1_{\{\frac{1}{\lambda^*} \log K - t_\delta < T_K < \frac{1}{\lambda^*} \log K + t_\delta\}} P \left(\frac{E_{i_0}}{f(i_0)} > \frac{1}{\lambda^*} \log K + t_\delta \right) \right] \\ = E \left[e^{-f^* \left(\frac{1}{\lambda^*} \log K + t_\delta \right)} 1_{\{\frac{1}{\lambda^*} \log K - t_\delta < T_K < \frac{1}{\lambda^*} \log K + t_\delta\}} \right] \\ = K^{-\frac{f^*}{\lambda^*}} e^{-f^* t_\delta} P(G_{3,K}).$$

Next, we will show that $P(G_{3,K})$ and $P(G_{1,K} \cap G_{2,K} \cap G_{3,K})$ are lower bounded by some positive constants uniformly in K . We start with

$$\begin{aligned} \limsup_{K \rightarrow \infty} P(G_{1,K}^c) &= \limsup_{K \rightarrow \infty} P(BP_f^{(1)}(\frac{1}{\lambda^*} \log(\delta K)) < (\delta K)/2) \\ &= \limsup_{K \rightarrow \infty} P\left(e^{-\lambda^* (\frac{1}{\lambda^*} \log(\delta K))} BP_f^{(1)}(\frac{1}{\lambda^*} \log(\delta K)) < \frac{1}{2}\right) \\ &\leq P(W_\infty \leq 1/2). \end{aligned}$$

Next, as $\sum_{i=1}^{\lfloor \delta K/2 \rfloor} 1_{\{\xi_{1,i}(t_\delta) \geq i_0\}} \sim \text{Binomial}(\lfloor \delta K/2 \rfloor, p)$, $\lim_{K \rightarrow \infty} P(G_{2,K}^c) = 0$ by standard results on Binomial random variables. Lastly, using (3.5),

$$\begin{aligned} \limsup_{K \rightarrow \infty} P(G_{3,n}^c) &\leq \limsup_{K \rightarrow \infty} P(T_K > \frac{1}{\lambda^*} \log K + t_\delta) + \limsup_{K \rightarrow \infty} P(T_K < \frac{1}{\lambda^*} \log K - t_\delta) \\ &\leq P\left(W_\infty \leq \exp(-\frac{1}{2} \log(1/\delta))\right) + P\left(W_\infty \geq \exp(\frac{1}{2} \log(1/\delta))\right), \end{aligned}$$

where both term can be arbitrarily small when δ is small since $P(W_\infty > 0) = 1$ and $EW_\infty < \infty$. Choosing and fixing δ sufficiently small and collecting all the terms gives

$$\begin{aligned} \liminf_{K \rightarrow \infty} P(G_{1,n} \cap G_{2,K} \cap G_{3,K}) &\geq 1 - \limsup_{K \rightarrow \infty} (P(G_{1,K}^c) + P(G_{2,K}^c) + P(G_{3,K}^c)) \\ &\geq 1 - P(W_\infty \leq 1/2) - P\left(W_\infty \leq \exp(-\frac{1}{2} \log(1/\delta))\right) - P\left(W_\infty \geq \exp(\frac{1}{2} \log(1/\delta))\right) > 0, \end{aligned}$$

and

$$\liminf_{K \rightarrow \infty} P(G_{3,K}) \geq \liminf_{K \rightarrow \infty} P(G_{1,K} \cap G_{2,K} \cap G_{3,K}) > 0.$$

In the above, we have used $P(W_\infty \leq 1/2) < 1$ as W_∞ is supported on all of \mathbb{R}_+ , by Theorem 3.4.

It thus follows from (7.1) that, for this fixed choice of $\delta \in (0, 1)$, there exists some constant $C > 0$ so that for all $K \geq 1$,

$$P(\mathcal{A}_{\delta p/3, i_0, K}) \geq CK^{-\frac{f_*}{\lambda^*}}.$$

Now, by Lemma 5.1 with $d^* = i_0$ and $\beta = \delta p/3$, we obtain

$$\liminf_{n \rightarrow \infty} P(v_0 \notin H_{\lfloor \beta K/2 \rfloor, \Psi}(\mathcal{G}_n)) \geq \frac{1}{2} P(\mathcal{A}_{\delta p/3, i_0, K}) \geq \frac{C}{2} K^{-\frac{f_*}{\lambda^*}},$$

which leads to the lower bound

$$K_\Psi(\varepsilon) \geq \left(\frac{C}{2\varepsilon}\right)^{\frac{\lambda^*}{f_*}}. \quad \square$$

Lemma 7.2. *Under the assumptions in Theorem 2.3 with $m = 1$. For any $\delta \in (0, 1)$, we have*

$$K_\Psi(\varepsilon) \leq C_\delta'' \varepsilon^{-\frac{\lambda^*}{(1-\delta)f_*}},$$

where C_δ'' is a constant depending on δ .

Proof. Let A_n, E_n be the events defined in (6.1), (6.2). The proof follows essentially the same strategy as the proof in Section 6.2. A slight difference here is we choose $\alpha_n = \eta \log n$ instead of $M(\log n)^{1/2}$ for some small $\eta > 0$. The same calculation that leads to (6.3) and (6.4) yields,

$$P(E_n \cap A_n) \leq C \exp(-C' \alpha_n^2) + C \exp\left(-C' \sqrt{f_*([\Phi_1^{-1}(\alpha_n)])} \alpha_n\right).$$

Fix any $M > f_*/\lambda^*$. Since $f_*([\Phi_1^{-1}(\alpha_n)]) \rightarrow \infty$ as $n \rightarrow \infty$, we have

$$P(E_n \cap A_n) \leq \hat{C}_\eta n^{-M}, \quad n \geq 1. \tag{7.2}$$

for some constant $\hat{C}_\eta > 0$ depending on η .

The calculations for Step 2, though following the same broad strategy, rely on more general properties of the point process ξ defined in (1.3), which we now present. To simplify notation let $\sigma''_n = \frac{1}{\lambda^*} \log n - \gamma_n$. Recall that $\xi(t)$ is the point process that encodes the reproduction times of an individual in the continuous time embedding of our discrete tree process. We can write

$$P(A_n^c) = P(\Phi_1(d_0(n)) < \alpha_n) = P(\Phi_1(\xi(T_n)) < \alpha_n) \leq P(T_n < \sigma''_n) + P(\Phi_1(\xi(\sigma''_n)) < \alpha_n).$$

First we will estimate $P(\Phi_1(\xi(\sigma''_n)) < \alpha_n)$. Recall $S_1(\cdot)$ from (1.2). Note that

$$\begin{aligned} P(\Phi_1(\xi(\sigma''_n)) < \alpha_n) &= P(\xi(\sigma''_n) < \Phi_1^{-1}(\alpha_n)) \leq P(S_1([\Phi_1^{-1}(\alpha_n)]) \geq \sigma''_n) \\ &= P(S_1([\Phi_1^{-1}(\alpha_n)]) - \Phi_1([\Phi_1^{-1}(\alpha_n)]) \geq \sigma''_n - \alpha_n) \\ &\leq P\left(S_1([\Phi_1^{-1}(\alpha_n)]) - \Phi_1([\Phi_1^{-1}(\alpha_n)] + 1) \geq \sigma''_n - \alpha_n - \frac{2}{f_*([\Phi_1^{-1}(\alpha_n)])}\right). \end{aligned} \tag{7.3}$$

For any $l \in \mathbb{N}$ and $\theta \in (0, f_*)$, we have

$$\begin{aligned} \log E[\exp(\theta(S_1(l) - \Phi_1(l+1)))] &= -\theta\Phi_1(l+1) - \sum_{i=1}^{l-1} \log\left(1 - \frac{\theta}{f(i)}\right) \\ &= \sum_{k=2}^{\infty} \frac{\theta^k}{k} \Phi_k(l+1) \leq \theta^2 \Phi_2(l+1) \sum_{k=2}^{\infty} \left(\frac{\theta}{f_*}\right)^{k-2} \\ &= \frac{f_*}{f_* - \theta} \theta^2 \Phi_2(l+1). \end{aligned}$$

It follows that for any $x > 0$, $l \in \mathbb{N}$ and $\theta \in (0, f_*)$,

$$P(S_1(l) - \Phi_1(l+1) \geq x) \leq e^{-\theta x} \exp\left[\frac{f_*}{f_* - \theta} \theta^2 \Phi_2(l+1)\right] \leq e^{-\theta x} \exp\left[\frac{f_*}{f_* - \theta} \theta^2 \Phi_2(\infty)\right]. \tag{7.4}$$

Hence, by (7.3) and (7.4), we obtain for any $\theta \in (0, f_*)$,

$$P(\Phi_1(\xi(\sigma''_n)) < \alpha_n) \leq e^{2\theta/f_*} e^{-\theta(\sigma''_n - \alpha_n)} \exp\left[\frac{f_*}{f_* - \theta} \theta^2 \Phi_2(\infty)\right] = C_\theta^* e^{\theta(\alpha_n + \gamma_n)} n^{-\theta/\lambda^*},$$

for any $n \geq 1$, where $C_\theta^* := e^{2\theta/f_*} \exp\left[\frac{f_*}{f_* - \theta} \theta^2 \Phi_2(\infty)\right]$. Recall that $\alpha_n = \eta \log n$. Since θ can be chosen arbitrarily close to f_* , for any $\delta > 0$ we can choose $\theta = (1 - \delta/4)f_*$ so that, writing $C'_\delta := C_{(1-\delta/4)f_*}^*$,

$$P(\Phi_1(\xi(\sigma''_n)) < \alpha_n) \leq C'_\delta e^{f_*(\alpha_n + \gamma_n)} n^{-\frac{(1-\delta/4)f_*}{\lambda^*}} = C'_\delta e^{f_*\gamma_n} n^{-\frac{(1-\lambda^*\eta - \delta/4)f_*}{\lambda^*}}, \quad n \geq 1.$$

Now, choosing $\gamma_n = \frac{\delta}{4\lambda^*} \log n$ and any integer $k = k_\delta > 4(1 - \delta)f_*/(\delta\lambda^*)$, (3.9) yields

$$P(T_n < \sigma''_n) \leq C_{k_\delta} n^{-\delta f_* k_\delta / 4} \leq C_{k_\delta} n^{-\frac{(1-\delta)f_*}{\lambda^*}}.$$

Therefore, taking $\eta = \delta/(2\lambda^*)$,

$$P(A_n^c) \leq C'_\delta e^{f_*\gamma_n} n^{-\frac{(1-\lambda^*\eta - \delta/4)f_*}{\lambda^*}} + C_{k_\delta} n^{-\frac{(1-\delta)f_*}{\lambda^*}} = \tilde{C}_\delta n^{-\frac{(1-\delta)f_*}{\lambda^*}},$$

where $\tilde{C}_\delta := C'_\delta + C_{k_\delta}$. Therefore, using (7.2) and recalling $M > f_*/\lambda^*$, we obtain for all $n \geq 1$,

$$P(E_n) \leq P(E_n \cap A_n) + P(A_n^c) \leq \hat{C}'_\delta n^{-M} + \tilde{C}'_\delta n^{-\frac{(1-\delta)f_*}{\lambda^*}} \leq \tilde{C}'_\delta n^{-\frac{(1-\delta)f_*}{\lambda^*}},$$

where $\hat{C}'_\delta := \hat{C}_{\delta/(2\lambda^*)}$ and $\tilde{C}'_\delta := \hat{C}'_\delta + \tilde{C}_\delta$. The above bound gives the following upper bound on $K_\Psi(\varepsilon)$:

$$K_\Psi(\varepsilon) \leq \left\lceil \left(\frac{\tilde{C}'_\delta}{\varepsilon} \right)^{\frac{\lambda^*}{(1-\delta)f_*}} \right\rceil,$$

which proves the lemma. □

7.2 For general $m > 1$

Lower bound.

Lemma 7.3. *Under the assumptions in Theorem 2.3 with $m > 1$, there exists some $C_1 > 0$ so that*

$$K_\Psi(\varepsilon) \geq C_1 \varepsilon^{-\frac{f_*}{mf(m)}}.$$

Proof. The proof is essentially the same as that of Lemma 6.1 with a different lower bound on $P(\mathcal{A}_K)$ for the event $\mathcal{A}_K = \{v_0 \text{ has degree } m \text{ in } \mathcal{G}_K\}$. Let $i_0 \geq 1$ be such that $\frac{f(m)}{f_*(i_0+1)} \leq 1/2$.

$$\begin{aligned} P(\mathcal{A}_K) &= E(1\{\mathcal{A}_{K-1}\}P(\mathcal{A}_K|\mathcal{G}_{K-1})) \\ &= E\left(1\{\mathcal{A}_{K-1}\} \prod_{i=1}^m \left(1 - \frac{f(m)}{\sum_{j=0}^{K-1} f(d_j(s_{K-1} + i - 1))}\right)\right) \\ &\geq E\left(1\{\mathcal{A}_{K-1}\} \left(1 - \frac{f(m)}{Kf_*}\right)^m\right) \geq P(\mathcal{A}_{i_0}) \prod_{i=i_0}^{K-1} \left(1 - \frac{f(m)}{(i+1)f_*}\right)^m \\ &= P(\mathcal{A}_{i_0}) \exp\left(m \sum_{i=i_0}^{K-1} \log\left(1 - \frac{f(m)}{(i+1)f_*}\right)\right) \geq CK^{-\frac{mf(m)}{f_*}}. \end{aligned}$$

Again, applying Lemma 5.1 with the above estimate gives the desired the result. □

7.2.1 Upper bound

Lemma 7.4. *Under the assumptions in Theorem 2.3 with $m > 1$, for any $\delta \in (0, 1)$, there exists $\bar{C}_\delta > 0$ such that for all $\varepsilon \in (0, 1)$,*

$$K_\Psi(\varepsilon) \leq \bar{C}_\delta \varepsilon^{-\frac{2C_f}{(1-\delta)f_*}}.$$

Proof. We will use the construction and terminology of the collapsed branching process embedding defined in Section 3. Let $\xi^{(0)}(t)$ denote the ‘continuous time’ degree of the root at time t . Let $\{\sigma_n^\circ\}_{n \geq 1}$, $\{\alpha_n^\circ\}_{n \geq 1}$ be increasing sequences to be specified later. We follow the same strategy as in Section 6.2. Define the events $A_n = \{\Phi_1(d_0(s_n)) \geq \alpha_n^\circ\}$ and $E_n = \{d_n(k) \geq d_0(k) \text{ for some } k \geq s_n\}$. We begin by upper bounding $P(A_n^c)$,

$$\begin{aligned} P(A_n^c) &= P(\Phi_1(d_0(s_n)) < \alpha_n^\circ) = P(\Phi_1(\xi^{(0)}(\tau_n)) < \alpha_n^\circ) \\ &\leq P(\Phi_1(\xi^{(0)}(\sigma_n^\circ)) < \alpha_n^\circ) + P(\tau_n < \sigma_n^\circ). \end{aligned}$$

Take any $l \geq 1$. Recall that $\tau_{l+1} = \tau_l + \Theta_{l+1}$, where Θ_{l+1} is the first time since τ_l when there are m newly born individuals $v_{l+1,1}, \dots, v_{l+1,m}$. Let $\Theta_{l+1,i}$ denote the arrival time

of $v_{l+1,i}$. Also, denote by \mathcal{F}_l the filtration generated by the collapsed branching process till time τ_l . It is not hard to see that, conditional on \mathcal{F}_l , the distribution of $\Theta_{l+1,1}$ follows the exponential distribution with rate $\sum_{i=0}^l f(d_i(s_l))$. By Assumption (A1), $f(i) \leq C_f i$ for all $i \geq m$, so $\sum_{i=0}^l f(d_i(s_l)) \leq C_f \sum_{i=0}^l d_i(s_l) = 2s_l C_f$. That is, $\Theta_{l+1,1}$ stochastically dominates an exponential random variable with rate $2s_l C_f$. Let $\{E_{l,i} : 1 \leq i \leq m, l \geq 1\}$ denote a collection of i.i.d. exponential random variables with rate 1. Then

$$\Theta_{l+1} = \Theta_{l+1,m} \geq \sum_{i=1}^m \frac{E_{l,i}}{2s_l C_f} = \sum_{i=1}^m \frac{E_{l,i}}{2m C_f l}.$$

Hence, for $\theta_n = \log n$,

$$\begin{aligned} Ee^{-\theta_n \tau_n} &= E \exp\left(-\theta_n \sum_{l=2}^n \Theta_l\right) \leq E \exp\left(-\theta_n \sum_{l=2}^n \sum_{i=1}^m \frac{E_{l,i}}{2m C_f l}\right) \\ &= \exp\left(-m \sum_{l=2}^n \log\left(1 + \frac{\theta_n}{2m C_f l}\right)\right) \leq C \exp\left(-m \sum_{l=\lfloor \theta_n / (m C_f) \rfloor + 1}^n \frac{\theta_n}{2m C_f l}\right) \\ &\leq C \exp\left(-\frac{\theta_n}{2C_f} (\log n - \log \log n - C')\right), \end{aligned}$$

where we used $\log(1+x) \geq x - x^2$ for $x \in [0, 1/2]$ to obtain the second inequality above. Take any $\delta \in (0, 1)$. Let $\sigma_n^\circ = \frac{(1-\delta/4) \log n}{2C_f}$. Then

$$\begin{aligned} P(\tau_n < \sigma_n^\circ) &\leq e^{\theta_n \sigma_n^\circ} Ee^{-\theta_n \tau_n} \\ &\leq C e^{\theta_n \sigma_n^\circ} \exp\left(-\frac{\theta_n}{2C_f} (\log n - \log \log n - C')\right) \\ &\leq C'' n^{-\theta_n \delta / 16 C_f} \leq \tilde{C} n^{-\frac{f_*}{2C_f}} \quad n \geq n_\delta, \end{aligned}$$

for positive constants C'', \tilde{C} , when n_δ is chosen sufficiently large. The same calculation as in (7.3) and (7.4) implies for any $\theta \in (0, f_*)$,

$$P(\Phi_1(\xi(\sigma_n^\circ)) < \alpha_n^\circ) \leq e^{2\theta/f_*} e^{-\theta(\sigma_n^\circ - \alpha_n^\circ)} \exp\left[\frac{f_*}{f_* - \theta} \theta^2 \Phi_2(\infty)\right] = C_\theta^* e^{\theta \alpha_n^\circ} n^{-(1-\delta/4)\theta/2C_f},$$

for any $n \geq 1$, where $C_\theta^* := e^{2\theta/f_*} \exp\left[\frac{f_*}{f_* - \theta} \theta^2 \Phi_2(\infty)\right]$. Choose $\alpha_n^\circ = (\delta/(8C_f)) \log n$. Since θ can be chosen arbitrarily close to f_* , for any $\delta > 0$ we can choose $\theta = (1 - \delta/4)f_*$ so that, writing $C'_\delta := C_{(1-\delta/4)f_*}^*$,

$$P(\Phi_1(\xi(\sigma_n^\circ)) < \alpha_n^\circ) \leq C'_\delta n^{\frac{\delta f_*}{8C_f} - \frac{(1-\delta/4)^2 f_*}{2C_f}} \leq C'_\delta n^{-\frac{(1-\delta)f_*}{2C_f}}, \quad n \geq 1.$$

Again, using (7.2) and recalling $M > f_*/(2C_f)$ gives that for all $n \geq 1$,

$$P(E_n) \leq P(E_n \cap A_n) + P(A_n^c) \leq \hat{C}'_\delta n^{-M} + C'_\delta n^{-\frac{(1-\delta)f_*}{2C_f}} \leq \tilde{C}'_\delta n^{-\frac{(1-\delta)f_*}{2C_f}},$$

where $\hat{C}'_\delta := \hat{C}_{\delta/(8C_f)}$ and $\tilde{C}'_\delta := \hat{C}'_\delta + C'_\delta$. The above bound gives the following upper bound on $K_\Psi(\varepsilon)$:

$$K_\Psi(\varepsilon) \leq \left\lceil \left(\frac{\tilde{C}'_\delta}{\varepsilon}\right)^{\frac{2C_f}{(1-\delta)f_*}} \right\rceil$$

which proves the lemma. □

Proof of Theorem 2.3. The lower and upper bounds in part (i) ($m_i \equiv m = 1$ case) follow respectively from Lemma 7.1 and Lemma 7.2. Those for part (ii) ($m_i \equiv m > 1$ case) respectively follow from Lemma 7.3 and Lemma 7.4. □

8 Non-persistent regime: proof of Theorem 2.5

This section is dedicated to the proof of Theorem 2.5. We will assume throughout that $m_i \equiv m = 1$. For $n \in \mathbb{N}$, denote by $v_{\max}(n)$ the youngest vertex with the maximal degree in \mathcal{G}_n .

The following lemma quantifies the probability of $v_{\max}(n)$ lying within a certain distance of the root in \mathcal{G}_n .

Lemma 8.1. *Let $r_n = c_1 \lambda^* \mathcal{K}(\frac{1}{\lambda^*} \log n)$ where c_1 is defined as in Lemma 3.12. There exist $C, C' > 0$ so that for all $n \geq 1$,*

$$P(v_{\max}(n) \notin B_n(v_0, r_n)) \leq C \exp\left(-C' \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right).$$

Proof. Let $a_n = \exp\left(\frac{\lambda^{*2}}{2} \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right)$, $t_n = \frac{1}{\lambda^*} \log a_n$ and $r_n = 2c_1 t_n$. Recall $\mathcal{I}_c^*(t)$ as defined in (3.13). Using the continuous time embedding of the discrete tree process,

$$P(v_{\max}(n) \in B_n(v_0, r_n)) \geq P(H(\mathcal{I}_c^*(T_n)) \leq r_n) \geq P(\mathcal{I}_c^*(T_n) \leq 2t_n, H(2t_n) \leq r_n).$$

Hence,

$$P(v_{\max}(n) \notin B_n(v_0, r_n)) \leq P(\mathcal{I}_c^*(T_n) \geq 2t_n) + P(H_{2t_n} \geq r_n). \quad (8.1)$$

By Lemma 3.8 with $\varepsilon = \lambda^*/2$, there exists $C, C' > 0$ so that

$$P(\mathcal{I}_c^*(T_n) \geq 2t_n) \leq C \exp(-C' \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)).$$

Applying Lemma 3.12 to the second term in the bound (8.1),

$$P(H_{2t_n} \geq r_n) = P(H_{2t_n} \geq 2c_1 t_n) \leq e^{-t_n} \leq \exp\left(-\frac{\lambda^*}{2} \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right).$$

The lemma follows on using the above two bounds in (8.1). \square

The following lemma introduces a key functional $\tilde{\alpha}^*(\cdot)$ which arises in quantifying the exponential growth rate of the number of vertices within a certain radius of the root (see (8.5)). We will require understanding of $\tilde{\alpha}^*(\cdot)$ near zero, as described in the next lemma.

Lemma 8.2. *Define $\tilde{\alpha}^*(x) = \inf_{0 \leq \theta \leq \lambda^*} \{x \log \hat{\mu}(\theta) + \theta\}$. Suppose $f(i) \leq C_0 i^\alpha$ for some constant C_0 and $\alpha \in (0, 1/2]$. For any $L > 0$, there exists some constant $C > 0$ depending on α, L so that*

$$\tilde{\alpha}^*(x) \leq Cx - \frac{2}{1-\alpha} x \log x, \quad x \in (0, L]. \quad (8.2)$$

Proof. All constants in the proof will depend on α, λ^*, f_* but not θ . For $x \in (0, 1)$, since $\log(1-x) \leq -x$,

$$\hat{\mu}(\theta) = \sum_{k=1}^{\infty} \prod_{i=1}^k \frac{f(i)}{\theta + f(i)} = \sum_{k=1}^{\infty} \exp\left(\sum_{i=1}^k \log\left(1 - \frac{\theta}{\theta + f(i)}\right)\right) \leq \sum_{k=1}^{\infty} \exp\left(-\sum_{i=1}^k \frac{\theta}{\theta + f(i)}\right).$$

Since we are considering $\theta \in [0, \lambda^*]$, $\theta + f(i) \leq \frac{\lambda^*}{f_*} f(i) + f(i) = \frac{f_* + \lambda^*}{f_*} f(i)$. It follows that there exists a positive constant C_1 such that

$$\hat{\mu}(\theta) \leq \sum_{k=1}^{\infty} \exp\left(-\frac{\theta f_*}{f_* + \lambda^*} \sum_{i=1}^k \frac{1}{f(i)}\right) \leq \sum_{k=1}^{\infty} \exp(-C_1 \theta k^{1-\alpha}). \quad (8.3)$$

Observe that there exists some constant $C' > 0$ so that

$$\exp(-C_1\theta k^{1-\alpha}) \leq \exp(-C_1k^{(1-\alpha)/2}) \leq \frac{C'}{k^2}$$

for all $k > \theta^{-\frac{2}{1-\alpha}}$. Hence,

$$\begin{aligned} \hat{\mu}(\theta) &\leq \sum_{k=1}^{\lfloor \theta^{-\frac{2}{1-\alpha}} \rfloor} \exp(-C_1\theta) + \sum_{k>\theta^{-\frac{2}{1-\alpha}}} \exp(-C_1\theta k^{1-\alpha}) \\ &\leq \theta^{-\frac{2}{1-\alpha}} \exp(-C_1\theta) + \sum_{k>\theta^{-\frac{2}{1-\alpha}}} \frac{C'}{k^2} \\ &\leq \theta^{-\frac{2}{1-\alpha}} \exp(-C_1\theta) + C''\theta^{\frac{2}{1-\alpha}} \leq C_2\theta^{-\frac{2}{1-\alpha}} \exp(-C_1\theta), \end{aligned}$$

where $C_2 = 1 + C''(\lambda^*)^{4/(1-\alpha)}e^{C_1\lambda^*}$. Therefore,

$$\begin{aligned} \inf_{0 \leq \theta \leq \lambda^*} \{x \log \hat{\mu}(\theta) + \theta\} &\leq \inf_{0 \leq \theta \leq \lambda^*} \left\{ x \log \left(C_2 e^{-C_1\theta} \cdot \frac{1}{\theta^{\frac{2}{1-\alpha}}} \right) + \theta \right\} \\ &= \inf_{0 \leq \theta \leq \lambda^*} \left\{ x \left(\log C_2 - C_1\theta - \frac{2}{1-\alpha} \log \theta \right) + \theta \right\}. \end{aligned}$$

Let $h(\theta) = (1 - C_1x)\theta - \frac{2x}{1-\alpha} \log \theta$. The minimum for the function h is achieved at $\hat{\theta} = \frac{2x}{(1-\alpha)(1-C_1x)}$. If $x \in [0, \frac{1}{2C_1} \wedge \frac{(1-\alpha)\lambda^*}{4}]$, $\hat{\theta} \in [0, \lambda^*]$. Hence, for $x \in (0, \frac{1}{2C_1} \wedge \frac{(1-\alpha)\lambda^*}{4} \wedge 1)$,

$$\begin{aligned} \tilde{\alpha}^*(x) &\leq x \log C_2 + \frac{2x}{1-\alpha} - \frac{2x}{1-\alpha} \log \left(\frac{2x}{(1-\alpha)(1-C_1x)} \right) \\ &\leq \left(\log C_2 + \frac{2}{1-\alpha}(1 - \log 2) \right) x - \frac{2}{1-\alpha} \cdot x \log x \\ &= C_3x - \frac{2}{1-\alpha}x \log x, \end{aligned}$$

where $C_3 = \log C_2 + \frac{2}{1-\alpha}(1 - \log 2)$. This proves (8.2) for $x \in (0, \frac{1}{2C_1} \wedge \frac{(1-\alpha)\lambda^*}{4} \wedge 1)$. The constant C in (8.2) can be chosen suitably large so that (8.2) continues to hold for $x \in [\frac{1}{2C_1} \wedge \frac{(1-\alpha)\lambda^*}{4} \wedge 1, L]$. This proves the lemma. \square

The next lemma quantifies tail probabilities for the distribution of the number of vertices in a ball of radius r_n around the root. In the following, $|B(v_0, R_n)|$ denotes the number of vertices in $B(v_0, R_n)$.

Lemma 8.3. *Suppose $f(i) \leq C_0i^\alpha$ for some constant C_0 and $\alpha \in (0, 1/2]$. Let the sequence $\{R_n\}_{n \geq 1}$ be given by $R_n = 2r_n$ where r_n is defined in Lemma 8.1. Let $b_n := \exp\left(\frac{4}{1-\alpha}R_n \log\left(\frac{\log n}{\lambda^*R_n}\right)\right)$. Then there exists $C > 0$ such that for all $n \geq 1$,*

$$P(|B_n(v_0, R_n)| \geq b_n) \leq C \exp\left(-\frac{f^*}{2}\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right).$$

Proof. Recall the embedding of $\{\mathcal{G}_l\}_{l \geq 1}$ into a continuous time branching process $\widetilde{BP}_f(\cdot)$ described in Lemma 3.2. Let $F(t) := E[\xi[0, t]]$ and, for $n \geq 0$, let $Z_n(t)$ denote the number of vertices in generation n at time t that are descendants of v_0 . Equation (1.1) in [Big77] gives for $n \geq 0$,

$$F^{n*}(t) = E[Z_n(t)],$$

where F^{n*} is the n -fold Stieltjes convolution of the increasing function F . For $n = 0$ we let $F^{0*}(t) \equiv 1$. It follows from Theorem 1 in [Kin75] that for all $\theta > \underline{\lambda}$,

$$\int_0^\infty e^{-\theta t} dE[Z_n(t)] = \hat{\mu}(\theta)^n,$$

which gives for any $t > 0, n \geq 1$,

$$e^{-\theta t} F^{n*}(t) \leq \int_0^t e^{-\theta s} dE[Z_n(s)] \leq \hat{\mu}(\theta)^n.$$

Let $\check{\sigma}_n = \frac{1}{\lambda^*} \log n + \gamma_n$ where $\gamma_n = o(\log n)$ and will be specified later. Define $x_n = R_n / \check{\sigma}_n$. By (3.23), $x_n \rightarrow 0$ as $n \rightarrow \infty$. By (8.3), $\underline{\lambda} := \inf\{\lambda > 0 : \hat{\mu}(\lambda) < \infty\} = 0$. For any $\theta > 0$, writing $Z_{\leq n}(t)$ for the number of vertices in generation at most n at time t that are descendants of v_0 ,

$$EZ_{\leq \lfloor R_n \rfloor}(\check{\sigma}_n) = \sum_{k=0}^{\lfloor R_n \rfloor} F^{k*}(\check{\sigma}_n) \leq \sum_{k=0}^{\lfloor R_n \rfloor} e^{\theta \check{\sigma}_n} \hat{\mu}(\theta)^k = e^{\theta \check{\sigma}_n} \cdot \frac{\hat{\mu}(\theta)^{\lfloor R_n \rfloor + 1} - 1}{\hat{\mu}(\theta) - 1}. \quad (8.4)$$

Let $\beta := -\hat{\mu}'(\lambda^*)$. By the representation (5.40) of [HJ16], $\beta > 0$. It is not hard to show that for $x \in (0, \beta^{-1})$,

$$\theta_x := \arg \inf_{0 \leq \theta \leq \lambda^*} \{\theta + x \log \hat{\mu}(\theta)\} \in (0, \lambda^*),$$

see the proof of (13.133) in [HJ16]. Hence, for $x \in (0, \beta^{-1})$, $\hat{\mu}(\theta_x) > 1$. Obtain $n_0 \in \mathbb{N}$ such that $x_n < \beta^{-1}$ for all $n \geq n_0$. For $n \geq n_0$, taking $\theta = \theta_{x_n}$ in (8.4) gives

$$\begin{aligned} EZ_{\leq \lfloor R_n \rfloor}(\check{\sigma}_n) &\leq e^{\theta_{x_n} \check{\sigma}_n} \cdot \frac{\hat{\mu}(\theta_{x_n})^{\lfloor R_n \rfloor + 1} - 1}{\hat{\mu}(\theta_{x_n}) - 1} \leq \frac{\hat{\mu}(\theta_{x_n})}{\hat{\mu}(\theta_{x_n}) - 1} \exp(\check{\sigma}_n (\theta_{x_n} + x_n \log \hat{\mu}(\theta_{x_n}))) \\ &= c(x_n) \exp(\tilde{\alpha}^*(x_n) \check{\sigma}_n), \end{aligned} \quad (8.5)$$

where $c(x) := \frac{\hat{\mu}(\theta_x)}{\hat{\mu}(\theta_x) - 1}$. Note that $\theta_{x_n} \rightarrow 0$ as $x_n \rightarrow 0$. Otherwise, we could obtain a subsequence $\{n_j\}$ such that $\liminf_j \theta_{x_{n_j}} > 0$. Then,

$$\liminf_j \tilde{\alpha}^*(x_{n_j}) = \liminf_j [\theta_{x_{n_j}} + x \log \hat{\mu}(\theta_{x_{n_j}})] \geq \liminf_j [\theta_{x_{n_j}} + x \log \hat{\mu}(\lambda^*)] = \liminf_j \theta_{x_{n_j}} > 0,$$

which is a contradiction to (8.2). Thus, $\hat{\mu}(\theta_{x_n}) \rightarrow \infty$ as $x_n \rightarrow 0$, which gives $\lim_{x_n \rightarrow 0} c(x_n) = 1$. That is, $C_0 := \sup\{c(x_n) : n \in \mathbb{N}\} < \infty$, which gives

$$EZ_{\leq \lfloor R_n \rfloor}(\check{\sigma}_n) \leq C_0 \exp(\tilde{\alpha}^*(x_n) \check{\sigma}_n).$$

Therefore,

$$\begin{aligned} P(Z_{\leq \lfloor R_n \rfloor}(T_n) \geq b_n/2) &\leq P(Z_{\leq \lfloor R_n \rfloor}(T_n) \geq b_n/2, T_n \leq \check{\sigma}_n) + P(T_n > \check{\sigma}_n) \\ &\leq P(Z_{\leq \lfloor R_n \rfloor}(\check{\sigma}_n) \geq b_n/2) + P(T_n > \check{\sigma}_n) \\ &\leq \frac{2C_0 \exp(\tilde{\alpha}^*(x_n) \check{\sigma}_n)}{b_n} + P(T_n > \check{\sigma}_n). \end{aligned}$$

Taking $\gamma_n = \mathcal{K}(\frac{1}{\lambda^*} \log n)$, Lemma 3.7 (i) with $a = f_*/(2\lambda^*)$ gives an upper bound on the second term,

$$P(T_n > \check{\sigma}_n) \leq C e^{-\frac{f_*}{2} \gamma_n} = C \exp\left(-\frac{f_*}{2} \mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right).$$

As $\lim_{l \rightarrow \infty} x_l = 0$, therefore, $\sup_l x_l < \infty$. Thus, it follows from Lemma 8.2 upon taking $L = \sup_l x_l$ that there exists some constant $C' > 0$ so that for all $n \geq 1$,

$$\begin{aligned} \exp(\tilde{\alpha}^*(x_n)\check{\sigma}_n) &\leq \exp\left(CR_n + \frac{2}{1-\alpha}R_n \log\left(\frac{\check{\sigma}_n}{R_n}\right)\right) \\ &\leq C' \exp\left(\frac{3}{1-\alpha}R_n \log\left(\frac{\log n}{\lambda^*R_n}\right)\right). \end{aligned}$$

Hence, recalling $b_n = \exp\left(\frac{4}{1-\alpha}R_n \log\left(\frac{\log n}{\lambda^*R_n}\right)\right)$,

$$\begin{aligned} P(Z_{\leq \lfloor R_n \rfloor}(T_n) \geq b_n/2) &\leq C \exp\left(-\frac{f^*}{2}\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right) + 2C_0C' \exp\left(-\frac{1}{1-\alpha}R_n \log\left(\frac{\log n}{\lambda^*R_n}\right)\right) \\ &\leq C'' \exp\left(-\frac{f^*}{2}\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right). \end{aligned}$$

Let $\bar{Z}_{\leq \lfloor R_n \rfloor}(t)$ denote the number of descendants of v_1 by time t that are within generation $\lfloor R_n \rfloor$. It is easy to see $\bar{Z}_{\leq \lfloor R_n \rfloor}(t)$ has the same law as $Z_{\leq \lfloor R_n \rfloor}(t)$. The lemma now follows upon noting

$$P(|B_n(v_0, R_n)| \geq b_n) \leq P(Z_{\leq \lfloor R_n \rfloor}(T_n) + \bar{Z}_{\leq \lfloor R_n \rfloor}(T_n) \geq b_n) \leq 2P(Z_{\leq \lfloor R_n \rfloor}(T_n) \geq b_n/2). \quad \square$$

Proof of Theorem 2.5. Define the event $\mathcal{E}^{(n)} := \{|B_n(v_{\max}(n), r_n)| \leq b_n, v_0 \in B_n(v_{\max}(n), r_n)\}$. Then

$$\begin{aligned} P(\mathcal{E}^{(n)}) &= P(|B_n(v_{\max}(n), r_n)| \leq b_n, v_{\max}(n) \in B_n(v_0, r_n)) \\ &\geq P(|B_n(v_0, 2r_n)| \leq b_n, v_{\max}(n) \in B_n(v_0, r_n)). \end{aligned}$$

It follows from Lemma 8.3 and Lemma 8.1 that there are $a, C_1, C_2, C_3, C, C' > 0$, so that

$$\begin{aligned} P((\mathcal{E}^{(n)})^c) &\leq P(|B_n(v_0, 2r_n)| \geq b_n) + P(v_{\max}(n) \notin B_n(v_0, r_n)) \\ &\leq C_1 \exp\left(-\frac{f^*}{2}\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right) + C_2 \exp(-C_3\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)) \\ &\leq C \exp\left(-C'\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)\right). \end{aligned}$$

For any $\varepsilon \in (0, 1)$, the above bound is $\leq \varepsilon$ when $n \geq \exp\left(\lambda^*\mathcal{K}^{-1}\left(\frac{\log(C/\varepsilon)}{C'}\right)\right)$. This proves the theorem. \square

References

- [AK68] Krishna B. Athreya and Samuel Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.*, 39(6):1801–1817, 12 1968. MR0232455
- [All10] Linda JS Allen. *An introduction to stochastic processes with applications to biology*. CRC Press, 2010. MR2560499
- [BB21] Sayan Banerjee and Shankar Bhamidi. Persistence of hubs in growing random networks. *Probability Theory and Related Fields*, 180(3-4):891–953, 2021. MR4288334
- [BB22] Sayan Banerjee and Shankar Bhamidi. Root finding algorithms and persistence of jordan centrality in growing random trees. *The Annals of Applied Probability*, 32(3):2180–2210, 2022. MR4430011
- [BBC⁺12] Christian Borgs, Michael Brautbar, Jennifer Chayes, Sanjeev Khanna, and Brendan Lucier. The power of local information in social networks. In *International Workshop on Internet and Network Economics*, pages 406–419. Springer, 2012.

- [BBC22] Sayan Banerjee, Shankar Bhamidi, and Iain Carmichael. Fluctuation bounds for continuous time branching processes and nonparametric change point detection in growing networks. *The Annals of Applied Probability (to appear)*, 2022.
- [BDL17] Sébastien Bubeck, Luc Devroye, and Gábor Lugosi. Finding Adam in random growing trees. *Random Structures & Algorithms*, 50(2):158–172, 2017. MR3607120
- [BDOC23] Sayan Banerjee, Prabhanka Deka, and Mariana Olvera-Cravioto. Local weak limits for collapsed branching processes with random out-degrees. *arXiv preprint arXiv:2302.00562*, 2023.
- [BEMR17] Sébastien Bubeck, Ronen Eldan, Elchanan Mossel, and Miklós Z Rácz. From trees to seeds: on the inference of the seed from large trees in the uniform attachment model. *Bernoulli*, 23(4A):2887–2916, 2017. MR3648049
- [BG79] John D Biggins and DR Grey. Continuity of limit random variables in the branching random walk. *Journal of Applied Probability*, pages 740–749, 1979. MR0549554
- [Big77] JD Biggins. Chernoff’s theorem in the branching random walk. *Journal of Applied Probability*, 14(3):630–636, 1977. MR0464415
- [BK10] Michael Brautbar and Michael J Kearns. Local algorithms for finding interesting individuals in large networks. *Innovations in Theoretical Computer Science (ITCS)*, 2010.
- [BMR15] Sébastien Bubeck, Elchanan Mossel, and Miklós Z Rácz. On the influence of the seed graph in the preferential attachment model. *IEEE Transactions on Network Science and Engineering*, 2(1):30–39, 2015. MR3361606
- [BRST01] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001. MR1824277
- [BV14] Paolo Boldi and Sebastiano Vigna. Axioms for centrality. *Internet Mathematics*, 10(3-4):222–262, 2014. MR3259267
- [CDKM15] Nicolas Curien, Thomas Duquesne, Igor Kortchemski, and Ioan Manolescu. Scaling limits and influence of the seed graph in preferential attachment trees. *Journal de l’École polytechnique-Mathématiques*, 2:1–34, 2015. MR3326003
- [DM09] Steffen Dereich and Peter Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electronic Journal of Probability*, 14:1222–1267, 2009. MR2511283
- [DR18] Luc Devroye and Tommy Reddad. On the discovery of the seed in uniform attachment trees. *arXiv preprint arXiv:1810.00969*, 2018. MR3934290
- [FP17] Alan Frieze and Wesley Pegden. Looking for vertex number one. *Annals of Applied Probability*, 27(1):582–630, 2017. MR3619796
- [Fre77] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.
- [Gal13] Pavel Galashin. Existence of a persistent hub in the convex preferential attachment model. *arXiv preprint arXiv:1310.7513*, 2013. MR3529340
- [GvdH18] 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. MR3877316
- [HJ16] C Holmgren and S Janson. Fringe trees, crump–mode–jagers branching processes and m-ary search trees. preprint. *arXiv preprint arXiv:1601.03691*, 2016. MR3626585
- [JL16] Varun Jog and Po-Ling Loh. Analysis of centrality in sublinear preferential attachment trees via the Crump-Mode-Jagers branching process. *IEEE Transactions on Network Science and Engineering*, 4(1):1–12, 2016. MR3625951
- [JL18] Varun Jog and Po-Ling Loh. Persistence of centrality in random growing trees. *Random Structures & Algorithms*, 52(1):136–157, 2018. MR3731614
- [JN84] Peter Jagers and Olle Nerman. The growth and composition of branching populations. *Advances in Applied Probability*, 16(2):221–259, 1984. MR0742953
- [Kin75] John Frank Charles Kingman. The first birth problem for an age-dependent branching process. *The Annals of Probability*, 3(5):790–801, 1975. MR0400438

- [KL16] Justin Khim and Po-Ling Loh. Confidence sets for the source of a diffusion in regular trees. *IEEE Transactions on Network Science and Engineering*, 4(1):27–40, 2016. MR3625953
- [Liu01] Quansheng Liu. Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stochastic processes and their applications*, 95(1):83–107, 2001. MR1847093
- [LP19] Gabor Lugosi and Alan S Pereira. Finding the seed of uniform attachment trees. *Electronic Journal of Probability*, 24, 2019. MR3925458
- [MR19] Tamás F Móri and Sándor Rokob. Moments of general time dependent branching processes with applications. *Acta Mathematica Hungarica*, 159(1):131–149, 2019. MR4003699
- [Ner81] Olle Nerman. On the convergence of supercritical general (CMJ) branching processes. *Probability Theory and Related Fields*, 57(3):365–395, 1981. MR0629532
- [New05] Mark E.J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.
- [NK11] Saket Navlakha and Carl Kingsford. Network archaeology: uncovering ancient networks from present-day interactions. *PLoS computational biology*, 7(4), 2011. MR2805381
- [PRR17] Erol Peköz, Adrian Röllin, and Nathan Ross. Joint degree distributions of preferential attachment random graphs. *Advances in Applied Probability*, 49(2):368–387, 2017. MR3668381
- [RTV07] Anna Rudas, Bálint Tóth, and Benedek Valkó. Random trees and general branching processes. *Random Structures & Algorithms*, 31(2):186–202, 2007. MR2343718
- [Sén21] Delphin Sénizergues. Geometry of weighted recursive and affine preferential attachment trees. *Electronic Journal of Probability*, 26:1–56, 2021. MR4269210
- [SZ11] Devavrat Shah and Tauhid Zaman. Rumors in a Network: Who’s the Culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011. MR2849111
- [SZ12] Devavrat Shah and Tauhid Zaman. Rumor centrality: a universal source detector. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems*, pages 199–210, 2012.

Acknowledgments. We are grateful to Shankar Bhamidi for continued discussions throughout the project. We thank Pratima Hebbar for very helpful advice. We also thank an anonymous referee and an associate editor whose suggestions substantially improved this article.

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System: <https://vtex.lt/services/ejms-peer-review/>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <https://imstat.org/shop/donation/>