CHANGE POINT DETECTION IN NETWORK MODELS: PREFERENTIAL ATTACHMENT AND LONG RANGE DEPENDENCE

BY SHANKAR BHAMIDI[1], JIMMY JIN[2] AND ANDREW NOBEL[3]

University of North Carolina

Inspired by empirical data on real world complex networks, the last few years have seen an explosion in proposed generative models to understand and explain observed properties of real world networks, including power law degree distribution and "small world" distance scaling. In this context, a natural question is how to understand the effect of *change points*—how abrupt changes in parameters driving the network model change structural properties of the network. We study this phenomenon in one popular class of dynamically evolving networks: preferential attachment models. We derive asymptotic properties of various functionals of the network including the degree distribution as well as maximal degree asymptotics, in essence showing that the change point does effect the degree distribution but does *not* change the degree exponent. This provides evidence for long range dependence and sensitive dependence of the evolution of the network on the initial evolution of the process. We propose an estimator for the change point and prove consistency properties of this estimator. The methodology developed highlights the effect of the nonergodic nature of the evolution of the network on classical change point estimators.

1. Introduction. The increasing availability of, and interest in, relational data for real world systems, has motivated the study of theoretical models for complex networks. The aim of these models is to explain structural features observed in the data (e.g., power law degree distribution or "small world" connectivity), and to understand and predict the behavior of dynamic processes on these networks. Dynamic processes on networks include disease contact networks, search algorithms, random walks, evolution and dissolution of communities and a variety of related processes [2, 10, 19, 23, 25, 42, 43, 58]. Among this research activity, the study of temporal, or time varying, networks has been particularly active; see the recent surveys [9, 31] and the references therein for methodological developments, as

well as applications in a wide array of fields ranging from social networks, online communication and cell biology.

Many dynamic network models involve a set of parameters that describe the time evolution of networks under the model. A natural question in this context is how to identify and understand the effect of *change points*—in particular, the effect of abrupt changes in parameters on the evolution and structural properties of the network. First, consider the simplest version of the classical (offline) change point detection problem for independent observations. Let F and G be (unknown but different) distribution functions, and let $\gamma \in (0, 1)$ be a change point parameter. We observe $\{X_i : 1 \leq i \leq n\}$ where X_i are i.i.d. with distribution F for $i \leq \lfloor n\gamma \rfloor$ and X_i are i.i.d. with distribution G (and independent of the initial segment) for $i > \lfloor n\gamma \rfloor$. The simplest version of the classical change point problem is to consistently estimate γ as $n \to \infty$.

In this spirit, this paper has two goals:

(a) We begin with a variant of the standard preferential attachment model for evolving networks that incorporates a change point. We study the effect of the change point on the structural properties of the network, including the scale-free (heavy tailed) behavior of the limiting degree distribution, and the asymptotics of the maximal degrees.

(b) We then propose and study the consistency of an offline estimation procedure to detect the location of the change point from observed data. Our analysis provides insight into the effects of the nonstationarity of the network evolution on existing methods for estimation in the i.i.d. and ergodic settings.

1.1. *Organization of the paper.* Over the last few decades there has been a substantial amount of work on change point detection and preferential attachment models. We defer a fuller discussion of related work, and its relevance to the paper, until Section 3. Our change point preferential attachment model is defined in the next section. In Section 1.3, we introduce notation required for the main results. The main results are contained in Section 2, beginning with Section 2.1, which describes the asymptotics for several functionals of the networks, including the degree distribution and maximal degree. Section 2.2 describes change point estimators and establishes their consistency. Proofs for the results on asymptotics of network functionals can be found in Section 4. Section 5 develops a functional central limit theorem for a specific functional of the network, which is used in Section 6 to establish consistency of the proposed estimator.

1.2. *Model formulation.* We begin by describing the original model of preferential attachment with no change point [4, 53, 60]. There are many variants of this model. Throughout the paper, we will consider the simplest case where the network at each stage is a tree, but we note that our methodology can be extended to the general network setup. Begin with a single vertex at time $m = 1$. This vertex

will be referred to as the *root* or the original progenitor of the process and denoted by ρ. Fix a parameter $\alpha \geq 0$. At each integer time $1 < m \leq n$ a new vertex enters the system with a *single edge* that is connected in a stochastic fashion to a pre-existing vertex. In particular, the edge connects to a pre-existing vertex v with probability proportional to the current degree of v *plus* α. Let $\mathcal{T}_m$ denote the graph at time m and let $\{\mathcal{T}_m : 1 \leq m \leq n\}$ be the graph valued process over n times steps. Since each new vertex is connected to an existing vertex, for each $m \geq 1$, $\mathcal{T}_m$ is a tree rooted at ρ. Thus for $m > 1$, the degree of every vertex is at least 1. If we regard the existing vertex to which a new vertex attaches as the *parent* of this vertex, then one can view this process as generating a directed tree with edges pointed from parents to children.

Our analysis is based on a continuous-time version of the preferential attachment model for which a slight variant of the above discrete-time process is more natural. Note that for directed, rooted trees, the degree of every vertex other than the root is $1 +$ out-degree of that vertex; for the root, the degree and the out-degree coincide. Fix a single vertex at time $m = 1$ and a parameter $\alpha > 0$. The variant of preferential attachment considered in this paper is as follows: at each stage $m > 1$, a new vertex enters the system and connects to a pre-existing vertex $v \in \mathcal{T}_{m-1}$ with probability proportional to $1 + \alpha$ plus the out-degree of v in $\mathcal{T}_{m-1}$. This model differs from the original only in the attachment probability to the root, and has all the same asymptotic properties as the original model but is slightly easier to deal with rigorously.

The preferential attachment model has been studied extensively and in particular it is known [12] that the degree distribution converges in the large network limit. Precisely, if $N_n(k)$ is the number of vertices with degree k in $\mathcal{T}_n$, then for each fixed $k \geq 1$:

$$(1.1) \qquad \frac{N_n(k)}{n} \xrightarrow{\text{a.e.}} p_\alpha(k) \qquad \text{where } p_\alpha(k) := (2 + \alpha) \frac{\prod_{j=1}^{k-1}(j + \alpha)}{\prod_{j=3}^{k+2}(j + 2\alpha)}.$$

Here, we use "a.e." to denote convergence almost everywhere and in the above expression, for $k = 1$, we use the notation $\prod_{j=1}^{k-1} = 1$. Writing D_α for a random variable having the above distribution, it is easy to check that there exists a constant $c := c(\alpha) > 0$ such that

$$(1.2) \qquad \mathbb{P}(D_\alpha \geq k) \sim \frac{c}{k^{\alpha+2}} \qquad \text{as } k \to \infty.$$

Further arranging the degrees of the vertices in $\mathcal{T}_n$ in decreasing order as $M_n(1) \geq M_n(2) \geq \cdots \geq M_n(n)$, it is known [6, 41] that for any fixed $k \geq 1$, there exists a nondegenerate probability distribution v_k^α on $\mathbb{R}_+^k$ such that

$$(1.3) \qquad \left(n^{-\frac{(1+\alpha)}{(2+\alpha)}} M_n(j) : 1 \leq j \leq k\right) \xrightarrow{\text{w}} v_\alpha^k,$$

where we use $\xrightarrow{\text{w}}$ to denote weak convergence.

1.2.1. *Model with change point.*   Now fix two attachment parameters α, $\beta > 0$, a change point parameter $\gamma \in (0, 1)$, and a system size $n > 1$. The model is as before, but now the attachment dynamics change at time $\lfloor n\gamma \rfloor$:

  (a) For time $0 < m \leq \lfloor n\gamma \rfloor$, the new vertex entering the system at time m connects to pre-existing vertices with probability proportional to their current out-degree plus $1 + \alpha$.
  (b) For time $\lfloor n\gamma \rfloor < m \leq n$, the new vertex connects to pre-existing vertices with probability proportional to their current out-degree plus $1 + \beta$.

Let $\boldsymbol{\theta} = (\alpha, \beta, \gamma)$ be the parameters of the model. Let $\mathcal{T}_{\boldsymbol{\theta},m}$ denote the rooted tree at time m, and let $\{\mathcal{T}_{\boldsymbol{\theta},m} : 1 \leq m \leq n\}$ denote the entire graph valued process. When the context is clear, for ease of notation we suppress the dependence on $\boldsymbol{\theta}$ and write $\{\mathcal{T}_m : 1 \leq m \leq n\}$. This model is the main object of interest for the rest of the paper.

1.3. *Preliminary notation.*   To state our main results, we will need to define some additional objects. Recall that $\boldsymbol{\theta} := (\alpha, \beta, \gamma)$ is the parameter set used to construct the model. Let $\{E_\alpha(k) : k \geq 1\}$ be a sequence of independent exponential random variables such that $E_\alpha(k)$ has rate $k + \alpha$ for each fixed $k \geq 1$. We view the random variables $E_\alpha(k)$ as the inter-arrival times of a point process $\mathcal{P}_\alpha$ on $\mathbb{R}_+$. In detail, define

$$L_\alpha(m) = E_\alpha(1) + \cdots + E_\alpha(m), \qquad m \geq 1,$$

and define the point process

(1.4) $$\mathcal{P}_\alpha := \big(L_\alpha(1), L_\alpha(2), \ldots\big).$$

In an analogous fashion, define random variables $\{E_\beta(k) : k \geq 1\}$ and $\{L_\beta(k) : k \geq 1\}$, and the corresponding point process $\mathcal{P}_\beta$. For $t \geq 0$, let $N_\alpha(t) := \mathcal{P}_\alpha[0, t]$ denote the number of points in $\mathcal{P}_\alpha$ which falls in the interval $[0, t]$.

  We will need a variant of the second point process. For $j \geq 1$, let $\mathcal{P}_\beta^j$ be the point process generated from the random variables $\{E_\beta(m) : m \geq j\}$. Thus the first point arrives at exponential rate $j + \beta$, the second point arrives at rate $j + 1 + \beta$ after the first point and so forth. Let $N_\beta^j(\cdot)$ be the corresponding counting process and note that $N_\beta^1(\cdot) = N_\beta(\cdot)$. Define the constant

(1.5) $$a = \frac{1}{2 + \beta} \log \frac{1}{\gamma},$$

and define the "truncated" exponential distribution the interval $[0, a]$ via the cumulative distribution function

(1.6) $$G_a(s) = \frac{1 - \exp(-(2 + \beta)s)}{1 - \exp(-(2 + \beta)a)}, \qquad s \in [0, a].$$

Write Age for a random variable with distribution G_a (the reason for this terminology will become clear in the proof). Generate a counting process $N_\beta(\cdot)$ as above

(independent of Age) so that $N_\beta[0, \mathsf{Age}]$ is the number of points that occur before the random time Age.

We are now in a position to define the limiting degree distribution under the change point model. Let D_α have distribution (1.1), namely the limiting degree distribution *without* change point. Define an integer valued random variable D_θ as follows:

(a) With probability $1 - \gamma$, $D_\theta = 1 + N_\beta[0, \mathsf{Age}]$.

(b) With probability γ, $D_\theta = D_\alpha + N_\beta^{D_\alpha}[0, a]$, where the point process $N_\beta^{D_\alpha}$ is generated conditional on the value of D_α.

Write $\mathbf{p}_\theta = (p_\theta(k) : k \geq 1)$ for the probability mass function of D_θ, namely

(1.7) $$p_\theta(k) = \mathbb{P}(D_\theta = k), \qquad k \geq 1.$$

2. Results. Let us now describe our main results. We state results about the asymptotic degree distribution in Section 2.1. Change point estimates and associated consistency results are presented in Section 2.2.

2.1. *Asymptotics for the degree distribution.* Fix $\boldsymbol\theta \in \mathbb{R}_+ \times \mathbb{R}_+ \times (0, 1)$. For fixed $k \geq 1$, let $N_n(k)$ denote the number of vertices with degree k in the random tree $\mathcal{T}_n$ constructed in the change point model of Section 1.2.1. The random variable D_θ in the following result is as defined in (1.7).

THEOREM 2.1.   *Let $k \geq 1$ be fixed. The degree distribution satisfies*

$$\frac{N_n(k)}{n} \overset{P}{\longrightarrow} \mathbb{P}(D_\theta = k) \qquad as \ n \to \infty.$$

Further, for $\alpha \neq \beta$ and $\gamma \in (0, 1)$, $\mathbf{p}_\theta \neq \mathbf{p}_\alpha$. However, there exist constants $0 < c < c'$ such that for all $k \geq 1$

(2.1) $$\frac{c}{k^{\alpha+2}} \leq \mathbb{P}(D_\theta \geq k) \leq \frac{c'}{k^{\alpha+2}}.$$

REMARK 1.   This theorem says that the network *does* feel the effect of the change point in the empirical degree distribution if $\alpha \neq \beta$ and $\gamma \in (0, 1)$. However, comparing (2.1) with (1.2), we see that for any fixed $\gamma \in (0, 1)$ the tail behavior of the degree distribution is the same with and without the change point. This is a little surprising as one might expect that the tail of the degree distribution would scale like $k^{-(2+\beta)}$, especially when γ is close to zero and $\beta < \alpha$ so that the degree distribution of the post-change process has a heavier tail. Nevertheless, the theorem shows that the tail behavior of the pre-change process governs the tail behavior of the final process, regardless of the other parameters.

REMARK 2.   The techniques developed in this paper easily extend to the setting of multiple change points. We describe these extensions in Theorem 3.1.
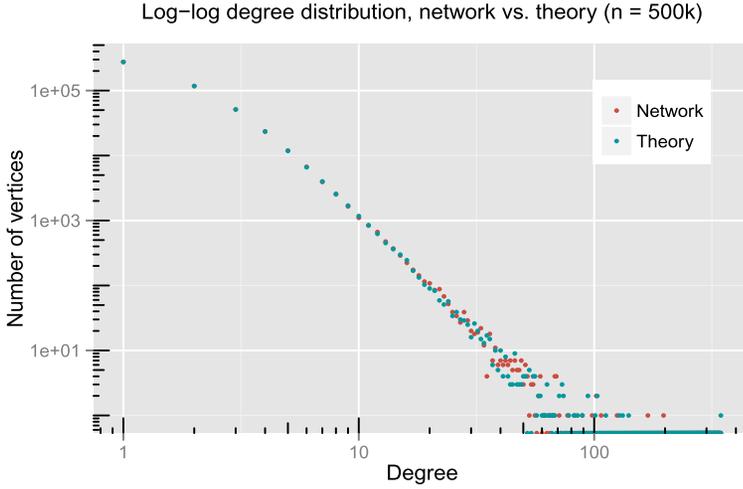
FIG. 1. *Log log plot showing the limiting degree distribution (red) and simulated network degree distribution (blue) with network size n = 500,000 and a corresponding sample of the same size from the predicted degree distribution. The model parameters are taken as $\alpha = 6$, $\beta = 1$ and the change point $\gamma = 0.5$. We discuss other values of the parameters in Section 3.*

The next result deals with maximal degree asymptotics. As before, arrange the degrees in $\mathcal{T}_n$ in decreasing order as $M_n(1) \geq M_n(2) \geq \cdots \geq M_n(n)$.

THEOREM 2.2. *Fix $k \geq 1$ and consider the k maximal degrees $(M_n(j) : 1 \leq j \leq k)$. Then the sequence of $\mathbb{R}_+^k$ valued random variables defined by*

$$\mathbb{M}_n(k) := \left(n^{-\frac{(1+\alpha)}{(2+\alpha)}} M_n(j) : 1 \leq j \leq k\right), \qquad n \geq 1$$

is tight and bounded away from zero.

REMARK 3. Comparing the scaling of the maximal degrees above to the setting of no change point in (1.3), we see that the order of magnitude of the maximal degrees is not affected by the change point. We conjecture that $\{\mathbb{M}_n(k) : n \geq 1\}$ converges weakly to a nondegenerate distribution on $\mathbb{R}_+^k$, but do not pursue this further in this paper.

2.2. *Change point detection.* In this section, we formulate a nonparametric estimator for the change point based on observations of the network and establish its consistency. While one could use the explicit *linear* nature of the attachment scheme to devise parametric or likelihood-based estimators of the change point, our aim is to develop more flexible methods that may work in settings where the precise form of the attachment model before and after the change point is not known. Extensions of the methodology to these more general settings are currently under study. The plan of the rest of this section is as follows. Our estimator tracks

the proportion of leaves as the process evolves and uses this functional to formulate a nonparametric estimator. Thus we start by describing a functional central limit for the proportion of leaves (Theorem 2.3). Then we formulate the actual estimator based on this functional. Theorem 2.3 is then used to establish the consistency result (Theorem 2.4) for the proposed estimator.

We begin with some notation and definitions. For fixed $k \geq 1$, let $N_n(k, m)$ denote the number of vertices with degree k in the tree $\mathcal{T}_m$ at the time of appearance of the mth vertex. Rescaling time by n, for $0 \leq t \leq 1$, let $\hat{N}_n(k, t) = N_n(k, \lfloor nt \rfloor)$ and let

$$(2.2) \qquad \hat{p}^n(k, t) = \frac{\hat{N}_n(k, t)}{nt}, \qquad 0 \leq t \leq 1,$$

be the proportion of vertices with degree k at time nt. The $k = 1$ case corresponds to the number of leaves in the tree. To ease notation in the displays below, we will write $\hat{p}^n(1, t) = \hat{p}_t^n$. Define the continuous function

$$(2.3) \qquad p_t^{(\infty)} = \begin{cases} \dfrac{2+\alpha}{3+2\alpha} & \text{if } 0 < t \leq \gamma, \\ \dfrac{2+\beta}{3+2\beta}\left(1 - \left(\dfrac{\gamma}{t}\right)^{\frac{3+2\beta}{2+\beta}}\right) + \dfrac{2+\alpha}{3+2\alpha}\left(\dfrac{\gamma}{t}\right)^{\frac{3+2\beta}{2+\beta}} & \text{if } \gamma \leq t \leq 1. \end{cases}$$

We will prove in Section 5.1 that for each fixed $t \in (0, 1]$, the value $p_t^{(\infty)}$ is the limiting proportion of leaves in $\mathcal{T}_{nt}$. Note that $p_t^{(\infty)} = p_\gamma^{(\infty)}$ for $t \leq \gamma$, namely the function is constant until time γ. To simplify notation, define $\delta : \mathbb{R}_+ \to [0, 1]$ by

$$(2.4) \qquad \delta_u := \frac{1+u}{2+u}, \qquad u \geq 0,$$

and define the (positive) function $\sigma_M : [0, 1] \to (0, \infty)$ via

$$(2.5) \qquad \sigma_M^2(t) := \begin{cases} t^{2\delta_\alpha}\left[\delta_\alpha p_\gamma^{(\infty)}\left(1 - \delta_\alpha p_\gamma^{(\infty)}\right)\right] & \text{if } 0 \leq t \leq \gamma, \\ [3pt]\gamma^{2\delta_\alpha}\left(\dfrac{t}{\gamma}\right)^{2\delta_\beta}\delta_\beta p_t^{(\infty)}\left(1 - \delta_\beta p_t^{(\infty)}\right) & \text{if } \gamma < t \leq 1. \end{cases}$$

Finally, define the functions

$$(2.6) \qquad \sigma^2(t) := \begin{cases} \left[\delta_\alpha p_\gamma^{(\infty)}\left(1 - \delta_\alpha p_\gamma^{(\infty)}\right)\right] & \text{if } 0 \leq t \leq \gamma, \\ [4pt]\delta_\beta p_t^{(\infty)}\left(1 - \delta_\beta p_t^{(\infty)}\right) & \text{if } \gamma < t \leq 1, \end{cases}$$

and

$$(2.7) \qquad \mu(t) := \begin{cases} \dfrac{\delta_\alpha}{t^{\delta_\alpha+1}}, & 0 < t \leq \gamma, \\ [2pt]-\dfrac{\delta_\beta \gamma^{\delta_\beta-\delta_\alpha}}{t^{\delta_\beta+1}}, & \gamma < t \leq 1. \end{cases}$$

Let $\{B(u) : u \geq 0\}$ be standard Brownian motion on $\mathbb{R}_+$, and define the diffusion $\{M(t) : 0 \leq t \leq 1\}$ via the prescription

$$(2.8) \qquad dM(t) = \sigma_M(t)\, dB(t), \qquad 0 \leq t \leq 1.$$

Then M is a deterministic time change of $B(\cdot)$, in the sense that

$$(2.9) \quad \{M(t) : 0 \leq t \leq 1\} \overset{d}{=} \{B(\phi(t)) : 0 \leq t \leq 1\} \qquad \text{where } \phi(t) = \int_0^t \sigma_M^2(s)\, ds,$$

and in particular, $M(\cdot)$ is a Gaussian process on $[0, 1]$. Define

$$(2.10) \qquad g(t) := \begin{cases} \dfrac{1}{t^{\delta_\alpha}} & \text{if } 0 < t \leq \gamma, \\[2ex] \dfrac{\gamma^{\delta_\beta - \delta_\alpha}}{t^{\delta_\beta}} & \text{if } \gamma < t \leq 1, \end{cases}$$

and the associated process

$$(2.11) \qquad G(t) = g(t)M(t), \qquad 0 < t \leq 1.$$

By Itô's formula, $G(\cdot)$ solves the SDE

$$(2.12) \qquad dG(t) = \mu(t)M(t)\, dt + \sigma(t)\, dB(t),$$

where $\sigma(\cdot)$ and $\mu(\cdot)$ are as in (2.6) and (2.7), respectively. Then we have the following result.

THEOREM 2.3. *Consider the process of re-centered and normalized number of leaves*

$$(2.13) \qquad G_n(t) := \frac{\hat{N}_n(1, t) - nt p_t^{(\infty)}}{\sqrt{n}}, \qquad 0 \leq t \leq 1.$$

Then as $n \to \infty$, $G_n \overset{w}{\longrightarrow} G$. Here, G is the diffusion defined in (2.12) $\overset{w}{\longrightarrow}$ denotes weak convergence on $D([0, 1])$ equipped with the Skorohod metric.

For the rest of this section, let $p_n(m)$ denote the proportion of leaves (degree one vertices) in $\mathcal{T}_m$, and let $\varepsilon > 0$ be fixed. Define two functions on $[\varepsilon, 1]$ as follows:

$$(2.14) \qquad {}_t h^{(n)} = \frac{1}{n(t - \varepsilon)} \sum_{m=n\varepsilon}^{nt} p_n(m), \qquad \varepsilon \leq t \leq 1$$

and

$$(2.15) \qquad h_t^{(n)} = \frac{1}{n(1 - t)} \sum_{m=nt+1}^{n} p_n(m), \qquad \varepsilon \leq t \leq 1.$$

In words, $_t h^{(n)}$ represents the average proportion of leaves in the process between time $n\varepsilon$ and nt, while $h_t^{(n)}$ represents the same quantity after time nt. Let

$$(2.16) \qquad D_n(t) := (1-t) \big|_t h^{(n)} - h_t^{(n)} \big|, \qquad t \in [\varepsilon, 1]$$

be the scaled absolute difference of the leaf proportions before and after time nt, and let $\mathcal{M}_n$ be the collection of $t \in [\varepsilon, 1]$ such that $D_n(t)$ is close to its maximum value $D_n^* = \max_{t \in [\varepsilon, 1]} D_n(t)$. In detail,

$$(2.17) \qquad \mathcal{M}_n := \left\{ t \in [\varepsilon, 1] : \big| D_n(t) - D_n^* \big| \leq \frac{\log n}{\sqrt{n}} \right\}.$$

Finally, let

$$(2.18) \qquad \hat{\gamma}_n := \max\{t : t \in \mathcal{M}_n\}.$$

The functionals D_n^*, $\mathcal{M}_n$, and $\hat{\gamma}_n$ all depend on ε but we suppress this dependence to ease exposition below.

THEOREM 2.4. *Assume that the change point $\gamma > \varepsilon$. Then $\hat{\gamma}_n \overset{P}{\longrightarrow} \gamma$, and, in fact,*

$$(2.19) \qquad |\hat{\gamma}_n - \gamma| = O_P\left(\frac{\log n}{\sqrt{n}}\right).$$

Thus $\hat{\gamma}_n$ is a consistent estimator for change points $\gamma > \epsilon$.

REMARK 4. Consideration of $t \geq \varepsilon > 0$ serves to control the factor t in the denominator of (2.14). Technically, one should be able to choose a sequence $\varepsilon_n \downarrow 0$ slowly enough so that the resulting estimators would be consistent for any $\gamma > 0$. However, even proving the above result turns out to be nontrivial owing to the nonergodic nature of the evolution of the process after the change point. Thus we restrict ourselves to proving the above result. We hope to address distributional convergence and sharper estimators for the above problem in future work.

REMARK 5. If we replace the threshold $\log n / \sqrt{n}$ in (2.17) by a threshold of the form $\omega_n / \sqrt{n}$, where ω_n tends to infinity arbitrarily slowly, then the corresponding estimator satisfies (2.19) with bound $O_P(\omega_n / \sqrt{n})$.

REMARK 6. See Figure 2 for a figure based on simulations for the function $D_n(t)$ with ε taken to be zero.

3. Discussion. In this section, we discuss the relevance of our results, their connections to the existing change point literature and possible extensions.
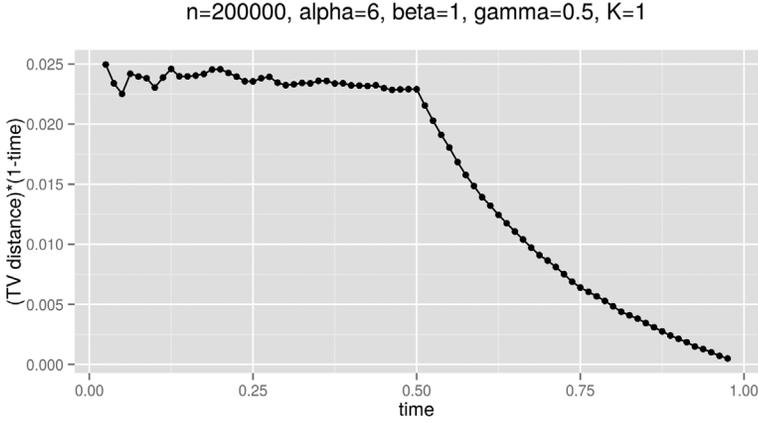
FIG. 2.   *The function $D_n(t)$ with network size $n = 200{,}000$ and model parameters $\alpha = 6, \beta = 1$ and the change point $\gamma = 0.5$ as in Figure 1.*

3.1. *Multiple change points.*   The proof techniques here carry over in a straightforward fashion to the general setting of multiple change points. Fix time points $0 < \gamma_1 < \gamma_2 < \cdots < \gamma_k < 1$ and parameters $\alpha, (\beta_i)_{1 \leq k}$. As before, write $\boldsymbol{\theta} = (\alpha, (\beta_i)_{1 \leq i \leq k}, (\gamma_i)_{1 \leq i \leq k})$ for the parameter set. Consider the random tree $\mathcal{T}_n = \mathcal{T}_{\boldsymbol{\theta}, n}$ where:

(i) In the interval $\{1 < t \leq \gamma_1 n\}$, vertices use the attachment scheme driven by α (namely each new vertex attaches to an existing vertex with probability proportional to out-degree plus $1 + \alpha$).

(ii) In subsequent intervals $\{(\gamma_j)n < t \leq (\gamma_{j+1})n\}$ where $1 \leq j \leq k$, vertices perform the attachment scheme driven by the parameter β_j. Here, we use the convention $\gamma_0 = 0, \gamma_{k+1} = 1$.

As in Section 1.3 define the point processes $\mathcal{P}_\alpha, \mathcal{P}_{\beta_i}$ and for fixed $j \geq 1$, the point processes $\mathcal{P}_\alpha^j, \mathcal{P}_{\beta_i}^j$. To simplify notation, for any $t \geq 0$ and point process $\mathcal{P}$, set $\mathcal{P}[0, t]$ for the *number* of points in the interval $[0, t]$. Define the constants

$$(3.1) \qquad \pi_j = \gamma_{j+1} - \gamma_j, \qquad a_j = \frac{1}{2 + \beta_j} \log \frac{\gamma_{j+1}}{\gamma_j}.$$

Note that $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_k)$ is a probability mass function. Write Epoch for a random variable with distribution $\boldsymbol{\pi}$ [i.e., $\mathbb{P}(\mathsf{Epoch} = i) = \pi_i$ for $0 \leq i \leq k$]. Using the constants $\{a_i : 1 \leq i \leq k\}$ let G_{a_i} denote corresponding truncated exponential distributions as in (1.6) and let Age_i denote a random variable with distribution G_{a_i}. Now construct the random variable TimeAlive as follows:

(a) Generate a collection of independent random variables Epoch and $\{\mathsf{Age}_i : 1 \leq i \leq k\}$ with distributions specified as above.

(b) Conditional on $\mathsf{Epoch} = i$, let

$$\mathsf{TimeAlive} = \mathsf{Age}_i + \sum_{j=i+1}^{k} a_j,$$

where again by convention, if $\mathsf{Epoch} = 0$, $\mathsf{Age}_0 = 0$ and so $\mathsf{TimeAlive} = \sum_{j=1}^{k} a_i$.

Construct a positive integer valued random variable D_θ as follows:

(i) Generate $\mathsf{Epoch} \sim \pi$ as above and the corresponding random variable $\mathsf{TimeAlive}$.

(ii) If Epoch takes a nonzero value $1 \le i \le k$, conditional on $\mathsf{Epoch} = i$, generate the switching point process $\mathcal{P}_\star$ on the interval $[0, \mathsf{TimeAlive}]$ as follows:

(a) *Initialization*: In the interval $[0, \mathsf{Age}_i]$, start with $\mathcal{P}_\star = \mathcal{P}_{\beta_i}$. Suppose by time Age_i, $\mathcal{P}_\star[0, \mathsf{Age}_i] = k$. Now generate a point process $\mathcal{P}_{\beta_{i+1}}^{k+1}$ and let $\mathcal{P}_\star[0, \mathsf{Age}_i + a_{i+1}] = \mathcal{P}_\star[0, \mathsf{Age}_i] + \mathcal{P}_{\beta_{i+1}}^k[0, a_{i+1}]$.

(b) *Recursion*: For each subsequent interval $[a_j, a_{j+1}]$ with $j > i$, conditional on $\mathcal{P}_\star[0, \mathsf{Age}_i + a_{i+1} + \cdots + a_j] = k_j$, generate the point process $\mathcal{P}_{\beta_{j+1}}^{k_j+1}$. Define

$$\mathcal{P}_\star[0, \mathsf{Age}_i + a_{i+1} + \cdots + a_{j+1}] = \mathcal{P}_\star[0, \mathsf{Age}_i + a_{i+1} + \cdots + a_j] + \mathcal{P}_{\beta_{j+1}}^{k_j+1}[0, a_{j+1}].$$

Iterate until the last interval resulting in $\mathcal{P}_\star[0, \mathsf{TimeAlive}]$.

Now define $D_\theta = 1 + \mathcal{P}_\star[0, \mathsf{TimeAlive}]$.

(iii) If $\mathsf{Epoch} = 0$, so that $\mathsf{TimeAlive} = a_1 + \cdots + a_k$, generate a random variable D_α with distribution $\mathbf{p}_\alpha$ as in (1.1). Conditional on D_α, generate $\mathcal{P}_\star$ in the interval $[0, a_1]$ with distribution $\mathcal{P}_{\beta_1}^{D_\alpha}$ and then sequentially proceed as in (ii). In this case, define $D_\theta = D_\alpha + \mathcal{P}_\star[0, \mathsf{TimeAlive}]$.

Write $p_\theta(\cdot)$ for the p.m.f. of D_θ. As before for $k \ge 1$, let $N_n(k)$ denote the number of vertices with degree k in $\mathcal{T}_n$. Then we have the following result.

THEOREM 3.1. *As $n \to \infty$, we have*

$$\frac{N_n(k)}{n} \xrightarrow{P} p_\theta(k).$$

Further there exist constants $0 < c < c'$ such that for all $k \ge 1$

(3.2) $$\frac{c}{k^{\alpha+2}} \le \mathbb{P}(D_\theta \ge k) \le \frac{c'}{k^{\alpha+2}}.$$

3.2. *Possible extensions.* Motivated by a number of illuminating comments by the anonymous referees and the associate editor, here we describe possible extensions of the treatment in this paper. First, note that we use information on leaf

densities in the large network $n \to \infty$ limit to develop change point estimators. As in [48], one should be able to build on the functional CLT for leaf counts to establish a joint functional CLT for $\{\hat{N}_n(k, t) : 1 \le k \le K, 0 \le t \le 1\}$ after proper normalization and re-centering for any fixed $K \ge 1$. Modifying the estimator in Section 6 should enable one to get estimators that perform better for finite n. We are currently in the process of establishing functional central limit theorems, which we then hope to use to develop change point estimators in order to rigorously understand the increase in asymptotic mean-square efficiency using more degree information.

Second, note that we have only considered the change point estimation problem in this paper. As pointed out by the referee a natural extension to explore is extending this treatment to a hypothesis testing framework in the context of testing the existence (or lack thereof) of a change point. We are currently in the process of studying the (properly normalized) behavior of the functionals proposed in this paper including the function $D_n(\cdot)$ defined in (2.16) under the null hypothesis of no change point.

A final natural extension is understanding what happens if one does not have data on every time point but rather a filtering (subset) of time points.

3.3. *Change point detection.* This problem of change point detection has a long history owing to its importance in applications in fields ranging from quality control and reliability of industrial processes (e.g., quick detection of process failure in production) to signal processing (e.g., automatic segmentation of signals into stationary segments via identification of change points, etc.). We direct the interested reader to [5, 13, 16, 17, 21, 51, 52, 54] and the references therein for an overview of some of the statistical methodology and applications.

Recall the motivating example of an independent stream of data $\{X_i : 1 \le i \le n\}$ with a change point in the distribution from F to G at time $n\gamma$ described in Section 1. Let $_t H^{(n)}(\cdot)$ and $H_t^{(n)}$ denote the empirical distribution of the data before and after t, namely

$$_t H^{(n)} := \frac{1}{nt} \sum_{i=1}^{nt} \delta_{X_i}, \qquad H_t^{(n)} := \frac{1}{n(1-t)} \sum_{i=nt+1}^{n} \delta_{X_i}, \qquad 0 < t < 1.$$

Now define

$$D_n(t) := t(1-t) \, \mathrm{dist}\big(_t H^{(n)}, H_t^{(n)}\big),$$

where dist is any standard notion of distance between probability distributions on $\mathbb{R}$, for example, Kolmogorov–Smirnov or total variation distance. Finally, define

$$\hat{\gamma}_n = \underset{t \in [0,1]}{\mathrm{argmax}} \, D_n(t).$$

In [16], it is shown that $\hat{\gamma}_n$ is a consistent estimator of γ. In contrast to the classical setting above, the statistics $D_n(t)$ defined in [see (2.16)] is not symmetric in t, reflecting the nonergodic nature of the evolution under the change point model.

3.4. *Temporal networks and change points.* As described in the Introduction, the availability of data on real world networks over the last few years has motivated development of mathematical methodology in a wide array of fields. Work similar in spirit to change point detection includes segmentation and boundary detection, for example, [37, 57], the detection of anomalous subgraphs and motifs *within* networks, for example, [1, 27, 28, 44], and the detection of anomalous edges via link prediction algorithms [32]. For a survey of this work, see [18]. There is also an active literature concerning the detection of change points in temporal (time-varying) network data, and in particular, structural properties of these such networks. An algorithmic approach to understanding evolving communities in social networks based on minimum description length is given in [55], while more statistically grounded approaches can be found in [24, 30, 39, 40, 47, 50, 61]. See [46] for an overview of the state of the art regarding change point detection in networks; this paper also develops new statistical methodology using a generalized hierarchical random graph model (GHRG) and various likelihood ratio based test statistics to detect existence of change points via online detection algorithms and studies the performance of these algorithms on synthetic as well as real data. A rigorous analysis of dynamic models, in which each time slice of the model is assumed to be an Erdős–Rényi random graph, is given in [38, 59].

3.5. *Preferential attachment.* The preferential attachment model has become one of the standard workhorses in the complex networks community, based in part on the fact that it exhibits the power law/heavy tailed degree distribution observed in an array of real world systems. As the literature on preferential attachment is large and very broad, we focus on work that is close in spirit to the work in this paper. The preferential attachment model was introduced in the combinatorics community in [56] and was brought to the attention of the networks community in [4]. The papers [42] and [23] give survey-level treatments of a wide array of related models, while [12] gives the first rigorous results on the asymptotic degree distribution. More general models and results can be found in [11, 20, 25, 49] and the references therein.

We are not aware of other analyses of the effect of change point in structural properties of such network models. There has been a lot of recent interest in understanding and detecting the "initial seed" [14, 15, 22]. Here, one starts with an initial "seed graph" at time $m = 0$ and then performs preferential attachment started from that seed. The aim is then to estimate this initial seed based on an observation of the network at some large time n. While different from this paper, this body of work again emphasizes the sensitive dependence on initial conditions for such network models.

3.6. *Proof techniques.* A number of techniques have been developed to rigorously analyze functionals such as asymptotic degree distributions (see [25, 58] for nice pedagogical treatment). The standard technique involves writing down

recursions for the expected degree distribution $\mathbb{E}(N_n(k))$ using the prescribed dynamics of the process to show that these expectations (normalized by n) converge in the limit, and then showing that the deviations $|N_n(k) - \mathbb{E}(N_n(k))|$ are small via concentration inequalities.

In this paper, for understanding structural properties we use a different technique, essentially embedding the discrete-time model in a corresponding continuous-time branching process $\{\mathsf{BP}_{\boldsymbol{\theta}}^n(t) : t \geq 0\}$ (based on the Athreya–Karlin embedding of urn processes [3]). This explains the various point processes that arise in the description of the limiting degree distribution. While mathematically more involved, this technique gives more insight into the results as it elucidates the natural time scale of the process. In various other settings, this technique has resulted in the study of much more general functionals of the process such as the spectral distribution of the adjacency matrix [6] and has been used to derive asymptotic results in "nonlocal" preferential attachment models [7]. In this paper, the technique also allows one to intuitively understand why the degree exponent does not change.

We advise the reader to come back to the text below after going through the proofs but let us explain the basic intuition here. In the continuous-time version, the process grows exponentially and in particular takes time $\tau_{\gamma n} \approx \frac{1}{2+\alpha} \log \gamma n + O_P(1)$ to get to size $n\gamma$. At this time, there is a change in the evolution where each vertex adopts attachment dynamics driven by the parameter β. However, owing to the exponential growth rate, the time for the process to get to size n is $\tau_n \approx \tau_{\gamma n} + a$ where a is as in (1.5). It turns out that this is not enough time for the dynamics with attachment parameter β to change the degree exponent [since we only have to wait an $O(1)$ extra units of time to get to system size n from γn]. These ideas are made mathematically rigorous in the next few sections. For the interested reader, much of the foundational work on continuous-time branching processes relevant for this paper can be found in [33–35]. For functional central limit theorems in these settings, we generalize techniques developed in [48] for the setting with no change point.

3.7. *Empirical dependence of the convergence on parameter values.* Recall that the Gaussian process defined in (2.12) underlying the main consistency result Theorem 2.4 depends on $\boldsymbol{\theta} = (\alpha, \beta, \gamma)$. One consequence of this dependence is that when the parameter values α and β are close, the change point becomes harder to detect in the sense that larger n is required to get good estimates. This is most easily seen in terms of the fluctuations of the proportion of leaves in the graph.

In both Figures 3 and 4, the preferential attachment process starts with $\alpha = 6$ and decreases, to $\beta = 1$ in 3 and $\beta = 5$ in 4. Furthermore, the predicted behavior (red line) is almost the same: the proportion of leaves is constant up to the change point $\gamma = 0.5$ and then increases, consistent with a *decrease* in the attachment parameter.
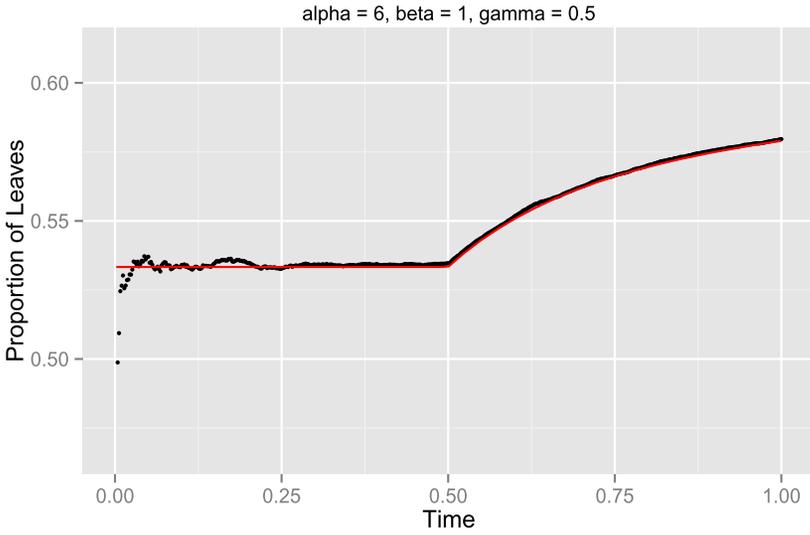
FIG. 3. *Empirical proportion of leaves in a simulation with $n = 200{,}000$, $\alpha = 6$, $\beta = 1$, $\gamma = 0.5$. The red line represents the theoretical predictions in* (2.3).

Despite the sizes of the final graphs in both simulations being $n = 200{,}000$ vertices, at first glance the fluctuations appear much greater in the latter case. On closer examination, however, this is simply an illusion of the axes. In essence, when the shift in parameters is smaller, the change in the proportion of leaves pre- and post-γ is smaller compared to the natural fluctuations in the proportion
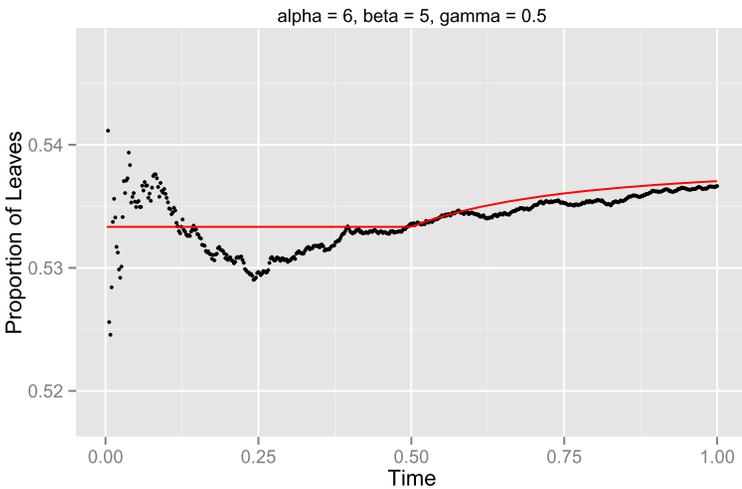

FIG. 4. *Empirical proportion of leaves in a simulation with $n = 200{,}000$, $\alpha = 6$, $\beta = 5$, $\gamma = 0.5$. The red line represents the theoretical predictions in* (2.3).

of leaves which is of order $\sqrt{n}$ (Theorem 2.3). Therefore, any difference is more difficult to detect for same n. This is not surprising, but worth noting in practice.

4. Proofs. As described in Section 3.6, the main conceptual idea is a continuous-time embedding of the discrete-time process. We start in Section 4.1 by describing this embedding and deriving simple properties. Then in Section 4.2, we prove Theorem 2.1. Section 4.3 proves the assertion that the degree exponent does not change. Section 4.4 analyzes asymptotics for the maximal degrees. Section 5 contains an in-depth analysis of the density of leaves and proves Theorem 2.3. Section 6 then uses this theorem to prove the consistency of the estimator namely Theorem 2.4.

4.1. *Preliminaries.* We start with the following definition. To ease notation, for the rest of the paper we use γn instead of $\lfloor \gamma n \rfloor$.

DEFINITION 4.1 (Continuous-time branching process). Fix $\alpha > 0$. We let $\{\mathsf{BP}_\alpha(t) : t \geq 0\}$ be a continuous-time branching process driven by the point process $\mathcal{P}_\alpha$ defined in (1.4). Precisely:

(a) At time $t = 0$, we start with one individual called the root ρ with an offspring point process with distribution $\mathcal{P}_\alpha^\rho \overset{d}{=} \mathcal{P}_\alpha$. The times of this point process represent times of birth of new offspring of ρ.

(b) Every new vertex v that is born into the system is given its own offspring point process $\mathcal{P}_\alpha^v \overset{d}{=} \mathcal{P}_\alpha$, independent across vertices.

Label vertices using integer labels according to the order in which they enter BP_α so that the root is labelled as 1, the next vertex to be born labeled by 2 and so on. For fixed $t \geq 0$, we will view $\mathsf{BP}_\alpha(t)$ as a (random) labelled tree representing the genealogical relationships between all individuals in the population present at time t. See Figures 5 and 6. Write $|\mathsf{BP}_\alpha(t)|$ for the number of individuals in the tree by time t. Fix $m \geq 1$ and define the stopping time

$$(4.1) \qquad \tau_m := \inf\{t : |\mathsf{BP}_\alpha(t)| = m\}.$$

Since there are no deaths and each individual reproduces at rate at least $1 + \alpha$, the stopping times $\tau_m < \infty$ a.s. for all $m \geq 1$. Now consider the original preferential attachment model where there is no change point. Using properties of the exponential distribution, the following lemma is easy to check as a special case of the famous Athreya–Karlin embedding [3].

LEMMA 4.2. *Viewed as random rooted trees on vertex set* $[n]$ *one has* $\mathsf{BP}_\alpha(\tau_n) \overset{d}{=} \mathcal{T}_n$. *In fact, the two processes of growing random trees have the same distribution, namely*

$$\{\mathsf{BP}_\alpha(\tau_n) : n \geq 1\} \overset{d}{=} \{\mathcal{T}_n : n \geq 1\}.$$
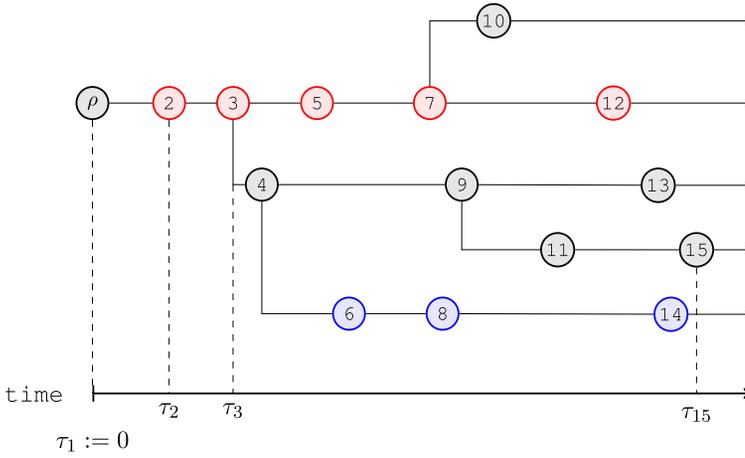
FIG. 5. *The process* $\mathsf{BP}_\alpha(\cdot)$ *in continuous time starting from the root* ρ *and stopped at* τ_{15}.

To construct the variant $\mathcal{T}_n$ where one has a change point, we run $\mathsf{BP}_\alpha(\cdot)$ until time $\tau_{\gamma n}$ (when the original process reaches size γn) and then every vertex changes the way it reproduces. More precisely, after this stopping time, an individual with k children would have reproduced at rate $k + 1 + \alpha$ in the original model but in the change point model this vertex reproduces at rate $k + 1 + \beta$ and uses the parameter β instead of α for each subsequent offspring times. Each new vertex v produced after time $\tau_{\gamma n}$ reproduces according to an independent copy of the point process $\mathcal{P}_\beta$. Call the resulting process $\mathsf{BP}_{\boldsymbol{\theta}}^n(\cdot)$ and run the process until time τ_n when the continuous-time process has n individuals. Analogous to (4.1), define the collection of stopping times $\{\tau_m : 1 \le m \le n\}$ by replacing BP_α with $\mathsf{BP}_{\boldsymbol{\theta}}^n$. The following is a simple extension of the previous lemma.
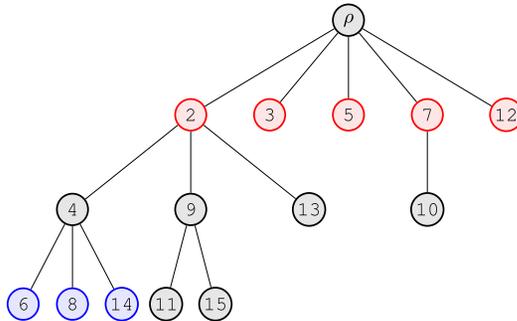

FIG. 6. *The corresponding discrete tree containing only the genealogical information of vertices in* $\mathsf{BP}_\alpha(\tau_{15})$.

LEMMA 4.3.   *Recall the family of random trees* $\{\mathcal{T}_{\theta,m} : 1 \leq m \leq n\}$ *generated using the change point preferential attachment model in Section* 1.2.1. *Then*

$$\{\mathsf{BP}_{\theta}^{n}(\tau_m) : 1 \leq m \leq n\} \overset{d}{=} \{\mathcal{T}_{\theta,m} : 1 \leq m \leq n\}.$$

REMARK 7.   Note that the processes $\{\mathcal{T}_{\theta,m} : 1 \leq m \leq n\}$ when one has a change point are *not* nested in a nice manner as growing trees for *different values* of n. Compare this with the original model (without change point) where we can view the entire sequence $\{\mathcal{T}_n : n \geq 1\}$ as an increasing family of random trees. In the above construction, it will be convenient to couple the processes across different n by using a *single* common branching process BP_{α} to generate the tree before the change point $\tau_{\gamma n}$ and then let the process evolve independently after the change point for different n using the prescribed dynamics modulated by the attachment parameter β. Further, it will be convenient to allow the process BP_{θ}^{n} to continue to grow after time τ_n as opposed to stopping it exactly at time τ_n.

For future reference, for each vertex v, we will use T_v for the time of birth of this vertex into the system. For fixed time t and a vertex v born before time t (namely $T_v \leq t$), we write $d_v(t)$ for the number of children of this vertex by time t. Note that for all $v \neq \rho \in \mathsf{BP}_{\theta}^{n}(t)$, the full degree of v by time t is $d_v(t) + 1$.

We will need some simple stochastic calculus calculations below to derive martingales related to processes of interest. Given a process $\{Z(t) : t \geq 0\}$ adapted to a filtration $\{\mathcal{F}(t) : t \geq 0\}$, we write $\mathbb{E}(dZ(t)|\mathcal{F}(t)) = a(t)\,dt$ for an adapted process $a(\cdot)$ if $Z(t) - \int_0^t a(s)\,ds$ is a (local) martingale. Similarly, write $\mathrm{Var}(dZ(t)|\mathcal{F}(t)) = b(t)\,dt$ if the process

$$V(t) := \left(Z(t) - \int_0^t a(s)\,ds\right)^2 - \int_0^t b(s)\,ds, \qquad t \geq 0,$$

is a local martingale.

Now recall that $\mathsf{BP}_{\alpha}(\tau_{\gamma n})$ is the random tree before the change point. These random trees are distributed as the original preferential attachment model without change point using attachment dynamics with parameter α. Using (1.1) and recalling that $N_n(k, \gamma n)$ denotes the number of vertices with degree k results in the following.

LEMMA 4.4.   *For each fixed* $k \geq 1$, *we have* $N_n(k, \gamma n)/\gamma n \xrightarrow{a.e.} p_{\alpha}(k)$, *as* $n \to \infty$ *where* $p_{\alpha}(\cdot)$ *is the probability mass function in* (1.1).

Recall that the branching process BP_{α} is driven by the offspring point process $\mathcal{P}_{\alpha}$ and $\mathcal{P}_{\alpha}(t) := \mathcal{P}_{\alpha}[0, t]$ is the number of points in $[0, t]$. Define the process

(4.2) $$M_{\alpha}(t) := e^{-t}\mathcal{P}_{\alpha}(t) - (1 + \alpha)(1 - e^{-t}), \qquad t \geq 0.$$

LEMMA 4.5. *The process $\{M_\alpha(t) : t \geq 0\}$ is a martingale with respect to the natural filtration of $\mathcal{P}_\alpha$. In particular,*

$$\mathbb{E}(\mathcal{P}_\alpha(t)) = (1 + \alpha)(e^t - 1). \tag{4.3}$$

PROOF. Write $\{\mathcal{F}(t) : t \geq 0\}$ for the natural filtration of the process. It is enough to show for all $t \geq 0$, $\mathbb{E}(dM_\alpha(t)|\mathcal{F}(t)) = 0$. By construction,

$$\mathbb{E}(d\mathcal{P}_\alpha(t)|\mathcal{F}(t)) = (1 + \alpha + \mathcal{P}_\alpha(t)) \, dt.$$

Further,

$$\mathbb{E}(dM_\alpha(t)|\mathcal{F}(t)) = e^{-t}\mathbb{E}(d\mathcal{P}_\alpha(t)|\mathcal{F}(t)) - e^{-t}\mathcal{P}_\alpha(t) \, dt + (1 + \alpha)e^{-t} \, dt.$$

Elementary algebra completes the proof. The final assertion regarding (4.3) follows using the martingale property of M_α and the initial condition $\mathcal{P}_\alpha(0) = 0$. □

The starting point in the analysis of continuous-time branching processes is the so-called Malthusian rate of growth parameter $\lambda > 0$, which solves the equation

$$\int_0^\infty \lambda e^{-\lambda t} \mathbb{E}(\mathcal{P}_\alpha(t)) \, dt = 1. \tag{4.4}$$

Using Lemma 4.5 now implies

$$\lambda = 2 + \alpha. \tag{4.5}$$

Let T_λ be an exponential random variable with parameter λ independent of $\mathcal{P}_\alpha$ and consider the integer valued random variable $\mathcal{P}_\alpha(T_\lambda)$. Note that (4.4) is equivalent to $\mathbb{E}(\mathcal{P}_\alpha(T_\lambda)) = 1$. Recall that D_α is a random variable with the (nonchange point) degree distribution (1.1). It is easy to check that $D_\alpha - 1 \overset{d}{=} \mathcal{P}_\alpha(T_\lambda)$. In particular for $\alpha \geq 0$,

$$\mathbb{E}(\mathcal{P}_\alpha(T_\lambda) \log^+ \mathcal{P}_\alpha(T_\lambda)) < \infty.$$

Using standard Jagers–Nerman stable age-distribution theory for branching processes [34, 35] now implies the following.

PROPOSITION 4.6. *There exists an integrable a.s. positive random variable W_α such that*

$$e^{-(2+\alpha)t}|\mathsf{BP}_\alpha(t)| \overset{a.e.,\mathbb{L}^1}{\longrightarrow} W_\alpha.$$

In particular,

$$\tau_{\gamma n} - \frac{1}{2 + \alpha} \log n \overset{a.e.}{\longrightarrow} W'_\alpha, \tag{4.6}$$

for a finite random variable W'_α.

We conclude this section with asymptotics for the amount of "continuous time" where the attachment dynamics using β is valid, namely $\tau_n - \tau_{\gamma n}$. Recall the constant a from (1.5). We will also write $\{\mathcal{F}_n(t) : t \geq 0\}$ for the natural filtration of the process $\{BP_{\boldsymbol{\theta}}^n(t) : t \geq 0\}$.

LEMMA 4.7. *Let $\Upsilon_n = \tau_n - \tau_{\gamma n}$ denote the time after the change point in the continuous-time embedding. Then*

$$\sqrt{n}(\Upsilon_n - a) \xrightarrow{w} \frac{1}{2+\beta}\sqrt{\frac{1-\gamma}{\gamma}} Z,$$

as $n \to \infty$. Here, Z is a standard normal random variable.

PROOF.   Note that $BP_{\boldsymbol{\theta}}^n(\cdot)$ is a Markov process. Further, for $t \geq \tau_{\gamma n}$ conditional on $BP_{\boldsymbol{\theta}}^n(t)$, the rate at which a new individual is born into the system is given by

$$\lambda(t) := \sum_{v \in BP_{\boldsymbol{\theta}}^n(t)} \big(d_v(t) + 1 + \beta\big)$$

(4.7)

$$= (2+\beta)\big|BP_{\boldsymbol{\theta}}^n(t)\big| - 1.$$

In particular,

(4.8)
$$\Upsilon_n \overset{d}{=} \sum_{j=\lfloor \gamma n \rfloor}^{n-1} \frac{E_j}{(2+\beta)j - 1},$$

where $\{E_i : i \geq 1\}$ is a sequence of i.i.d. rate one exponential random variables. Using Lyapunov's central limit theorem now completes the proof. □

Using the distributional characterization in (4.8) and standard concentration inequalities for sums of independent random variables, one can show the following tail bound on Υ_n. We omit the proof.

LEMMA 4.8.   *For any $\kappa > 0$, there exists $N = N(\kappa) < \infty$ such that for all $n > N(\kappa)$,*

$$\mathbb{P}\left(|\Upsilon_n - a| > \frac{1}{n^{1/3}}\right) \leq \frac{1}{n^\kappa}.$$

In particular, by Borel–Cantelli, $\mathbb{P}(|\Upsilon_n - a| \leq n^{-1/3} \text{ eventually}) = 1$.

Here, the bound $n^{-1/3}$ was arbitrary. An upper bound of $n^{-(1/2-\delta)}$ with any $\delta > 0$ would result in the identical result—we fix $n^{-1/3}$ for definiteness. We end this section by defining the Yule process. Properties of this process will be needed in the next few sections.

DEFINITION 4.9 (Rate ν Yule process). Fix $\nu > 0$. A rate ν Yule process is a pure birth process $\{Y_\nu(t) : t \geq 0\}$ with $Y_\nu(0) = 1$ and where the rate of birth of new individuals is proportional to size of the current population. More precisely,

$$\mathbb{P}\big(Y_\nu(t+) - Y_\nu(t)|\mathcal{F}(t)\big) := \nu Y_\nu(t)\,dt + o(dt),$$

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

The following is a standard property of the Yule process; see, for example, [45], Section 2.5.

LEMMA 4.10. *Fix time $t > 0$ and rate $\nu > 0$. Then the random variable $Y_\nu(t)$, namely the number of individuals in the population by time t has a Geometric distribution with parameter $p = e^{-\nu t}$, namely*

$$\mathbb{P}(Y_\nu(t) = k) = e^{-\nu t}\big(1 - e^{-\nu t}\big)^{k-1}, \qquad k \geq 1.$$

4.2. *Convergence of the degree distribution.* In this section, we will prove Theorem 2.1. Recall the description of the limit random variable D_θ in Section 1.3. It will be easier to deal with the random variable $D_\theta^{\mathsf{out}} := D_\theta - 1$. Then the distribution of D_θ^{out} can be written succinctly as:

(a) with probability γ, $D_\theta^{\mathsf{out}} := Y_{\mathsf{BC}}$ where $Y_{\mathsf{BC}} := D_\alpha - 1 + N_\beta^{D_\alpha}[0, a]$;
(b) with probability $1 - \gamma$, $D_\theta^{\mathsf{out}} = Y_{\mathsf{AC}}$ where $Y_{\mathsf{AC}} := N_\beta[0, \mathsf{Age}]$.

Now recall that for any time t and vertex v born before time t, $d_v(t)$ denotes the number of children (out-degree) of vertex v at time t. For fixed $k \geq 0$, define

(4.9) $$\bar{N}_n^{\mathsf{BC}}(k) := \sum_{v \in \mathsf{BP}_\theta(\tau_n)} \mathbb{1}\big\{T_v \leq \tau_{\gamma n}, d_v(\tau_n) \geq k\big\}$$

and

(4.10) $$\bar{N}_n^{\mathsf{AC}}(k) := \sum_{v \in \mathsf{BP}_\theta(\tau_n)} \mathbb{1}\big\{T_v > \tau_{\gamma n}, d_v(\tau_n) \geq k\big\}.$$

In words, $\bar{N}_n^{\mathsf{BC}}(k)$ denotes the number of vertices that were born before the change point and have out-degree at least k by time τ_n (thus in the tree $\mathcal{T}_{\theta,n}$) whilst $\bar{N}_n^{\mathsf{AC}}(k)$ is defined similarly but for vertices born after the change point $\tau_{\gamma n}$. The following proposition is equivalent to Theorem 2.1.

PROPOSITION 4.11. *Fix $k \geq 0$. Then we have*

(4.11) $$\frac{\bar{N}_n^{\mathsf{BC}}(k)}{n} \xrightarrow{P} \gamma \mathbb{P}(Y_{\mathsf{BC}} \geq k), \qquad \frac{\bar{N}_n^{\mathsf{AC}}(k)}{n} \xrightarrow{P} (1 - \gamma)\mathbb{P}(Y_{\mathsf{AC}} \geq k),$$

as $n \to \infty$.

The rest of this section deals with proving this proposition.

4.2.1. *Analysis of* $\bar{N}_n^{\mathsf{BC}}(\cdot)$. We start with the easier case. We will need some more notation. For fixed $0 \leq j, k$, define $\bar{N}_n^{\mathsf{BC}}(j:k)$ for the number of vertices that were born before the change point $\tau_{\gamma n}$ with out-degree exactly j at time $\tau_{\gamma n}$ that end up with at least k children by time τ_n. Note that

$$\sum_{j \geq k} \bar{N}_n^{\mathsf{BC}}(j:k) = N_n(k+1, \gamma n)$$

namely the number of vertices with total degree $k + 1$ (thus out-degree k) in the tree before change point $\mathcal{T}_{\gamma n}$. Recall that by Lemma 4.4, the asymptotic degree distribution of $\mathcal{T}_{\gamma n}$ is D_α, and thus the asymptotic out-degree distribution of the tree $\mathcal{T}_{\gamma n}$ is $D_\alpha^{\mathsf{out}} = D_\alpha - 1$. Using the form of Y_{BC}, it is thus enough to show for each fixed $0 \leq j \leq k$,

$$(4.12) \qquad \frac{\bar{N}_n^{\mathsf{BC}}(j:k)}{n} \xrightarrow{\text{a.e.}} \gamma \mathbb{P}(D_\alpha^{\mathsf{out}} = j) \mathbb{P}(\mathcal{P}_\beta^{j+1}[0, a] \geq k - j).$$

We start with the following simple lemma.

LEMMA 4.12. *Fix $0 < p, q < 1$, a sequence of non-negative integer valued random variables $\{N_n : n \geq 1\}$ and a sequence $\{q_n : n \geq 1\} \in [0, 1]$. Conditional on N_n, let S_n be a Binomial(N_n, q_n) random variable. Further suppose that*

$$\frac{N_n}{n} \xrightarrow{\text{a.e.}} p, \qquad q_n \to q.$$

Then $S_n/n \xrightarrow{\text{a.e.}} pq$.

PROOF. We assume we work on a rich enough probability space where we can couple $\{S_n : n \geq 1\}$ with a sequence $\{\tilde{S}_n : n \geq 1\}$ where $\tilde{S}_n$ is Binomial(np, q_n) such that $|S_n - \tilde{S}_n| \leq |N_n - np|$. Standard exponential tail bounds for the Binomial distribution coupled with Borel–Cantelli and the hypothesis of the lemma imply that $\tilde{S}_n/n \xrightarrow{\text{a.e.}} pq$. Since $|S_n - \tilde{S}_n|/n \leq |N_n/n - p|$, again using the hypothesis of the lemma completes the proof. $\square$

We now proceed with the proof. Recall the definition of the random variable $\bar{N}_n^{\mathsf{BC}}(j:k)$ at the beginning of this section. In the same vein, for each $s \geq 0$ define $\bar{Z}_n^{\mathsf{BC}}((j:k), s)$ for the number of vertices born before the change point $\tau_{\gamma n}$ such that at $\tau_{\gamma n}$ they have out-degree exactly j and further by time $\tau_{\gamma n} + s$ they have degree at least k. Then note that conditional on the information at time $\tau_{\gamma n}$:

$$(4.13) \qquad \bar{Z}_n^{\mathsf{BC}}((j:k), s) \stackrel{d}{=} \text{Bin}(N_n(j+1, \gamma n), \mathbb{P}(\mathcal{P}_\beta^{j+1}[0, s] \geq k - j)).$$

Further, the random variables of interest $\bar{N}_n^{\mathsf{BC}}(j:k) = \bar{Z}_n^{\mathsf{BC}}((j:k), \Upsilon_n)$ where Υ_n is as in Lemma 4.7. Thus writing $a_n^+ = a + n^{-1/3}$ and $a_n^- = a - n^{-1/3}$ and using Lemma 4.8,

$$(4.14) \qquad \bar{Z}_n^{\mathsf{BC}}((j:k), a_n^-) \leq \bar{N}_n^{\mathsf{BC}}(j:k) \leq \bar{Z}_n^{\mathsf{BC}}((j:k), a_n^+) \qquad \text{eventually a.s.}$$

Using the Binomial convergence Lemma 4.12 and noting that by Lemma 4.4 and choice of a_n^+, a_n^-, the hypothesis of this lemma are satisfied, implies that

$$\frac{\bar{Z}_n^{\mathsf{BC}}((j:k), a_n)}{n} \xrightarrow{\text{a.e.}} \gamma \mathbb{P}(D_\alpha^{\mathsf{out}} = j)\mathbb{P}(\mathcal{P}_\beta^{j+1}[0, a] \geq k - j),$$

where take a_n as either a_n^+ or a_n^-. Now using (4.14) proves (4.12). This completes the analysis of $\bar{N}_n^{\mathsf{BC}}(\cdot)$. $\square$

4.2.2. *Analysis of $\bar{N}_n^{\mathsf{AC}}(\cdot)$.* We start by setting up some notation. Fix $k \geq 0$ and define the function

$$(4.15) \qquad g_k(u) := \mathbb{P}(\mathcal{P}_\beta[0, u] \geq k), \qquad u \geq 0.$$

Here, $\mathcal{P}_\beta$ is the offspring point process with attachment parameter β. Then writing out the form of the distribution of Y_{AC} more explicitly [and using the definition of a from (1.5)], to prove the second assertion of (4.11), we want to show

$$(4.16) \qquad \frac{\bar{N}_n^{\mathsf{AC}}(k)}{n} \xrightarrow{\text{P}} \gamma(2 + \beta) \int_0^a e^{(2+\beta)u} g_k(a - u)\, du.$$

For $s \geq 0$, define $\bar{Z}_n^{\mathsf{AC}}(k, s)$ for the number of individuals born in the interval $[\tau_{\gamma n}, \tau_{\gamma n} + s]$ such that by time $\tau_{\gamma n} + s$, these vertices have at least k children. Then note that $\bar{N}_n^{\mathsf{AC}}(k) = \bar{Z}_n^{\mathsf{AC}}(k, \Upsilon_n)$. Mimicking the proof of $N_n^{\mathsf{BC}}(k)$, it is enough to show that

$$(4.17) \qquad \frac{\bar{Z}_n^{\mathsf{AC}}(k, a_n)}{n} \xrightarrow{\text{P}} \gamma(2 + \beta) \int_0^a e^{(2+\beta)u} g_k(a - u)\, du,$$

where a_n is either the sequence $a_n^- = a - n^{-1/3}$ or $a_n^+ = a + n^{-1/3}$. To ease notation, we will just work with the sequence $a_n = a$. The entire proof goes through by replacing a in the steps below by a_n.

We start with a few preliminary results. The first result describes strong concentration results of the growth of the number of individuals in BP_θ^n in the interval $[\tau_{\gamma n}, \tau_{\gamma n} + s]$. Define the process

$$(4.18) \qquad \mathscr{Z}_n(u) := \left|\mathsf{BP}_\theta^n(\tau_{\gamma n} + u)\right|, \qquad 0 \leq u \leq a.$$

PROPOSITION 4.13. *There exists a constant $C < \infty$ such that for all n,*

$$\mathbb{P}\left(\sup_{0 \leq u \leq a} |\mathscr{Z}_n(u) - n\gamma e^{(2+\beta)u}| > \sqrt{n \log n}\right) \leq \frac{C}{\log n}.$$

PROOF. The plan is to use Doob's L^2-maximal inequality for continuous-time martingales (see, e.g., [36], Chapter 1.9). For this, we will need to derive martingales related to the process $\mathscr{Z}_n(\cdot)$. Throughout we will write $\{\mathcal{F}_t^n : 0 \leq t \leq a\}$ for the filtration $\{\mathsf{BP}_\theta(\tau_{\gamma n} + t) : 0 \leq t \leq a\}$.

Recall from the rate description in (4.7) that $\mathscr{Z}_n(\cdot)$ is a pure birth process such for any $t \geq 0$, conditional on $\mathcal{F}_t^n$, $\mathscr{Z}_n(t) \rightsquigarrow \mathscr{Z}_n(t) + 1$ at rate $(2 + \beta)\mathscr{Z}_n(t) - 1$. Arguing as in the proof of Lemma 4.5, it is easy to check that the process

$$(4.19) \qquad M_1(t) := \left(e^{-(2+\beta)t} \mathscr{Z}_n(t) - n\gamma\right) - \frac{e^{-(2+\beta)t} - 1}{2 + \beta}, \qquad 0 \leq t \leq a,$$

is a mean-zero martingale. This in particular gives that

$$(4.20) \qquad e^{-(2+\beta)t} \mathbb{E}\left(\mathscr{Z}_n(t)\right) = n\gamma + \frac{e^{-(2+\beta)t} - 1}{2 + \beta}, \qquad 0 \leq t \leq a.$$

By Doob's L^2-maximal inequality applied to the process $M_1(\cdot)$, we have for any $\lambda > 0$,

$$(4.21) \qquad \mathbb{P}\left(\sup_{0 \leq t \leq a}\left|\left(e^{-(2+\beta)t} \mathscr{Z}_n(t) - n\gamma\right) - \frac{e^{-(2+\beta)t} - 1}{2 + \beta}\right| \geq \lambda\right) \leq \frac{\mathbb{E}(M_1^2(a))}{\lambda^2}.$$

If we can show there exists a constant $C < \infty$ such that $\mathbb{E}(M_1^2(a)) \leq Cn$, using $\lambda = 0.5\sqrt{n \log n}$ and algebraic manipulation of (4.21) completes the proof. So let us now derive this bound on $\mathbb{E}(M_1^2(a))$.

First, squaring the expression in (4.19), expanding and using (4.20) gives for $t \geq 0$,

$$(4.22) \qquad \mathbb{E}(M_1^2(t)) = \mathbb{E}\left(e^{-(2+\beta)t} \mathscr{Z}_n(t) - n\gamma\right)^2 - \left(\frac{e^{-(2+\beta)t} - 1}{2 + \beta}\right)^2.$$

Thus we need to understand the evolution of the process $\mathscr{Z}_n^2(\cdot)$. Again using the rate description of $\mathscr{Z}_n$, this process undergoes a change

$$\Delta \mathscr{Z}_n^2(t) := \mathscr{Z}_n^2(t+) - \mathscr{Z}_n^2(t) = \left(1 + 2\mathscr{Z}_n(t)\right),$$

at rate $(2 + \beta)\mathscr{Z}_n(t) - 1$. Using this, one may check that the following process on $[0, a]$:

$$(4.23) \qquad M_2(t) := e^{-2(2+\beta)t} \mathscr{Z}_n^2(t) - \int_0^t e^{-2(2+\beta)s} \beta \mathscr{Z}_n(s)\, ds - \frac{e^{-2(2+\beta)t}}{2(2 + \beta)},$$

is also a martingale. In particular, since first moments are conserved,

$$(4.24) \quad \mathbb{E}\left(e^{-2(2+\beta)} \mathscr{Z}_n^2(t)\right) = n^2\gamma^2 + \int_0^t \beta e^{-2(2+\beta)s} \mathbb{E}\left(\mathscr{Z}_n(s)\right) ds - \frac{e^{-2(2+\beta)t} - 1}{2(2 + \beta)}.$$

Using (4.20) shows that there exists a constant C such that

$$(4.25) \qquad \left|\mathbb{E}\left(e^{-2(2+\beta)} \mathscr{Z}_n^2(t)\right) - n^2\gamma^2\right| \leq n\gamma.$$

Expanding the first bracket in (4.22), using (4.20) and (4.25) shows that $\mathbb{E}(M_1^2(a)) \leq Cn$ for some constant C. This completes the proof. $\quad\square$

Now divide the interval $[\tau_{\gamma n}, \tau_{\gamma n} + a]$ into $\lceil an^{1/3} \rceil$ intervals of length $n^{-1/3}$:

$$\left\{ \left[\tau_{\gamma n}, \tau_{\gamma n} + \frac{1}{n^{1/3}} \right], \left[\tau_{\gamma n} + \frac{1}{n^{1/3}}, \tau_{\gamma n} + \frac{2}{n^{1/3}} \right], \ldots, \right.$$

$$\left. \left[\tau_{\gamma n} + \frac{an^{1/3} - 1}{n^{1/3}}, \tau_{\gamma n} + \frac{an^{1/3}}{n^{1/3}} \right] \right\},$$

To ease notation, write the above collection as $\{\mathcal{I}_i : 0 \le i \le an^{1/3} - 1\}$. Further, let $\tau_i^n = \tau_{\gamma n} + i/n^{1/3}$ with $\tau_0^n = \tau_{\gamma n}$ so that $\mathcal{I}_i = [\tau_i^n, \tau_{i+1}^n]$.

Now write Birth_i for the collection of vertices that were born in interval $\mathcal{I}_i$ (i.e., the collection of vertices v with birth times $T_v \in \mathcal{I}_i$) and write

$$\mathscr{Z}_n(\mathcal{I}_i) := |\mathsf{Birth}_i| = \mathscr{Z}_n(\tau_{i+1}^n) - \mathscr{Z}_n(\tau_i^n),$$

for the number of individuals born in this interval. Then the following is an easy corollary of Proposition 4.13.

COROLLARY 4.14. *We have*

$$\mathbb{P}\left(\bigcap_{i=0}^{an^{1/3}-1} \left\{ \left| \mathscr{Z}_n(\mathcal{I}_i) - (2+\beta)\gamma n^{2/3} e^{\frac{(2+\beta)i}{n^{1/3}}} \right| < 2\sqrt{n \log n} \right\} \right) \to 1,$$

as $n \to \infty$.

For future reference, write $\mathscr{G}_n$ for the event above. Namely,

$$(4.26) \qquad \mathscr{G}_n := \bigcap_{i=0}^{an^{1/3}-1} \left\{ \left| \mathscr{Z}_n(\mathcal{I}_i) - (2+\beta)\gamma n^{2/3} e^{\frac{(2+\beta)i}{n^{1/3}}} \right| < 2\sqrt{n \log n} \right\}.$$

Now for each interval $\mathcal{I}_i$, we will partition the vertices born in this interval into two classes:

(a) The collection of good vertices $\mathcal{G}_i$: This consists of all $v \in \mathsf{Birth}_i$ such that they produce *no* children by the end of the interval, that is, vertices v with $T_v \in [\tau_{\gamma n} + i/n^{1/3}, \tau_{\gamma n} + (i+1)/n^{1/3}]$ such that by time $\tau_{\gamma n} + (i+1)/n^{1/3}$, vertex v still has no children. Note that since the intervals are of time length $n^{-1/3}$, one expects a large proportion of vertices born in the interval $\mathcal{I}_i$ to be good. Write $\mathscr{Z}_n^{\mathsf{good}}(\mathcal{I}_i) = |\mathcal{G}_i|$ for the number of good vertices in $\mathcal{I}_i$.

(b) The collection of bad vertices $\mathcal{B}_i := \mathsf{Birth}_i \setminus \mathcal{G}_i$: This consists of all vertices born in $\mathcal{I}_i$ which produce at least one child by time $\tau_{\gamma n} + i/n^{1/3}$. Write $\mathscr{Z}_n^{\mathsf{bad}}(\mathcal{I}_i) = |\mathcal{B}_i|$ for the number of such bad vertices in $\mathcal{I}_i$. Write

$$\mathscr{Z}_n^{\mathsf{bad}} := \sum_{i=0}^{an^{1/3}-1} \mathscr{Z}_n^{\mathsf{bad}}(\mathcal{I}_i)$$

for the total number of bad vertices.

Fix a constant C and define the event $B_i^n = \{\mathscr{Z}_n^{\mathsf{bad}}(\mathcal{I}_i) \geq Cn^{1/3}\log n\}$. These events depend on C but we suppress this in the notation.

PROPOSITION 4.15.   *The constant $C < \infty$ can be chosen large enough such that $\mathbb{P}(\bigcup_{i=1}^{an^{1/3}} B_i^n) \to 0$ as $n \to \infty$. In particular, for the total number of bad vertices we have $\mathscr{Z}_n^{\mathsf{bad}} = O_P(n^{2/3}\log n)$.*

PROOF.   Fix an interval $\mathcal{I}_i$. Note that every bad vertex is one of two types:

(a) A vertex that is a direct child of a vertex born before this time interval. Write $\mathcal{D}_n^{\mathsf{bad}}$ for these *direct* bad vertices and write $\mathscr{D}_n^{\mathsf{bad}}(\mathcal{I}_i) = |\mathcal{D}_n^{\mathsf{bad}}|$ for the number of such vertices. Further, write $\mathscr{D}_{n,\star}^{\mathsf{bad}}(\mathcal{I}_i)$ for the total number of descendants of direct bad vertices born in the interval $\mathcal{I}_i$ (including the direct bad vertices).

(b) A vertex that is bad and is a child of a vertex born in $\mathcal{I}_i$. Thus the parent of this vertex is necessarily bad.

Thus in particular we have that $\mathscr{Z}_n^{\mathsf{bad}}(\mathcal{I}_i) \leq \mathscr{D}_{n,\star}^{\mathsf{bad}}(\mathcal{I}_i)$. Now note that direct bad vertices in $\mathcal{D}_n^{\mathsf{bad}}$ are created via the following steps:

(i) A descendant (maybe good or bad) of a vertex born before $\mathcal{I}_i$ is born into the system. The number of such individuals $\mathscr{R}_n(\mathcal{I}_i) \leq \mathscr{Z}_n(\mathcal{I}_i)$, the total number of individuals born in the interval $\mathcal{I}_i$. Using Corollary 4.14, there exists a constant C such that whp as $n \to \infty$, *for all* the intervals $0 \leq i \leq an^{1/3} - 1$, $\mathscr{R}_n(\mathcal{I}_i) \leq Cn^{2/3}$.

(ii) Conditional on *all these* descendants of vertices born before $\mathcal{I}_i$, such a descendant has to give birth to one individual in the interval $[i/n^{1/3}, (i+1)/n^{1/3}]$. Recall that the time to give birth to the first child is an exponential random variable E_1 with rate $(2 + \beta)$. Thus the probability of birthing this first child is bounded by

$$p_n = \mathbb{P}(E_1 \leq n^{-1/3}) \sim \frac{2 + \beta}{n^{1/3}}.$$

Further, by construction, none of these vertices can have a parent child relationship, and thus their offspring lineages evolve independently.

In particular, conditional on all descendants of vertices born before time interval $\mathcal{I}_i$,

$$(4.27) \qquad\qquad \mathscr{D}_n^{\mathsf{bad}}(\mathcal{I}_i) \leq_{\mathsf{st}} \mathrm{Bin}(\mathscr{R}_n(\mathcal{I}_i), p_n).$$

Here, st denotes stochastic domination. Thus using Corollary 4.14, (4.27) and standard tail bounds for the binomial distribution implies that there exists a constant $C < \infty$ such that

$$(4.28) \qquad \mathbb{P}(\mathscr{D}_n^{\mathsf{bad}}(\mathcal{I}_i) \leq Cn^{1/3}\log n \text{ for all } 0 \leq i \leq (an^{1/3} - 1)) \to 1,$$

as $n \to \infty$.

Let us now complete the analysis of $\mathscr{D}_{n,\star}^{\mathsf{bad}}(\mathcal{I}_i)$. Let us start with the evolution of descendants of a single bad *direct* vertex after it gives birth to its child. This

process then starts reproducing at rate $2 + \beta + 1 + \beta = 3 + 2\beta$. Further, whenever a new vertex is added to the system, the rate of production increases by at most $2 + \beta$. Thus writing $K = \lfloor 3 + 2\beta \rfloor$ and $\nu = 2 + \beta$, the number of descendants of such a bad vertex can be bounded by a rate ν Yule process (see Definition 4.9) that starts with K individuals at time zero. Write $\{Y_\nu^K(t) : t \geq 0\}$ for such a process. Thus the number of descendants of such a bad vertex in the time interval $[\tau_{\gamma n} + i/n^{1/3}, \tau_{\gamma n} + (i+1)/n^{1/3}]$ can be stochastically bounded by $Y_\nu^K(n^{-1/3})$. In particular, conditional on $\mathcal{D}_n^{\mathsf{bad}}(\mathcal{I}_i)$,

$$(4.29) \qquad \mathscr{D}_{n,\star}^{\mathsf{bad}}(\mathcal{I}_i) \leq_{\mathsf{st}} \sum_{j=1}^{\mathscr{D}_n^{\mathsf{bad}}(\mathcal{I}_i)} Y_\nu^{K,(j)}(n^{-1/3}).$$

Here, $\{Y_\nu^{K,(j)}(\cdot) : j \geq 1\}$ are an i.i.d. collection of Yule processes with distribution $Y_\nu^K(\cdot)$. Using the explicit distribution of the Yule process at a fixed time (Lemma 4.10), it is easy to check that given constant $C > 0$ we can find $A > 0$ such that

$$(4.30) \quad \mathbb{P}(\mathscr{D}_{n,\star}^{\mathsf{bad}}(\mathcal{I}_i) \geq 10KCn^{1/3} \log n | \mathscr{D}_n^{\mathsf{bad}}(\mathcal{I}_i) \leq Cn^{1/3} \log n) \leq \exp(-An^{1/3}).$$

Using this exponential bound with (4.28) completes the proof. $\quad\square$

We now proceed with the proof of (4.17). For $0 \leq i \leq an^{1/3} - 1$, let $Z_n^{\mathsf{good}}(k, a : \mathcal{I}_i)$ be the number of *good* vertices in Birth_i which have at least k children by time a. Then note that conditional on $\mathsf{BP}_\theta^n(\tau_{i+1}^n)$,

$$(4.31) \qquad Z_n^{\mathsf{good}}(k, a : \mathcal{I}_i) \stackrel{d}{=} \mathrm{Bin}\left(\mathscr{Z}_n^{\mathsf{good}}(\mathcal{I}_i), g_k\left(a - \frac{i+1}{n^{1/3}}\right)\right).$$

Define the events

$$G_i^n := \left\{\left|Z_n^{\mathsf{good}}(k, a : \mathcal{I}_i) - \gamma(2+\beta)n^{2/3} e^{\frac{(2+\beta)i}{n^{1/3}}} g_k\left(a - \frac{i+1}{n^{1/3}}\right)\right| < Cn^{1/3} \log n\right\}.$$

PROPOSITION 4.16. *There exists a constant $C < \infty$ such that* $\mathbb{P}(\bigcap_{i=1}^{an^{1/3}} G_i^n) \to 1$ *as $n \to \infty$.*

PROOF. Note that $\mathscr{Z}_n^{\mathsf{good}}(\mathcal{I}_i) = \mathscr{Z}_n(\mathcal{I}_i) - \mathscr{Z}_n^{\mathsf{bad}}(\mathcal{I}_i)$. Combining Corollary 4.14 with Proposition 4.15 implies that

$$\mathbb{P}\left(\bigcap_{i=0}^{an^{1/3}-1} \{|\mathscr{Z}_n^{\mathsf{good}}(\mathcal{I}_i) - (2+\beta)\gamma n^{2/3} e^{\frac{(2+\beta)i}{n^{1/3}}}| < 3\sqrt{n \log n}\}\right) \to 1.$$

Now using the distributional identity (4.31) and standard tail bounds for the Binomial distribution completes the proof. $\quad\square$

We are finally in a position to complete the proof of (4.17), and thus (4.11). First, note that

$$(4.32) \qquad \sum_{i=0}^{an^{1/3}-1} Z_n^{\mathsf{good}}(k, a : \mathcal{I}_i) \leq \bar{Z}_n^{\mathsf{AC}}(k, a) \leq \sum_{i=0}^{an^{1/3}-1} Z_n^{\mathsf{good}}(k, a : \mathcal{I}_i) + \mathscr{Z}_n^{\mathsf{bad}}.$$

Using Proposition 4.15, $n^{-1} \mathscr{Z}_n^{\mathsf{bad}} \xrightarrow{\mathrm{P}} 0$. Using Proposition 4.16,

$$\frac{\sum_{i=1}^{an^{1/3}} Z_n^{\mathsf{good}}(k, a : \mathcal{I}_i)}{n} \sim \frac{\gamma(2+\beta)}{n^{1/3}} \sum_{i=0}^{an^{1/3}-1} e^{\frac{(2+\beta)i}{n^{1/3}}} g_k\left(a - \frac{i+1}{n^{1/3}}\right)$$

$$\to \gamma(2+\beta) \int_0^a e^{(2+\beta)u} g_k(a-u)\, du.$$

This completes the proof of the convergence of the degree distribution of the model to the asserted limit in Theorem 2.1. □

We conclude this section with a related result regarding the evolution of the degree distribution. This follows by directly modifying the proof above. Recall the definitions of $N_n(k, m)$ and $\hat{N}_n(k, t)$ from Section 2.2. For future use, define for each $k \geq 1$ and $0 \leq t \leq 1$

$$(4.33) \qquad N_{n,\geq}(k, m) = \sum_{j \geq k} N_n(j, m), \qquad \hat{N}_{n,\geq}(k, t) = \sum_{j \geq k} \hat{N}_n(j, t),$$

namely the number of vertices with degree at least k respectively at discrete time m and at time t when we rescale time by n. Write $\hat{q}_{\geq}^{(n)}(k, t) = \hat{N}_{n,\geq}(k, t)/n$. Note that since we divide by n and not nt in this expression we have $\sum_{k=1}^{\infty} \hat{q}_{\geq}^{(n)}(k, t) = t$. Now note that by Lemma 4.4 we have for each fixed $0 < t \leq \gamma$,

$$(4.34) \qquad \hat{p}^{(n)}(k, t) \xrightarrow{\mathrm{P}} p_\alpha(k) = p^{(\infty)}(k, \gamma),$$

where $p_\alpha(k)$ as in (1.1) is the limiting degree distribution with no change point. For $\gamma \leq t \leq 1$, akin to the definition of a in (1.5) define

$$(4.35) \qquad\qquad\qquad a(t) := \frac{1}{2+\beta} \log \frac{t}{\gamma}.$$

Analogous to the definition of D_θ in Section 1.3, define $D_\theta(t)$ by replacing a by $a(t)$ throughout the construction. Thus $D_\theta = D_\theta(1)$. Let

$$(4.36) \qquad\qquad p^{(\infty)}(k, t) := \mathbb{P}(D_\theta(t) = k), \qquad k \geq 1, \gamma \leq t \leq 1.$$

Let $p_{\geq}^{(\infty)}(k, t) = \mathbb{P}(D_\theta(t) \geq k)$. For $0 \leq t \leq 1$, let $q_{\geq}^{(\infty)}(k, t) = t p_{\geq}^{(\infty)}(k, t)$.

PROPOSITION 4.17.   *For all $k \geq 1$, we have*

$$\sup_{0 \leq t \leq 1} |\hat{q}_{\geq}^{(n)}(k, t) - q_{\geq}^{(\infty)}(k, t)| \xrightarrow{P} 0,$$

as $n \to \infty$.

PROOF.   For fixed $t \geq \gamma$, define the stopping time

$$\tau_{tn} = \inf\{s : |\mathsf{BP}_{\boldsymbol{\theta}}^n(s)| = tn\},$$

namely the first time that the continuous-time embedding reaches size tn. Note that at this time, the corresponding tree has distribution $\mathcal{T}_{tn}$. Write $\Upsilon_n(t) = \tau_{tn} - \tau_{\gamma n}$ for the amount of (continuous) time it takes for the process to reach this size after the change point. Then note that by Proposition 4.13 we can choose an appropriate constant $C < \infty$ such that

$$(4.37) \qquad \mathbb{P}\left(\sup_{\gamma \leq t \leq 1} |\Upsilon(t) - a(t)| \leq C\sqrt{\frac{\log n}{n}} \right) \to 1,$$

as $n \to \infty$, where $a(t)$ is as defined in (4.35). Repeating the above proof for the convergence of degree distribution and replacing a by $a(t)$ throughout the argument shows that for each $t \geq \gamma$ $\hat{N}_{n,\geq}(k,t)/nt \xrightarrow{\text{P}} \mathbb{P}(D_{\boldsymbol{\theta}}(t) \geq k)$. Combining this with (4.34) implies that we have pointwise convergence $\hat{q}_{\geq}^{(n)}(k,t) \to q_{\geq}^{(\infty)}(k,t)$. Now note that for each fixed n, the function $\hat{q}_{\geq}^{(n)}(k,\cdot)$ is nondecreasing on $[0,1]$ while the limit function is also monotonically increasing and continuous (and thus uniformly continuous). Given $\varepsilon > 0$, fix $\delta > 0$ such that for any $t, s \in [0,1]$ with $|t - s| < \delta$,

$$\left| q_{\geq}^{(\infty)}(k,t) - q_{\geq}^{(\infty)}(k,s) \right| < \frac{\varepsilon}{4}.$$

Divide $[0,1]$ into intervals $\{[i\delta,(i+1)\delta]\}$ for $1 \leq i \leq 1/\delta$ of length δ. Via the pointwise convergence above, get $n_0 < \infty$ large such that for all $n > n_0$:

$$(4.38) \qquad \mathbb{P}\left(\sup_{1 \leq i \leq \frac{1}{\delta}} |\hat{q}_{\geq}^{(n)}(k,i\delta) - q_{\geq}^{(\infty)}(k,i\delta)| < \frac{\varepsilon}{4} \right) \geq 1 - \varepsilon.$$

Write $G_n(\varepsilon,\delta)$ for the event in the above equation. Then on this event, by the choice of δ, for all i we have $|\hat{q}_{\geq}^{(n)}(k,i\delta) - \hat{q}_{\geq}^{(n)}(k,(i+1)\delta)| \leq \varepsilon/2$. Using monotonicity, for any $t \in [i\delta,(i+1)\delta]$, $|\hat{q}_{\geq}^{(n)}(k,i\delta) - \hat{q}_{\geq}^{(n)}(k,t)| \leq \varepsilon/2$. By the triangle inequality on $G_n(\varepsilon,\delta)$, for all $t \in [0,1]$ and $n > n_0$,

$$\left| \hat{q}_{\geq}^{(n)}(k,t) - q_{\geq}^{(\infty)}(k,t) \right| \leq \left| \hat{q}_{\geq}^{(n)}(k,t) - \hat{q}_{\geq}^{(n)}(k,i\delta) \right| + \left| \hat{q}_{\geq}^{(n)}(k,i\delta) - q_{\geq}^{(\infty)}(k,i\delta) \right|$$

$$+ \left| q_{\geq}^{(\infty)}(k,i\delta) - q_{\geq}^{(\infty)}(k,t) \right|$$

$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon.$$

Since n_0 is independent of t, this completes the proof. $\quad\square$

4.3. *Proof of the tail exponent for the limiting degree distribution.*   The aim of this section is to prove the asserted tail bound, namely (2.1). First, note that the

lower tail bound is obvious since with probability γ, D_{θ} stochastically dominates D_{α} and by (1.2), D_{α} has the asserted tail behavior. The main crux is then proving the upper bound, namely

$$(4.39) \qquad \mathbb{P}(D_{\theta} \geq x) \leq c'/x^{2+\alpha}.$$

Recall Definition 4.9 of the Yule process and in particular Lemma 4.10 on finite-time marginal distribution of the Yule process.

Now note that in the description of the limit random variable D_{θ}, with probability $1 - \gamma$, $D_{\theta} = N_{\beta}[0, \mathsf{Age}] \leq_{\mathsf{st}} N_{\beta}[0, a]$ where as before $\leq_{\mathsf{st}}$ represents stochastic domination. Now define

$$(4.40) \qquad \nu = 2 + \beta, \qquad K = \lfloor 1 + \beta \rfloor.$$

As before, let Y_{ν}^{K} be a rate ν Yule process started with K individuals at time zero. Comparing the rate of production of new individuals in the point process $\mathcal{P}_{\beta}$ with Y_{ν}^{K}, we get that $N_{\beta}[0, a] \leq_{\mathsf{st}} Y_{\nu}^{K}(a)$. By Lemma 4.10, $Y_{\nu}^{K}(a)$ is the sum of K independent geometric random variables. Using the fact that a geometric random variable has finite moment generating function in a neighborhood of zero and an elementary Chernoff bound implies that there exist constants $\kappa, \kappa' > 0$ such that for all $x \geq 1$, we have an exponential tail bound:

$$(4.41) \qquad \mathbb{P}\big(N_{\beta}[0, \mathsf{Age}] > x\big) \leq \mathbb{P}\big(Y_{\nu}^{K}(a) > x\big) \leq \kappa' \exp(-\kappa x).$$

Thus when, with probability $1 - \gamma$, $D_{\theta} = N_{\beta}[0, \mathsf{Age}]$, then the corresponding random variable has an exponential tail. Thus the main contribution to the tail arises when with probability γ, $D_{\theta} = D_{\alpha} + N_{\beta}^{D_{\alpha}}[0, a]$. Arguing as above (and assuming $\beta \geq 1$), conditional on $D_{\alpha} = k$, we have

$$N_{\beta}^{D_{\alpha}}[0, a] \leq_{\mathsf{st}} \sum_{j=1}^{k} Y_{\nu}^{K,(j)}(a),$$

where, as in (4.29), $\{Y_{\nu}^{K,(j)}(\cdot) : j \geq 1\}$ are a collection of independent rate ν Yule processes each started at time zero with K individuals and independent of D_{α}. The following elementary lemma completes the proof.

LEMMA 4.18. *Let $D \geq 1$ be a nonnegative integer valued random variable with $\mathbb{P}(D \geq x) \leq c/x^{\gamma}$ for all $x \geq 1$, for two constants $c, \gamma > 0$. Let $\{Y_i : i \geq 1\}$ be a sequence of independent and identically distributed positive integer valued random variables, independent of D. Consider the random variable $D^* := \sum_{j=1}^{D} Y_i$. If Y_1 has finite moment generating function in a neighborhood of zero then there exists a constant $c' > 0$ such that for all $x \geq 1$,*

$$\mathbb{P}\big(D^* \geq x\big) \leq c'/x^{\gamma}.$$

PROOF. For the rest of the proof, write $\mu = \mathbb{E}(Y_1) < \infty$. Then note that

$$\mathbb{P}(D^* \geq x) \leq \sum_{j=1}^{\frac{x}{2\mu}} \mathbb{P}(D = j)\mathbb{P}\left(\sum_{i=1}^{j} Y_i \geq x\right) + \mathbb{P}\left(D \geq \frac{x}{2\mu}\right)$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{\frac{x}{2\mu}} Y_i \geq x\right) + \frac{c}{x^\gamma},$$

where the second equation follows using the fact that $Y_i \geq 1$ for all i and the tail bound for D from the hypothesis of the lemma. To complete the proof, note that standard large deviation bounds imply (since Y_i has a finite moment generating function about zero) imply that there exist constants κ, κ' such for all large x

$$\mathbb{P}\left(\sum_{i=1}^{\frac{x}{2\mu}} Y_i \geq x\right) \leq \kappa' \exp(-\kappa x).$$

This completes the proof. $\square$

The only item left to complete the proof of Theorem 2.1 is to show that the change point *does* change the degree distribution from the original (no change point) model. In Section 5, we will carry out a detailed analysis of the density of leaves which in particular will show that the asymptotic density of leaves $p_\theta(1) \neq p_\alpha(1)$.

4.4. *Analysis of the maximal degree.* The aim of this section is to prove Theorem 2.2. First, note that, for any fixed $k \geq 1$, writing $M_{\gamma n}(k)$ for the kth maximal degree of a vertex in $\mathcal{T}_{\gamma n}$ namely in the tree just before the change point, using (1.3) implies that $M_{\gamma n}(k)/n^{1/(2+\alpha)}$ converges weakly to a strictly positive random variable. Since $M_n(k) \geq M_{\gamma n}(k)$, this implies that given any $\varepsilon > 0$ and any fixed $k \geq 1$, there exists a constant $K'_\varepsilon > 0$ such that

$$\liminf_{n\to\infty} \mathbb{P}\left(\frac{M_n(k)}{n^{1/(2+\alpha)}} > K'_\varepsilon\right) > 1 - \varepsilon.$$

Thus to complete the proof of Theorem 2.2, we need to show, given any $\varepsilon > 0$, there exists $K_\varepsilon < \infty$ such that

$$(4.42) \qquad \limsup_{n\to\infty} \mathbb{P}\left(\frac{M_n(1)}{n^{1/(2+\alpha)}} < K_\varepsilon\right) \geq 1 - \varepsilon.$$

For any vertex $v \in [n]$ time point $m \in [n]$, write $\deg(v, m)$ for the degree of vertex v in $\mathcal{T}_m$ with the obvious convention that $\deg(v, k) = 0$ if $k < v$. Then note that $M_n(1) = \max(M_{\text{pre}}(n), M_{\text{post}}(n))$ where

$$(4.43) \qquad M_{\text{pre}}(n) := \max_{v\in[1,n\gamma]} \deg(v, n), \qquad M_{\text{post}}(n) := \max_{v\in[n\gamma+1,n]} \deg(v, n).$$

Let us first analyze the maximal degree of vertices that appeared after the change point. Recall the constants a from (1.5) and ν, K from (4.40).

LEMMA 4.19. *We have* $\mathbb{P}(M_{\text{post}}(n) > 2K e^{\nu(a+1)} \log n) \to 0$ *as* $n \to \infty$.

PROOF. We will assume $\beta \geq 1$ below, else replace β with 1 in the rest of the argument below. For simplicity, write $k_n = 2K e^{\nu(a+1)} \log n$. Recall that in the continuous-time embedding, T_v represents the time of birth of vertex v and further for $v \in [\gamma n + 1, n]$, each such vertex is equipped with a offspring point process $\mathcal{P}_\beta^v$. As in Section 4.3, $1 + \mathcal{P}_\beta \leq_{\text{st}} Y_v^K$ where Y_v^K is a rate ν Yule process started with K individuals at time zero. Now note that via our continuous-time embedding,

$$M_{\text{post}}(n) := \max_{v \in [\gamma n + 1, n]} (1 + \mathcal{P}_\beta^v(0, \tau_n - T_v)),$$

since by time τ_n, a vertex born after the change time has been alive for $\tau_n - T_v \leq \tau_n - \tau_{\gamma n} := \Upsilon_n$ units of time. Now

$$
\begin{aligned}
\mathbb{P}(M_{\text{post}}(n) > k_n) &\leq \mathbb{P}(M_{\text{post}}(n) > k_n, \Upsilon_n < a + 1) \\
&\quad + \mathbb{P}(\Upsilon_n > a + 10) \\
&\leq \mathbb{P}\Big(\max_{v \in [\gamma n + 1, n]} (1 + \mathcal{P}_\beta^v(0, a + 1)) > k_n \Big) \\
&\quad + \mathbb{P}(\Upsilon_n > a + 1).
\end{aligned}
$$

(4.44)

Using Lemma 4.7, we have $\limsup_{n \to \infty} \mathbb{P}(\Upsilon_n > a + 1) = 0$. Let $\{Y_{v,v}^K : v \in [\gamma n + 1, n]\}$ be a family of independent rate ν Yule processes started with K individuals at time zero. Using Lemma 4.10, a simple union bound and the choice of k_n implies $\mathbb{P}(\max_{v \in [\gamma n + 1, n]} Y_\beta^v(a + 1) > k_n) \to 0$. $\square$

Thus the above lemma implies that the maximal degree amongst vertices that arrive after the change point is $O_P(\log n)$. To complete the proof of (4.42), it is enough to show that (4.42) holds with $M_n(1)$ replaced by $M_{\text{pre}}(1)$. Thus fix $\varepsilon \in (0, 1)$. Using Proposition 4.6, fix $A = A_\varepsilon$ such that

$$\limsup_{n \to \infty} \mathbb{P}\Big(\tau_{\gamma n} - \frac{1}{2 + \alpha} \log \gamma n > A \Big) \leq \varepsilon/2.$$

(4.45)

Now consider the following process $\mathsf{BP}_{\theta,\star}^n$:

(a) Run the process BP_α until time $t_n(A) := \frac{1}{2+\alpha} \log \gamma n + A$.

(b) At this time: all vertices in $\mathsf{BP}_\alpha(t_n)$ switch to the dynamics with parameter β. Namely, each vertex now reproduces at rate proportional to its out-degree $+ 1 + \beta$.

(c) Run this process for an additional $a + 1$ units of time where a is as in (1.5).

Abusing notation, let $M^\star_{\text{pre},A}(1)$ denote the maximal degree by time $t_n + a + 1$ of all vertices born *before* time t_n. We can obviously couple the original process $\mathsf{BP}^n_{\boldsymbol{\theta}}$ and $\mathsf{BP}^n_{\boldsymbol{\theta},\star}$ such that on the set $\{\tau_{\gamma n} - \frac{1}{2+\alpha}\log\gamma n \le A, \Upsilon_n \le a + 1\}$ we have $M_{\text{pre}}(1) \le M^\star_{\text{pre},A}(1)$. Further note that for any fixed K we have

$$\mathbb{P}\big(M_{\text{pre}}(1) > Kn^{1/(2+\alpha)}\big)$$

$$\le \mathbb{P}\bigg(M_{\text{pre}}(1) > Kn^{1/(2+\alpha)}, \Upsilon_n < a + 1, \tau_{\gamma_n} < \frac{1}{2+\alpha}\log\gamma n + A\bigg)$$

$$+ \mathbb{P}(\Upsilon_n > a + 1) + \mathbb{P}\bigg(\tau_{\gamma_n} > \frac{1}{2+\alpha}\log\gamma n + A\bigg).$$

First, choosing A appropriately as in (4.45) and using Lemma 4.7 we get that for any fixed K,

$$\limsup_{n\to\infty} \mathbb{P}\big(M_{\text{pre}}(1) > Kn^{1/(2+\alpha)}\big) \le \limsup_{n\to\infty} \mathbb{P}\big(M^\star_{\text{pre},A}(1) > Kn^{1/(2+\alpha)}\big) + \varepsilon/2.$$

The following lemma completes the proof of (4.42).

LEMMA 4.20. *Fix $A > 0$. Given any $\varepsilon > 0$, we can choose $K = K(A, \varepsilon) < \infty$ such that*

$$\limsup_{n\to\infty} \mathbb{P}\big(M^\star_{\text{pre},A}(1) > Kn^{1/(2+\alpha)}\big) \le \varepsilon.$$

PROOF. First, note that until time $t_n(A)$, the process $\mathsf{BP}^n_{\boldsymbol{\theta},\star}$ is a the continuous-time version of a (nonchange point) preferential attachment model with attachment parameter α. This continuous-time embedding was used to derive asymptotics for the maximal degree in [6, 7]. In particular, the bounds derived in these papers imply the following for a fixed A: Write $\tilde{M}_n(1)$ for the maximal degree exactly at time $t_n(A)$. Then there exists $L = L(A, \varepsilon) < \infty$ such that

$$(4.46) \qquad \limsup_{n\to\infty} \mathbb{P}\big(\tilde{M}_n(1) > Ln^{1/(2+\alpha)}\big) \le \varepsilon/2.$$

Now note that on the event $\{\tilde{M}_n(1) \le Ln^{1/(2+\alpha)}\}$ at time $t_n + a + 1$, the degree of every fixed vertex in the system is stochastically dominated by a rate ν Yule process started with $Ln^{1/(2+\alpha)}$ vertices at time zero and run for time $a + 1$ where ν is as in (4.40). Write D_n for such a random variable and note that by the description of the dynamics of the Yule process and Lemma 4.10, we have that

$$(4.47) \qquad D_n \stackrel{d}{=} \sum_{j=1}^{Ln^{1/(2+\alpha)}} Y_{\nu,j}(a + 1),$$

where $\{Y_{\nu,j}(a + 1) : j \ge 1\}$ are i.i.d. Geometric random variables with $p = e^{-\nu(a+1)}$. Further note that using Proposition 4.6 on the size of the branching process, we can choose C such that

$$(4.48) \qquad \limsup_{n\to\infty} \mathbb{P}\big(|\mathsf{BP}^n_{\boldsymbol{\theta},\star}(t_n)| > Cn\big) \le \varepsilon/2.$$

Thus on the "good" event

$$\mathcal{G}_n := \{|\mathsf{BP}^n_{\theta,\star}(t_n)| \le Cn, \tilde{M}_n(1) \le Ln^{1/(2+\alpha)}\},$$

we have that

$$M^\star_{\mathrm{pre},A}(1) \le_{\mathrm{st}} \max_{1 \le v \le Cn} D^v_n := \mathcal{M}_n,$$

where $\{D^v_n : v \ge 1\}$ is an i.i.d. sequence with distribution (4.47). Note that $\mathbb{E}(Y_{v,i}(a+1)) = e^{v(a+1)}$. Let $K := 10Le^{v(a+1)}$. Then standard large deviations for the geometric distribution implies that there exists a constant $C' > 0$ such that for all $n \ge 1$

$$\mathbb{P}(D_n \ge Kn^{1/(1+\alpha)}) \le \exp(-C'n^{1/(1+\alpha)}).$$

Thus by the union bound,

$$(4.49) \qquad \mathbb{P}(\mathcal{M}_n > Kn^{1/(1+\alpha)}) \le Cn \exp(-C'n^{1/(1+\alpha)}) \to 0,$$

as $n \to \infty$. Finally,

$$\limsup_{n\to\infty} \mathbb{P}(M^\star_{\mathrm{pre},A}(1) > Kn^{1/(2+\alpha)})$$

$$\le \limsup_{n\to\infty} \mathbb{P}(\mathcal{G}^c_n) + \limsup_{n\to\infty} \mathbb{P}(\mathcal{M}_n > Kn^{1/(2+\alpha)}) \le \varepsilon,$$

using (4.46), (4.48) and (4.49). This completes the proof of the lemma, and thus the analysis of the maximal degree asymptotics. □

5. Analysis of the proportion of leaves.

The aim of this section is to prove Theorem 2.3. In the next section, we will use the proportion of leaves (degree one vertices) to construct consistent estimators of the change point γ. We start in Section 5.1 by deriving strong error bounds between the expected proportion of leaves and the asserted limits in (2.3). Then in Section 5.2, we complete the proof of the functional central limit theorem. We start with some preliminary notation. For the rest of the proof, to ease notation, we will write $N_n(m) := N_n(1, m)$ for the number of leaves in $\mathcal{T}_m$ and let $\hat{N}_n(t) = N_n(nt)$. Recall the asserted limiting proportion $\{p^{(\infty)}_t : 0 \le t \le 1\}$ from (2.3). For each $n \ge 2$, define the collection of real numbers $\mathbf{w}_n = \{w_m : 2 \le m \le n - 1\}$:

$$(5.1) \qquad w_m = \begin{cases} \left(1 - \dfrac{1+\alpha}{(2+\alpha)m - 1}\right) & \text{if } 2 \le m \le n\gamma - 1, \\[3mm] \left(1 - \dfrac{1+\beta}{(2+\beta)m - 1}\right) & \text{if } n\gamma \le m \le n - 1. \end{cases}$$

5.1. *Expectation error bounds.* The following proposition is the main result of this section.

PROPOSITION 5.1. *There exists a constant $C < \infty$ independent of n such that the expectations satisfy*

$$(5.2) \qquad \sup_{n \geq 1} \sup_{0 \leq t \leq 1} \left| \mathbb{E}(\hat{N}_n(t)) - nt p_t^{(\infty)} \right| \leq C.$$

REMARK 8. Note that by Proposition 4.17, we know there exists a function $p^{(\infty)}(0, \cdot)$ such that $\hat{p}^{(n)}(0, t) \to p^{(\infty)}(0, t)$ for $0 < t \leq 1$. By the bounded convergence theorem, $\mathbb{E}(\hat{p}^{(n)}(0, t)) \to p^{(\infty)}(0, t)$. Thus the above proposition implies that $p^{(\infty)}(0, t) = p_t^{(\infty)}$. In particular, it shows that the degree distribution owing to the change point is *different* from the degree distribution without change point. This is the final nail in proving Theorem 2.1.

REMARK 9. A similar result was shown in the context of no change point in [58], Section 8.6, and [25] (not just for leaves but for all fixed $k \geq 1$). Our proof uses slightly different ideas starting from the same point as in [58]. While we do not consider higher-degree vertices, as in [58], the result above can be used as a building block to show identical error bounds for expectations of the number of higher degree vertices about limit constants.

PROOF OF PROPOSITION 5.1. To ease notation, write $\vartheta_n(m) = \mathbb{E}(N_n(m))$. The main crux of the proof is studying a recursion relation for $\vartheta_n(m + 1)$ in terms of $\vartheta_n(m)$. We will give a careful analysis of the time period before the change point and then describe how the same ideas give the result for after the change point.

For each $1 < m \leq n$, write $\mathcal{L}_{m+1}$ for the event that vertex $m + 1$ connects to a leaf vertex in $\mathcal{T}_m$. Then note that conditioning on $\mathcal{T}_m$, when $m < n\gamma$ we have

$$\mathbb{E}(N_n(m + 1) | \mathcal{T}_m) = N_n(m) + 1 - \mathbb{P}(\mathcal{L}_{m+1} | \mathcal{T}_m)$$

$$(5.3)$$

$$= N_n(m) + 1 - \frac{(1 + \alpha) N_n(m)}{(2 + \alpha)m - 1}.$$

When $m \geq n\gamma$, we have the same recursion as above but with α replaced by β. Taking full expectations and simplifying gives the following recursion:

$$(5.4) \qquad N_n(m + 1) = 1 + w_m N_n(m), \qquad \vartheta_n(m + 1) = 1 + w_m \vartheta_n(m),$$

where $\{w_m : 2 \leq m \leq n\}$ are as defined in (5.1).

Before the change point: Repeatedly using this recursion and using the boundary condition $\vartheta_n(2) = 1$ gives for $m + 1 \leq n\gamma$,

$$(5.5) \qquad \vartheta_n(m + 1) = \sum_{s=2}^{m} \prod_{k=s}^{m} \left(1 - \frac{(1 + \alpha)}{(2 + \alpha)k - 1} \right).$$

Now fix $s_0 \geq 1$ large enough such that the following three conditions hold:

(i) For all $k \geq s_0$,

$$\log k + \gamma \leq \sum_{i=1}^{k} \frac{1}{i} \leq (\log k + \gamma) + \frac{1}{k}.$$

Here, γ is the Euler–Mascheroni constant. See [8].

(ii) For all $k \geq s_0$, $1 - \frac{(1+\alpha)}{(2+\alpha)k - 1} \geq 1/2$.

(iii) We may choose a constant $C < \infty$ such that for all $k \geq 1$,

(5.6)
$$\sum_{i=k}^{\infty} \frac{1}{((2+\alpha)k - 1)^2} \leq \frac{C}{k}.$$

Further there exists a constant C' such that for all $s > s_0$, $|\exp(C/s) - 1| \leq C'/s$ and

$$\left| \left(1 - \frac{(1+\alpha)}{(2+\alpha)s - 1} \right) - e^{-\frac{(1+\alpha)}{(2+\alpha)s - 1}} \right| \leq \frac{C'}{s^2}.$$

To ease notation, for the rest of the proof let $\delta = (1+\alpha)/(2+\alpha)$. Using the elementary inequality $1 - x \leq e^{-x}$ for $x \in (0, 1)$ and the choice of s_0 above, the following inequalities with a constant $C = C(s_0, \alpha) < \infty$ are readily verified:

(A) For all $m \geq s \geq s_0$,

(5.7)
$$\left| e^{-\sum_{i=s}^{m} \frac{\delta}{i}} - \left(\frac{s}{m} \right)^{\delta} \right| \leq C \frac{s^{\delta - 1}}{m^{\delta}}.$$

(B) For all $m \geq s \geq s_0$,

(5.8)
$$\left| e^{-\sum_{i=s}^{m} \frac{(1+\alpha)}{(2+\alpha)i}} - e^{-\sum_{i=s}^{m} \frac{(1+\alpha)}{(2+\alpha)i - 1}} \right| \leq C \frac{s^{\delta - 1}}{m^{\delta}}.$$

(C) For all $m \geq s \geq s_0$,

(5.9)
$$\prod_{k=s}^{m} \left(1 - \frac{(1+\alpha)}{(2+\alpha)k - 1} \right) \leq C \left(\frac{s}{m} \right)^{\delta}.$$

Now note that by the "Lindeberg" trick, for any $s \leq m$ and two collections of nonnegative numbers $\{w_k : s \leq k \leq m\}$ and $\{z_k : s \leq k \leq m\}$ we have

(5.10)
$$\left| \prod_{k=s}^{m} w_k - \prod_{k=s}^{m} z_k \right| \leq \sum_{k=s}^{m} |w_k - z_k| \prod_{s \leq l < k} z_k \prod_{l > k} w_k.$$

Using this with $w_k = 1 - \frac{(1+\alpha)}{(2+\alpha)k - 1}$ and $z_k = e^{-\frac{(1+\alpha)}{(2+\alpha)k - 1}}$ and using (5.7), (5.8) and (5.9) gives the following lemma.

LEMMA 5.2. *Fix s_0 as above. Writing $\delta = (1+\alpha)/(2+\alpha)$ there exists a constant $C < \infty$ such that, for all $m \geq s \geq s_0$,*

$$\left| \prod_{k=s}^{m} \left(1 - \frac{(1+\alpha)}{(2+\alpha)k - 1} \right) - \left(\frac{s}{m} \right)^{\delta} \right| \leq C \frac{s^{\delta - 1}}{m^{\delta}}.$$

Now using the form of the expectation $\vartheta_n(m)$ in (5.5), the error bound in the above lemma and the integral comparison

$$\frac{1}{m^\delta} \int_{s_0}^{m-1} x^\delta \, dx \leq \sum_{s_0+1}^{m} \left(\frac{s}{m}\right)^\delta \leq \frac{1}{m^\delta} \int_{s_0+2}^{m+1} x^\delta \, dx,$$

shows that there exists a constant C such that for $m \leq n\gamma$

$$(5.11) \qquad \left| \vartheta_n(m) - \frac{m}{\delta} \right| \leq C.$$

This is the assertion for the expected number of leaves before the change point.

After the change point: We now describe the evolution of $\vartheta_n(m)$ for $n\gamma < m \leq n$. We only give the basic idea as the details are the same as before the change point. First, note that by the above analysis, there exists a constant C such that $|\vartheta_n(n\gamma) - n\gamma/\delta| \leq C$. Now the evolution of the process after γn is as in (5.3) with α replaced by β. Thus starting at $m > n\gamma$ and using the argument above we get

$$(5.12) \qquad \begin{aligned} \vartheta_n(m+1) &:= \sum_{s=n\gamma+1}^{m} \prod_{j=s}^{m} \left(1 - \frac{1+\beta}{(2+\beta)j - 1}\right) \\ &\quad + \vartheta_n(n\gamma) \prod_{j=n\gamma}^{m} \left(1 - \frac{1+\beta}{(2+\beta)j - 1}\right). \end{aligned}$$

Simplifying notation and writing $m = nt$ where $\gamma \leq t \leq 1$ and repeating the arguments above it is easy to check that there exists a constant C independent of n such that

$$(5.13) \qquad \left| \vartheta_n(nt) - nt p_t^{(\infty)} \right| \leq C,$$

where $p_t^{(\infty)}$ is as in (2.3). This completes the proof. $\quad \square$

5.2. *Proof of Theorem* 2.3. A central limit theorem for the number of leaves $N_n(n)$ [in fact all degree counts $N_n(k, n)$] at time n in the setting of no change point was established in [48]. We will extend this to a functional central limit theorem in the change point setting. First, recall the function δ_α from (2.4). Define the stochastic process

$$(5.14) \qquad M_n^*(t) = \begin{cases} t^{\delta_\alpha} \dfrac{(N_n(nt) - \vartheta_n(nt))}{\sqrt{n}} & \text{if } t \leq \gamma, \\ \gamma^{\delta_\alpha} \left(\dfrac{t}{\gamma}\right)^{\delta_\beta} \dfrac{(N_n(nt) - \vartheta_n(nt))}{\sqrt{n}} & \text{if } t \geq \gamma. \end{cases}$$

Recall the process $M(\cdot)$ in (2.8) and the relationship between M and G. Using Proposition 5.1 and the continuous mapping theorem, it is enough to show the following result.

PROPOSITION 5.3.   *We have $M_n^*(\cdot) \xrightarrow{w} M(\cdot)$ on $D[0, 1]$ as $n \to \infty$.*

PROOF.   The main idea is to study martingales associated with the $\{N_n(m) : 2 \le m \le n\}$ and then use the martingale functional central limit theorem. There are an enormous number of variants of such functional limit theorems under a multitude of conditions. We quote the specific form relevant to this setting. Recall the function $\phi(\cdot)$ and the corresponding diffusion $M(\cdot)$ defined in (2.9).

THEOREM 5.4 ([26, 29]).   *For each $n \ge 1$, let $\{M_n(m) : 1 \le m \le n\}$ be a mean zero martingale with finite second moments adapted to a filtration $\{\mathcal{F}_n(m) : 1 \le m \le n\}$. Write $\{X_n(m) : 1 \le m \le n\}$ for the associated martingale difference sequence namely $X_n(m) = M_n(m) - M_n(m-1)$ with $M_n(0) = 0$. Assume the following two hypotheses*:

(i) *For each $0 \le t \le 1$,*

$$(5.15) \qquad V_n(nt) := \sum_{m=1}^{nt} \mathbb{E}\big([X_n(m)]^2 | \mathcal{F}_n(m-1)\big) \xrightarrow{P} \phi(t) \qquad as\ n \to \infty.$$

(ii) *For each fixed $\varepsilon > 0$,*

$$(5.16) \qquad \sum_{m \le n} \mathbb{E}\big([X_n(m)]^2 \mathbb{1}\{|X_n(m)| > \varepsilon\} | \mathcal{F}_n(m-1)\big) \xrightarrow{P} 0.$$

Then defining the process $\bar{M}_n(t) := M_n(nt)$, one has $\bar{M}_n \xrightarrow{w} M$ in $D[0, 1]$.

For our example (following [48]), define the process

$$(5.17) \qquad N_n^*(m) = \frac{N_n(m) - \vartheta_n(m)}{\prod_{j=2}^{m-1} w_j}, \qquad 2 \le m \le n.$$

Here, w_j is as in (5.1). Using the recursion (5.4) results in the following lemma.

LEMMA 5.5.   *The process N_n^* is a martingale with respect to the filtration generated by $\{\mathcal{T}_m : 2 \le m \le n\}$.*

Now define the corresponding martingale differences $d_n(m) = N_n^*(m) - N_n^*(m-1)$. Define $\Delta_n(m) = \mathbb{1}\{m+1$ connects to a nonleaf vertex in $\mathcal{T}_{m-1}\}$. Then simple algebra and (5.4) implies that for $m \le n\gamma$

$$(5.18) \qquad d_n(m) = \frac{1}{\prod_{j=2}^{m-1} w_j}\left[\Delta_n(m) + N_n(m-1)\frac{(1+\alpha)N_n(m-1)}{(2+\alpha)(m-1)-1} - 1\right]$$

and

$$(5.19) \qquad \mathbb{E}(\Delta_n(m)|\mathcal{T}_{m-1}) = 1 - \frac{(1+\alpha)N_n(m-1)}{(2+\alpha)(m-1)-1}.$$

For $m \geq n\gamma$, we have identical formulae as (5.18) and (5.19) but now α is replaced by β. For the rest of the argument, we will replace the denominator for the second term $(2 + \alpha)(m - 1) - 1$ by $(2 + \alpha)(m - 1) - 1$. It is easy to check that the error is negligible and will ease presentation.

Now use Proposition 4.17 which allows us to uniformly approximate $N_n(m-1)/(m-1)$ by $p_{m/n}^{(\infty)}$. Further the asymptotics of $\prod_{j=2}^{m} w_j$ derived in the previous section implies that for $m \leq n\gamma$, $\prod_{j=2}^{m} w_j \sim m^{-\delta_\alpha}$ while for $m > n\gamma$, $\prod_{j=2}^{m} w_j \sim (n\gamma)^{-\delta_\alpha}(m/n\gamma)^{-\delta_\beta}$ where $\delta_\alpha, \delta_\beta$ as defined in (2.4). Taking conditional expectations in (5.18), using (5.19) and using the above approximations results in

$$
\mathbb{E}\big([d_n(m)]^2 | \mathcal{T}_{m-1}\big)
$$

$$
(5.20) \qquad \sim \begin{cases} m^{2\delta_\alpha}\big[\delta_\alpha p_{m/n}^{(\infty)}\big(1 - \delta_\alpha p_{m/n}^{(\infty)}\big)\big] & \text{if } m \leq n\gamma, \\[2mm] (n\gamma)^{2\delta_\alpha}\left(\dfrac{m}{n}\right)^{2\delta_\beta}\big[\delta_\beta p_{m/n}^{(\infty)}\big(1 - \delta_\beta p_{m/n}^{(\infty)}\big)\big] & \text{if } m \geq n\gamma. \end{cases}
$$

Now consider the martingale

$$
(5.21) \qquad M_n(m) := \frac{1}{n^{\delta_\alpha + 1/2}} \frac{N_n(m) - \vartheta_n(m)}{\prod_{j=2}^{m-1} w_j}, \qquad 2 \leq m \leq n.
$$

We will apply Theorem 5.4 to this martingale. Let $\{X_n(m) : 2 \leq m \leq n\}$ denote the corresponding martingale differences. First, fix $t \leq \gamma$ and recall the definition of the cumulative conditional variance $V_n(nt)$ until time t in (5.15). Using the first expression in (5.20), we get

$$
V_n(nt) \sim \frac{1}{n^{2\delta_\alpha + 1}} \sum_{j=1}^{nt} j^{2\delta_\alpha}\big[\delta_\alpha p_{m/n}^{(\infty)}\big(1 - \delta_\alpha p_{m/n}^{(\infty)}\big)\big]
$$

$$
\rightarrow \int_0^t s^{2\delta_\alpha}\big[\delta_\alpha p_s^{(\infty)}\big(1 - \delta_\alpha p_s^{(\infty)}\big)\big]\,ds = \phi(t),
$$

as $n \to \infty$. Thus (5.15) is satisfied for $t \leq \gamma$. A similar calculation now incorporating the second expression in (5.20) implies that (5.15) is satisfied for all $t \in [0, 1]$ with ϕ as in (2.9). Now let us check the second condition namely (5.16). Note that for $m \leq n\gamma$, $X_n(m) \geq \varepsilon$ implies that $3m^{\delta_\alpha} \geq \varepsilon n^{\delta_\alpha + 1/2}$. For large n, this is impossible for all $m \leq n\gamma$. A similar calculation for $m > n\gamma$ completes the proof of (5.16). Using Theorem 5.4, we get that $M_n(n\cdot) \xrightarrow{\text{w}} M(\cdot)$ in $D[0, 1]$. Using the asymptotics for $\prod_{j=2}^{m} w_j$ derived in Section 5.1 in (5.21) now completes the proof of Proposition 5.3, and thus Theorem 2.3. $\square$

6. Consistency of the estimator. The aim of this section is to prove Theorem 2.4. Fix a truncation level $\varepsilon > 0$ from zero as in the theorem. Recall the time-averaged proportion of leaves before and after each time t, namely (2.15)

and (2.14). Also recall the expression for the limiting proportion of leaves from (2.3). For any fixed interval $[s, t] \subseteq [0, 1]$, define $H[s, t]$ by

$$(6.1) \qquad H[s, t] := \frac{1}{t - s} \int_s^t p_u^{(\infty)} \, du.$$

The interpretation is as follows: the above gives the expected proportion of leaves in the large network limit if one were to sample a time point $U \in [s, t]$ uniformly at random. Now define the two functions ${}_t h^{(\infty)}$ and $h_t^{(\infty)}$ via:

(a) *Case 1*: For $\varepsilon \le t \le \gamma$,

$$_t h^{(\infty)} := p_\gamma^{(\infty)}, \qquad h_t^{(\infty)} := \frac{\gamma - t}{1 - t} p_\gamma^{(\infty)} + \frac{1 - \gamma}{1 - t} H[\gamma, 1].$$

(b) *Case 2*: For $t > \gamma$,

$$_t h := \frac{\gamma - \varepsilon}{t - \varepsilon} p_\gamma + \frac{t - \gamma}{t - \varepsilon} H([\gamma, t]), \qquad h_t := H([t, 1]).$$

In similar vein to (2.16), define the function

$$(6.2) \qquad D(t) := (1 - t) |_t h^{(\infty)} - h_t^{(\infty)}|, \qquad t \in [\varepsilon, 1].$$

Routine algebra shows that

$$(6.3) \qquad D(t) := \begin{cases} (1 - \gamma) |p_\gamma^{(\infty)} - H[\gamma, 1]| & \text{for } \varepsilon \le t \le \gamma, \\ (1 - \varepsilon) |H[\varepsilon, t] - H[\varepsilon, 1]| & \text{for } t > \gamma. \end{cases}$$

Using the form of the limit proportion $p_t^{(\infty)}$ from (2.3) the following result is easy to check.

LEMMA 6.1. *Fix $\varepsilon < \gamma$ and assume $\alpha \ne \beta$. Then $D(\cdot)$ is a continuous function on $[\varepsilon, 1]$ such that $D(\cdot)$ is constant on the interval $[\varepsilon, \gamma]$ and then is strictly monotonically decreasing on the interval $[\gamma, 1]$ with $D(t) \to 0$ as $t \to 1$. Further the function has a strictly negative right derivative at γ, namely*

$$(6.4) \qquad \partial_+ D(\gamma) := \lim_{t \downarrow \gamma} \frac{D(t) - D(\gamma)}{t - \gamma} < 0.$$

Now Theorem 2.3 immediately results in the following result.

LEMMA 6.2. *Fix $\varepsilon > 0$. Then*

$$\sup_{t \in [\varepsilon, 1]} |D_n(t) - D(t)| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Combining Lemmas 6.1 and 6.2 completes the proof. $\quad\square$

REFERENCES

[1] AKOGLU, L., MCGLOHON, M. and FALOUTSOS, C. (2012). Oddball: Spotting anomalies in weighted graphs. In *Advances in Knowledge Discovery and Data Mining* 410–421. Springer, Berlin.

[2] ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** 47–97. MR1895096

[3] ATHREYA, K. B. and KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Stat.* **39** 1801–1817. MR0232455

[4] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

[5] BASSEVILLE, M. and NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes*: *Theory and Application*. Prentice Hall, Englewood Cliffs, NJ. MR1210954

[6] BHAMIDI, S., EVANS, S. N. and SEN, A. (2012). Spectra of large random trees. *J. Theoret. Probab.* **25** 613–654. MR2956206

[7] BHAMIDI, S., STEELE, J. M. and ZAMAN, T. (2015). Twitter event networks and the superstar model. *Ann. Appl. Probab.* **25** 2462–2502. MR3375881

[8] BOAS, R. P. JR. (1977). Partial sums of infinite series, and how they grow. *Amer. Math. Monthly* **84** 237–258. MR0440240

[9] BOCCALETTI, S., BIANCONI, G., CRIADO, R., DEL GENIO, C. I., GÓMEZ-GARDEÑES, J., ROMANCE, M., SENDIÑA-NADAL, I., WANG, Z. and ZANIN, M. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* **544** 1–122. MR3270140

[10] BOLLOBÁS, B. (1985). *Random Graphs*. Academic Press, London. MR0809996

[11] BOLLOBÁS, B. and RIORDAN, O. M. (2003). Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks* 1–34. Wiley-VCH, Weinheim. MR2016117

[12] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. and TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. MR1824277

[13] BRODSKY, B. E. and DARKHOVSKY, B. S. (1993). *Nonparametric Methods in Change-Point Problems. Mathematics and Its Applications* **243**. Kluwer Academic, Dordrecht. MR1228205

[14] BUBECK, S., DEVROYE, L. and LUGOSI, G. (2017). Finding Adam in random growing trees. *Random Structures Algorithms* **50** 158–172. MR3607120

[15] BUBECK, S., MOSSEL, E. and RÁCZ, M. Z. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Trans. Netw. Sci. Eng.* **2** 30–39. MR3361606

[16] CARLSTEIN, E. (1988). Nonparametric change-point estimation. *Ann. Statist.* **16** 188–197. MR0924865

[17] CARLSTEIN, E., MÜLLER, H.-G. and SIEGMUND, D., eds. (1994). *Change-Point Problems* (*South Hadley, MA*, 1992). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23**. IMS, Hayward, CA. MR1477909

[18] CHANDOLA, V., BANERJEE, A. and KUMAR, V. (2009). Anomaly detection: A survey. *ACM Comput. Surv.* **41** Article No. 15.

[19] CHUNG, F. and LU, L. (2006). *Complex Graphs and Networks. CBMS Regional Conference Series in Mathematics* **107**. Amer. Math. Soc., Providence, RI. MR2248695

[20] COOPER, C. and FRIEZE, A. (2003). A general model of web graphs. *Random Structures Algorithms* **22** 311–335. MR1966545

[21] CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester. MR2743035

[22] CURIEN, N., DUQUESNE, T., KORTCHEMSKI, I. and MANOLESCU, I. (2015). Scaling limits and influence of the seed graph in preferential attachment trees. *J. Éc. Polytech. Math.* **2** 1–34. MR3326003

[23] DOROGOVTSEV, S. N. and MENDES, J. F. F. (2003). *Evolution of Networks*. Oxford Univ. Press, Oxford. MR1993912

[24] DUAN, D., LI, Y., JIN, Y. and LU, Z. (2009). Community mining on dynamic weighted directed graphs. In *Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledge Management* 11–18.

[25] DURRETT, R. (2007). *Random Graph Dynamics. Cambridge Series in Statistical and Probabilistic Mathematics* **20**. Cambridge Univ. Press, Cambridge. MR2271734

[26] DURRETT, R. and RESNICK, S. I. (1978). Functional limit theorems for dependent variables. *Ann. Probab.* **6** 829–846. MR0503954

[27] EAGLE, N. and PENTLAND, A. (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing* **10** 255–268.

[28] EBERLE, W. and HOLDER, L. (2007). Discovering structural anomalies in graph-based data. In *Seventh IEEE International Conference on Data Mining Workshops* (*ICDMW* 2007) 393–398.

[29] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York. MR0838085

[30] HEARD, N. A., WESTON, D. J., PLATANIOTI, K. and HAND, D. J. (2010). Bayesian anomaly detection methods for social networks. *Ann. Appl. Stat.* **4** 645–662. MR2758643

[31] HOLME, P. and SARAMÄKI, J. (2012). Temporal networks. *Phys. Rep.* **519** 97–125.

[32] HUANG, Z. and ZENG, D. D. (2006). A link prediction approach to anomalous email detection. In 2006 *IEEE International Conference on Systems*, *Man and Cybernetics* 1131–1136.

[33] JAGERS, P. (1975). *Branching Processes with Biological Applications*. Wiley-Interscience, London. MR0488341

[34] JAGERS, P. and NERMAN, O. (1984). The growth and composition of branching populations. *Adv. in Appl. Probab.* **16** 221–259. MR0742953

[35] JAGERS, P. and NERMAN, O. (1984). Limit theorems for sums determined by branching and other exponentially growing processes. *Stochastic Process. Appl.* **17** 47–71. MR0738768

[36] LIPTSER, R. and SHIRYAYEV, A. N. (1989). *Theory of Martingales. Mathematics and Its Applications* (*Soviet Series*) **49**. Kluwer Academic, Dordrecht. MR1022664

[37] MA, W.-Y. and MANJUNATH, B. S. (1997). Edge flow: A framework of boundary detection and image segmentation. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 744–749.

[38] MARANGONI-SIMONSEN, D. and XIE, Y. (2015). Sequential changepoint approach for online community detection. *IEEE Signal Process. Lett.* **22** 1035–1039.

[39] MCCULLOH, I. and CARLEY, K. M. (2011). Detecting change in longitudinal social networks. DTIC Document.

[40] MORENO, S. and NEVILLE, J. (2013). Network hypothesis testing using mixed Kronecker product graph models. In 2013 *IEEE* 13*th International Conference on Data Mining* 1163–1168.

[41] MÓRI, T. F. (2007). Degree distribution nearby the origin of a preferential attachment graph. *Electron. Commun. Probab.* **12** 276–282. MR2342706

[42] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256. MR2010377

[43] NEWMAN, M. E. J. (2010). *Networks*: *An Introduction*. Oxford Univ. Press, Oxford. MR2676073

[44] NOBLE, C. C. and COOK, D. J. (2003). Graph-based anomaly detection. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 631–636.

[45] NORRIS, J. R. (1998). *Markov Chains*. *Cambridge Series in Statistical and Probabilistic Mathematics* **2**. Cambridge Univ. Press, Cambridge. Reprint of 1997 original. MR1600720

[46] PEEL, L. and CLAUSET, A. (2014). Detecting change points in the large-scale structure of evolving networks. Available at arXiv:1403.0989.

[47] PRIEBE, C. E., CONROY, J. M., MARCHETTE, D. J. and PARK, Y. (2005). Scan statistics on enron graphs. *Comput. Math. Organ. Theory* **11** 229–247.

[48] RESNICK, S. I. and SAMORODNITSKY, G. (2016). Asymptotic normality of degree counts in a preferential attachment model. *Adv. in Appl. Probab.* **48** 283–299. MR3539311

[49] RUDAS, A., TÓTH, B. and VALKÓ, B. (2007). Random trees and general branching processes. *Random Structures Algorithms* **31** 186–202. MR2343718

[50] SHARPNACK, J., RINALDO, A. and SINGH, A. (2012). Changepoint detection over graphs with the spectral scan statistic. Available at arXiv:1206.0773.

[51] SHIRYAEV, A. N. (2008). *Optimal Stopping Rules*. *Stochastic Modelling and Applied Probability* **8**. Springer, Berlin. MR2374974

[52] SIEGMUND, D. (1985). *Sequential Analysis*: *Tests and Confidence Intervals*. Springer, New York. MR0799155

[53] SIMON, H. A. (1955). On a class of skew distribution functions. *Biometrika* **42** 425–440. MR0073085

[54] ŠIRJAEV, A. N. (1963). Optimal methods in quickest detection problems. *Teor. Verojatnost. i Primenen.* **8** 26–51. MR0155708

[55] SUN, J., FALOUTSOS, C., PAPADIMITRIOU, S. and YU, P. S. (2007). Graphscope: Parameter-free mining of large time-evolving graphs. In *Proceedings of the* 13*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 687–696.

[56] SZYMAŃSKI, J. (1987). On a nonuniform random recursive tree. In *Random Graphs '85* (*Poznań*, 1985). *North-Holland Math. Stud.* **144** 297–306. North-Holland, Amsterdam. MR0930497

[57] TSYBAKOV, A. B. (1994). Multidimensional change-point problems and boundary estimation. In *Change-Point Problems* (*South Hadley*, *MA*, 1992). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23** 317–329. IMS, Hayward, CA. MR1477933

[58] VAN DER HOFSTAD, R. (2016). *Random Graphs and Complex Networks*, *Vol.* 1. Cambridge Univ. Press, Cambridge. Available at http://www.win.tue.nl/rhofstad/NotesRGCN.pdf. MR3617364

[59] YUDOVINA, E., BANERJEE, M. and MICHAILIDIS, G. (2015). Changepoint inference for Erdős–Rényi random graphs. In *Stochastic Models*, *Statistics and Their Applications*. *Springer Proc. Math. Stat.* **122** 197–205. Springer, Cham. MR3336439

[60] YULE, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos. Trans. R. Soc. Lond. Ser. B* **213** 21–87.

[61] ZHANG, W., PAN, G., WU, Z. and LI, S. (2013). Online community detection for large complex networks. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* 1903–1909.

S. BHAMIDI
J. JIN
A. NOBEL
DEPARTMENT OF STATISTICS AND
    OPERATIONS RESEARCH
UNIVERSITY OF NORTH CAROLINA
304 HANES HALL
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: bhamidi@email.unc.edu
            jimmyjin@live.unc.edu
            nobel@email.unc.edu