FLUCTUATION BOUNDS FOR CONTINUOUS TIME BRANCHING PROCESSES AND NONPARAMETRIC CHANGE POINT DETECTION IN GROWING NETWORKS

SAYAN BANERJEE, SHANKAR BHAMIDI, AND IAIN CARMICHAEL

ABSTRACT. Motivated by applications, both for modeling real world systems as well as in the study of probabilistic systems such as recursive trees, the last few years have seen an explosion in models for dynamically evolving networks. The aim of this paper is two fold: (a) develop mathematical techniques based on continuous time branching processes (CTBP) to derive quantitative error bounds for functionals of a major class of these models about their large network limits; (b) develop general theory to understand the role of abrupt changes in the evolution dynamics of these models using which one can develop non-parametric change point detection estimators. In the context of the second aim, for fixed final network size \( n \) and a change point \( \tau(n) < n \), we consider models of growing networks which evolve via new vertices attaching to the pre-existing network according to one attachment function \( f \) till the system grows to size \( \tau(n) \) when new vertices switch their behavior to a different function \( g \) till the system reaches size \( n \). With general non-explosivity assumptions on the attachment functions \( f, g \), we consider both the standard model where \( \tau(n) = \Theta(n) \) as well as the quick big bang model when \( \tau(n) = n^{\gamma} \) for some \( 0 < \gamma < 1 \). Proofs rely on a careful analysis of an associated inhomogeneous continuous time branching process. Techniques developed in the paper are robust enough to understand the behavior of these models for any sequence of change points \( \tau(n) \to \infty \). This paper derives rates of convergence for functionals such as the degree distribution; the same proof techniques should enable one to analyze more complicated functionals such as the associated fringe distributions.

1. INTRODUCTION

1.1. Motivation. Driven by the explosion in the amount of data on various real world networks, the last few years have seen the emergence of many new mathematical network models. Motivations behind these models are diverse including (a) extracting unexpected patterns as densely connected regions in the network (e.g. community detection); (b) understand properties of dynamics on these real world systems such as the spread of epidemics, the efficacy of random walk search algorithms etc; (c) most relevant for this study, understanding mechanistic reasons for the emergence of empirically observed properties of these systems such as heavy tailed degree distribution or the small world property. We refer the interested reader to [1, 14, 24, 35, 36, 44] and the references therein for a starting point to the vast literature on network models. A small but increasingly important niche is the setting of dynamic network models, networks that evolve over time. In the context of probabilistic combinatorics, in particular in the study of growing random trees, these models have been studied for decades in the vast field of recursive trees, see [10, 23, 25, 32] and the references therein. To fix ideas, consider one of the standard examples: start with a base graph \( G_0 \) (e.g. two vertices connected by an edge) and an attachment function \( f : \mathbb{Z}_+ \to [0, \infty) \) where \( \mathbb{Z}_+ := \{0, 1, 2, \ldots\} \). For each fixed time \( n \geq 1 \), having constructed the network \( G_{n-1} \) at time \( n-1 \), the network transitions to \( G_n \) as follows: a new vertex enters the system and attaches to a pre-existing vertex \( v \in G_{n-1} \) with probability proportional to \( f(\text{deg}(v)) \) where \( \text{deg}(v) \) is the current degree of this vertex. The case of \( f(\cdot) \equiv 1 \) corresponds to the famous class of random recursive trees [41]. The specific case of \( f(k) = k \) for \( k \geq 0 \) was considered in [8] where they showed, via non-rigorous arguments,
that the resulting graph has a heavy tailed degree distribution with exponent 3 in the large \( n \to \infty \) limit; this was rigorously proved in [15].

1.2. **Informal description of our aims and results.** This paper has the following two major aims:

(a) In the context of models described above, asymptotics in the large network limit for a host of random tree models as well as corresponding functionals have been derived ranging from the degree distribution to the so-called fringe distribution [2][11][27] of random trees. One of the major drivers of research has been proving convergence of the empirical distribution of these functionals to limiting (model dependent) constants. Establishing (even suboptimal) rates of convergence for these models has been non-trivial other than for models related to urn models e.g. see the seminal work of Janson [31]. The aim of this paper is to develop robust methodology for proving such error bounds for general models. Our results will not be optimal owing to the generality of the model considered in the paper; however using the techniques in this paper coupled with higher moment assumptions can easily lead to more refined results for specific models. To keep the paper to manageable length, we focus on the degree distribution but see Section 4 for our work in progress of using the methodology in this paper for more general functionals.

(b) Consider general models of network evolution as described in the above paragraph but wherein, beyond some point, new individuals entering the system change their evolution behavior. This is reflected via a change in the the attachment function \( f \) to a different attachment function \( g \).

(i) We first aim to understand the effect of change points on structural properties of the network model and the interplay between the time scale of the change point and the nature of the attachment functions before and after the change point. Analogous to classical change point detection, we start by considering models which evolve for \( n \) steps with a change point at time \( \gamma n \) for \( 0 < \gamma < 1 \); we call this the **standard model**. Counter-intuitively, we find that irrespective of the value of \( \gamma \), structural properties of the network such as the tail of the degree distribution are determined by model parameters before the change point; motivated by this we consider other time scales of the change point (which we call the **quick big bang model**) to see the effect of the long range dependence phenomenon in the evolution of the process.

(ii) We then develop nonparametric change point detection techniques for the standard model when one has no knowledge of the attachment functions, pre or post change point.

1.3. **Model definition.** Fix \( J \geq 0 \). For each \( 0 \leq j \leq J \), fix functions \( f_j : \mathbb{Z}_+ \to \mathbb{R}_+ \), which we will refer to as attachment functions. Let us start by describing the model when \( J = 0 \), and we have one attachment function \( f_0 \). This setting will be referred to as **nonuniform random recursive trees [42]** or attachment model. We will grow a sequence of random trees \( \{ T_j : 2 \leq j \leq n \} \) as follows:

(i) For \( n = 2 \), \( T_2 \) consists of two vertices attached by a single edge. Label these using \( \{ 1, 2 \} \) and call the vertex \( v = 1 \) as the “root” of the tree. We will think of the tree as directed with edges being pointed away from the root (from parent to child).

(ii) Fix \( n > 2 \). Let the vertices in \( T_{n-1} \) be labeled by \( \{ n - 1 \} \). For each vertex \( v \in T_{n-1} \) let out-deg(\( v \)) denote the out-degree of \( v \). A new vertex labelled by \( n \) enters the system. Conditional on \( T_{n-1} \), this new vertex attaches to a currently existing vertex \( v \in \{ n - 1 \} \) with probability proportional to \( f_0(\text{out-deg}(v)) \). Call the vertex that \( n \) attaches to, the “parent” of \( n \) and direct the edge from this parent to \( n \) resulting in the tree \( T_n \).

**Model with change point:** Next we define the model with \( J \geq 1 \) distinct change points. Fix attachment functions \( f_0 \neq f_1 \neq f_2 \cdots \neq f_J \). For \( n \geq J + 2 \) fix \( J \) distinct times \( 2 \leq \tau_1 < \tau_2 < \cdots < \tau_J < n \in [n] \). Let \( f = (f_j : 0 \leq j \leq J) \) and \( \tau = (\tau_j : 1 \leq j \leq J) \) and write \( \theta = (f, \tau) \) for the driving parameters of the process. For notational convenience, let \( \tau_0 = 2 \) and \( \tau_{J+1} = n \). Consider a sequence of random trees \( \{ T_j^\theta : 2 \leq i \leq n \} \) constructed as follows. For \( 2 \leq i \leq \tau_1 \), the process evolves as in the non-change point model using the attachment function \( f_0 \). We will call this the **initializer** function. Then or each change point index \( 1 \leq j \leq J \) and time \( i \in (\tau_j, \tau_{j+1}] \) the process evolves according to the function \( f_j \) i.e. each new vertex entering
2.3. Branaching processes. Fix an attachment function $f$ as above. We can construct a point process $\xi_f$ on $\mathbb{R}_+$ as follows: Let $\{E_i : i \geq 0\}$ be a sequence of independent exponential random variables with $E_i \sim \exp(f(i))$. Now define $L_i := \sum_{j=0}^{i-1} E_i$ for $i \geq 1$. The point process $\xi_f$ is defined via,

$$\xi_f := (L_1, L_2, \ldots).$$

(2.4)

Abusing notation, we write for $t \geq 0$,

$$\xi_f[0, t] := \# \{i : L_i \leq t\}, \quad \mu_f[0, t] := \mathbb{E}(\xi_f[0, t]).$$

(2.5)

This object is often referred to as the Malthusian rate of growth parameter.
Here we view $\mu_f$ as a measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. We will need a variant of the above objects: for fixed $k \geq 0$, let $\xi_f^{(k)}$ denote the point process where the first inter-arrival time is $E_k$ namely define the sequence,

$$L_i^{(k)} = E_k + E_{k+1} + \cdots + E_{k+i-1}, \quad i \geq 1.$$ 

Then define,

$$\xi_f^{(k)} := (L_1^{(k)}, L_2^{(k)}, \ldots), \quad \mu_f^{(k)} [0, t] := \mathbb{E}(\xi_f^{(k)} [0, t]).$$ 

(2.6)

As above, $\xi_f^{(k)} [0, t] := \# \{ i : L_i^{(k)} \leq t \}$. We abbreviate $\xi_f [0, t]$ as $\xi_f (t)$ and similarly $\mu_f (t), \xi_f^{(k)} (t), \mu_f^{(k)} (t)$.

**Definition 2.2** (Continuous time Branching process (CTBP)). Fix attachment function $f$ satisfying Assumption 2.1(ii). A continuous time branching process driven by $f$, written as $\text{BP}_f (t) : t \geq 0$, is defined to be a branching process started with one individual at time $t = 0$ and such that every individual born into the system has an offspring distribution that is an independent copy of the point process $\xi_f$ defined in (2.4).

We refer the interested reader to [4, 28] for general theory regarding continuous time branching processes. We will also use $\text{BP}_f (t)$ to denote the collection of all individuals at time $t \geq 0$. For $x \in \text{BP}_f (t) : t \geq 0$, denote by $\sigma_x$ the birth time of $x$. Let $Z_f (t)$ denote the size (number of individuals born) by time $t$. Note in our construction, by our assumption on the attachment function, individuals continue to reproduce forever. Write $m_f (\cdot)$ for the corresponding expectation i.e.,

$$m_f (t) := \mathbb{E}(Z_f (t)), \quad t \geq 0,$$ 

(2.7)

Under Assumption 2.1(iii), it can be shown [28, Chapter 3] that for all $t > 0$, $m_f (t) < \infty$, is strictly increasing with $m_f (t) \uparrow \infty$ as $t \uparrow \infty$. In the sequel, to simplify notation we will suppress dependence on $f$ on the various objects defined above and write $\mathbb{P}(\cdot), \mathbb{E}(\cdot)$ etc. The connection between CTBP and the discrete random tree models in the previous section is given by the following result which is easy to check using properties of exponential distribution (and is the starting point of the Athreya-Karlin embedding [3]).

**Lemma 2.3.** Fix attachment function $f$ consider the sequence of random trees $\{ T_m : 2 \leq m \leq n \}$ constructed using attachment function $f$. Consider the continuous time construction in Definition 2.2 and define for $m \geq 1$ the stopping times $\tau_m := \inf \{ t \geq 0 : | \text{BP}_f (t) | = m \}$. Then viewed as a sequence of growing random labelled rooted trees we have, $| \text{BP}_n (\tau_m) : 2 \leq m \leq n | \overset{d}{=} | T_m : 2 \leq m \leq n |$.

### 3. MAIN RESULTS

#### 3.1. Convergence rates for model without change point.

Consider a continuous time branching process with attachment function $f$ and Malthusian rate $\lambda^*$. For each $k \geq 0, t \geq 0$, denote by $D(k, t)$ the number of vertices in $\text{BP}_f (t)$ of degree $k$ and abbreviate $Z_f (t)$ to $Z (t)$. Let $\lambda^* = \lambda^* (f)$ be as in (2.3). Define the probability mass function $p(f) := \{ p_k : k \geq 0 \}$ via,

$$p_k = p_k (f) := \frac{\lambda^*}{\lambda^* + f(i)} \prod_{j=0}^{k-1} \frac{f(i)}{\lambda^* + f(i)}, \quad k \geq 0.$$ 

(3.1)

Here for $k = 0$, the $\prod_{j=0}^{k-1}$ is by convention taken to be 1. Verification that the above is an honest probability mass function can be found in [40, Theorem 2]. Following the seminal work of [28, 29, 34, 40], it follows that for each $k \geq 0$ that

$$\frac{D(k, t)}{Z(t)} \xrightarrow{p} p_k \quad \text{as} \quad t \to \infty.$$ 

However, to obtain consistent change point estimators, we need to strengthen the above convergence to a sup-norm convergence on a time interval whose size goes to infinity with growing $t$ and also, a quantitative rate for this convergence. Such results have been obtained for very specific attachment functions via functional central limit theorems (e.g. see [31] for models whose degree evolution can be reduced to the evolution of urn processes satisfying regularity conditions and [39] for the linear preferential attachment model), but do not extend to the general setting. We make the following assumptions throughout this section.
Assumption 3.1. There exists $C^* \geq 0$ such that $\lim_{k \to \infty} f(k)/k = C^*$.

Assumption 3.2. $\text{Var} \left( \int_0^\infty e^{-\lambda t} \xi_f(dt) \right) < \infty$.

Remark 1. Assumption 3.2 is implied by $\sum_{k=0}^\infty k^2 p_k(f) < \infty$ since

$$
\mathbb{E} \left( \int_0^\infty e^{-\lambda t} \xi_f(dt) \right)^2 = \mathbb{E} \left( \int_0^\infty \lambda^* e^{-\lambda t} \xi_f(t) dt \right)^2 \leq \mathbb{E} \left( \int_0^\infty \lambda^* e^{-\lambda t} \xi_f^2(t) dt \right) = \sum_{k=1}^\infty k^2 p_k(f_0) < \infty.
$$

Fix a sequence of growing trees $\{T_m : m \geq 2\}$ and recall that for any $N \geq 2$ and $k \geq 0$, $D_N(k)$ denotes the number of vertices with out-degree $k$. The main theorem of this section is

Theorem 3.3. Consider a continuous time branching process with attachment function $f$ that satisfies Assumptions 2.1, 3.1 and 3.2. Let $(p_1, p_2, \ldots)$ denote the limiting degree distribution. There exist $\omega^*, \epsilon^{**} \in (0, 1)$, such that for any $\epsilon \leq \epsilon^{**}$,

$$
n^{\omega^*} \sum_{k=0}^\infty 2^{-k} \left( \sup_{t \in [0, 2 \epsilon \log n]} \left| \frac{D(k, \frac{1-\epsilon}{\lambda} \log n + t)}{Z(\frac{1-\epsilon}{\lambda} \log n + t)} - p_k \right| \right) \to 0.
$$

Thus for a sequence of nonuniform recursive trees $\{T_m : m \geq 2\}$ grown using attachment function $f$,

$$
n^{\omega^*} \sum_{k=0}^\infty 2^{-k} \left( \sup_{n^{1-\epsilon} \leq N \leq n^{1+\epsilon}} \left| \frac{D_N(k)}{N} - p_k \right| \right) \to 0.
$$

Remark 2. In the notation of Jagers and Neuman [30][34], the result above is stated for the “characteristic” corresponding to degree (see the discussion below). We believe our proof techniques are robust enough to generalize to more complex functionals such as the fringe distribution [2][27]. We will pursue this in a separate paper. However below we describe one of the key estimates derived in this paper of more general relevance.

Remark 3. For special cases such as the uniform or linear preferential attachment, stronger results are obtainable via Janson’s “superball” argument [31] as well as application of the Azuma-Hoeffding inequality [15][44]. However these do not appear to work for the general model considered in this paper.

Recall from [34] that a characteristic $\phi$ is a non-negative random process $\{\phi(t) : t \in \mathbb{R}\}$, assigning some kind of score to the typical individual at age $t$. We assume $\phi(t) = 0$ for every $t < 0$. For this article, we will be interested in the following class of characteristics:

$$
\mathcal{C} := \{ \phi \text{ with càdlàg paths : } \exists b_\phi > 0 \text{ such that } \phi(t) \leq b_\phi (\xi_f(t) + 1) \text{ for all } t \geq 0 \}. \tag{3.2}
$$

For any characteristic $\phi$, define $Z^\phi_f(t) := \sum_{x \in \text{BP}_f(t)} \phi(t - \sigma_x)$. This can be thought of as the sum of $\phi$-scores of all individuals in BP$_f(t)$. Write $m^\phi_f(t) = \mathbb{E}(Z^\phi_f(t))$ and $M^\phi_f(t) = \mathbb{E}(e^{-\lambda t} Z^\phi_f(t))$. For fixed $k \geq 0$ and for the specific characteristic $\phi(t) = 1(\xi(t) = k)$, write $m^k_f(\cdot) := m^\phi_f(\cdot)$.

It is easy to check that for a general (integrable) characteristic $\phi$, $M^\phi_f(t)$ satisfies the renewal equation

$$
M^\phi_f(t) = e^{-\lambda t} \mathbb{E}(\phi(t)) + \int_0^t M^\phi_f(t-s) e^{-\lambda s} \mu_f(ds). \tag{3.3}
$$

Write $M^\phi_f(\infty) = \lim_{t \to \infty} M^\phi_f(t)$ when the limit exists. Following [34], we write $x = (x', i)$ to denote that $x$ is the $i$-th child of $x'$ and define for any $t \geq 0$,

$$
\mathcal{I}(t) = \{ x = (x', i) : \sigma_x \leq t \text{ and } t < \sigma_x < \infty \}.
$$

Write $W_t = \sum_{x \in \mathcal{I}(t)} e^{-\lambda^* \sigma_x}$ By Corollary 2.5 of [34], $W_t$ converges almost surely to a finite random variable $W_\infty$ as $t \to \infty$. By Theorem 3.1 of [34], $e^{-\lambda^* t} Z^\phi_f(t) \overset{P}{\to} W_\infty M^\phi_f(\infty)$ for any $\phi \in \mathcal{C}$. An important technical contribution of this paper is the following result.
Theorem 3.4. Consider a continuous time branching process with attachment function $f$ that satisfies Assumptions 2.1 and 3.1. There exist positive constants $C_1, C_2$ such that for any $b_\phi > 1$ and any characteristic $\phi \in \Phi$ satisfying $|\phi(t)| \leq b_\phi(\xi_f(t) + 1)$ for all $t \geq 0$,

$$E|e^{-\lambda^* t} Z^\Phi_f(t) - W_{\infty} M^\phi_f(\infty)| \leq C_1 b_\phi e^{-C_2 t}.$$

Remark 4. The constants $\phi^*$ in Theorem 3.3 and $C_1, C_2$ in Theorem 3.4 are explicitly computable from our proof techniques. However, they depend on the Malthusian rate and thus we have not tried to derive an explicit form of these objects.

3.2. Change point detection: Sup-norm convergence of degree distribution for the standard model.

Fix $J \geq 1$. We start by studying the model under the following assumption which we refer to as the "standard" model owing to the analogous assumptions for change point methodology in time series:

Assumption 3.5. Fix $J \geq 1$ and assume there exist $0 < \gamma_1 < \cdots < \gamma_J < 1$ such that for all $1 \leq j \leq J$, the $j^{th}$ change point is $\tau_j = [n\gamma_j]$.

To simplify notation we will drop $[\cdot]$. Recall the sequence of random trees $\{\mathcal{T}_m^n : 2 \leq m \leq n\}$. For any $0 < t \leq 1$ and $k \geq 0$, write $D_n(k, nt)$ for the number of vertices with out-degree $k$. We will sometimes abuse notation and write $D_n(k, \mathcal{T}_m^n) := D_n(k, nt)$ to explicitly specify the dependence of this object on the underlying tree. In this section we mainly consider the case where there is exactly one change point at time $n\gamma_1$ for fixed $0 < \gamma_1 < 1$. In Section 3.3 we describe the general result for multiple change points. The notation is cumbersome so this general case can be skipped over on an initial reading. We also give the proof for the single change point case; the general case follows via straightforward extensions. Fix initializer attachment function $f_0$ and let $\lambda_0^* = \lambda^*(f_0)$ be as in (2.3). Define the probability mass function $\{p_k^n : k \geq 0\}$ via (3.1) with $(\lambda_0^*, f_0)$ in place of $(\lambda^*, f)$. As before let the attachment function after change point be $f_1$. Recall from (2.6), for fixed $k \geq 0$, the function $\mu_{f_1}^k(0, \cdot)$ and the function $m_{f_1}(\cdot)$ from (2.7). Also recall that, for fixed $k \geq 0$,

$$m_{f_1}^k(t) = E\left(\sum_{x \in \mathcal{B} f_1(t)} \mathbb{1}\{\xi_{f_1}(t - \sigma_x) = k\}\right).$$

It can be checked (using the continuity estimates obtained in Lemmas 6.2 and 6.9 that for any $k \geq 0$, $t \geq 0$, $m_{f_1}^k(t) = f^*_0(\mathbb{P}(\xi_{f_1}(u) = k) m_{f_1}(t - du)$. For $\ell, k \geq 0$, define

$$\lambda_{\ell}(t) = 1 + \int_0^t m_{f_1}(t - s) \mu_{f_1}^* (d s), \quad \lambda_{\ell}^*(t) = \mathbb{P}\left(\xi_{f_1}(t) = k - \ell\right) + \int_0^t m_{f_1}^k(t - s) \mu_{f_1}^*(d s).$$

(3.4)

Let $\mathcal{P}$ denote the collection of all probability measures on $\mathbb{N} \cup \{0\}$. For each $a > 0$, consider the functional $\Phi_a : \mathcal{P} \rightarrow \mathcal{P}$ given by

$$\Phi_a(p) = \left(\sum_{\ell=0}^\infty p \lambda_{\ell}^*(a) \sum_{\ell=0}^\infty p \lambda_{\ell}^*(a) \right)_{k \geq 0}
$$

(3.5)

where $p = (p_0, p_1, \ldots) \in \mathcal{P}$. Write $\{\Phi_a(p)\}_k$ for the $k$-th coordinate of the above map. Let $p^i = p(f_i) := (p_{i,0}, p_{i,1}, \ldots)$ for $i = 0, 1$ denote the degree distribute or a random recursive tree grown with attachment function $f_i$ (i.e. without any change point). Corollary 7.2 shows that for each $t > \gamma$, there is a unique $0 < a_t < \infty$ such that

$$\sum_{k=0}^\infty p_k^u \left[\int_0^{a_t} m_{f_i}(a_t - s) \mu_{f_i}^k(d s)\right] = \frac{t - \gamma}{\gamma}.$$ 

(3.6)

Define $a_t = 0$ for $t \leq \gamma$. Now, we are ready to state our main theorem on sup-norm convergence of degree distributions post-change point.

Theorem 3.6. Suppose $f_0, f_1$ satisfy Assumption 2.1. For any $k \geq 0$ and $s \in [\gamma, 1]$

$$\sup_{t \in [\gamma, s]} \left|\frac{D_n(k, nt)}{nt} - (\Phi_{a_t}(p_0^0))_k\right| \rightarrow 0.$$
There is a probabilistic way to view the limit which we now describe at the end of the construction of the process namely \( t = 1 \). Write \( \alpha \) for \( a_1 \). Construct an integer valued random variable \( D_\theta \) using the following auxiliary random variables:

**Construction 3.7 (X_{BC}).** Generate \( D \sim \{p_k^0 : k \geq 1\} \). Conditional on \( D = k \), generate point process \( \xi^{(k)}_{f_1} \) and let \( \mathcal{C} = \xi^{(k)}_{f_1}[0, \alpha] \), with \( \alpha \) as in (3.6). Now set \( X_{BC} = D + \mathcal{C} \).

**Construction 3.8 (X_{AC}, Age).** (a) Generate \( D \sim \{p_k^0 : k \geq 1\} \).
(b) Conditional on \( D = k \), generate random variable \( \text{Age} \) supported on the interval \([0, \alpha]\) with distribution

\[
\mathbb{P}(\text{Age} > u) := \frac{\int_0^u m_{f_1}(\alpha - u - s)d\mu_{f_1}^{(k)}(ds)}{\int_0^\alpha m_{f_1}(\alpha - s)d\mu_{f_1}^{(k)}(ds)}, \quad 0 \leq u \leq \alpha.
\]  
(3.7)
(c) Conditional on \( D \) and \( \text{Age} \), let \( X_{AC} = \xi_{f_1}[0, \text{Age}] \), where as in (2.5), \( \xi_{f_1} \) is the point process constructed using attachment function \( f_1 \).

Now let \( \theta = (f_0, f_1, \gamma) \). Let \( D_{\theta} \) be the integer valued random variable defined as follows: with probability \( \gamma \), \( D_{\theta} = X_{BC} \) and with probability \( 1 - \gamma \), \( D_{\theta} = X_{AC} \). The following is a restatement of the convergence result implied by Theorem 3.6 for time \( t = 1 \).

**Theorem 3.9 (Standard model, \( J = 1 \)).** As in Section 2, fix \( k \geq 0 \) and let \( D_n(k) \) denote the number of vertices with out-degree \( k \) in the tree \( \mathcal{T}_n \). Under Assumption 2.1 on the attachment functions \( f_0, f_1 \) and Assumption 3.5 on the change point \( \gamma \), we have that

\[
\frac{D_n(k)}{n} \xrightarrow{p} \mathbb{P}(D_{\theta} = k).
\]

Write \( p(\theta) \) for the pmf of \( D_{\theta} \). The next result, albeit intuitively reasonable is non-trivial to prove in the generality of the models considered in the paper.

**Corollary 3.10.** Assume that \( p^0 \neq p^1 \). Then for any \( 0 < \gamma < 1 \) one has \( p^0 \neq p(\theta) \). Thus the change point always changes the degree.

The next result describes the tail behavior of the ensuing random variable.

**Corollary 3.11 (Initializer always wins).** The initializer function \( f_0 \) determines the tail behavior of \( D_{\theta} \) in the sense that

(i) If in the model without change point using \( f_0 \), the degree distribution has an exponential tail then so does the model with change point irrespective of \( \gamma > 0 \) and \( f_1(\cdot) \).
(ii) If in the model without change point using \( f_0 \), the degree distribution has a power law tail with exponent \( k > 0 \) then so does model with change point irrespective of \( \gamma > 0 \) and \( f_1(\cdot) \).

**Corollary 3.12 (Maximum degree).** Suppose the initializer function is linear with \( f_0(i) = i + 1 + \alpha \) for \( i \geq 0 \). For fixed \( k \geq 1 \), let \( M_n(k) \) denote the size of the \( k \)-th maximal degree. Then as long as the function \( f_1 \) satisfies Assumption 2.1 \( M_n(k)/n^{1/(\alpha+2)} \) is a tight collection of random variables bounded away from zero as \( n \to \infty \).

**Remark 5.** Without change point, it is known [33] that for each fixed \( k \), \( M_n(k)/n^{1/(\alpha+2)} \xrightarrow{d} X_{k}(\alpha) \) for a non-degenerate distribution. Thus the above result shows that irrespective of the second attachment function \( f_0 \), the maximal degree asymptotics for linear preferential attachment remain unaffected. The proof of the above result follows via analogous arguments as [13, Theorem 2.2] and thus we will not prove it in this paper.
3.3. Multiple change points. Fix $J \geq 1$, $\gamma := (\gamma_1, \gamma_2, \ldots, \gamma_J)$ with $0 < \gamma_1 < \gamma_2 < \cdots < \gamma_J < 1$ and let $\gamma_0 = 0, \gamma_{J+1} = 1$. Further fix attachment functions $f_0, f_1, \ldots, f_j$ satisfying Assumption 2.1 and let $f := (f_0, f_1, \ldots, f_J)$. We start with the following recursive construction of a sequence of probability mass functions $\{p^j : 0 \leq j \leq J\}$ and positive constants $\alpha := \{\alpha_j : 1 \leq j \leq J\}$.

(a) Initialiation: For $j = 0$, let $p^0 := \{p^0_k : k \geq 0\}$ as in (3.1).

(b) Pre-epoch distribution: For $1 \leq j \leq J + 1$, define the random variable $X_{PE}^{J-1} \sim p^{J-1}$.

(c) $\alpha$ recursion: For $1 \leq j \leq J$, define $\alpha_j > 0$ as the unique root of the equation:

$$\sum_{k=0}^\infty p^j_k \bigg[ \int_0^{\alpha_j} m_{f_j}(\alpha - s) d\mu_{f_j}^k(s) \bigg] := \frac{\gamma_{j+1} - \gamma_j}{\gamma_j}. \tag{3.8}$$

(d) Epoch age distribution: Fix $1 \leq j \leq J$. Generate $X_{PE}^{j-1}$ as above. Conditional on $X_{PE}^{j-1} = k$, generate random variable Epoch $j$ supported on the interval $[0, \alpha_j]$ with distribution

$$\mathbb{P}(\text{Epoch } j > u) := \frac{\int_0^{\alpha_j-u} m_{f_j}(\alpha - u - s) d\mu_{f_j}^k(ds)}{\int_0^{\alpha_j} m_{f_j}(\alpha - s) d\mu_{f_j}^k(ds)}, \quad 0 \leq u \leq \alpha_j. \tag{3.9}$$

(e) Alive after epoch degree distribution: Conditional on the random variables in (d) let $X_{AE}^{j} := \xi_{f_j}[0, \text{Epoch } j]$ where as before $\xi_{f_j}$ is the point process with attachment function $f_j$.

(f) Alive before epoch distribution: Fix $j \geq 1$. For $k \geq 0$, let $\xi_{f_j}^{(k)}$ be the point process 2.6 using attachment function $f_j$. Generate $X_{PE}^{j-1}$ as in (b). Conditional on $X_{PE}^{j-1} = k$, let $\xi_{f_j} := \xi_{f_j}^{(k)}[0, \alpha_j]$ with $\alpha_j$ as in (3.8). Define the random variable $X_{BE}^{j} := X_{PE}^{j-1} + \xi_{f_j}$.

(g) Mixture distribution: Finally define $X_{BE}^{j}$ as the following mixture: with probability $\gamma_{j}/\gamma_{j+1}$ $X_{BE}^{j} = X_{BE}^{j}$, with probability $(\gamma_{j+1} - \gamma_{j})/\gamma_{j+1}$, let $X_{BE}^{j} = X_{AE}^{j}$.

(h) Let $p^j$ be the probability mass function of $X_{BE}^{j}$.

With $\theta := (\gamma, f)$, write $D_{\theta} := X_{BE}^{j}$.

**Theorem 3.13** (Standard model, multiple change points). As in Section 2 fix $k \geq 0$ and let $D_n(k)$ denote the number of vertices with out-degree $k$ in the tree $\mathcal{T}_n^\theta$ with $\theta$ as above. Under Assumption 2.1 on the attachment functions $f$ we have that

$$\frac{D_n(k)}{n} \xrightarrow{P} \mathbb{P}(D_{\theta} = k).$$

Further the assertions of Corollaries 3.11 and 3.12 continue to hold in this regime.

3.4. The quick big bang model. Now we consider the case where the change point happens “early” in the evolution of the process, where the change point scales like $o(n)$. To simplify notation, we specialize to the case $J = 1$, however our methodology is easily extendable to the general regime. Let $\{p^1_k : k \geq 0\}$ denote the probability mass function as in (3.1) but using the function $f_1$ to construct $\lambda^*$ in (2.3) and then $f_1$ in place of $f_0$ in (3.1).

Define for $\alpha > 0$ and any non-negative measure $\mu$,

$$\hat{\mu}(\alpha) := \int_0^\infty \alpha e^{-\alpha t} \mu(dt).$$

We will work under the following assumption.

**Assumption 3.14.** $\mathbb{E}(\hat{\xi}_f(\lambda^*) \log(\hat{\xi}_f(\lambda^*))) < \infty$.

**Remark 6.** Assumption 3.14 is weaker than Assumption 3.2 as seen by considering the linear preferential attachment model with attachment function $f(i) = i + 1, i \geq 0$. In this case, $E(\hat{\xi}_f(\lambda^*))^2 = \infty$ but $E(\hat{\xi}_f(\lambda^*))^\beta < \infty$ for any $1 < \beta < 2$ (see 11 Proposition 53 (a)).
Recall that in the previous section, one of the messages was that the initializer function \( f_0 \) determined various macroscopic properties of the degree distribution for the standard model.

**Theorem 3.15.** Suppose \( \tau_1 = n^\gamma \) for fixed \( 0 < \gamma < 1 \). If \( f_0, f_1 \) satisfy Assumptions 2.1, 3.1 and 3.14, the degree distribution does not feel the effect of the change point or the initializer attachment function \( f_0 \) in the sense that for any fixed \( k \geq 0 \),

\[
\frac{D_n(k)}{n} \xrightarrow{p} p_{k^*}^1 \quad \text{as } n \to \infty.
\]

**Remark 7.** The form \( \tau_1 := n^\gamma \) was assumed for simplicity. We believe the proof techniques are robust enough to handle any \( \tau_1 = \omega_n \), where \( \omega_n = o(n) \) and \( \omega_n \uparrow \infty \). We defer this to future work.

The next result implies that the maximal degree does feel the effect of the change point. Instead of proving a general result we will consider the following special cases. Throughout \( M_n(1) \) denotes the maximal degree in \( \mathcal{T}^\theta_n \).

**Theorem 3.16 (Maximal degree under quick big bang).** Once again assume \( \tau_1 = n^\gamma \). Consider the following special cases:

(a) **Uniform \( \sim \) Linear:** Suppose \( f_0 \equiv 1 \) whilst \( f_1(k) = k + 1 + \alpha \) for fixed \( \alpha > 0 \). Then with high probability as \( n \to \infty \), for any sequence \( \omega_n \uparrow \infty \),

\[
\frac{n^{\frac{1}{1-\gamma}} \log n}{\omega_n} \ll M_n(1) \ll n^{\frac{1}{\gamma}} (\log n)^2.
\]

(b) **Linear \( \sim \) Uniform:** Suppose \( f_0(k) = k + 1 + \alpha \) for fixed \( \alpha > 0 \) whilst \( f_1(\cdot) \equiv 1 \). Then with high probability as \( n \to \infty \), for any sequence \( \omega_n \uparrow \infty \),

\[
\frac{n^{\frac{1}{1-\gamma}} \log n}{\omega_n} \ll M_n(1) \ll n^{\frac{1}{\gamma}} (\log n)^2.
\]

(c) **Linear \( \sim \) Linear:** Suppose \( f_0(k) = k + 1 + \alpha \) whilst \( f_1(k) = k + 1 + \beta \) where \( \alpha \neq \beta \). Then \( M_n(1) / n^{\eta(\alpha, \beta)} \) is tight and bounded away from zero where

\[
\eta(\alpha, \beta) := \frac{\gamma(2 + \beta) + (1 - \gamma)(2 + \alpha)}{(2 + \alpha)(2 + \beta)}.
\]

**Remark 8.** It is instructive to compare the above results to the setting without change point. For the uniform \( f \equiv 1 \) model, it is known [22, 43] that the maximal degree scales like \( \log_2(n) \) whilst for the linear preferential attachment, the maximal degree scales like \( n^{1/(\alpha + 2)} \) [33]. Thus for example, (b) of the above result coupled with Theorem 3.15 implies that the limiting degree distribution in this case is the same as that of the uniform random recursive tree (URRT) namely Geometric with parameter 1/2; however the maximal degree scales polynomially in \( n \) and not like \( \log n \) as in the URRT.

**Remark 9.** For any \( \tau_1 \to \infty \), the initial segment should always leave its signature in some functional of the process. See for example [17, 18, 20] where the evolution of the system (using typically linear preferential attachment albeit [17] also considered the uniform attachment case) starting from a fixed “seed” tree was considered and the aim was to detect (up to some level of accuracy) this seed tree after observing the tree \( \mathcal{T}_n \). Similar heuristics suggest that in the context of our model, the initial segment of the process however small should show its signature at some level. We discuss this aspect further in Section 4.

Proofs of results for the quick big bang model are given in Section 4.

3.5. **Change point detection.** In this Section, we discuss the statistical issues of actual change point detection from an observation of the network. We will only consider the standard model and one change point \( (J = 1) \). We do not believe the estimator below is “optimal” in terms of rates of convergence, however the motivation behind proving the sup-norm convergence result Theorem 3.6 is to provide impetus for further research in obtaining better estimators.
Consider any two sequences \( h_n \to \infty, b_n \to \infty \) satisfying \( \frac{\log h_n}{\log n} \to 0, \frac{\log b_n}{\log n} \to 0 \) as \( n \to \infty \). We define the following change point estimator:

\[
\hat{T}_n = \inf \left\{ t \geq 1 : \sum_{k=0}^{\infty} 2^{-k} \frac{1}{h_n} \left| D_n(k, \mathcal{T}_{[nt]}^\theta) - D_n(k, \mathcal{T}_{[n/h_n]}^\theta) \right| n/h_n > b_n \right\}.
\]

The following theorem establishes the consistency of the above estimator.

**Theorem 3.17.** Assume that \( p^0 \neq p^1 \). Suppose \( f_0 \) satisfies Assumptions 2.1, 3.1 and 3.2 and \( f_1 \) satisfies Assumption 2.1. Then \( \hat{T}_n \xrightarrow{P} \gamma \).

**Remark 10.** From a practical point of view, for the proposed estimator to be close to the change point even for moderately large \( n \), we should select \( h_n, b_n \) satisfying the above hypotheses so that \( h_n \) grows as slowly as possible (which ensures that we look at the evolving tree not too early, before the ‘law of large numbers’ effect has set in) and \( b_n \) grows as quickly as possible (to ensure that the detection threshold is sufficiently close to zero to capture the change in degree distribution close to the change point). One reasonable choice is \( h_n = \log \log n \) and \( b_n = n^{1/\log \log n} \).

Theorem 3.17 is proved in Section 10. Figure 3.1 shows the result of computing the change point estimator for a network with a single change point. We plot the function:

\[
d_n(m) := \sum_{k=0}^{\infty} 2^{-k} \left| D_n(k, \mathcal{T}_m^\theta) - D_n(k, \mathcal{T}_{[n/h_n]}^\theta) \right| m/n > \frac{n}{\log \log n} < m.
\]

**Figure 3.1.** The function \( d_n(\cdot) \) for one network. (A) \( d_n(\cdot) \) for one network. (B) Mean, 10/90 th percentiles from 100 simulations.
4. Discussion

(i) **Random recursive trees:** Random recursive trees have now been studied for decades, motivated by a wide array of fields including convex hull algorithms, linguistics, epidemiology and first passage percolation and recently in the study of various coalescent processes. See [21, 23, 26, 32, 41] and the references therein for starting points to this vast literature. For specific examples such as the uniform attachment or the linear attachment model with \( f(i) = i + 1 \), one can use the seminal work of Janson [31] via a so-called “super ball” argument to obtain functional central limit theorems for the degree distribution. Obtaining quantitative error bounds let alone weak convergence results in the general setting considered in this paper is much more non-trivial. Regarding proof techniques, we proceed via embedding the discrete time models into continuous time branching processes and then using martingale/renewal theory arguments for the corresponding continuous time objects to read off corresponding results for the discrete models; this approach goes back all the way to [3]. Limit results for the corresponding CTBPs in the setting of interest for this paper were developed in the seminal work of Jagers and Nerman [28, 30, 34]. One contribution of this work is to derive quantitative versions for this convergence, a topic less explored but required to answer questions regarding statistical estimation of the change point.

(ii) **Fringe convergence of random trees:** A second aim of this work (albeit not developed owing to space) is understanding rates of convergence of the fringe distribution. We briefly describe the context, referring the interested reader to [2, 27] for general theory and discussion of their importance in computer science. Let \( T \) denote the space of all rooted (unlabelled) finite trees (with \( \emptyset \) denoting the empty tree). Fix a finite non-empty rooted tree \( T \in T \) with root \( \rho \). For each \( v \in T \) let \( f(v, T) \) denote the sub-tree consisting of the set of vertices “below” \( v \) namely vertices for which the shortest path from \( \rho \) needs to pass through \( v \). View \( f(v, T) \) as an element in \( T \) via rooting it at \( v \). The fringe distribution of \( T \) is the probability distribution on \( T \):

\[
\pi_T(t) := \frac{1}{|T|} \sum_{v \in T} \mathbb{1}\{f(v, T) = t\}, \quad t \in T.
\]

If \( \{T_n : n \geq 1\} \) is a sequence of random trees, one now obtains a sequence of random probability measures. Aldous in [2] shows that convergence of the associated fringe measures implies convergence of the associated random trees locally to limiting infinite random trees with a single infinite path; this then implies convergence of a host of global functionals such as the empirical spectral distribution of the adjacency matrix, see e.g. [12]. For a number of discrete random tree models, embedding these in continuous time models and using results of [30, 34] has implied convergence of this fringe distribution; however establishing rates of convergence has been non-trivial [27]. While many of the results in this paper are all formulated in terms of the degree distribution, the results and most of the proofs in Section 9 extend to more general characteristics such as the fringe distribution. To keep the paper to manageable length, this is deferred to future work.

(iii) **General change point:** Change point detection especially in the context of univariate time series has also matured into a vast field, see [16, 19]. Even in this context, consistent estimation especially in the setting of multiple change points is non-trivial and requires specific assumptions on the nature of the change see e.g. [45] for work in estimating the change in mean of a sequence of independent observations from the normal distribution; in the context of econometric time series settings including linear regression see for example [5–7]; for recent applications in the biological sciences [38, 46]. The only pre-existing work on change point in the context of evolving networks formulated in this paper that we are aware of was carried out in [13] where one assumed linear attachment functionals of the form \( f(k) = k + \alpha \) for some parameter \( \alpha \geq 0 \). In this context, specialized computations specific to this model enabled one to derive change point detection estimators that were \( \log n / \sqrt{n} \) consistent. Unfortunately these techniques do not extend to the general case considered in this paper.
(iv) **Open questions:** In the context of rates of convergence, one natural question is to understand if one can obtain tighter bounds than those in Theorem 3.3 and in particular prove a functional central limit theorem (FCLT) with $\sqrt{n}$ scaling as in [31]. In fact, a more general FCLT for the model with change point of the following form should hold: there exists a Gaussian process $\{G_{\infty}(t)\}_{t \in [0,1]}$ such that for any $\epsilon \in (0,1)$,

$$
\left\{ \sqrt{n} \left( \sum_{k=0}^{\infty} 2^{-k} \frac{D_n(k, nt)}{nt} - (\Phi_{a_\epsilon}(p_0^\epsilon))_t \right) \right\} \overset{d}{\to} \{G_{\infty}(t)\}_{t \in [\epsilon,1]}
$$

in $C[\epsilon,1]$. This will directly imply $\log n/\sqrt{n}$ consistency for the proposed change point estimator. One of the major issues that this paper does not address is the question of consistently estimating multiple change points. In the context of univariate change point detection, one is able to often use methodology for estimating a single change point to sequentially estimate multiple change points. However the non-ergodic nature of evolution of the model considered in this paper after the first change point does not lend itself easily to this scheme of analysis. A second line of work that we are currently exploring is extending the above techniques to general network (i.e. non-tree) models.

5. **Initial Embeddings and Constructions**

The rest of the paper is devoted to proofs of the main results.

5.1. **Road map for proofs of the main results.** The rest of this section is devoted to some preliminary estimates and constructions that will then be repeatedly used in the proofs. Although the results about convergence rates for the model without change point are stated before the change point results, the proof of Theorem 3.3 is quite technical and an essential ingredient is a "sup-norm estimate" given in Lemma 7.11 which is proven more generally in the context of a change point. Thus, we defer the proof of Theorem 3.3 to Section 9. Section 6 deals with a continuous time version of the change point model analyzed for a fixed time $a$ after the change point. Theorem 6.1 proved here estimates for a general characteristic $\phi \in \mathcal{C}$ the $L^1$-error in approximating the aggregate $\phi$-score at time $a$ of all individuals born after the change point with a weighted linear combination of the degree counts at the change point. This estimate, apart from directly yielding a law of large numbers (see second part of Theorem 6.1), turns out to be crucial in most subsequent proofs. The estimates derived in Section 6 are then used in Section 7 to analyze the standard model and prove the main theorems in this setting (Theorems 3.6 and 3.9) as well as Corollary 3.11 on the initializer always winning. Corollary 3.10 follows directly from Lemma 10.3 and requires an in-depth analysis of the fluid limits derived in Theorem 3.6 and is postponed to Section 10. Section 8 contains proofs of the quick big bang model. We note here that all the estimates obtained in Sections 6 and 7 to analyze the model for a fixed time $a$ after the change point explicitly exhibited the dependence on $a$. This turns out to be crucial in Section 8 where we take $a = \eta_0 \log n$ and the estimates above still hold if $\eta_0$ is sufficiently small. Roughly speaking, we partition the interval $[T_{n'}', T_n]$ into finitely many subintervals of size at most $\eta_0 \log n$ and ‘bootstrap’ the estimates obtained above to prove Theorem 3.15. We conclude in Section 10 with the proof of Theorem 3.17 on the change point detection estimator.

5.2. **Initial constructions.** Fix $n \geq 3$, and $1 < r_n < n$ ($r_n$ will later assume the value $\gamma n$ or $n^\gamma$), two attachment functions $f_0, f_1$ satisfying Assumption 2.1.

**Definition 5.1 (CTBP with change point).** Recall that $\{BP_{f_0}(t) : t \geq 0\}$ denotes a continuous-time branching process driven by the point process $\xi_{f_0}$ defined in (2.4). Now for $n, r_n$ and two attachment functions $f_0, f_1$ as above define $\{BP_n(t) : t \geq 0\}$ as follows:

(a) Generate a process $BP_{f_0}(\cdot)$ as above. For $0 \leq t \leq T_{r_n}$ let $BP_n(\cdot) \equiv BP_{f_0}(\cdot)$.

(b) At time $T_{r_n}$ all existing vertices change their reproduction, so that for any fixed $k \geq 0$, a vertex with $k$ children in $BP_{f_0}(T_{r_n})$ now uses offspring distribution $\xi_{f_1}^{(k)}$ for all subsequent offspring. Each new vertex born into the system has offspring point process with distribution $\xi_{f_1}$, independent across vertices. Label vertices as above according to the time order they enter the system.
The following is the analog of Lemma 2.3 in the change point setting.

**Lemma 5.2.** Let \( \theta = (f_0, f_1, r_n) \) and consider the sequence of random trees \( \{ T_m^\theta : 2 \leq m \leq n \} \) with one change point at \( r_1 = r_n \). Consider the continuous time construction in Definition 5.1 and define for \( m \geq 1 \) the stopping times \( T_m := \inf \{ t \geq 0 : |BP_n(t)| = m \} \). Then viewed as a sequence of growing random labelled rooted trees we have, \( |BP_n(T_m) : 2 \leq m \leq n | \overset{d}{=} \{ T_m : 2 \leq m \leq n \} \).

The next few Lemmas deal with properties of one important class of offspring point processes that arise in the study of linear preferential attachment.

**Definition 5.3** (Rate \( \nu \) Yule process). Fix \( \nu > 0 \). A rate \( \nu \) 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

\[
\mathbb{P}(Y_\nu(t+) - Y_\nu(t)|\mathcal{F}(t)) = \nu Y_\nu(t) dt + o(dt),
\]

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 e.g. [37, Section 2.5].

**Lemma 5.4.** Fix \( t > 0 \) and rate \( \nu > 0 \). Then \( Y_\nu(t) \) has a Geometric distribution with parameter \( p = e^{-\nu t} \). Precisely,

\[
\mathbb{P}(Y_\nu(t) = k) = e^{-\nu t}(1 - e^{-\nu t})^{k-1}, \quad k \geq 1.
\]

The process \( \{ Y_\nu(t) \exp(-\nu t) : t \geq 0 \} \) is an \( \mathbb{L}^2 \) bounded martingale and thus there exists a strictly positive random variable \( W \) such that \( Y_\nu(t) \exp(-\nu t) \overset{a.s.}{=} W \). Further \( W = \exp(1) \).

Next we derive moment bounds for the attachment point processes for linear preferential attachment.

**Lemma 5.5.** Fix \( \nu > 0, \kappa \geq 0 \). Let \( \xi_{\nu,\kappa}(t) \) be the offspring distribution of a linear preferential attachment process with attachment function \( f(i) = \nu(i + 1) + \kappa \). Then with respect to the natural filtration the following two processes are martingales:

\[
M_1(t) := e^{-\nu t} \xi_{\nu,\kappa}(t) - \frac{\nu + \kappa}{\nu}(e^{\nu t} - 1), \quad t \geq 0
\]

and

\[
M_2(t) := e^{-2\nu t} \xi_{\nu,\kappa}(t)^2 - \int_0^t (2\kappa + 3\nu) \xi_{\nu,\kappa}(s)e^{-2\nu s} ds - \frac{\nu + \kappa}{2\nu}(1 - e^{-2\nu t}), \quad t \geq 0.
\]

In particular,

\[
\mathbb{E}\xi_{\nu,\kappa}(t) = \frac{\nu + \kappa}{\nu}(e^{\nu t} - 1), \quad \text{and} \quad \mathbb{E}(\xi_{\nu,\kappa}(t))^2 = \frac{(2\kappa + 3\nu)(\nu + \kappa)}{2\nu^2}(e^{\nu t} - 1)^2 + \frac{\nu + \kappa}{2\nu}(e^{2\nu t} - 1).
\]

**Proof.** We sketch the proof. Let \( \mathcal{F}(t) \) be the natural filtration corresponding to the continuous time branching process with attachment function \( f \). Note that \( \xi_{\nu,\kappa}(t) \sim \xi_{\nu,\kappa}(t + 1) + \kappa \). This can be used to check \( \mathbb{E}[dM_1(t)|\mathcal{F}(t)] = 0 \) showing \( M_1(t) \) is a martingale. Similarly, \( \xi_{\nu,\kappa}(t)^2 \sim \xi_{\nu,\kappa}(t)^2 + 2\xi_{\nu,\kappa}(t) + 1 \) at rate \( v(\xi_{\nu,\kappa}(t) + 1) + \kappa \). This expression can similarly be used to check \( M_2(t) \) is a martingale. The first expectation claimed in the lemma follows immediately by setting the expectation of \( M_1(t) \) equal to zero. The second expectation follows by computing the expectation of \( M_2(t) \) and then using the expectation of \( \xi_{\nu,\kappa}(t) \). \( \square \)

The next result derives moment bounds for a particular class of CTBP.

**Definition 5.6** (Rate \( \nu \) Affine \( \kappa \) PA model). Fix \( \nu > 0, \kappa \geq 0 \). A branching process whose offspring distribution is given by an offspring distribution constructed using attachment function \( f(i) = \nu(i + 1) + \kappa \) will be called a linear PA branching process with rate \( \nu \) and affine parameter \( \kappa \). Denote this as \( \{ PA_{\nu,\kappa}(t) : t \geq 0 \} \).

We will now derive expressions for moments of the process \( PA_{\nu,\kappa} \) that will be useful in the sequel. To simplify notation, when possible we will suppress dependence on \( \nu, \kappa \) and write the above as \( PA(\cdot) \).
Proposition 5.7. Fix $\nu > 0, \kappa \geq 0$. With respect to the natural filtration, the following processes are martingales:

$$M_1(t) := e^{-(2\nu + \kappa)t}(|PA_{V,\kappa}(t)| - 1) - \frac{\nu + \kappa}{2\nu + \kappa} (1 - e^{-(2\nu + \kappa)t}), \quad t \geq 0$$

and

$$M_2(t) := (|PA_{V,\kappa}(t)| - 1)^2 - \int_0^t ((4\nu + 2\kappa)(|PA_{V,\kappa}(s)| - 1)^2 + (4\nu + 3\kappa)(|PA_{V,\kappa}(s)| - 1) + (\nu + \kappa))ds, \quad t \geq 0.$$

In particular, for any fixed $a > 0$, there exists $C$ (dependent on $\nu$ and $\kappa$ but not on $a$) such that for $0 \leq t \leq a$

$$E(|PA_{V,\kappa}(t)| - 1) \leq Ce^{(2\nu + \kappa)a}t; \quad E((|PA_{V,\kappa}(t)| - 1)^2) \leq Ce^{(4\nu + 2\kappa)a}t.$$

Proof. Write $\{x \in \mathcal{P}_{\nu}\}$ for any $\nu > 0$ and $\kappa \geq 0$. Also recall for any $\nu > 0$, $\kappa \geq 0$, and any $\alpha > 0$.

$$\hat{x}(\alpha) := \int_0^\infty a e^{-\alpha t} \chi(t)dt.$$

Recall that $\lambda^*$ was the Malthusian rate of growth and $\mu_f$ denoted the mean measure of the offspring distribution. Let $m^* := \int_{\mathbb{R}_+} u e^{-\lambda^* u} \mu_f(du)$. For any fixed characteristic $\chi \in \mathcal{C}$ and any $\alpha > 0$, define

$$\hat{x}(\alpha) := \int_0^\infty a e^{-\alpha t} \chi(t)dt.$$

Also recall for $\alpha > 0$,

$$\hat{\mu}(\alpha) := \int_0^\infty a e^{-\alpha t} \mu(t)dt.$$

A useful fact is that for any $\alpha > 0$, recalling $\hat{\rho}$ from Assumption 2.1 (iii),

$$\hat{\rho}(\alpha) = \hat{\mu}_f(\alpha) = \int_0^\infty e^{-\alpha t} \mu_f(dt).$$

Recall $Z_f^\chi(t) = \sum_{x \in \mathcal{P}_{\nu}\chi(t - \sigma_x)}$ and $M_f^\chi(t) = \mathbb{E}(e^{-\lambda^* t} Z_f^\chi(t))$. Recall $Z_f(t)$ is the total number of vertices at time $t$ and $M_f(t) = \mathbb{E}(e^{-\lambda^* t} Z_f(t))$. The following Lemma is a consequence of 34 Theorem 6.3.

Lemma 5.8. (i) Under Assumption 2.1 (iii), for any characteristic $\chi \in \mathcal{C}$,

$$\frac{Z_f^\chi(t)}{Z_f(t)} \overset{a.s.}{\longrightarrow} \mathbb{E}(\hat{x}(\lambda^*))$$

(ii) Under Assumptions 2.1 and 3.14, there exists a strictly positive random variable $W_\infty$ with $\mathbb{E}(W_\infty) = 1$ such that for characteristics $\chi \in \mathcal{C}$,

$$e^{-\lambda^* t} Z_f^\chi(t) \overset{a.s., \text{.d.}}{\longrightarrow} \frac{\mathbb{E}(\hat{x}(\lambda^*))}{\lambda^* m^*} W_\infty.$$

Proof. (i) We will apply 34 Theorem 6.3 with characteristics $\chi$ and $\psi$ defined by $\psi(t) := 1 \{t \geq 0\}$ by verifying Conditions 6.1 and 6.2 in 34. Condition 6.1 holds for $\xi_f$ by Assumption 2.1 (iii). Condition 6.2 requires there exist $\beta < \lambda^*$ such that $\mathbb{E}(\sup_t (e^{-\beta t} \chi(t))) < \infty$ and $\mathbb{E}(\sup_t (e^{-\beta t} \psi(t))) < \infty$. For $\psi$ this condition holds for any $\beta$ since $\mathbb{E}(\sup_t (e^{-\beta t} \psi(t))) = \mathbb{E}(\sup_t (e^{-\beta t})) = 1$. To verify Condition 6.2 for $\chi$ note that for any $\beta \geq 0$, using the fact that $\chi \in \mathcal{C}$,

$$\sup_{t \in [0, \infty)} (e^{-\beta t} \chi(t)) \leq \sum_{j=0}^\infty \sup_{t \in [j, j+1)} (e^{-\beta t} \chi(t)) \leq b \sum_{j=0}^\infty e^{-\beta j} (\xi_f(j + 1) + 1).$$
From Assumption 2.1(iii), there exists $\beta_0 < \lambda^*$ such that $\mu_f(\beta_0) < \infty$ which implies there exists $C > 0$ such that $\mathbb{E}\{\xi_f(t + 1) + 1\} \leq Ce^{\beta_0 t}$ for all $t \geq 0$. Using this and setting $\beta = (\beta_0 + \lambda^*)/2$, we get

$$\mathbb{E}\left(\sup_{t \in [0,\infty)} e^{-\beta t} \chi(t)\right) \leq Cb_\delta \sum_{j=0}^\infty e^{-\beta j e^{\beta_0 j}} < \infty \quad (5.2)$$

which verifies Condition 6.2 for $\chi$. It is easy to check condition (2.6) in [34] using the fact that $\mathbb{E}\{\chi(t)\} \leq Cb_\delta e^{\beta_0 t}$ and $\psi(t) \leq 1$. Thus, Proposition 2.2 of [34] implies $M_f^1(\infty) = \mathbb{E}(\chi(\lambda^*))/\mathbb{E}(\xi_f(\lambda^*) + \lambda_m^* + 1)$ and $M_f(\infty) = 1/(\lambda^* m^*)$ and this, along with Theorem 6.3 from [34] implies (i).

(ii) To show almost sure convergence, we will verify Conditions 5.1 and 5.2 of [34]. From (2.2), we obtain $\beta_0 < \lambda^*$ such that $f_0^\infty e^{-\beta_0 t} \mu(d t) < \infty$ and this implies Condition 5.1 with $g(t) = e^{-(\lambda^* - \beta_0) t}$ (see the remark following Condition 5.1 of [34]). Condition 5.2 for $\chi \in \mathcal{C}$ follows from (5.2) with $h(t) = e^{-(\lambda^* - \beta_0) t}$.

We will eventually take limits as $t \to \infty$ i.e. $t = s$ corresponds to actual time $T_{\beta_1} + s$ for any $s \in [0, a]$. For $t \geq 0$, $BP_n(t)$ will denote the branching process at time $t$ (i.e. time $t$ after the change point).

6. CHANGE POINT MODEL FOR FIXED TIME $a$: POINT-WISE CONVERGENCE FOR GENERAL CHARACTERISTICS

In this section we consider growing the tree for a constant time $a$ after the change point i.e. for $t \in [T_{\beta_1}, T_{\beta_1} + a]$ using the second attachment function, $f_i$. Consider the class of characteristics $\mathcal{C}$ defined in (3.2). We will count vertices born after the change point according to a general characteristic $\phi \in \mathcal{C}$ and prove a law of large numbers for this aggregate $\phi$-score at time $a$ as $n \to \infty$ (see Theorem 6.1). This will be a key tool in the rest of the paper. For notational convenience we will consider the time to start at $t = 0$ (i.e. $t = s$ corresponds to actual time $T_{\beta_1} + s$ for any $s \in [0, s]$).

6.1. Notation. Let $\lambda^*$ denote the Malthusian parameter for the branching process with attachment function $f_i$. For the branching process (without change point) with attachment function $f_i$, and for any characteristic $\phi$, recall $Z_i^\phi(t) = \sum_{x \in BP_i(t)} \phi_x(t - \sigma_x)$. When $\phi(t) = 1 \{ t \geq 0 \}$, we will write $Z_i^\phi$ for $Z_i^{\phi,i}$. Let $m_i^{\phi}(t) := \mathbb{E}Z_i^\phi(t)$ and let $v_i^{\phi}(t) = \text{Var}\left(Z_i^\phi(t)\right)$. For $\phi \in \mathcal{C}$, an easy computation implies there exists $c > 0$ such that $Z_i^{\phi,i}(t) \leq 2cZ_i^{\phi,i}(t)$ for every $t \geq 0$ and hence,

$$\sup_{t \in [0,a]} v_i^{\phi}(t) \leq 4c^2 \mathbb{E}(Z_i^{\phi,i}(a)) \leq Ce^{C'a} \quad (6.1)$$

where $C, C'$ do not depend on $a$. This follows by Assumption 2.1(ii) on $f_i$ that implies $BP_i(\cdot)$ is stochastically dominated by a rate $C$ PA branching processes (see Definition 5.6) and then by appealing to (5.1).

6.2. Definitions. Next we define various constructs which will be used in this section. Divide the interval $[0, \alpha] := \cup_{i=0}^{n^\delta - 1} [i n^\delta, (i + 1)n^\delta]$ into subintervals of size $a/n^\delta$. We will eventually take limits as $\delta \to \infty$.

(i) System at change point: Recall the construction of the change point model in continuous time via Lemma 5.2. Let $\mathcal{F}_n(0)$ denote the $\sigma$-field containing the information till $T_{\beta_1}$, the change point. Define the filtration $\mathcal{F}_n(t) := \{\sigma(BP_n(t)) : t \geq 0\}$. We will first work conditional on $\mathcal{F}_n(0)$. For fixed $k \geq 0$, to specify dependence on time, we write $D_n(k, t)$ to be the set of vertices with (out-)degree $k$ at time $t$ and let $D_n(k, t) := \{D_n(k, t)\}$. The initial set $D_n(k, 0)$ which arose from the pre-change point dynamics will play a special role. Label the vertices in $D_n(k, 0)$ in the order they were born into $BP_n(0)$ as $D_n(k, 0) = \{v_{j,1}^{(k)}(k), v_{j,2}^{(k)}(k), \ldots, v_{j,D_n(k,0)}^{(k)}(k)\}$.

(ii) Descendants in small intervals: For $0 \leq i \leq n^\delta - 1$ and $v_{j}^{(k)} \in D_n(k, 0)$, we track evolution of descendants of this vertex in the various subintervals. Let $\mathcal{F}_n(i, j)$ denote the set of children born
in the interval \( \left[ \frac{ia}{n^a}, \frac{(i+1)a}{n^a} \right] \) to \( \nu^j(i) \). Let \( N^{(i)}(i, j) := |V^{(i)}(i, j)| \) be the number of such vertices. Write \( N^{(i)}(i) := \sum_{j=1}^{D_n(k,0)} N^{(i)}(i, j) \).

(iii) **Good and bad vertices:** Call a vertex in \( \bar{V}^{(i)}(i, j) \) a **good** vertex if it does **not** give birth to any children by \( \frac{(i+1)a}{n^a} \). Let \( \bar{V}^{(i)}(i, j) \subseteq V^{(i)}(i, j) \) denote the set of good children born in the interval \( \left[ \frac{ia}{n^a}, \frac{(i+1)a}{n^a} \right] \) born to \( \nu^j(i) \). Let \( \bar{N}^{(i)}(i, j) := |\bar{V}^{(i)}(i, j)| \) be the number of such vertices. As above, write \( \bar{N}^{(i)}(i) := \sum_{j=1}^{D_n(k,0)} \bar{N}^{(i)}(i, j) \) be the total number of **good** children born to vertices which originally had degree \( k \) at the change point. Let \( \bar{R}^{(i)}(i, j) := \bar{V}^{(i)}(i, j) \setminus \bar{V}^{(i)}(i, j) \) be the collection of bad children namely those in \( \bar{V}^{(i)}(i, j) \) who have reproduced by time \( \frac{(i+1)a}{n^a} \). Let \( B^{(i)}(i, j) = |\bar{R}^{(i)}(i, j)| \). Let \( \bar{R}^{(i)}(i, j) \) be the set of descendants born in \( \left[ \frac{ia}{n^a}, \frac{(i+1)a}{n^a} \right] \) to vertices in \( \bar{R}^{(i)}(i, j) \) and let \( R^{(i)}(i, j) := |\bar{R}^{(i)}(i, j)| \).

(iv) **Vertices counted by a characteristic:** Let \( Z^{(i,j)}(i, j, x) \) be the number of descendents, satisfying characteristic \( \phi \), at time \( a \), born to \( x \in \bar{V}^{(i)}(i, j) \). Write \( Z^{(i,j)}(i, j, x) = \sum_{i=1}^{D_n(k,0)} \sum_{j=0}^{n^a-1} \sum_{x \in \bar{V}^{(i)}(i, j)} Z^{(i,j,\phi)}(i, j, x) \). Let \( Z^{(i,j,\phi)} = \sum_{i=1}^{D_n(k,0)} \sum_{j=0}^{n^a-1} \sum_{x \in \bar{V}^{(i)}(i, j)} Z^{(i,j,\phi)}(i, j, x) \). Let \( \xi^{(i,j)}(s, t) \) denote the distribution of the number of children born in the time interval \( [s, t] \) to a vertex who had degree \( k \) at time \( 0 \) with attachment function \( f_i \). Write \( \xi^{(i,j)}(t) \) for \( \xi^{(i,j)}(0, t) \).

(v) **Technical conditioning tool:** Define the following \( \sigma \)-field

\[
\mathcal{G}_n = \sigma \{ \mathcal{F}_n(0) \cup \{ \text{life history of } v \in BP_n(0) \text{ till time } a \} \}
\]

\[
\cup \left\{ \text{all vertices born in } \left[ \frac{ia}{n^a}, \frac{(i+1)a}{n^a} \right] \text{ and their life history till time } \left( \frac{j+1}{n^a} \right) \right\}
\]

(vi) **Mean of characteristics emanating from degree \( k \) parent:** Let \( \lambda_k^{(i,j)}(t) = \int_t^s m_{f_i}^{(i)}(t-s) \mu^{(k)}(d s) \) for \( t \leq a \).

For notational simplicity since \( a \) is fixed in this Section, we will write \( \lambda_k^{(i,j)} = \lambda_k^{(i,j)}(a) \).

The following is the main result we prove in this section.

**Theorem 6.1.** Fix any \( \phi \in \mathcal{C} \). There exist deterministic positive constants \( C, C' < \infty \) (not dependent on \( a \)) such that for every \( a > 0 \) and \( n \geq 2 \),

\[
E \left[ Z^{(i,j)} - \sum_{k=0}^{\infty} D_n(k,0) \lambda_k^{(i,j)} \right] \mathcal{F}_n(0) \right] \leq Ce^{C'a} \sqrt{n}.
\]

In particular, as \( n \to \infty \),

\[
\frac{Z^{(i,j)}}{n} \to \gamma \sum_{k=0}^{\infty} D_k^{(i,j)} \lambda_k^{(i,j)}(a).
\]

6.3. **Proof of Theorem 6.1.** We fix a characteristic \( \phi \in \mathcal{C} \) throughout the proof. The main tools in order to prove this result are Lemmas 6.10, 6.11 below. In order to prove these results we will need a number of supporting results which we now embark upon. First we start with a technical lemma controlling the number of children a vertex with degree \( k \) at change point can produce within a fixed interval. For the rest of this section we write \( C_1, C_2, C_3, C_4, C_5, k \) for constants which are independent of \( a, n, \delta, k \).

**Lemma 6.2.** For any interval \( [b, b + \eta] \subseteq [0, a] \),

\[
E \left[ \zeta^{(i,j)}(b, b + \eta) \right] \leq C_1 e^{C_2 a}(k + 1) \eta^2, \quad E \left[ \zeta^{(i,j)}(b, b + \eta)^2 \right] \leq C_3 e^{C_4 a} \{(k + 1)^2 \eta^2 + (k + 1) \eta \}.
\]

**Proof.** By Assumption 6.1(ii), the process \( \{ U(t) := \dot{\xi}^{(i,j)}(t/C) : t \geq 0 \} \) is stochastically dominated by the offspring distribution of a linear preferential attachment (PA) \( \{ P_k(t) : t \geq 0 \} \) point process started at \( k + 1 \),
namely a point process constructed using attachment function \( f^{(k)}(i) = k + 1 + i \) for \( i \geq 0 \) with initial condition \( P_k(0) := 0 \). From the first moment bound in Lemma 5.5 (with \( \nu = 1 \) and \( k = k \)) we find

\[
E(P_k(t)) = (1 + k)(e^t - 1)
\]

(6.2)

We show how to use the first moment of \( P_k(\cdot) \) to obtain the first assertion in the Lemma. The second assertion follows from the same argument using the second moment of \( P_k(\cdot) \) which is also obtained from Lemma 5.5. Conditioning on \( \xi^{(k)}_f(b) \) and using the Markov property we get,

\[
E \xi^{(k)}_{f_1}(b, b + \eta) = \sum_{d=0}^{\infty} P\left( \xi^{(k)}_{f_2}(b) = d \right) E \xi^{(k+d)}_{f_1}(\eta)
\]

(6.3)

Now for any fixed \( k \geq 0 \) and \( t \leq a \), using domination by the corresponding PA process, we get

\[
E[ \xi^{(k)}_{f_1}(t) ] \leq E(P_k(tC)) = e^{tC}(1 + k)(1 - e^{-tC}) \leq e^{Ca}(1 + k)t.
\]

(6.4)

Using this bound twice in (6.3) gives,

\[
E \xi^{(k)}_{f_1}(b, b + \eta) \leq Ce^{Ca} \sum_{d=0}^{\infty} P\left( \xi^{(k)}_{f_1}(b) = d \right) (1 + k + d) = Ce^{Ca} \eta(1 + k + E(\xi^{(k)}_{f_1}(b)))
\]

\[
\leq Ce^{Ca} \eta(1 + k + Cbe^{Ca}(1 + k)) \leq C'e^{Ca}(k + 1)\eta
\]

(6.5)

where \( C', C'' \) are constants that do not depend on \( k, a \). This completes the proof.

Recall that conditional on the initial \( \sigma \)-field \( \mathcal{F}_n(0) \), the random variable \( N^{(k)}_n(i, j) \) is given by \( \xi^{(k)}_f \left[ \frac{i}{\sqrt{n}}, \frac{j}{\sqrt{n}} \right] \). Using Lemma 6.2 now gives the following result.

**Corollary 6.3.** For all \( 1 \leq j \leq D_n(k, 0) \), \( E(N^{(k)}_n(i, j) | \mathcal{F}_n(0)) \leq C_1 e^{C_2 a}(k + 1)n^{-\delta} \) and \( E\left[ N^{(k)}_n(i, j)^2 | \mathcal{F}_n(0) \right] \leq C_3 e^{C_4 a} \left\{ (k + 1)^2 n^{-2\delta} + (k + 1) n^{-\delta} \right\} \).

The next Lemma bounds the number of “bad” vertices and their descendants born within small intervals. For the rest of this section, unless specified otherwise we always work conditional on \( \mathcal{F}_n(0) \) so that expectation operations such as \( E(\cdot) \) and \( \text{Var}(\cdot) \) in the ensuing results mean \( E(\cdot | \mathcal{F}_n(0)) \) and \( \text{Var}(\cdot | \mathcal{F}_n(0)) \).

**Lemma 6.4.** For any \( k, i, j \),

\[
E(R^{(k)}_n(i, j)) \leq C_1 e^{C_2 a} \frac{(k + 1)}{n^{2\delta}}, \quad E\left[ R^{(k)}_n(i, j) \right]^2 \leq C_3 e^{C_4 a} \frac{(k + 1)}{n^{2\delta}} + \frac{(k + 1)^2}{n^{4\delta}}.
\]

**Proof.** For every child \( u \in V^{(k)}_n(i, j) \), write \( BP(; u) \) for the branching process lineage emanating from \( u \). Conditional on \( V^{(k)}_n(i, j) \), using Assumption 2.1[ii] on \( f_1 \), generate a collection of independent rate \( C \) PA branching processes (see Definition 5.6) \{ \( Y_\ell : 1 \leq \ell \leq |V^{(k)}_n(i, j)| \} \) such that \( \text{|BP(; u)} | \leq |Y_\ell(\cdot)| \). Now note that \( X_\ell(t) := Y_\ell(t) - 1 \) is the number of descendants of the root for this branching process by time \( t \).

Using this construction we have the trivial inequality \( R^{(k)}_n(i, j) \leq \sum_{\ell=1}^{N^{(k)}_n(i, j)} X_\ell \left[ 0, \frac{a}{n^\delta} \right] \). This implies

\[
E(R^{(k)}_n(i, j)) \leq E(N^{(k)}_n(i, j)) E\left( X_1 \left[ 0, \frac{a}{n^\delta} \right] \right),
\]

and

\[
E\left[ R^{(k)}_n(i, j) \right]^2 \leq E(N^{(k)}_n(i, j)) E\left( \left[ X_1 \left[ 0, \frac{a}{n^\delta} \right] \right]^2 \right) + E\left( \left[ N^{(k)}_n(i, j) \right]^2 \right) E\left( X_1 \left[ 0, \frac{a}{n^\delta} \right] \right)^2.
\]

Corollary 6.3 for moments of \( N^{(k)}_n(i, j) \) and (5.1) for moments of \( X_1 \left[ 0, \frac{a}{n^\delta} \right] \) completes the proof.

The next Lemma bounds fluctuations of good descendants of degree \( k \) ancestors at the change point counted according to a characteristic.
Lemma 6.5. For any $k \geq 0$, \( \text{Var} \left( Z_{n}^{(k),\phi} \right) \leq Ce^{C\alpha} \left( (k+1)^2n^{-\delta} + (k+1) \right) D_n(k,0) \).

Proof. By construction we have
\[
\text{Var} \left( Z_{n}^{(k),\phi} \right) = \text{Var} \left( \sum_{j=1}^{n^{\alpha} - 1} \sum_{x \in \tilde{Y}_{n}^{(k)}(i,j)} Z_{n}^{(k),\phi}(i,j,x) \right) = D_n(k,0) \text{Var} \left( \sum_{i=0}^{n^{\alpha} - 1} \sum_{x \in \tilde{Y}_{n}^{(k)}(i,1)} Z_{n}^{(k),\phi}(i,1,x) \right). \tag{6.6}
\]

We analyze the variance term on the right by first conditioning on \( \mathcal{G}_n \). First note that,
\[
\mathbb{E} \left[ \text{Var} \left( \sum_{i=0}^{n^{\alpha} - 1} \sum_{x \in \tilde{Y}_{n}^{(k)}(i,1)} Z_{n}^{(k),\phi}(i,1,x) \Big| \mathcal{G}_n \right) \right] = \mathbb{E} \left[ \sum_{i=0}^{n^{\alpha} - 1} \tilde{N}_{n}^{(k)}(i,1) m_{fi}^{\phi}(a - \frac{(i+1)a}{n^{\delta}}) \right] \leq C_{1} e^{C_{2} \alpha}(k+1)n^{-\delta} \mathbb{E} \left( Z_{fi}^{2} \right) (a) \leq C e^{C\alpha}(k+1) \tag{6.7}
\]
where \( C, C' \) do not depend on \( k, a, n, \delta \). The first equality comes from noting \( \tilde{Z}_{n}^{(k)}(i,1) \) is \( \mathcal{G}_n \)-measurable, the collection \( \{ Z_{n}^{(k),\phi}(i,1,x) \mid x \in \tilde{Y}_{n}^{(k)}(i,1), 1 \leq i \leq n^{\alpha} - 1 \} \) are independent and further for each \( 0 \leq i \leq n^{\alpha} - 1 \) and \( x \in \tilde{Y}_{n}^{(k)}(i,1) \), \( Z_{n}^{(k),\phi}(i,1,x) \) is distributed as \( Z_{fi}^{\phi}(a - \frac{(i+1)a}{n^{\delta}}) \), since \( x \) has no children by time \( \frac{(i+1)a}{n^{\delta}} \).

The second inequality follows by using Corollary 6.3 for \( \tilde{N}_{n}^{(k)}(i,1) \) and (6.1). Similarly
\[
\text{Var} \left( \mathbb{E} \left( \sum_{i=0}^{n^{\alpha} - 1} \sum_{x \in \tilde{Y}_{n}^{(k)}(i,1)} Z_{n}^{(k),\phi}(i,1,x) \Big| \mathcal{G}_n \right) \right) = \text{Var} \left( \sum_{i=0}^{n^{\alpha} - 1} \tilde{N}_{n}^{(k)}(i,1) m_{fi}^{\phi}(a - \frac{(i+1)a}{n^{\delta}}) \right) \leq 4c^{2} \mathbb{E} \left( Z_{fi}^{2} \right) (a)^{2} \sum_{i=0}^{n^{\alpha} - 1} \mathbb{E} \left( \tilde{N}_{n}^{(k)}(i,1)^{2} \right) \leq C e^{C\alpha} \left( (k+1)^2n^{-\delta} + (k+1) \right) \tag{6.8}
\]
where \( C, C' \) do not depend on \( k, a, n, \delta \). Here we use Corollary 6.3 in the second inequality. Using (6.7) and (6.8) to bound the variance term in the right of (6.6) completes the proof.

The next Lemma provides tight bounds on expectations of descendants of good vertices counted according to \( \phi \). Recall \( \mu_{i}^{(k)} \) denotes the mean measure of a vertex which had degree \( k \) at the change point.

Lemma 6.6. For any \( k \geq 0 \),
\[
\epsilon_{n} := \mathbb{E} \left[ \tilde{Z}_{n}^{(k),\phi} \right] - D_n(k,0) \sum_{i=0}^{n^{\alpha} - 1} m_{fi}^{\phi} \left( a - \frac{(i+1)a}{n^{\delta}} \right) \mu_{i}^{(k)} \left[ \frac{ia}{n^{\delta}}, \frac{(i+1)a}{n^{\delta}} \right] \leq C e^{C\alpha} \frac{(k+1)D_n(k,0)}{n^{\delta}}.
\]

Proof. First note,
\[
\mathbb{E} \left[ \tilde{Z}_{n}^{(k),\phi} \right] = \sum_{i=0}^{n^{\alpha} - 1} D_n(k,0) \sum_{j=1}^{n^{\alpha} - 1} \mathbb{E} \left[ \sum_{x \in \tilde{Y}_{n}^{(k)}(i,j)} Z_{n}^{(k),\phi}(i,j,x) \Big| \mathcal{G}_n \right] \]
\[
= \sum_{i=0}^{n^{\alpha} - 1} D_n(k,0) \mathbb{E} \left[ \sum_{x \in \tilde{Y}_{n}^{(k)}(i,1)} Z_{n}^{(k),\phi}(i,1,x) \Big| \mathcal{G}_n \right] \]
\[
= D_n(k,0) \sum_{i=0}^{n^{\alpha} - 1} m_{fi}^{\phi} \left( a - \frac{(i+1)a}{n^{\delta}} \right) \mathbb{E} \left[ \tilde{N}_{n}^{(k)}(i,1) \right].
\]

Here the third equality follows from \( \tilde{Y}_{n}^{(k)}(i,1) \) is \( \mathcal{G}_n \)-measurable and for fixed \( i \), and for each \( x \in \tilde{Y}_{n}^{(k)}(i,1) \), conditional on \( \mathcal{G}_n \), \( Z_{n}^{(k),\phi}(i,j,x) \overset{d}{=} Z_{fi}^{\phi}(a - \frac{(i+1)a}{n^{\delta}}) \). Applying equation (6.1), the error term \( \epsilon_{n} \) in the statement of the Lemma can be bounded as,
\[
\epsilon_{n} \leq 2cD_n(k,0) m_{fi}^{\phi}(a) \sum_{i=0}^{n^{\alpha} - 1} \mathbb{E} \left[ \tilde{N}_{n}^{(k)}(i,1) - \tilde{N}_{n}^{(k)}(i,1) \right]. \tag{6.9}
\]
Next using that the total number of descendants of bad vertices in an interval bounds the number of bad vertices in this interval since each bad vertex has at least one child, we get using Lemma 6.4

\[ 0 \leq \mathbb{E} \left[ N_n^{(k)}(i, 1) - \tilde{N}_n^{(k)}(i, 1) \right] = \mathbb{E}[B_n^{(k)}(i, 1)] \leq \mathbb{E}[\tilde{B}_n^{(k)}(i, j)] \leq C_1 e^{C_2 a (k + 1) / n^{2\delta}}. \]

Using this and (6.1) in (6.9) completes the proof.

Lemma 6.7. There exist a constant \( a_0 < \infty \) independent of \( n, \delta \) such that for any \( k \geq 0 \), whenever \( a \leq \frac{\delta}{a_0} \log n \),

\[ \mathbb{E} \left[ Z_n^{(k),\phi} - \bar{Z}_n^{(k),\phi} \right] \leq C e^{C_0 k \delta} n^{-\delta (k + 1)} D_n(k, 0) \]

Proof.

\[ \mathbb{E} \left[ Z_n^{(k),\phi} - \bar{Z}_n^{(k),\phi} \right] \leq \mathbb{E} \left[ \sum_{j=1}^{D_n(k, 0)} \sum_{x \in \mathcal{V}_n^{(i, j)}} Z_n^{(k),\phi}(i, j, x) \mathbb{1} \{ B_x \} \right] = D_n(k, 0) \sum_{i=0}^{n-k-1} \mathbb{E} \left[ \sum_{x \in \mathcal{V}_n^{(i, 1)}} Z_n^{(k),\phi}(i, 1, x) \mathbb{1} \{ B_x \} \right], \quad (6.10) \]

where \( B_x \) is the event that a vertex is bad namely has one or more descendants in the interval that it was born. Now note that for a fixed \( i \), conditional on the number of births \( N_n^{(i)}(i, 1) \), we have

\[ \sum_{x \in \mathcal{V}_n^{(i, 1)}} Z_n^{(k),\phi}(i, 1, x) \leq \sum_{i=1}^{N_n^{(i)}(i, 1)} \sum_{j=1}^{N_n^{(k)}(i, 1)} 2c |\text{PA}_j[i, a]| \mathbb{1} \{ B_i \}, \quad (6.11) \]

where \( \{ \text{PA}_j[i, a] : l \geq 1 \} \) is a collection of PA branching processes with parameters \( \nu = C \) and \( \kappa = 0 \) (independent of \( N_n^{(k)}(i, 1) \)) and

\[ \bar{B}_i := \left\{ \left| \text{PA}_j[i, a] \right| \geq 2, \right\}, \]

namely the root of \( \text{PA}_j[i, a] \) has at least one child by time \( a / n^\delta \). Using this in (6.10) implies,

\[ \mathbb{E} \left[ Z_n^{(k),\phi} - \bar{Z}_n^{(k),\phi} \right] \leq 2c D_n(k, 0) \sum_{i=1}^{n-k-1} \mathbb{E}(N_n^{(k)}(i, 1)) \mathbb{E}(\text{PA}_j[i, a] \mathbb{1} \{ \bar{B}_i \}). \quad (6.12) \]

Conditioning on the number of births \( Y(a / n^\delta) \) of the root of \( \text{PA}_j[i, a] \) in \( [0, a / n^\delta] \) and using the Markov property,

\[ \mathbb{E}(\text{PA}_j[i, a] \mathbb{1} \{ \bar{B}_i \}) \leq \sum_{j=1}^\infty \mathbb{P} \left( Y \left( \frac{a}{n^\delta} \right) = j \right) \mathbb{E}(\text{PA}_j[i, a]), \]

where \( \text{PA}_j[i, a] \) is a modified PA process with \( \nu = C, \kappa = 0 \) with the modification that the offspring distribution of the root of \( \text{PA}_j[i, a] \) is constructed using attachment function \( f(i) := C(j + i + 1) \) for \( i \geq 0 \). Comparing rates, it is easy to see that for each \( j \geq 1 \), \( \text{PA}_j[i, a] \leq_{st} U_j(a) \), where \( U_j(a) \) is constructed by first running a PA processes \( \text{PA}_{\nu, \kappa} \) with \( \nu = C \) and \( \kappa = C j \) and then setting \( U_j(a) = |\text{PA}_{\nu, \kappa}[0, a]| \). By Lemma 5.4 for \( Y(a / n^\delta) \) and Proposition 5.7 for \( \mathbb{E}(U_j(a)) \), we get \( a_0 > 0 \) such that whenever \( a \leq \frac{\delta}{a_0} \log n \),

\[ \mathbb{E}(\text{PA}_j[i, a] \mathbb{1} \{ \bar{B}_i \}) \leq \sum_{j=1}^\infty \left( \frac{C a}{n^\delta} \right)^j e^{a(2C + C) j} \leq C e^{C a n^{-\delta}} \quad (6.13) \]

where \( C, C' \) do not depend on \( k, a, n, \delta \). In (6.12), using this bound and using Corollary 6.3 for \( \mathbb{E}(N_n^{(k)}(i, 1)) \) completes the proof.
Lemma 6.8. For any $k \geq 0$, whenever $a \leq \frac{a}{a_0} \log n$,
\[
\omega_n := \mathbb{E} \left[ Z_n^\phi - \sum_{k=0}^\infty D_n(k,0) \sum_{i=0}^{n^\delta-1} m_f^\phi \left( a - \frac{(i+1)a}{n} \right) \mu_f \left[ n \right] \right] \leq C e^{C^a} \left( n^{1-\delta} + \sqrt{n} + n^{-\delta/2} \left( \sum_{k=1}^\infty (k+1)^2 D_n(k,0) \right)^{1/2} \right).
\]

Proof. The term above can be written as $\omega_n := \omega_n^{(1)} + \omega_n^{(2)} + \omega_n^{(3)}$, where $\omega_n^{(1)} := Z_n^\phi - \bar{Z}_n^\phi$, $\omega_n^{(2)} := \bar{Z}_n^\phi - \mathbb{E}(\bar{Z}_n^\phi)$ and $\omega_n^{(3)} := \mathbb{E}(\bar{Z}_n^\phi) - \sum_{k=0}^\infty D_n(k,0) \sum_{i=0}^{n^\delta-1} m_f^\phi \left( a - \frac{(i+1)a}{n} \right) \mu_f \left[ n \right]$. Now fix $\epsilon > 0$. Using Lemma 6.7 we get,
\[
\mathbb{E}(\omega_n^{(1)}) \leq \frac{Ce^{C^a}}{n^\delta} \sum_{k=0}^\infty (k+1) D_n(k,0) \leq 2\gamma C e^{C^a} n^{1-\delta},
\]

since $\sum_{k=1}^\infty (k+1) D_n(k,0) = 2\gamma n - 1$ for tree $\mathcal{F}_n$. Next using Lemma 6.5 and Jensen’s inequality,
\[
\mathbb{E}(\omega_n^{(2)}) \leq Ce^{C^a} \left( \sum_{k=1}^\infty (k+1)^2 D_n(k,0) \right)^{1/2} \leq C e^{C^a} \left( n^{-\delta/2} \left( \sum_{k=1}^\infty (k+1)^2 D_n(k,0) \right)^{1/2} + \sqrt{n} \right).
\]

Finally using Lemma 6.6 gives,
\[
|\omega_n^{(3)}| \leq Ce^{C^a} \sum_{k=0}^\infty \frac{(k+1)D_n(k,0)}{n^\delta} \leq C e^{C^a} n^{1-\delta}.
\]

Combining (6.14), (6.15) and (6.16) completes the proof.

The next lemma establishes Lipschitz continuity of $m_f^\phi(t)$ in $t$ for any $\phi \in \mathcal{C}$.

Lemma 6.9. For any $k \geq 0$ and any $\eta \in [0,1]$,
\[
\sup_{t \in [0,a]} |m_f^\phi(t+\eta) - m_f^\phi(t)| \leq C e^{C^a} \eta.
\]

Proof. Let $\tau_1$ be the time of the first birth for the branching process with attachment function $f_1$. For any $t \in [0,a]$ and $\eta \in [0,1]$, using the Markov property at time $\eta$, we obtain
\[
m_f^\phi(t+\eta) = \mathbb{E} \left[ Z_f^\phi(t+\eta) \right] = \mathbb{E} \left[ Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 > \eta) + Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 \leq \eta) \right] = \mathbb{E} \left[ Z_f^\phi(t) \right] + \mathbb{E} \left[ Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 \leq \eta) \right] = m_f^\phi(t)(1 - \mathbb{P}(\tau_1 > \eta)) + \mathbb{E} \left[ Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 = \eta) \right].
\]

Using the strong Markov property at $\tau_1$, we can write the second term above as $\mathbb{E} \left[ Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 \leq \eta) \right] = \mathbb{E} \left[ Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 \leq \eta) \right]$, where $\mathcal{F}_{\tau_1}$ denotes the associated stopped sigma field. Now that at time $\tau_1$, there are two vertices, one with out-degree one and the other with out-degree zero. Thus, conditional on $\mathcal{F}_{\tau_1}$, for $i = 1,2$, if $U_i(t)$ is distributed as the size of the PA process $PA_{\nu,\kappa_i}$ with $\nu = C$ and $\kappa_i = C(i-1)$ at time $t$ (where $C$ is the same constant appearing in Assumption 2.1(ii)), we have
\[
\mathbb{E} \left[ Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 \leq \eta) \right] \leq 2c \mathbb{E} \left( Z_f^\phi(t+\eta) \mathbb{I}(\tau_1 \leq \eta) \right) \leq 2c \mathbb{E}(U_1(a+1) + U_2(a+1)) \leq Ce^{C^a}
\]
Lemma 6.10. For any \( k \geq 0 \), whenever \( a \leq \frac{\delta}{a_0} \log n \),
\[
\mathbb{E} \left[ Z_n^\phi - \sum_{k=0}^{\infty} D_n(k,0) \lambda_k^\phi \right] \leq C e^{C' a} \left( n^{1-\delta} + \sqrt{n} + n^{-\delta/2} \left( \sum_{k=1}^{\infty} (k+1)^2 D_n(k,0) \right) \right)^{1/2}.
\]

Proof. Owing to Lemma 6.8, it is enough to show, for a positive constants \( C, C' \) not depending on \( a, n, \delta \) such that
\[
\omega_n^* := \sum_{k=0}^{\infty} D_n(k,0) \lambda_k^\phi - \sum_{k=0}^{\infty} D_n(k,0) \sum_{i=0}^{n-1} m_{fi}^\phi \left( a - \frac{(i+1)a}{n} \right) \mu^{(k)}_{fi} \left( \frac{ia}{n^\delta}, \frac{(i+1)a}{n^\delta} \right) \leq C e^{C' a} n^{1-\delta}. \tag{6.18}
\]
Using Lemma 6.9,
\[
\omega_n^* \leq \sum_{k=0}^{\infty} D_n(k,0) \int_0^{n^{-\delta}} \sum_{i=0}^{n-1} m_{fi}^\phi (a-s) - m_{fi}^\phi \left( a - \frac{(i+1)a}{n} \right) \left( s \in \left[ \frac{ia}{n^\delta}, \frac{(i+1)a}{n^\delta} \right] \right) \mu^{(k)}_{fi} (ds) \\
\leq C e^{C' a} n^{-\delta} \sum_{k=0}^{\infty} D_n(k,0) \int_0^{n^{-\delta}} \sum_{i=0}^{n-1} \left( s \in \left[ \frac{ia}{n^\delta}, \frac{(i+1)a}{n^\delta} \right] \right) \mu^{(k)}_{fi} (ds) = C e^{C' a} n^{-\delta} \sum_{k=0}^{\infty} D_n(k,0) \mu^{(k)}_{fi} [0,a] \\
\leq (C e^{C' a})^2 a n^{-\delta} \sum_{k=0}^{\infty} (k+1) D_n(k,0) = (C e^{C' a})^2 a n^{-\delta} (2\gamma n - 1)
\]
where the last inequality comes from Lemma 6.2 and the last equality uses \( \sum_{k=0}^{\infty} (k+1) D_n(k,0) = 2\gamma n - 1 \).

Lemma 6.11. Let \( \phi \in \mathcal{F} \) then \( n \to \infty \),
\[
\frac{\sum_{k=1}^{\infty} D_n(k,0)}{n} \lambda_k^\phi (a) \xrightarrow{a.s.} \gamma \sum_{k=1}^{\infty} p_k^0 \lambda_k^\phi (a).
\]

Proof. Let \( \chi \) be the characteristic \( \chi(t) = \sum_{k=0}^{\infty} \lambda_k^\phi (a) \mathbb{1} \{ \xi_j(t) = k \} \). Note by equation (6.1) and Lemma 6.2 \( \lambda_k^\phi (a) \leq C e^{C' a} (k+1) \) thus \( \chi \in \mathcal{C} \). Now apply Lemma 6.8 (i).

Completing the proof of Theorem 6.1: By letting \( \delta \to \infty \) keeping \( n \geq 2 \) fixed in Lemma 6.10 the first claim follows. Lemma 6.11 then gives the second claim.

7. Proofs: Sup-norm convergence of degree distribution for the standard model

7.1. Proof of Theorems 3.6 and 3.9: In this section, we will prove a convergence result for the empirical degree distribution post change-point. As before, we start time at the change point, i.e. \( t = 0 \) represents the time \( T_{\gamma n} \). Focus will be on the characteristic \( \phi(t) = \mathbb{1} \{ \xi_j(t) = k \} \) for \( k \geq 0 \) and we will denote the corresponding \( Z_n^\phi \) and \( m_{fi}^\phi \) by \( Z_n^{(k)} \) and \( m_{fi}^{(k)} \) respectively. \( B_P(n,t) \) will denote the branching process at time \( t \) (i.e. \( t \) units after the change point).
7.1.1. **Notation.** We will use the following notation for fixed $t \geq 0$ in this section.

(i) Recall that $n\gamma$ are the number of vertices born **before** the change point. Let $Z_{AC,n}(t) = \text{number of vertices at time } t \text{ who were born } \textbf{after} \text{ the change point. Let } Z_n(t) = n\gamma + Z_{AC,n}(t)$ be the total number of vertices at time $t$.

(ii) Let $\mathcal{D}_n^{BC}(k,t)$ be the set of vertices with degree $k$ at time $t$ who were born **before** the change point $T_{\gamma,n}$. Let $D_n^{BC}(k,t) = |\mathcal{D}_n^{BC}(k,t)|$. Similarly, let $\mathcal{D}_n^{AC}(k,t)$ be the set of vertices with degree $k$ at time $t$ who were born **after** the change point. Let $D_n^{AC}(k,t) = D_n^{BC}(k,t) + D_n^{AC}(k,t)$. Let $D_n(k,t) = D_n^{BC}(k,t) + D_n^{AC}(k,t)$ be the total number of vertices with degree $k$.

(iii) Let $\lambda_{\ell}^{AC}(t) = \int_{0}^{t} m_{\ell}(t-s)\mu_{\ell}^{(k)}(ds)$ and $\lambda_{\ell}^{AC,(k)}(t) = \int_{0}^{t} m_{\ell}^{(k)}(t-s)\mu_{\ell}^{(k)}(ds)$. Let $\lambda_{\ell}(t) = 1 + \lambda_{\ell}^{AC}(t)$ and $\lambda_{\ell}^{(k)}(t) = P(\xi_{\ell}^{(k)}(t) = k - \ell) + \lambda_{\ell}^{AC,(k)}(t)$.

(iv) Let $q_k(t) := P(\xi_{\ell}^{(k)}(t) > 1)$.

The following is the main theorem proven in this Section.

**Theorem 7.1.** For any $k \geq 0$, $a > 0$, $\epsilon > 0$,

$$P\left(\sup_{t \in [0,a]} \left| D_n(k,t) - n \sum_{\ell=0}^{\infty} \gamma p_{\ell}^{0} \lambda_{\ell}^{(k)}(t) \right| > \epsilon n \right) \rightarrow 0$$

and

$$P\left(\sup_{t \in [0,a]} \left| Z_n(t) - n \sum_{\ell=0}^{\infty} \gamma p_{\ell}^{0} \lambda_{\ell}(t) \right| > \epsilon n \right) \rightarrow 0.$$

Assuming the above result for the time being, we now describe how this (coupled with a technical continuity result, Lemma 7.4) is now enough to prove Theorems 3.6 and 3.9. Recall for $m \geq 1$, $T_m = \inf\{t \geq 0 : BP_n(t) = m\}$.

**Corollary 7.2.** Let $G(t) := \sum_{\ell=0}^{\infty} P_0^{0} \lambda_{\ell}^{AC}(t)$. For any $s \in [\gamma,1]$, let $a_s$ be the unique solution to $G(a_s) = \frac{s - \gamma}{t}$. Then $n \rightarrow \infty$, $\sup_{t \in [\gamma,\gamma]} |T_{[sn]} - a_s| \xrightarrow{P} 0$.

**Proof.** As $f_1$ is a strictly positive function, it is easy to see that $G(t)$ is strictly increasing in $t$ and $G(\gamma) = 0$. By Lemma 7.4 proved below, $G$ (and hence $G^{-1}$) is continuous. Moreover since $m_{\ell}(t) \geq 0$, $\lambda_{\ell}^{AC}(t) \geq \mu_{\ell}^{(t)}(t) \uparrow \infty$ and we see $G(t) \rightarrow \infty$ as $t \rightarrow \infty$. Therefore $G(a_s) = \frac{s - \gamma}{t}$ has a unique solution for $s \in [\gamma,1]$.

Next fix $s \in [\gamma,1]$ and let $a_s$ be as above. For any $\eta > 0$, choosing $\epsilon = \frac{G(a_s + \eta) - G(a_s)}{2\eta}$, the second assertion in Theorem 7.1 readily implies $P(Z_n(a_s + \eta) > sn + 1) \rightarrow 1$. Similarly, it follows that $P(Z_n(a_s - \eta) < sn - 1) \rightarrow 1$. Therefore, $T_{[sn]} \xrightarrow{P} a_s$. From this and Theorem 7.1 we conclude that

$$\frac{1}{n} \sup_{t \in [\gamma,\gamma]} \left| Z_n(t) - \gamma n (1 + G(t)) \right| \xrightarrow{P} 0$$

which implies

$$\sup_{t \in [\gamma,\gamma]} \left| \frac{t - \gamma}{\gamma} - G(T_{[sn]}) \right| \xrightarrow{P} 0.$$

By continuity of $G^{-1}$, this implies

$$\sup_{t \in [\gamma,\gamma]} \left| G^{-1} \left( \frac{t - \gamma}{\gamma} \right) - T_{[sn]} \right| \xrightarrow{P} 0$$

which proves the corollary.

**Proof of Theorem 3.6** Fix $s \in [\gamma,1]$. It follows from Lemma 7.4 and Corollary 7.6 proved below that $t \mapsto \Phi_{\ell}(\mathbf{p}^0)$ is continuous and hence, from Corollary 7.2 for each fixed $k \geq 0$,

$$\sup_{t \in [\gamma,\gamma]} \left| (\Phi_{T_{[sn]}(\mathbf{p}^0))}_k - (\Phi_{n}(\mathbf{p}^0))_k \right| \xrightarrow{P} 0. \quad (7.1)$$
Proof. We prove this inequality in two steps. First note Lemma 7.5. For \( k \geq \ell \) and \( t \), \( |\lambda_{\ell}(t + s) - \lambda_{\ell}(t)| \leq C e^{C_{\ell}(\ell + 1)s} \), \( |\lambda_{\ell}(t + s) - \lambda_{\ell}(t)| \leq C e^{C_{\ell}(\ell + 1)s} \). The second inequality uses Lemma 6.9 and the third inequality uses Lemma 6.2 and (6.1).

We will only prove the first inequality. The second one follows similarly. Lemma 7.3.

\[
q_{k}(t) \leq C(k + 1) t
\]

where \( C \) is the constant appearing in Assumption 2.1(ii) on \( f_{1} \).

Proof. Let \( \tau_{t}^{k} \) be the time of the first born to a vertex started with degree \( k \). Note \( \tau_{t}^{k} \sim \text{Exp}(f_{1}(k)) \). Thus

\[
\mathbb{P}(\tau_{t}^{k} < t) = 1 - e^{-f_{1}(k)t} \leq f_{1}(k)t \leq C(k + 1)t
\]

where the final inequality comes from Assumption 2.1(ii) on \( f_{1} \).

Lemma 7.4. For any \( \ell, k \geq 0 \) and \( t, t + s \leq a \),

\[
|\lambda_{\ell}(t + s) - \lambda_{\ell}(t)| \leq C e^{C_{\ell}(\ell + 1)s}, \quad |\lambda_{\ell}(t + s) - \lambda_{\ell}(t)| \leq C e^{C_{\ell}(\ell + 1)s}.
\]

Proof. We will only prove the first inequality. The second one follows similarly.

\[
|\lambda_{\ell}(t + s) - \lambda_{\ell}(t)| \leq \int_{0}^{t} m_{f_{1}}(t + s - x) - m_{f_{1}}(t - x) \mu_{f_{1}}(dx) + \int_{t}^{t+s} m_{f_{1}}(t + s - x) \mu_{f_{1}}(dx)
\]

where the second inequality uses Lemma 6.9 and the third inequality uses Lemma 6.2 and (6.1).

Lemma 7.5. For \( k \geq l \) and \( t, t + s \leq a \),

\[
|\mathbb{P}(\xi_{f_{1}}^{(\ell)}(t + s) = k - \ell) - \mathbb{P}(\xi_{f_{1}}^{(\ell)}(t) = k - \ell)| \leq C e^{C_{a}(k + 1)s}.
\]

Proof. We prove this inequality in two steps. First note
\[
\mathbb{P}\left(\ell_{f_1}^\ell(t+s) = k - \ell\right) = \sum_{d=0}^{k-\ell} \mathbb{P}\left(\ell_{f_1}^\ell(t) = d\right) \mathbb{P}\left(\ell_{f_1}^{d+\ell}(s) = k - \ell - d\right) \\
= \sum_{d=0}^{k-\ell-1} \mathbb{P}\left(\ell_{f_1}^\ell(t) = d\right) \mathbb{P}\left(\ell_{f_1}^{d+\ell}(s) = k - \ell - d\right) + \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \mathbb{P}\left(\ell_{f_1}^{k}(s) = 0\right) \\
\leq \sum_{d=0}^{k-\ell-1} \mathbb{P}\left(\ell_{f_1}^\ell(t) = d\right) \mathbb{P}\left(\ell_{f_1}^{d+\ell}(s) \geq 1\right) + \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \\
\leq \sum_{d=0}^{k-\ell-1} Ce^{C\alpha}(d + \ell + 1)s \mathbb{P}\left(\ell_{f_1}^\ell(t) = d\right) + \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \\
\leq Ce^{C\alpha}s \left[ \mathbb{E}\left(\ell_{f_1}^\ell(t)\right) + \ell + 1\right] + \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \\
\leq C'^s e^{2C\alpha}(\ell + 1)s + \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right).
\]

The first equality comes from the Markov property. The second inequality comes from Markov’s inequality. The third and fifth inequalities use Lemma 6.2. We now show the opposite inequality.

\[
\mathbb{P}\left(\ell_{f_1}^\ell(t+s) = k - \ell\right) \geq \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \mathbb{P}\left(\ell_{f_1}^k(s) = 0\right) = \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \left[1 - \mathbb{P}\left(\ell_{f_1}^k(s) \geq 1\right)\right]
\]

Thus

\[
\mathbb{P}\left(\ell_{f_1}^\ell(t+s) = k - \ell\right) - \mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \geq -\mathbb{P}\left(\ell_{f_1}^\ell(t) = k - \ell\right) \mathbb{P}\left(\ell_{f_1}^k(s) \geq 1\right) \\
\geq -\mathbb{E}\left(\ell_{f_1}^k(s)\right) \geq -Ce^{C\alpha}(k+1)s
\]

where the second inequality comes from Markov’s inequality and the last inequality comes from Lemma 6.2.

An immediate consequence of Lemmas 7.4 and 7.5 is

**Corollary 7.6.** For any \(k, \ell > 0\) and \(t, t+s < a\),

\[|\lambda_{\ell}^{(k)}(t+s) - \lambda_{\ell}^{(k)}(t)| \leq Ce^{C\alpha}(k + \ell + 2)s.\]

**Corollary 7.7.** For any \(k\) and \(t, t+s < a\),

\[\sum_{\ell=0}^{\infty} D_n(\ell, 0)|\lambda_{\ell}^{(k)}(t+s) - \lambda_{\ell}^{(k)}(t)| \leq Ce^{C\alpha}(k + 3)sn.\]

**Proof.** By the above Corollary 7.6 (with \(k\) fixed)

\[\sum_{\ell=0}^{\infty} D_n(\ell, 0)|\lambda_{\ell}^{(k)}(t) - \lambda_{\ell}^{(k)}(t+s)| \leq Ce^{C\alpha}s \sum_{\ell=0}^{\infty} (k + \ell + 2)D_n(\ell, 0) \leq Ce^{C\alpha}(k + 3)s\gamma n\]

since \(\sum_{\ell=0}^{\infty} D_n(\ell, 0) = \gamma n\) and \(\sum_{\ell=0}^{\infty} \ell D_n(\ell, 0) = \gamma n - 1\).

For the rest of this section, unless specified otherwise, we always work conditional on \(\mathcal{F}_n(0)\) so that expectation operations such as \(\mathbb{P}(\cdot), \mathbb{E}(\cdot)\) and \(\text{Var}(\cdot)\) in the ensuing results mean \(\mathbb{P}(\cdot|\mathcal{F}_n(0)), \mathbb{E}(\cdot|\mathcal{F}_n(0))\) and \(\text{Var}(\cdot|\mathcal{F}_n(0))\) respectively.

We will use Theorem 6.1 crucially in what follows for two significant characteristics. Taking \(\phi(t) = \mathbb{1} \{t \geq 0\}\) in Theorem 6.1 there exist deterministic positive constants \(C, C' < \infty\) independent of \(a, n\) such that for every \(n \geq 2\),

\[\sup_{t\in[0,a]} \mathbb{E}\left|Z_{\lambda, n}(t) - \sum_{k=0}^{\infty} D_n(k, 0)\lambda_{\ell}^{AC}(t)\right| < Ce^{C\alpha}\sqrt{n}.\]
Taking any \( k \geq 0 \) and setting \( \phi(t) = 1 \{ \xi_j(t) = k \} \) in Theorem 6.1, there exist deterministic positive constants \( C, C' < \infty \) independent of \( a, n, k \) such that for every \( n \geq 2 \),

\[
\sup_{t \in [0, a]} \mathbb{E} \left| D_n^{AC}(k, t) - \sum_{\ell=0}^{\infty} D_n(\ell, 0)A_{\ell}^{AC,k}(t) \right| < Ce^{C'a}\sqrt{n}. \tag{7.4}
\]

Take any \( \tilde{\theta} \in (0, 1/2) \). Take \( \omega \in (0, 1) \) such that \( \omega > \max\{1 - \tilde{\theta}, 1/2 + \tilde{\theta}\} \).

Now let \( \{t_i\}_{i=0}^{n^{\tilde{\theta}} - 1} \) be an equispaced partition of \([0, a]\) of mesh \( an^{-\tilde{\theta}} \).

**Lemma 7.8.** Let \( \{t_j\}, \tilde{\theta} \) and \( \omega \) be as above. Fix \( \epsilon \in (0, 1) \) and \( k \). Then we have

\[
\sum_{j=0}^{n^{\tilde{\theta}} - 1} \mathbb{P} \left( \sup_{t \in [t_j, t_{j+1}]} |D_n(k, t) - D_n(k, t_j)| > \epsilon n^\omega \right) \leq \frac{Ce^{C'a}}{\epsilon^2} \frac{1}{n^{\omega - \tilde{\theta} - 1/2}}.
\]

**Proof.** Condition on \( \mathcal{F}_n(t_j) \). Fix \( j \) and consider \( t \in [t_j, t_{j+1}] \). We clearly have the following lower bound on \( D_n(k, t) \):

\[
D_n(k, t) \geq D_n(k, t_j) - Y_1
\]

where \( Y_1 \) is the number of degree \( k \) vertices at time \( t_j \) which have given birth by time \( t_{j+1} \). Note that

\[
Y_1 \overset{d}{=} \text{Bin}\left(D_n(k, t_j), q_k(an^{-\tilde{\theta}})\right).
\]

We also have the following upper bound on \( D_n(k, t) \):

\[
D_n(k, t) \leq (Z_{AC,n}(t_{j+1}) - Z_{AC,n}(t_j)) + Y_2 + D_n(k, t_j)
\]

where \( Y_2 \) denotes the number of vertices existing at time \( t_j \) of degree less than \( k \) which have given birth by time \( t_{j+1} \). Note that

\[
Y_2 \overset{d}{=} \sum_{\ell=0}^{k-1} \text{Bin}\left(D_n(\ell, t_j), q_\ell\left(an^{-\tilde{\theta}}\right)\right).
\]

To see this upper bound, note that the degree \( k \) vertices at time \( t \) originate from vertices either existing at time \( t_j \) or new vertices born in the time interval \([t_j, t]\). The latter is bounded by \( Z_{AC,n}(t_{j+1}) - Z_{AC,n}(t_j) \), namely, the total number of new births in the time interval \([t_j, t_{j+1}]\). The former is bounded by the sum of the number of vertices which are of degree \( k \) at time \( t_j \) and have not given birth by time \( t \) (which, in turn, is bounded by \( D_n(k, t_j) \)) and the number of vertices of lower degree at time \( t_j \) which have grown to degree \( k \) at time \( t \) (which, in turn, is bounded by \( Y_2 \)).

These two bounds give the following

\[
|D_n(k, t) - D_n(k, t_j)| \leq (Z_{AC,n}(t_{j+1}) - Z_{AC,n}(t_j)) + Y_1 + Y_2.
\]

Note the right hand side does not depend on \( t \). We now have for all \( 0 \leq j \leq n^{\tilde{\theta}} - 1 \) and \( t \in [t_j, t_{j+1}] \),

\[
\sup_{j \leq n^{\tilde{\theta}} - 1} \mathbb{P} \left( \sup_{t \in [t_j, t_{j+1}]} |D_n(k, t) - D_n(k, t_j)| > \epsilon n^\omega \right)
\]

\[
\leq \sup_{j \leq n^{\tilde{\theta}} - 1} \left\{ \mathbb{P} \left( \sum_{\ell=0}^{k} \text{Bin}\left(D_n(\ell, t_j), q_\ell\left(an^{-\tilde{\theta}}\right)\right) > \frac{\epsilon}{2} n^\omega \right) + \mathbb{P} \left( |Z_{AC,n}(t_{j+1}) - Z_{AC,n}(t_j)| > \frac{\epsilon}{2} n^\omega \right) \right\}
\]

\[
\leq \frac{Ce^{C'a}}{\epsilon^2} \frac{1}{n^{\omega - \tilde{\theta} - 1/2}} + \frac{Ce^{C'a}}{\epsilon} \frac{1}{n^{\omega - 1/2}}
\]

where the second inequality comes from Lemmas 7.9 and 7.10 which are proved below. The result now follows after taking the sum of these terms.

\[\blacksquare\]
Lemma 7.9. Let \( \{t_j, \tilde{\theta} \text{ and } \omega \} \) be as above and let \( \varepsilon \in (0, 1) \). Then there exist constants \( C', n_0 \) such that for all \( n \geq n_0 \) and all \( \alpha \leq C' \log n \),

\[
\sup_{j \leq n^\alpha} \mathbb{P} \left( \sum_{\ell=0}^{k} \text{Bin} \left( D_n(\ell, t_j), q_\ell \left( an^{-\tilde{\theta}} \right) \right) > \frac{\varepsilon}{2} n^\omega \right) \leq \frac{Ce^{C'a}}{\varepsilon^2 n^{\tilde{\theta} + \omega - \frac{1}{2}}}. 
\]

Proof. Define the event \( A_j = \{ Z_n(t_j) < \left( \gamma + \frac{\varepsilon}{8} \right) n^{\tilde{\theta} + \omega} \} \). Note that as \( \sum_{\ell=0}^{\infty} (\ell+1)D_n(\ell, t_j) = 2Z_n(t_j) - 1 \), therefore on the event \( A_j \),

\[
\sum_{\ell=0}^{\infty} (\ell+1)D_n(\ell, t_j) < 2 \left( \gamma + \frac{\varepsilon}{8} \right) n^{\tilde{\theta} + \omega}. \tag{7.6}
\]

Applying Chebyshev’s inequality, on the event \( A_j \), we have

\[
\mathbb{P} \left( \sum_{\ell=0}^{k} \text{Bin} \left( D_n(\ell, t_j), q_\ell \left( an^{-\tilde{\theta}} \right) \right) > \frac{\varepsilon}{2} n^\omega \mid \mathcal{F}_n(t_j) \right) \leq \frac{4}{\varepsilon^2 n^{2\omega}} \sum_{\ell=0}^{k} \text{Var} \left( \text{Bin} \left( D_n(\ell, t_j), q_\ell \left( an^{-\tilde{\theta}} \right) \right) \right) \leq \frac{C'a}{\varepsilon^2 n^\omega} \leq \frac{C'a}{\varepsilon^2 n^\omega}. \tag{7.7}
\]

for \( C' \) not depending on \( j \), where the first inequality is from Chebyshev’s inequality the third inequality is a consequence of Lemma/7.3 and the fourth inequality follows from the definition of \( A_j \).

We now have

\[
\mathbb{P} \left( \sum_{\ell=0}^{k} \text{Bin} \left( D_n(\ell, t_j), q_\ell \left( an^{-\tilde{\theta}} \right) \right) > \frac{\varepsilon}{2} n^\omega \right) \leq \frac{C'a}{\varepsilon^2 n^\omega} + \mathbb{P} \left( Z_n(t_j) \geq \left( \gamma + \frac{\varepsilon}{8} \right) n^{\tilde{\theta} + \omega} \right). \tag{7.8}
\]

Now, we control the second term above. By Lemma/6.2 (and the fact the integral is over a bounded interval) \( \lambda_\ell(a) \leq Ce^{C'a}(\ell + 1) \). As \( \tilde{\theta} + \omega > 1 \), we can clearly choose \( C'', n_0 \) such that for all \( n \geq n_0 \) and all \( a \leq C'' \log n, \frac{\varepsilon}{16} n^{\tilde{\theta} + \omega} > (1 + \gamma)Ce^{C'a}n \). For such \( n, a \),

\[
\sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_j) \leq Ce^{C'a} \sum_{\ell=0}^{\infty} (\ell + 1)D_n(\ell, 0) = Ce^{C'a}(2\gamma n - 1) < \frac{\varepsilon}{16} n^{\tilde{\theta} + \omega}.
\]

Consequently,

\[
\mathbb{P} \left( Z_n(t_j) \geq \left( \gamma + \frac{\varepsilon}{8} \right) n^{\tilde{\theta} + \omega} \right) \leq \mathbb{P} \left( Z_n(t_j) - \gamma n \geq \left( \gamma + \frac{\varepsilon}{8} \right) n^{\tilde{\theta} + \omega} - \gamma n \right) \leq \mathbb{P} \left( Z_n(t_j) - \gamma n > \frac{\varepsilon}{8} n^{\tilde{\theta} + \omega} \right) 
= \mathbb{P} \left( Z_{AC,n}(t_j) > \frac{\varepsilon}{8} n^{\tilde{\theta} + \omega} \right) \leq \mathbb{P} \left( \left| Z_{AC,n}(t_j) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_j) \right| > \frac{\varepsilon}{16} n^{\tilde{\theta} + \omega} \right) 
\leq \frac{16}{\varepsilon} \frac{1}{n^{\tilde{\theta} + \omega}} \mathbb{E} \left| Z_{AC,n}(t_j) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_j) \right| \leq \frac{16}{\varepsilon} Ce^{C'a} \frac{1}{n^{\tilde{\theta} + \omega - \frac{1}{2}}}. \tag{7.9}
\]

for \( C, C' \) not depending on \( j \), where the last inequality comes from (7.3), (7.7) and (7.9) and the fact that \( \tilde{\theta} < 1/2 \). The result now follows.

Lemma 7.10. Let \( \{t_j, \tilde{\theta} \text{ and } \omega \} \) be as above and let \( \varepsilon > 0 \). Then

\[
\sup_{j \leq n^\omega - 1} \mathbb{P} \left( \left| Z_{AC,n}(t_{j+1}) - Z_{AC,n}(t_j) \right| > \frac{\varepsilon}{2} n^\omega \right) \leq \frac{Ce^{C'a}}{\varepsilon} \frac{1}{n^{\omega - \frac{1}{2}}}. 
\]
Proof. Applying the triangle inequality,
\[
|Z_{AC,n}(t_{j+1}) - Z_{AC,n}(t_j)| \leq |Z_{AC,n}(t_{j+1}) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_{j+1})| + |Z_{AC,n}(t_j) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_j)| + \sum_{\ell=0}^{\infty} D_n(\ell, 0) |\lambda_\ell(t_{j+1}) - \lambda_\ell(t_j)|.
\]

Note by Lemma 7.4 and the fact that \( t_{j+1} - t_j = an^{-\bar{\gamma}} \)
\[
\sum_{\ell=0}^{\infty} D_n(\ell, 0) |\lambda_\ell(t_{j+1}) - \lambda_\ell(t_j)| \leq C e^{C_1 a} \alpha n^{\bar{\gamma}} \sum_{\ell=0}^{\infty} D_n(\ell, 0)(\ell + 1) = C e^{C_1 a} \alpha n^{\bar{\gamma} - 1} \leq C'' ae^{C_1 a} n^{1-\bar{\gamma}}.
\] (7.10)

From equation (7.3) we get
\[
\sup_{j \leq n^{\delta-1}} \mathbb{E} \left[ |Z_n(t_j) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_j)| \right] \leq C e^{C_1 a} \sqrt{n}.
\]

Putting this all together, using (7.10) along with the fact that \( \omega > (1 - \bar{\gamma}) \) and applying Markov's inequality we get for sufficiently large \( n \)
\[
\mathbb{P} \left[ |Z_{AC,n}(t_{j+1}) - Z_{AC,n}(t_j)| > \frac{\epsilon}{2} n^\omega \right] = \mathbb{P} \left[ |Z_n(t_{j+1}) - Z_n(t_j)| > \frac{\epsilon}{2} n^\omega \right] 
\leq \mathbb{P} \left[ \left| \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_j)(\ell + 1) \right| > \frac{\epsilon}{4} n^\omega \right) 
\leq \frac{2}{\epsilon} n^{-\omega} \mathbb{E} \left[ |Z_n(t_j) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_j)| \right] + \mathbb{E} \left[ |Z_n(t_{j+1}) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t_{j+1})| \right] \leq \frac{2C e^{C_1 a}}{\epsilon n^{\omega-\frac{1}{2}}}
\]

for \( C, C' \) not depending on \( j \), which proves the lemma. 

Lemma 7.11. There exist positive constants \( C, C' \) such that for each \( k \) and \( \epsilon \in (0, 1) \),
\[
\mathbb{P} \left( \sup_{t \in [0, a]} |D_n(k, t) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell^{(k)}(t)| > \epsilon (k + 1) n^\omega \right) \leq \frac{C e^{C_1 a}}{\epsilon^2 n^{\omega-\bar{\gamma}-\frac{1}{2}}}
\]

and moreover,
\[
\mathbb{P} \left( \sup_{t \in [0, a]} |Z_n(t) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell(t)| > \epsilon n^\omega \right) \leq \frac{C e^{C_1 a}}{\epsilon^2 n^{\omega-\bar{\gamma}-\frac{1}{2}}}
\]

Proof. Fix \( k \) and \( \epsilon \in (0, 1) \). Note that
\[
\mathbb{P} \left( \sup_{t \in [0, a]} |D_n(k, t) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell^{(k)}(t)| > \epsilon n^\omega \right) 
\leq \sum_{\ell=0}^{\omega-1} \mathbb{P} \left( \sup_{t \in [t_j, t_{j+1}]} |D_n(k, t) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell^{(k)}(t)| > \epsilon n^\omega \right) 
\leq \sum_{\ell=0}^{\omega-1} \left[ \mathbb{P} \left( \sup_{t \in [t_j, t_{j+1}]} |D_n(k, t) - D_n(k, t_j)| > \frac{\epsilon}{3} n^\omega \right) + \mathbb{P} \left( \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell^{(k)}(t_j)| > \frac{\epsilon}{3} n^\omega \right) 
\right. 
\left. + \mathbb{P} \left( \sup_{t \in [t_j, t_{j+1}]} |\lambda_\ell^{(k)}(t) - \lambda_\ell^{(k)}(t_j)| > \frac{\epsilon}{3} n^\omega \right) \right] \). (7.11)

By Lemma 7.8
\[
\sum_{j=0}^{\omega-1} \mathbb{P} \left( \sup_{t \in [t_j, t_{j+1}]} |D_n(k, t) - D_n(k, t_j)| > \frac{\epsilon}{3} n^\omega \right) \leq \frac{C e^{C_1 a}}{\epsilon^2 n^{\omega-\bar{\gamma}-\frac{1}{2}}}.
\] (7.12)
By Corollary 7.7
\[
\sup_{j \leq n^\delta - 1} \sup_{t \in [t_j, t_{j+1}]} \sum_{\ell = 0}^{\infty} D_n(\ell, 0) \left| \lambda^{(k)}_\ell(t) - \lambda^{(k)}_{\ell+1}(t) \right| \leq C e^{C_{\epsilon}} (k + \gamma + 2) n^{1-\tilde{\theta}}
\]
and hence, as \( \omega > 1 - \tilde{\theta} \), there exists \( n_0 \) not depending on \( k \) such that for all \( n \geq n_0 \),
\[
\sum_{j=0}^{n^\delta - 1} \mathbb{P} \left( \sup_{t \in [t_j, t_{j+1}]} \sum_{\ell = 0}^{\infty} D_n(\ell, 0) \left| \lambda^{(k)}_\ell(t) - \lambda^{(k)}_{\ell+1}(t) \right| > \frac{\epsilon (k+1)}{3} n^\omega \right) = 0. \tag{7.13}
\]
Finally we control the second term appearing in the sum (7.11). It is sufficient to show
\[
\sup_{j \leq n^\delta} \mathbb{P} \left( D_n(k, t) - \sum_{\ell = 0}^{\infty} D_n(\ell, 0) \lambda^{(k)}_\ell(t_j) \right) > \frac{\epsilon}{3} n^\omega \right) \leq \frac{C e^{C_{\epsilon}a}}{\epsilon} \frac{1}{n^{\omega - \frac{1}{2}}}. \tag{7.14}
\]
By the triangle inequality and definitions of \( D_n(k, t) \), and \( \lambda^{(k)}_\ell(t) \), we see that for each fixed \( j, k, \)
\[
\left| D_n(k, t_j) - \sum_{\ell = 0}^{\infty} D_n(\ell, 0) \lambda^{(k)}_\ell(t_j) \right| \leq \left| D_n^{BC}(k, t_j) - \sum_{\ell = 0}^{k} D_n(\ell, 0) \mathbb{P} \left( \xi^{(k)}_{m_\ell}(t_j) = k - \ell \right) \right| + \left| D_n^{AC}(k, t_j) - \sum_{\ell = 0}^{\infty} D_n(\ell, 0) \lambda^{AC, (k)}_\ell(t_j) \right|.
\tag{7.15}
\]
By (7.4) and Markov’s inequality,
\[
\sup_{j \leq n^\delta} \mathbb{P} \left( \left| D_n^{AC}(k, t_j) - \sum_{\ell = 0}^{\infty} D_n(\ell, 0) \lambda^{AC, (k)}_\ell(t_j) \right| > \frac{\epsilon}{6} n^\omega \right) \leq \frac{6 C e^{C_{\epsilon}a}}{\epsilon} \frac{1}{n^{\omega - \frac{1}{2}}}. \tag{7.16}
\]
We now control the first term appearing in the bound in equation (7.15) by showing
\[
\sup_{t \in [0, a]} \mathbb{E} \left[ \left( D_n^{BC}(k, t) - \sum_{\ell = 0}^{k} D_n(\ell, 0) \mathbb{P} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) \right)^2 \right] \leq C n. \tag{7.17}
\]
Fix \( k \) and \( t \in [0, a] \). Define a collection of mutually independent random variables \( \{ \xi^{(k)}_{m_\ell}(t) \mid 1 \leq m \leq D_n(\ell, 0), 0 \leq \ell \leq k \} \) where \( \xi^{(k)}_{m_\ell}(t) \sim \xi^{(k)}_{m_\ell}(t) \). Note that
\[
D_n^{BC}(k, t) \overset{\text{d}}{=} \sum_{\ell = 0}^{k} \sum_{m=1}^{D_n(\ell, 0)} \mathbb{1} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right),
\]
i.e. a vertex that was born before the change point and was of degree \( \ell \) at the change point has to add \( k - \ell \) new births to reach degree \( k \) at time \( t \).
Therefore,
\[
\mathbb{E} \left[ \left( D_n^{BC}(k, t) - \sum_{\ell = 0}^{k} D_n(\ell, 0) \mathbb{P} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) \right)^2 \right] = \mathbb{E} \left[ \left( \sum_{\ell = 0}^{k} \sum_{m=1}^{D_n(\ell, 0)} \mathbb{1} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) - \sum_{\ell = 0}^{k} D_n(\ell, 0) \mathbb{P} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) \right)^2 \right]
\]
\[
= \mathbb{E} \left[ \left( \sum_{\ell = 0}^{k} D_n(\ell, 0) \left( \mathbb{1} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) - \mathbb{P} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) \right) \right)^2 \right].
\]
Note that
\[
\sum_{\ell = 0}^{k} \sum_{m=1}^{D_n(\ell, 0)} \left( \mathbb{1} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) - \mathbb{P} \left( \xi^{(k)}_{m_\ell}(t) = k - \ell \right) \right) \overset{\text{d}}{=} \sum_{\ell = 0}^{k} D_n(\ell, 0) Y_{\ell, m}
\]
The random variables \( \{ Y_{\ell, m} \mid 1 \leq m \leq D_n(\ell, 0), 0 \leq \ell \leq k \} \) are mutually independent, supported on \([-1, 1]\) and \( \mathbb{E} Y_{\ell, m} = 0 \). Thus,
Proof. Fix \( m \) and \( m \) fixed time points. Then, almost surely, there exists \( n_0 \geq 1 \) such that for all \( n \geq n_0 \),

\[
\sup_{1 \leq j \leq m} \left| \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell,0)\lambda_\ell^{(k)}(s_j) - \gamma \sum_{\ell=0}^{\infty} p_0^0 \lambda_\ell^{(k)}(s_j) \right| \leq \epsilon.
\]

Moreover,

\[
\sup_{1 \leq j \leq m} \left| \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell,0)\lambda_\ell(s_j) - \gamma \sum_{\ell=0}^{\infty} p_0^0 \lambda_\ell(s_j) \right| \leq \epsilon.
\]

Proof. Follows from Lemma 6.11 and the union bound.

Lemma 7.13. Let \( \{ p_k(f) : k \geq 0 \} \) as in (3.1) be the asymptotic degree distribution using attachment function \( f \) satisfying Assumption 2.1. Then \( \sum_{k=0}^{\infty} kp_k(f) = 1 \).

Proof. Recall that \( p_k(f) = t_{k-1} - t_k \) where \( t_k := \prod_{i=0}^{k} f(t_i) \) and \( \lambda^* \) is the Malthusian parameter for the corresponding preferential attachment branching process. Therefore, \( \sum_{k=1}^{\infty} kp_k(f) = \sum_{k=0}^{\infty} k(t_{k-1} - t_k) = \sum_{k=0}^{\infty} t_k \). By the definition of \( \lambda^* \) and \( t_k \) we see \( \sum_{k=1}^{\infty} t_k = 1 \), proving the lemma.

Lemma 7.14. For any \( k \geq 0 \),

\[
\sup_{t \in [0,a]} \left| \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell,0)\lambda_\ell^{(k)}(t) - \gamma \sum_{\ell=0}^{\infty} p_0^0 \lambda_\ell^{(k)}(t) \right| \xrightarrow{a.s.} 0.
\]

Moreover,

\[
\sup_{t \in [0,a]} \left| \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell,0)\lambda_\ell(t) - \gamma \sum_{\ell=0}^{\infty} p_0^0 \lambda_\ell(t) \right| \xrightarrow{a.s.} 0.
\]

Proof. Fix \( \epsilon > 0 \). Let 0 = \( s_1 < s_2 < \cdots < s_m = a \) be a partition such that \( |s_j+1 - s_j| \leq \epsilon \).

By Corollary 7.7,

\[
\sup_{1 \leq j \leq m} \sup_{t \in [s_j,s_{j+1}]} \left| \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell,0)\lambda_\ell^{(k)}(t) - \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell,0)\lambda_\ell^{(k)}(s_j) \right| \leq C e^{C\epsilon}(k+3)\epsilon.
\]

Similarly, using Corollary 7.6,

\[
\sup_{1 \leq j \leq m-1} \sup_{t \in [s_j,s_{j+1}]} \left| \gamma \sum_{\ell=0}^{\infty} p_0^0 \lambda_\ell^{(k)}(t) - \gamma \sum_{\ell=0}^{\infty} p_0^0 \lambda_\ell^{(k)}(s_j) \right| \leq \sup_{1 \leq j \leq k-1} \sup_{t \in [s_j,s_{j+1}]} \gamma \sum_{\ell=0}^{\infty} p_0^0 \lambda_\ell^{(k)}(t) - \lambda_\ell^{(k)}(s_j) \leq C e^{C\epsilon} \gamma \sum_{\ell=0}^{\infty} p_0^0 (k+\ell+2) = Ce^{C\epsilon} \gamma (k+3)\epsilon.
\]
By Corollary 7.12 almost surely, there exists $n_0 \geq 1$ such that for all $n \geq n_0$,
\[ \sup_{1 \leq j \leq m} \left| \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(k)}_\ell(s_j) - \gamma \sum_{\ell=0}^{\infty} p^0_\ell \lambda^{(k)}_\ell(s_j) \right| \leq \epsilon. \]

From the above, we now have that for $n \geq n_0$,
\[ \sup_{t \in [0, a]} \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(k)}_\ell(t) - \sum_{\ell=0}^{\infty} p^0_\ell \lambda^{(k)}_\ell(t) \leq \sup_{1 \leq j \leq m-1} \sup_{t \in [s_j, s_{j+1}]} \left| \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(k)}_\ell(t) - \frac{1}{n} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(k)}_\ell(s_j) \right| \]
\[ + \sup_{1 \leq j \leq m-1} \sup_{t \in [s_j, s_{j+1}]} \left| \gamma \sum_{\ell=0}^{\infty} p^0_\ell \lambda^{(k)}_\ell(t) - \gamma \sum_{\ell=0}^{\infty} p^0_\ell \lambda^{(k)}_\ell(s_j) \right| \leq C e^{C_1 a} (k + 3) \epsilon \]

which proves the first assertion of the lemma. The second assertion follows similarly using Corollary 7.12 and the first bound in Lemma 7.4.

**Proof of Theorem 7.1** The theorem follows from Lemmas 7.11 and 7.14.

7.2. **Proof of Corollary 3.11** The essential message of this Corollary 3.11 is that the tail of the distribution prescribed by the initializer function always wins. Recall that the limit random variable $D_0$ is a mixture of the distributions of $X_{BC}$ and $X_{AC}$.

**Lemma 7.15.** The random variable $X_{AC}$ always has an exponential tail.

**Proof:** By construction, note that $X_{AC} \preceq_{st} \xi_f [0, \alpha]$. Further our assumption on the attachment functions implies that there exists $\kappa > 0$ such that $\max(\ell_0(i), \ell_1(i)) \leq \kappa(i + 1)$ for all $i$. In particular $\xi_f [0, \alpha] \preceq_{st} Y_{\kappa} [0, \alpha]$ where $Y_{\kappa} (\cdot)$ is a rate $\kappa$ Yule process as in Definition 5.3. Using Lemma 5.4 now completes the proof.

Thus is is enough to consider $X_{BC}$ and show that this random variable has the same tail behavior as the random variable $D \sim \{p_\kappa^0 : k \geq 1\}$. Once again by construction,
\[ X_{BC} \preceq_{st} D + \sum_{i=1}^{D} Y_{\kappa, i} [0, \alpha], \]

where $\{Y_{\kappa, i} (\cdot) : i \geq 1\}$ is an infinite collection of independent Yule processes (independent of $D$). Let $\mu := \mathbb{E}(Y_{\kappa, i} [0, \alpha])$. Note $\mu > 1$. Now note that for $x \geq 1$,
\[
\Pr(X_{BC} > x) \leq \sum_{j=1}^{x/2\mu} \Pr(D = j) \Pr\left(\sum_{i=1}^{j} Y_{\kappa, i} [0, \alpha] > x - j\right) + \Pr(D > x/2\mu) \\
\leq \Pr\left(\sum_{i=1}^{x/2\mu} Y_{\kappa, i} [0, \alpha] > x\left(1 - \frac{1}{2\mu}\right)\right) + \Pr(D > x/2\mu). \tag{7.19}
\]

Standard large deviation bounds for the probability measure of $Y_{\kappa, i}$ implies that there exists constants $C_1, C_2$ such that for all $x$,
\[
\Pr\left(\sum_{i=1}^{x/2\mu} Y_{\kappa, i} [0, \alpha] > x\left(1 - \frac{1}{2\mu}\right)\right) \leq C_1 \exp(-C_2 x).
\]
Thus in the setting of Corollary 3.11, assuming $D$ has exponential tails, one finds using (7.19) that there exist finite constants $C_1, C_2$ such that

$$P(X_{BC} > x) \leq C_1 \exp(-C_2 x).$$

This completes the proof of Corollary 3.11 (ii). A similar argument using the obvious inequality $P(D > x) \leq P(X_{BC} > x)$ verifies Corollary 3.11 (ii).

$\blacksquare$

8. Proofs: Quick Big Bang

8.1. Proof of Theorem 3.15. Recall that in this section, we throughout work under Assumptions 2.1, 3.1 and 3.14 for $f_0, f_1$. For notational convenience, instead of considering the change point at $T_{n^*}$ and evolving the tree till $T_n$, we will consider the problem of the change point being at $T_n$ and evolving the tree till $T_{n^*+n^*+}$ for some $\theta > 0$ (where $\lambda^*_i$ is the Malthusian rate corresponding to $f_1$). For this section, $t = 0$ represents time $T_n$ (the smallest time the change point process has $n$ vertices). It is easy to see that Theorem 3.15 is equivalent to Theorem 8.13 proved below.

Recall the notation from Section 7. From Lemma 7.11 for every $k \geq 0$, there exists $\eta_0 > 0$ such that for all $\eta \leq \eta_0$,

$$\frac{1}{n} \sup_{t \in [0, n \log n]} \left| D_n(k, t) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(k)}(\ell)(t) \right| \overset{P}{\rightarrow} 0, \quad \text{as } n \rightarrow \infty. \quad (8.1)$$

Similarly, using Lemma 7.11 we obtain $\eta_0$ such that for all $\eta \leq \eta_0$,

$$\frac{1}{n} \sup_{t \in [0, n \log n]} \left| Z_n(t) - \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(0)}(\ell)(t) \right| \overset{P}{\rightarrow} 0, \quad \text{as } n \rightarrow \infty. \quad (8.2)$$

(8.1) and (8.2) immediately imply for any $\eta \leq \eta_0$,

$$\frac{1}{n^{1+\eta \lambda^*_i}} D_n(k, \eta \log n) - \frac{1}{n^{1+\eta \lambda^*_i}} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(k)}(\eta \log n) \overset{P}{\rightarrow} 0, \quad (8.3)$$

$$\frac{1}{n^{1+\eta \lambda^*_i}} Z_n(\eta \log n) - \frac{1}{n^{1+\eta \lambda^*_i}} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda^{(0)}(\eta \log n) \overset{P}{\rightarrow} 0$$

as $n \rightarrow \infty$. Define for each $\ell \geq 0$ and $\beta > 0$,

$$w_\ell(\beta) := \int_0^\infty e^{-\beta s} u^{(\ell)}(s) ds.$$ We will simply write $w_\ell$ for $w_\ell(\lambda^*_i)$. We will need the following technical lemmas. Recall from Assumption 2.1 (iii) that there exists $\beta_1 \in (0, \lambda^*_i)$ such that $R(\beta_1) < \infty$. Recall $C^*$ from Assumption 3.1 applied to $f_1$.

Lemma 8.1. $\beta_1 \geq C^*$.

Proof. If $C^* = 0$, there is nothing to prove. So we assume $C^* > 0$. For any $\epsilon \in (0, C^*)$, by Assumption 3.1 there exists $j_0 \geq 1$ such that for all $j \geq j_0$, $f_1(j) \geq (C^* - \epsilon) j$. Finiteness of $R(\beta_1)$ implies that

$$\sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \frac{f_1(i + j_0)}{\beta_1 + f_1(i + j_0)} < \infty. \quad (8.4)$$

For any $k \geq 1$, noting that $x \mapsto \frac{x}{\beta_1 + x}$ is a strictly increasing function and, $\log(1 + x) \leq x$ for any $x \geq 0$, and $\sum_{j=j_1}^{j_2} \frac{1}{j} \leq \int_{j_1-1}^{j_2} \frac{dx}{x}$ for any $j_2 \geq j_1 \geq 1$,

$$\log \left[ \prod_{i=0}^{k-1} \frac{f_1(i + j_0)}{\beta_1 + f_1(i + j_0)} \right] \geq \log \left[ \prod_{i=0}^{k-1} \frac{i + j_0}{\beta_1 f_1(i + j_0)} \right] = - \sum_{i=0}^{k-1} \log \left[ 1 + \frac{\beta_1}{(C^* - \epsilon)(i + j_0)} \right]$$

$$\geq - \frac{\beta_1}{C^* - \epsilon} \sum_{i=0}^{k-1} \frac{1}{i + j_0} \geq - \frac{\beta_1}{C^* - \epsilon} \int_{j_0+k-1}^{j_0+k-1} \frac{dx}{x} = - \frac{\beta_1}{C^* - \epsilon} \log \left( \frac{j_0 + k - 1}{j_0 - 1} \right)$$
and thus
\[
\prod_{i=0}^{k-1} \frac{f_i(i + j_0)}{\beta_i + f_i(i + j_0)} \geq \left( \frac{j_0 - 1}{j_0 + k - 1} \right)^{\frac{\epsilon}{C^* + \epsilon}}.
\]
Thus, (8.4) holds only if \( \beta_1 > C^* - \epsilon \). As \( \epsilon > 0 \) is arbitrary, this proves the lemma.

**Lemma 8.2.** For any \( \beta \in (\beta_1, \lambda_1^*) \), there exists a constant \( C(\beta) > 0 \) such that \( w_\ell(\beta) \leq C(\beta) (\ell + 1) \) for all \( \ell \geq 0 \).

**Proof.** Fix any \( \beta \in (\beta_1, \lambda_1^*) \) and \( \ell \geq 0 \). Since \( \int_0^\infty e^{-\beta s} \mu f_i (d s) = \sum_{k=1}^\infty \prod_{i=0}^{k-1} \frac{f_i(i)}{\beta + f_i(i)} \), the sum on the right hand side is finite. Note that
\[
w_\ell(\beta) = \int_0^\infty e^{-\beta s} \mu f_i (d s) = \sum_{k=1}^\infty \prod_{i=0}^{k-1} \frac{f_i(i)}{\beta + f_i(i)} = \sum_{k=1}^\infty \prod_{i=0}^{k-1} \frac{f_i(i)}{\beta + f_i(i)} < \infty.
\]
Choose and fix \( \epsilon > 0 \) such that \( C^* + 2\epsilon < \beta \) (which is possible by Lemma 8.1). By Assumption 3.1, there exists \( j_0 \geq 1 \) such that for all \( j \geq j_0 \), \( f_i(j) \leq (C^* + \epsilon) j \). For any \( \ell \geq j_0 \), using the facts that \( x \rightarrow \frac{x}{\beta + x} \) is a strictly increasing function and, \( \log(1 + x) \geq \frac{x}{1 + x} \) for any \( x \geq 0 \), and \( \sum_{j=0}^{j_2} \frac{1}{j_2} \geq \int_{j_2}^{j_2 + 1} \frac{dx}{x} \) for any \( j_2 \geq 1 \), we obtain for any \( \ell \geq j_0 \),
\[
\log \left( \prod_{i=0}^{\ell-1} \frac{f_i(i)}{\beta + f_i(i)} \right) \leq \log \left( \prod_{i=0}^{\ell-1} \frac{i}{\beta + \frac{\beta}{C^* + \epsilon}} \right) = - \sum_{i=0}^{\ell-1} \log \left( 1 + \frac{\beta}{(C^* + \epsilon)} \right) \leq - \sum_{i=0}^{\ell-1} \frac{\beta}{1 + \frac{\beta}{(C^* + \epsilon)}} \leq - \frac{\beta}{1 + \frac{\beta}{(C^* + \epsilon)}} \int_{\ell}^{\infty} \frac{dx}{x} = - \frac{\beta}{1 + \frac{\beta}{(C^* + \epsilon)}} \log 2.
\]
Take \( \ell_1 \geq j_0 \) such that \( \frac{\beta}{1 + \frac{\beta}{(C^* + \epsilon)}} \geq \frac{\beta}{C^* + 2\epsilon} \). From the above calculation, for all \( \ell \geq \ell_1 \), \( \prod_{i=0}^{\ell-1} \frac{f_i(i)}{\beta + f_i(i)} \leq 2^{-\frac{\beta}{C^* + 2\epsilon}} \).

Using this bound iteratively, we obtain for any \( j \geq 1 \),
\[
\prod_{i=0}^{\ell-1} \frac{f_i(i)}{\beta + f_i(i)} \leq 2^{-\frac{\beta}{C^* + 2\epsilon}}.
\]
Thus, for all \( \ell \geq \ell_1 \),
\[
w_\ell(\beta) = \sum_{k=1}^\infty \prod_{i=0}^{k-1} \frac{f_i(i)}{\beta + f_i(i)} \leq \ell + \sum_{j=0}^{\ell-1} \sum_{k=2j}^\infty \prod_{i=0}^{k-1} \frac{f_i(i)}{\beta + f_i(i)} \leq \ell + \sum_{j=0}^{\ell-1} 2^j \ell \prod_{i=0}^{k-1} \frac{f_i(i)}{\beta + f_i(i)} = \ell \left[ 1 + \sum_{j=0}^{\ell-1} 2^{1 - \frac{\beta}{C^* + 2\epsilon}} \right] = \left( \frac{2 - 2^{1 - \frac{\beta}{C^* + 2\epsilon}}}{1 - 2^{1 - \frac{\beta}{C^* + 2\epsilon}}} \right) \ell
\]
where the sum converges as \( C^* + 2\epsilon < \beta \). This proves the lemma.

Recall the class of characteristics \( \mathcal{C} \) defined in (3.2).

**Lemma 8.3.** Let \( \phi \in \mathcal{C} \) such that \( \lim_{t \to -\infty} e^{-\lambda_1^* t} m_{f_i}^\phi(t) = c_\phi \). For \( \ell \geq 0 \), define
\[
\lambda_{\ell}^\phi(t) = \lambda_{\ell}^\phi(0) + \int_0^t m_{f_i}^\phi(t-s) \mu_{f_i}^\ell(ds)
\]
where \( \lambda_{\ell}^\phi(0) \in [0,1] \) for each \( \ell \). There is a constant \( C > 0 \) for which the following holds: for any \( \epsilon > 0 \), there exists \( t(\epsilon) > 0 \) such that for any \( \ell \geq 0 \),
\[
\sup_{t \leq t(\epsilon)} \left| e^{-\lambda_{\ell}^\phi t} \lambda_{\ell}^\phi(t) - w_\ell c_\phi \right| \leq C \epsilon (\ell + 1).
\]
Lemma 8.4. Thus, there exists $\lambda^*_1 - \bar{\theta} > 0$. As $\lim_{t \to \infty} e^{-\lambda^*_1 t} m_{f_1}(t) = c_f$ and $\sup_{t < 0} e^{-\lambda^*_1 t} m_{f_1}(t) < \infty$ (which holds because the limit as $t \to \infty$ exists and as $\phi \in \mathcal{C}$, therefore for each $a > 0$, $\sup_{t \in [0, a]} m_{f_0}(t) \leq C \sup_{t \in [0, a]} m_{f_1}(t) < \infty$ by virtue of (6.1)), there exists $t_0 > 0$ such that for all $t \geq t_0,\{e^{-\lambda^*_1 t} m_{f_1}(t) - c_f\} \leq \epsilon$ and $e^{-\theta t} (\sup_{s \leq t} e^{-\lambda^*_1 s} m_{f_1}(z) + c_f) \leq \epsilon$. Thus, for any $t \geq 2t_0$,

$$\sup_{s \leq t} \left| e^{-\lambda^*_1 (t-s)} m_{f_1}^\phi(t-s) - c_f \right| \leq \epsilon.$$ 

Thus, applying Lemma 8.2 with $\bar{\theta} = \lambda^*_1 - \bar{\theta}$, we conclude that for any $t \geq 2t_0$,

$$\int_0^t \left| e^{-\lambda^*_1 (t-s)} m_{f_1}^\phi(t-s) - c_f \right| e^{-\lambda^*_1 s} \mu_{f_1}^{(\ell)}(ds) = \int_0^t e^{-\theta s} \left| e^{-\lambda^*_1 (t-s)} m_{f_1}^\phi(t-s) - c_f \right| e^{-\lambda^*_1 s} \mu_{f_1}^{(\ell)}(ds) \leq c \epsilon \sup_{t \geq t_0} e^{-\lambda^*_1 s} m_{f_1}^\phi(t-s) \leq C_c (\ell + 1).$$

Moreover, as $\int_0^\infty e^{-\lambda^*_1 t} \mu_{f_1}^{(\ell)}(ds) \leq C(\ell + 1)$, for $t \geq 0$,

$$c_f \int_0^\infty e^{-\lambda^*_1 s} \mu_{f_1}^{(\ell)}(ds) \leq C_c (\ell + 1) e^{-\theta t}.$$ 

Using these in (8.6) and recalling $\lambda^*_1(0) \in [0, 1]$ for each $\ell$, we obtain for $t \geq 2t_0$,

$$\left| e^{-\lambda^*_1 t} \lambda^*_1(0) - w^*_c c_f \right| \leq e^{-\lambda^*_1 t} + C_c (\ell + 1) e^{-\theta t} + C_c (\ell + 1).$$

Thus, there exists $t_1 \geq 2t_0$ such that for all $t \geq t_0$ and all $t \geq t_1$,

$$\left| e^{-\lambda^*_1 t} \lambda^*_1(0) - w^*_c c_f \right| \leq C_c (\ell + 1).$$

Lemma 8.4. Let $\phi \in \mathcal{C}$ such that $\lim_{t \to \infty} e^{-\lambda^*_1 t} m_{f_1}(t) = c_f$. For $t \geq 0$, let $\lambda^*_1(\cdot)$ be defined as in (8.5). Fix any $\eta > 0, a \in \mathbb{R}$. Then as $n \to \infty$,

$$\frac{1}{n^{1+\eta \lambda^*_1}} \sum_{\ell=0}^\infty D_n(\ell, 0) \lambda^*_1(\eta \log n + a) \xrightarrow{P} c_f e^{\lambda^*_1 a} \sum_{\ell=0}^\infty p_\ell^0 w^\ell.$$ 

Proof. In this proof, $C, C', C''$ will denote generic positive constants not depending on $n, t, \ell$ whose values might change from line to line. Note that

$$\left| \frac{1}{n^{1+\eta \lambda^*_1}} \sum_{\ell=0}^\infty D_n(\ell, 0) \lambda^*_1(\eta \log n + a) - c_f e^{\lambda^*_1 a} \sum_{\ell=0}^\infty p_\ell^0 w^\ell \right| \leq \frac{\sum_{\ell=0}^\infty D_n(\ell, 0)}{\eta \lambda^*_1} \frac{\lambda^*_1(\eta \log n + a)}{n^{\eta \lambda^*_1}} - w^*_c c_f e^{\lambda^*_1 a} + c_f e^{\lambda^*_1 a} \sum_{\ell=0}^\infty D_n(\ell, 0) w^\ell - \sum_{\ell=0}^\infty p_\ell^0 w^\ell.$$ 

(8.7)

To show that the second term goes to zero in probability, consider the characteristic $\chi(t) = \sum_{\ell=0}^\infty w^\ell \mathbb{1}\{\xi_f(t) = \ell\}$. By Lemma 8.2, $w^\ell \leq C(\ell + 1)$ and hence, $\chi \in \mathcal{C}$. Thus, by Lemma 5.8(ii),

$$\left| \sum_{\ell=0}^\infty D_n(\ell, 0) w^\ell - \sum_{\ell=0}^\infty p_\ell^0 w^\ell \right| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$ 

(8.8)
To show that the first term in the bound (8.7) goes to zero in probability, take any \( \epsilon > 0 \). Recalling \( \sum_{\ell=0}^{\infty} D_n(\ell, 0) = n \) and \( \sum_{\ell=0}^{\infty} (\ell + 1) D_n(\ell, 0) = 2n - 1 \), and taking \( t = \eta \log n + a \) for any \( n \geq e^{(t(e) - a)/\eta} \) in Lemma 8.3, we obtain

\[
\sum_{\ell=0}^{\infty} \frac{D_n(\ell, 0)}{n} \left| \frac{\lambda_\ell^*(\eta \log n + a)}{\eta^k \lambda_1^*} - n \right| \leq C'' \epsilon \sum_{\ell=0}^{\infty} (\ell + 1) D_n(\ell, 0) \leq 2C'' \epsilon.
\]

As \( \epsilon > 0 \) is arbitrary, this shows that the first term in (8.7) converges to zero as \( n \to \infty \) and completes the proof of the lemma.

Define \( m^* := \int_0^\infty u e^{-\lambda_1^* u} \mu_{f_1}(du) \).

**Corollary 8.5.** Fix any \( \eta > 0 \). Then

\[
\frac{1}{n^{1+\eta \lambda_1^*}} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell^*(\eta \log n) \xrightarrow{P} \frac{1}{\lambda_1^* m^*} \sum_{\ell=0}^{\infty} p_\ell^0 \omega_\ell
\]

and for each \( k \geq 0 \),

\[
\frac{1}{n^{1+\eta \lambda_1^*}} \sum_{\ell=0}^{\infty} D_n(\ell, 0) \lambda_\ell^{(k)}(\eta \log n) \xrightarrow{P} \frac{1}{\lambda_1^* m^*} \sum_{\ell=0}^{\infty} p_\ell^0 \omega_\ell
\]

as \( n \to \infty \).

**Proof.** Follows from Lemma 8.4 upon noting that

\[
\lambda_\ell(t) = 1 + \int_0^t m_{f_1}(t-s) \mu_{f_1}(ds), \quad \lambda_\ell^{(k)}(t) = \mathbb{P}\left( \xi_{f_1}^i(t) = k - \ell \right) + \int_0^t m_{f_1}(t-s) \mu_{f_1}^{(k)}(ds)
\]

and observing by Lemma 8.6 (ii)

\[
\lim_{t \to \infty} e^{-\lambda_1^* t} m_{f_1}(t) = \frac{1}{\lambda_1^* m^*}, \quad \lim_{t \to \infty} e^{-\lambda_1^* t} m_{f_1}^{(k)}(t) = \frac{p_k^1}{\lambda_1^* m^*}.
\]

**Lemma 8.6.** There exists \( \eta_0 > 0 \) such that for any \( \eta \leq \eta_0 \), the following limits hold as \( n \to \infty \):

(i) \( \frac{1}{n^{1+\eta \lambda_1^*}} Z_n(\eta \log n) \xrightarrow{P} \frac{1}{\lambda_1^* m^*} \sum_{\ell=0}^{\infty} p_\ell^0 \omega_\ell \),

(ii) For any \( k \geq 0 \), \( \frac{1}{n^{1+\eta \lambda_1^*}} D_n(k, \eta \log n) \xrightarrow{P} \frac{p_k^1}{\lambda_1^* m^*} \sum_{\ell=0}^{\infty} p_\ell^0 \omega_\ell \).

**Proof.** (i) and (ii) follow from (8.2) and (8.1) respectively along with Corollary 8.5.

**Corollary 8.7.** \( \sum_{\ell=0}^{\infty} p_\ell^1 \omega_\ell = \lambda_1^* m^* \).

**Proof.** Note that Lemma 8.6 (i) holds in the special case where \( f_0 = f_1 \) (the model without change point). In this case, \( p_\ell^0 = p_\ell^1 \) for all \( \ell \geq 0 \). By Lemma 8.6 (ii),

\[
\frac{Z_n(\eta \log n)}{e^{\lambda_1^* (T_n + \eta \log n)}} \xrightarrow{a.s.} \frac{W_\infty}{\lambda_1^* m^*}.
\]

Moreover, as \( Z(T_n) = n \), therefore, applying Lemma 5.8 (ii) again,

\[
\frac{e^{\lambda_1^* T_n}}{n} = \frac{1}{e^{-\lambda_1^* T_n} Z(T_n)} \xrightarrow{a.s.} \frac{\lambda_1^* m^*}{W_\infty}.
\]

Using these observations, we obtain

\[
\frac{1}{n^{1+\eta_0 \lambda_1^*}} Z_n(\eta_0 \log n) \xrightarrow{P} \frac{e^{\lambda_1^* T_n}}{n} \xrightarrow{a.s.} \frac{Z_n(\eta_0 \log n)}{e^{\lambda_1^* (T_n + \eta_0 \log n)}} \xrightarrow{a.s.} 1.
\]

Comparing this with Lemma 8.6 (ii) with \( f_0 = f_1 \) gives the result.
Recall that for any $k \geq 0$, $\xi^{(k)}(\cdot)$ is the point process denoting the distribution of birth times of children of a vertex which is of degree $k$ at time zero. The following lemma gives an estimate on the second moment of $\xi^{(k)}(t)$ under Assumption \ref{as:main}.

**Lemma 8.8.** There exists $C > 0$ and $\beta' < \lambda_1^*$ such that for any $k \geq 0, t \geq 0$,
\[
\mathbb{E}\left(\xi^{(k)}(t)\right)^2 \leq C(k+1)^2 e^{2\beta' t}.
\]

**Proof.** By Assumption \ref{as:main} and Lemma \ref{lem:att}, for any $\beta' \in (\beta_1, \lambda_1^*)$, there exists $\ell_0 \geq 0$ such that for all $\ell \geq \ell_0$, $f_1(\ell) \leq \beta' \ell$. Let $m = \max_{\ell \leq \ell_0} f_1(\ell)$. It is clear that $\xi^{(k)}(\cdot)$ is stochastically dominated by the offspring distribution of a continuous time branching process with attachment function $f^*(\ell) = \beta' \ell + 1 + (m + \beta' k), \ell \geq 0$, which we denote by $\xi_f^{(k)}(\cdot)$. Applying the second moment obtained in Lemma \ref{lem:2} with $\nu = \beta'$ and $\kappa = 1 + m + \beta' (k - 1)$ the lemma follows.

For $j \geq 0, \eta > 0$, let $D_n(k, j, \eta)$ denote the number of vertices of degree $k$ at time $(j+1)\eta \log n$ that were born before time $j\eta \log n$.

**Lemma 8.9.** For any $\eta > 0, j \geq 0$, as $n \to \infty$,
\[
\frac{\sum_{k=0}^\infty (k+1)D_n(k, j, \eta)}{Z_n(j\eta \log n) n^{\lambda_1^*} \eta} \xrightarrow{P} 0.
\]

**Proof.** We will condition on $\mathcal{F}_n(j\eta \log n)$ throughout the proof. Denoting by $(\xi^{(\ell)}_{f_i, m}(t))_{1 \leq m \leq D_n(\ell, j\eta \log n)}$ the degree at time $t + j\eta \log n$ of the $m$-th vertex of degree $\ell$ at time $j\eta \log n$, observe that
\[
\sum_{k=0}^\infty (k+1)D_n(k, j, \eta) = \sum_{k=0}^\infty (k+1) \sum_{\ell=0}^k \sum_{m=1}^{D_n(\ell, j\eta \log n)} \mathbb{1}\{\xi^{(\ell)}_{f_i, m}(\eta \log n) = k - \ell\}
\]
\[
= \sum_{\ell=0}^\infty \sum_{m=1}^{D_n(\ell, j\eta \log n)} (k+1) \sum_{k=\ell}^\infty \mathbb{1}\{\xi^{(\ell)}_{f_i, m}(\eta \log n) = k - \ell\} = \sum_{\ell=0}^\infty D_n(\ell, j\eta \log n) \sum_{m=1}^{D_n(\ell, j\eta \log n)} (\ell + 1 + \xi^{(\ell)}_{f_i, m}(\eta \log n))
\]
\[
= \sum_{\ell=0}^\infty (\ell + 1)D_n(\ell, j\eta \log n) + \sum_{\ell=0}^\infty \sum_{m=1}^{D_n(\ell, j\eta \log n)} \xi^{(\ell)}_{f_i, m}(\eta \log n)
\]
\[
= 2Z_n(j\eta \log n) - 1 + \sum_{\ell=0}^\infty \sum_{m=1}^{D_n(\ell, j\eta \log n)} \xi^{(\ell)}_{f_i, m}(\eta \log n).
\]

Thus, it suffices to show that as $n \to \infty$,
\[
\frac{1}{Z_n(j\eta \log n)} \sum_{\ell=0}^\infty \sum_{m=1}^{D_n(\ell, j\eta \log n)} \frac{1}{n^{\lambda_1^*} \eta} \xi^{(\ell)}_{f_i, m}(\eta \log n) \xrightarrow{P} 0.
\]

(8.10)

Note that using Lemma \ref{lem:2'}
\[
\text{Var}\left(\frac{1}{Z_n(j\eta \log n)} \sum_{\ell=0}^\infty \sum_{m=1}^{D_n(\ell, j\eta \log n)} \frac{1}{n^{\lambda_1^*} \eta} \xi^{(\ell)}_{f_i, m}(\eta \log n)\right)
\]
\[
\leq \frac{1}{Z^2(j\eta \log n) n^{2\lambda_1^*} \eta} \sum_{\ell=0}^\infty \sum_{m=1}^{D_n(\ell, j\eta \log n)} \mathbb{E}\left(\xi^{(\ell)}_{f_i, m}(\eta \log n)\right)^2 \leq \frac{Cn^2\beta' \eta}{Z^2(j\eta \log n) n^{2\lambda_1^*} \eta} \sum_{\ell=0}^\infty (\ell + 1)^2 D_n(\ell, j\eta \log n).
\]

Denoting the maximum out-degree at time $j\eta \log n$ of the branching process by $D^{\text{max}}$, note that $D^{\text{max}} + 1 \leq Z_n(j\eta \log n)$ and hence,
\[
\sum_{\ell=0}^\infty (\ell + 1)^2 D_n(\ell, j\eta \log n) \leq (D^{\text{max}} + 1) \sum_{\ell=0}^\infty (\ell + 1)D_n(\ell, j\eta \log n) \leq Z_n(j\eta \log n)(2Z_n(j\eta \log n) - 1).
Using this in the above variance bound, we get
\[
\text{Var}\left( \frac{1}{Z_n(j\eta \log n)} \sum_{\ell=0}^{\infty} D_n(\ell, j\eta \log n) \sum_{m=1}^{D_n(\ell, j\eta \log n)} \frac{1}{n^{1+\ell} \eta f_{i,m}(\eta \log n)} \right) \leq \frac{2Cn^{2\beta \eta} Z^2(j\eta \log n)}{Z^2(j\eta \log n)n^{2\lambda^*_\eta}} = \frac{2C}{n^{2(\lambda^*_\eta - \beta \eta)}} \to 0
\]
as \( n \to \infty \) and hence,
\[
\frac{1}{Z_n(j\eta \log n)} \sum_{\ell=0}^{\infty} D_n(\ell, j\eta \log n) \sum_{m=1}^{D_n(\ell, j\eta \log n)} \frac{1}{n^{1+\ell} \eta f_{i,m}(\eta \log n)} - \frac{1}{Z_n(j\eta \log n)} \sum_{\ell=0}^{\infty} D_n(\ell, j\eta \log n) \sum_{m=1}^{D_n(\ell, j\eta \log n)} \frac{1}{n^{1+\ell} \eta} \mathbb{E}\left( \xi_{f_{i,m}(\eta \log n)}^* \right) \xrightarrow{p} 0.
\]
(8.11)

By Lemma 8.2, we obtain \( \beta \in (\lambda^*_\eta - 1, \lambda^*_\eta) \) such that \( \eta \leq 1, \eta \log n \leq C(\beta)(\ell + 1) \). This implies for any \( m, \ell \),
\[
\mathbb{E}\left( \xi_{f_{i,m}(\eta \log n)}^* \right) \leq C(\beta)n^{\beta \eta}(\ell + 1)
\]
and consequently,
\[
\frac{1}{Z_n(j\eta \log n)} \sum_{\ell=0}^{\infty} D_n(\ell, j\eta \log n) \sum_{m=1}^{D_n(\ell, j\eta \log n)} \frac{1}{n^{1+\ell} \eta} \mathbb{E}\left( \xi_{f_{i,m}(\eta \log n)}^* \right) \leq \frac{C(\beta)}{n^{(\lambda^*_\eta - \beta \eta)}} \sum_{\ell=0}^{\infty} (\ell + 1)D_n(\ell, j\eta \log n) \leq \frac{2C(\beta)}{n^{(\lambda^*_\eta - \beta \eta)}} \to 0 \quad (8.12)
\]
as \( n \to \infty \). From (8.11) and (8.12), the proof of (8.10), and hence the lemma, is complete.

\textbf{Lemma 8.10.} Let \( \phi \in \mathcal{E} \) such that \( \lim_{t \to \infty} e^{-\lambda^*_t^t} m^\phi(t) = c_{\phi} \). For \( \ell \geq 0 \), let \( \lambda^\phi_t(\cdot) \) be defined as in (8.5). Fix any \( j \geq 0 \). There exists \( \eta_0 > 0 \) such that for any \( \eta \leq \eta_0 \) and any \( a \in \mathbb{R} \), the following limit holds as \( n \to \infty \):
\[
\frac{1}{n^{1+[(j+1)\eta_0 + \eta] \lambda^*_t}} \sum_{\ell=0}^{\infty} D_n(\ell, (j + 1)\eta_0 \log n) \lambda^\phi_{t,\ell}(\eta \log n + a) \xrightarrow{p} c_{\phi} e^{\lambda^*_t^t} a \sum_{\ell=0}^{\infty} p^\phi_{\ell} w_{\ell}.
\]

\textbf{Proof.} We will proceed by induction. Suppose we can show that for some \( j \geq 0 \), the assertion of the lemma holds for all \( j' \leq j \). Taking \( \phi(t) = \mathbb{1}_{\{t \geq 0\}} \) and \( \eta = \eta_0 \) and recalling \( \lim_{t \to \infty} e^{-\lambda^*_t^t} m_{f_{i}}(t) = \frac{1}{\lambda^*_t^t m^*} \), we obtain for any \( j' \leq j \) and \( a \in \mathbb{R} \),
\[
\frac{1}{n^{1+(j+1)\eta_0 \lambda^*_t}} Z_n((j+1)\eta_0 \log n + a) \xrightarrow{p} \frac{1}{\lambda^*_t^t m^*} e^{\lambda^*_t^t} a \sum_{\ell=0}^{\infty} p^\phi_{\ell} w_{\ell}.
\]
(8.13)

Fix any \( \phi \in \mathcal{E} \). Note that for any \( \eta \leq \eta_0 \),
\[
\left| \frac{1}{n^{1+[(j+1)\eta_0 + \eta] \lambda^*_t}} \sum_{\ell=0}^{\infty} D_n(\ell, (j + 1)\eta_0 \log n) \lambda^\phi_{t,\ell}(\eta \log n + a) - c_{\phi} e^{\lambda^*_t^t} a \sum_{\ell=0}^{\infty} p^\phi_{\ell} w_{\ell} \right|
\]
\[
\leq \sum_{\ell=0}^{\infty} D_n(\ell, (j + 1)\eta_0 \log n) \left| \lambda^\phi_{t,\ell}(\eta \log n + a) \right| \frac{n^{\lambda^*_t}}{n^{1+[(j+1)\eta_0 \lambda^*_t]}} - c_{\phi} e^{\lambda^*_t^t} a w_{\ell}
\]
\[
+ c_{\phi} e^{\lambda^*_t^t} a \sum_{\ell=0}^{\infty} D_n(\ell, (j + 1)\eta_0 \log n) \frac{n^{\lambda^*_t}}{n^{1+[(j+1)\eta_0 \lambda^*_t]}} w_{\ell} - \sum_{\ell=0}^{\infty} p^\phi_{\ell} w_{\ell}. \quad (8.14)
\]

For any \( \epsilon > 0 \), by Lemma 8.3, there exists \( n_0 \) such that for all \( n \geq n_0 \),
\[
\left| \frac{\lambda^\phi_{t,\ell}(\eta \log n + a)}{n^{\lambda^*_t}} - c_{\phi} e^{\lambda^*_t^t} a w_{\ell} \right| \leq C'' e^{\lambda^*_t^t} \epsilon (\ell + 1)
\]
and hence,
\[
\sum_{\ell=0}^{\infty} D_n(\ell, (j + 1)\eta_0 \log n) \left| \lambda^\phi_{t,\ell}(\eta \log n + a) \right| \frac{n^{\lambda^*_t}}{n^{1+[(j+1)\eta_0 \lambda^*_t]}} - c_{\phi} e^{\lambda^*_t^t} a w_{\ell}
\]
\[
\leq C'' e^{\lambda^*_t^t} \epsilon \sum_{\ell=0}^{\infty} \frac{(\ell + 1)D_n(\ell, (j + 1)\eta_0 \log n)}{n^{1+[(j+1)\eta_0 \lambda^*_t]}} \leq 2C'' e^{\lambda^*_t^t} \epsilon \frac{Z_n((j+1)\eta_0 \log n)}{n^{1+[(j+1)\eta_0 \lambda^*_t]}}.
\]
Therefore, using (8.13), the first term in the bound (8.14) converges to zero in probability. To estimate the second term in (8.14), consider the characteristic \( \chi(t) = \sum_{\ell=0}^{\infty} w_\ell 1 \{ \xi_{f_1}(t) = \ell \} \) and note that by Lemma 8.2, \( \chi \in \mathcal{C} \). Recall \( Z_n^k \) from Section 6.1 with \( \mathcal{F}_n(0) \) replaced by \( \mathcal{F}_n(j \eta_0 \log n) \) and time starting at \( T_n + j \eta_0 \log n \). As \( Z_n^k \) denotes the aggregate \( \chi \)-score of the children of all vertices born in the interval \([j \eta_0 \log n, (j + 1) \eta_0 \log n]\),

\[
\frac{1}{n^{1+(j+1)\eta_0 \lambda_1^*}} \sum_{\ell=0}^{\infty} D_n(\ell, (j+1)\eta_0 \log n) w_\ell - Z_n^k \leq \frac{C(\lambda_1^*)}{n^{1+(j+1)\eta_0 \lambda_1^*}} \sum_{\ell=0}^{\infty} (\ell + 1) D_n(\ell, j, \eta_0)
\]

\[
= \frac{Z_n(j \eta_0 \log n)}{n^{1+j\eta_0 \lambda_1^*}} \frac{C(\lambda_1^*)}{Z_n(j \eta_0 \log n)n^{\eta_0 \lambda_1^*}} \sum_{\ell=0}^{\infty} (\ell + 1) D_n(\ell, j, \eta_0) \xrightarrow{p} 0 \quad (8.15)
\]

as \( n \to \infty \) by (8.13) and Lemma 8.9 where \( C(\lambda_1^*) \) is the constant appearing in Lemma 8.2. By Theorem 6.1 (taking \( a = \eta_0 \log n \) and (8.13), if \( \eta_0 \) is chosen such that \( \frac{C e^{C \eta_0 \log n}}{\sqrt{n}} \to 0 \), where \( C, C' \) are the constants appearing in Theorem 6.1 (note that this condition on \( \eta_0 \) is independent of \( j \)),

\[
\frac{1}{n^{1+(j+1)\eta_0 \lambda_1^*}} \left| Z_n^k - \sum_{\ell=0}^{\infty} D_n(\ell, j \eta_0 \log n) \lambda_\ell(\eta_0 \log n) \right| \leq \frac{C e^{C \eta_0 \log n}}{n^{1+(j+1)\eta_0 \lambda_1^*}} \sqrt{Z_n(j \eta_0 \log n)} \leq \frac{C e^{C \eta_0 \log n}}{\sqrt{n}} \sqrt{Z_n(j \eta_0 \log n)} \xrightarrow{p} 0 \quad (8.16)
\]

where we recall \( \lambda_\ell(t) = \int_0^t m_{f_1}(s) \mu_{f_1}(d s) \). By (8.15) and (8.16), we obtain

\[
\left| \sum_{\ell=0}^{\infty} \frac{D_n(\ell, (j+1)\eta_0 \log n)}{n^{1+(j+1)\eta_0 \lambda_1^*}} w_\ell - \sum_{\ell=0}^{\infty} \frac{D_n(\ell, j \eta_0 \log n)}{n^{1+j\eta_0 \lambda_1^*}} \lambda_\ell(\eta_0 \log n) \right| \leq \frac{1}{n^{1+(j+1)\eta_0 \lambda_1^*}} \sum_{\ell=0}^{\infty} D_n(\ell, (j+1)\eta_0 \log n) w_\ell - Z_n^k \leq \frac{1}{n^{1+(j+1)\eta_0 \lambda_1^*}} \left| Z_n^k - \sum_{\ell=0}^{\infty} D_n(\ell, j \eta_0 \log n) \lambda_\ell(\eta_0 \log n) \right| \xrightarrow{p} 0 \quad (8.17)
\]

Next, we will show that

\[
e^{-\lambda_1^* t} m_{f_1}^X(t) \xrightarrow{d} 1 \quad \text{as } t \to \infty. \quad (8.18)
\]

To see this, first note that it follows from Assumption 2.1 (iii) that there exists \( \beta < \lambda_1^* \) such that \( \mathbb{E}(\xi_{f_1}(t)) \leq C e^{\beta t} \). Moreover, \( w_\ell \leq C(\ell + 1) \) for all \( \ell \geq 0 \). These observations imply

\[
\sum_{k=0}^{\infty} \sup_{t \in [k,k+1]} \left| e^{-\lambda_1^* t} E(\chi(t)) \right| \leq C \sum_{k=0}^{\infty} \sup_{t \in [k,k+1]} \left| e^{-\lambda_1^* t} \sum_{\ell=0}^{\infty} (\ell + 1) \mathbb{P}(\xi_{f_1}(t) = \ell) \right|
\]

\[
= C \sum_{k=0}^{\infty} \sup_{t \in [k,k+1]} \left| e^{-\lambda_1^* t} \mathbb{E}(\xi_{f_1}(t) + 1) \right| \leq C' \sum_{k=0}^{\infty} \sup_{t \in [k,k+1]} \left| e^{-\lambda_1^* t} e^{\beta t} \right| \leq C' e^{\beta} \sum_{k=0}^{\infty} e^{-(\lambda_1^* - \beta)k} < \infty
\]

where \( C, C' > 0 \) are constants. Thus, by Proposition 2.2 of [34] and Corollary 8.7, it follows that

\[
\lim_{t \to \infty} e^{-\lambda_1^* t} m_{f_1}^X(t) = \frac{1}{\lambda_1^* m_X} \sum_{\ell=0}^{\infty} w_\ell \lambda_1^* \int_0^{\infty} e^{-\lambda_1^* s} \mathbb{P}(\xi_{f_1}(s) = \ell) \, ds = \frac{1}{\lambda_1^* m_X} \sum_{\ell=0}^{\infty} w_\ell p_\ell^X = 1.
\]

Using this, the definition of \( \lambda_\ell^* \), the fact that \( \chi \in \mathcal{C} \) and the induction hypothesis, we obtain

\[
\frac{1}{n^{1+(j+1)\eta_0 \lambda_1^*}} \sum_{\ell=0}^{\infty} D_n(\ell, j \eta_0 \log n) \lambda_\ell^*(\eta_0 \log n) \xrightarrow{p} \sum_{\ell=0}^{\infty} p_\ell^0 w_\ell \quad \text{as } n \to \infty. \quad (8.19)
\]
From \([8.17]\) and \([8.19]\), we conclude that the second term in the bound \([8.14]\) goes to 0 as \(n \to \infty\) which proves that
\[
\left| \frac{1}{n^{1+(j+1)\eta_0+\eta}} \sum_{\ell=0}^{\infty} D_n(\ell, (j+1)\eta_0 \log n) \lambda_\ell^\phi (\eta \log n + a) - c_\phi e^{\lambda_\ell^* a} \sum_{\ell=0}^{\infty} p_\ell^0 w_\ell \right| \xrightarrow{P} 0
\]
establishing the induction hypothesis for \(j+1\). The induction hypothesis for \(j = 0\) is true by Lemma 8.4. Thus, the lemma is proved.

\[\textbf{Lemma 8.11.} \text{ For any } k \geq 0, \theta > 0 \text{ and } a \in \mathbb{R}, \text{ as } n \to \infty:\]
\[
\frac{1}{n^{1+\theta \lambda_1^*}} Z_n(\theta \log n + a) \xrightarrow{P} \frac{1}{\lambda_1^* m^*} e^{\lambda_1^* a} \sum_{\ell=0}^{\infty} p_\ell^0 w_\ell, \quad \frac{D_n(k, \theta \log n + a)}{Z_n(\theta \log n + a)} \xrightarrow{P} \frac{p_1^k}{n^{1+\lambda_1^*}}.
\]

**Proof.** The first assertion follows by the exact argument used to derive (8.13). To prove the second assertion, fix any \(k \geq 0\). Obtain \(\eta_0 > 0\) as in Lemma 8.10. Moreover, without loss of generality, assume \(\eta_0 < 1\) so that \(C \int^\infty e^{-\eta \log n} \frac{1}{n^{1+\theta \lambda_1^*}} \, d\nu \to 0\), where \(C, C', \omega, \bar{\theta}\) are as in Lemma 7.11. Let \(j \geq 0\), \(\eta \in [0, \eta_0]\) such that \(\theta = j\eta_0 + \eta\). Recall that the probability bound obtained in Lemma 8.10 (taking \(\phi(t) = 1 \{t \geq 0\}\)), \(Z_n(j\eta_0 \log n) \xrightarrow{P} 0\) if \(\eta > 0\) and \(Z_n(j\eta_0 \log n) / Z_n(\theta \log n + a) \xrightarrow{P} e^{-\lambda_1^* a}\) if \(\eta = 0\) and thus, multiplying both sides of the above by \(D_n(\ell, j\eta_0 \log n) \lambda_\ell^{(k)} (\eta \log n + a)\), we obtain
\[
\frac{D_n(k, \theta \log n + a)}{Z_n(\theta \log n + a)} - \frac{1}{Z_n(\theta \log n + a)} \sum_{\ell=0}^{\infty} D_n(\ell, j\eta_0 \log n) \lambda_\ell^{(k)} (\eta \log n + a) \xrightarrow{P} 0, \quad \text{as } n \to \infty. \tag{8.20}
\]

Taking \(\phi(t) = 1 \{\xi_\ell \geq k\}\), we see that \(\lambda_\ell^{(k)} = \lambda_\ell^*\) for each \(\ell \geq 0\). Moreover, recall from (8.9)
\[
\lim_{\ell \to \infty} e^{-\lambda_1^* t} m_{\ell, \eta_0 \log n}^{(k)} (t) = \frac{p_1^k}{\lambda_1^* m^*}.
\]

Thus, from Lemma 8.10
\[
\frac{1}{n^{1+\theta \lambda_1^*}} \sum_{\ell=0}^{\infty} D_n(\ell, j\eta_0 \log n) \lambda_\ell^{(k)} (\eta \log n + a) \xrightarrow{P} \frac{p_1^k}{\lambda_1^* m^*} e^{\lambda_1^* a} \sum_{\ell=0}^{\infty} p_\ell^0 w_\ell. \tag{8.21}
\]

Using (8.21) and the first assertion of the lemma in (8.20), the second assertion follows.

**Lemma 8.12.** \(T_n^0 - \theta \log n \xrightarrow{P} a_0\).

**Proof.** Follows immediately from the first assertion of Lemma 8.11.

**Theorem 8.13.** For any \(k \geq 0, \theta > 0\), as \(n \to \infty\),
\[
\frac{D_n(k, T_n^0)}{n^{1+\lambda_1^* \theta}} \xrightarrow{P} p_k^1.
\]
Proof. In the proof, we will abbreviate $z^* = \frac{1}{n+\epsilon} \sum_{\ell=0}^\infty p^\ell w_\ell$. Fix any $k \geq 0$, $\theta > 0$. Take any $\epsilon \in (0, 1)$. By the same argument as in the proof of Lemma 7.8

$$
\sup_{t \leq 2\epsilon} |D_n(k, \theta \log n + a_0 - \epsilon + t) - D_n(k, \theta \log n + a_0 - \epsilon)| \leq (Z_n(\theta \log n + a_0 + \epsilon) - Z_n(\theta \log n + a_0 - \epsilon)) + Y_n.
$$

(8.22)

where, conditionally on $\mathcal{F}_n(\theta \log n + a_0 - \epsilon)$, $Y_n$ is distributed as $\sum_{\ell=0}^k \text{Bin}(D_n(\ell, \theta \log n + a_0 - \epsilon), q_\ell(2\epsilon))$. Observe that by the first assertion in Lemma 8.11, for small enough $\epsilon$,

$$
\frac{Z_n(\theta \log n + a_0 + \epsilon) - Z_n(\theta \log n + a_0 - \epsilon)}{n^{1+\lambda_1^* \theta}} \xrightarrow{P} e^{\lambda_1^* \epsilon} - e^{-\lambda_1^* \epsilon} \leq 4\lambda_1^* \epsilon.
$$

(8.23)

Note that for any $C > 0$,

$$
P\left(Y_n > C\sqrt{\epsilon} n^{1+\lambda_1^* \theta}\right) \leq P\left(Y_n > C\sqrt{\epsilon} n^{1+\lambda_1^* \theta}, Z_n(\theta \log n + a_0 - \epsilon) \leq e^{-1/2} n^{1+\lambda_1^* \theta}\right)
$$

$$
+ P\left(Z_n(\theta \log n + a_0 - \epsilon) > e^{-1/2} n^{1+\lambda_1^* \theta}\right).
$$

(8.24)

For $\epsilon$ sufficiently small, by the first assertion of Lemma 8.11 as $n \to \infty$,

$$
P\left(Z_n(\theta \log n + a_0 - \epsilon) > e^{-1/2} n^{1+\lambda_1^* \theta}\right) \to 0.
$$

(8.25)

Let $\mathcal{H}_n := \mathcal{F}_n(\theta \log n + a_0 - \epsilon)$. Using Lemma 7.3

$$
\mathbb{E}(Y_n \mid \mathcal{H}_n) = \sum_{\ell=0}^k D_n(\ell, \theta \log n + a_0 - \epsilon) q_\ell(2\epsilon) \leq C' \epsilon \sum_{\ell=0}^k (\ell + 1) D_n(\ell, \theta \log n + a_0 - \epsilon)
$$

$$
\leq 2C' \epsilon Z_n(\theta \log n + a_0 - \epsilon).
$$

Thus, choosing $C > 4C'$, using Chebychev’s inequality, conditionally on $\mathcal{H}_n$ on the event $\{Z_n(\theta \log n + a_0 - \epsilon) \leq e^{-1/2} n^{1+\lambda_1^* \theta}\}$,

$$
P\left(Y_n > C\sqrt{\epsilon} n^{1+\lambda_1^* \theta} \mid \mathcal{H}_n\right) \leq P\left(Y_n - \mathbb{E}(Y_n \mid \mathcal{H}_n) > \frac{C}{2}\sqrt{\epsilon} n^{1+\lambda_1^* \theta} \mid \mathcal{H}_n\right)
$$

$$
\leq \frac{4\text{Var}(Y_n \mid \mathcal{H}_n)}{C^2 \epsilon n^{2(1+\lambda_1^* \theta)}} \leq \frac{4\sum_{\ell=0}^k D_n(\ell, \theta \log n + a_0 - \epsilon) q_\ell(2\epsilon) (1 - q_\ell(2\epsilon))}{C^2 \epsilon n^{2(1+\lambda_1^* \theta)}}
$$

$$
\leq \frac{4C' \epsilon \sum_{\ell=0}^k (\ell + 1) D_n(\ell, \theta \log n + a_0 - \epsilon)}{C^2 \epsilon n^{2(1+\lambda_1^* \theta)}} \leq \frac{8C' Z_n(\theta \log n + a_0 - \epsilon)}{C^2 \epsilon n^{2(1+\lambda_1^* \theta)}}
$$

$$
\leq \frac{8C'}{C^2 \sqrt{\epsilon} n^{1+\lambda_1^* \theta}} \to 0 \text{ as } n \to \infty.
$$

(8.26)

Using (8.25) and (8.26) in (8.24), we conclude

$$
P\left(Y_n > C\sqrt{\epsilon} n^{1+\lambda_1^* \theta}\right) \to 0 \text{ as } n \to \infty.
$$

(8.27)

Using (8.23), (8.27) and (8.22), we conclude that there exist $C_0 > 0, \epsilon_0 > 0$ such that for all $\epsilon \in (0, \epsilon_0)$,

$$
P\left(\sup_{t \leq 2\epsilon} |D_n(k, \theta \log n + a_0 - \epsilon + t) - D_n(k, \theta \log n + a_0 - \epsilon)| > C_0 \sqrt{\epsilon} n^{1+\lambda_1^* \theta}\right) \to 0 \text{ as } n \to \infty.
$$

(8.28)

From (8.28) and Lemma 8.12 as $n \to \infty$,

$$
P\left(|D_n(k, T_n^\theta) - D_n(k, \theta \log n + a_0 - \epsilon)| > C_0 \sqrt{\epsilon} n^{1+\lambda_1^* \theta}\right) \leq P\left(|T_n^\theta - \theta \log n - a_0| > 2\epsilon\right)
$$

$$
+ P\left(\sup_{t \leq 2\epsilon} |D_n(k, \theta \log n + a_0 - \epsilon + t) - D_n(k, \theta \log n + a_0 - \epsilon)| > C_0 \sqrt{\epsilon} n^{1+\lambda_1^* \theta}\right) \to 0.
$$

(8.29)
For any $\epsilon > 0$,
\[
\mathbb{P}\left(\left| \frac{D_n(k, T^\theta_n)}{n^{1+\lambda_1^\theta}} - p_k \right| > 2C_0\sqrt{\epsilon} \right) \leq \mathbb{P}\left(\left| \frac{D_n(k, T^\theta_n)}{n^{1+\lambda_1^\theta}} - \frac{D_n(k, \theta \log n + a_0 - \epsilon)}{n^{1+\lambda_1^\theta}} \right| > C_0\sqrt{\epsilon} \right) + \mathbb{P}\left(\left| \frac{D_n(k, \theta \log n + a_0 - \epsilon)}{n^{1+\lambda_1^\theta}} - p_k \right| > C_0\sqrt{\epsilon} \right) \quad (8.30)
\]
By Lemma 8.11,
\[
\frac{D_n(k, \theta \log n + a_0 - \epsilon)}{n^{1+\lambda_1^\theta}} = \frac{D_n(k, \theta \log n + a_0 - \epsilon)}{Z_n(\theta \log n + a_0 - \epsilon)} \frac{Z_n(\theta \log n + a_0 - \epsilon)}{n^{1+\lambda_1^\theta}} p_k^1 e^{-\lambda_1^\theta},
\]
and therefore, there is $\epsilon_1 > 0$ such that for all $\epsilon \in (0, \epsilon_1)$,
\[
\frac{D_n(k, \theta \log n + a_0 - \epsilon)}{n^{1+\lambda_1^\theta}} - p_k^1 \quad \frac{p_k^1 (1 - e^{-\lambda_1^\theta}) \leq p_k^1 \lambda_1^\theta \epsilon < C_0\sqrt{\epsilon}}{\leq C_0\sqrt{\epsilon}} \quad (8.31)
\]
For $\epsilon \in (0, \epsilon_1)$, using (8.29) and (8.31) in (8.30), we conclude
\[
\mathbb{P}\left(\left| \frac{D_n(k, T^\theta_n)}{n^{1+\lambda_1^\theta}} - p_k^1 \right| > 2C_0\sqrt{\epsilon} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty
\]
proving the theorem.

8.2. **Proof of Theorem 3.16** We will prove (a) of the Theorem. The remaining results follow via straightforward modifications of the arguments for (a). For (a) recall that we first grow the tree using the uniform attachment scheme with $f_0 \equiv 1$ till it is of size $n^\gamma$ and then use the preferential attachment scheme. We will assume that $\mathcal{T}_n^\theta$ has been constructed as follows:

(a) Generate the genealogical tree according to a rate one Yule process $\{\mathcal{T}^{\text{Yule}}(t) : t \geq 0\}$ as in Definition 5.3 run for ever.

(b) To obtain $\mathcal{T}_n^\theta$, let $\mathcal{T}_n^\theta = \mathcal{T}^{\text{Yule}}(T_n^\theta)$. Now every vertex in $\mathcal{T}_n^\theta$ switches to offspring dynamics giving birth to children at rate corresponding to the number of children $+1 + \alpha$ (thus modulated by the function $f_1$). Write $\text{BP}_n(\cdot)$ for the combined process and stop this process at time $T_n$ and let $\mathcal{T}_n^\theta = \text{BP}_n(T_n)$.

The following describes asymptotics for the above continuous time construction.

**Proposition 8.14.** For the process $\text{BP}_n(\cdot)$ as constructed above:

(a) The stopping time $T_n^\theta$ satisfies,
\[
T_n^\theta - \gamma \log n \xrightarrow{a.e.} \tilde{W},
\]
where $\tilde{W} = -\log W$ and $W = \exp(1)$.

(b) Let $\omega_n \rightarrow \infty$ arbitrarily slowly. Then there exists a constant $C > 0$ independent of $\omega_n$ such that
\[
\mathbb{P}\left(\sup_{t \geq 0} \left| e^{-(2+\alpha)t} \frac{\text{BP}_n(t + T_n^\theta)}{n^\gamma} - 1 \right| > \frac{\omega_n}{n^{\gamma/2}} \right) \leq \frac{C}{\omega_n^2}.
\]
In particular whp as $n \rightarrow \infty$,
\[
T_n - \frac{1 - \gamma}{2 + \alpha} \log n \leq \frac{\omega_n}{n^{\gamma/2}}.
\]

**Proof.** Part(a) follows from Lemma 5.4. To prove (b), recall that for $t > T_n^\theta$, all individuals switch to offspring dynamics modulated by $f_1$. For the rest of the proof, we proceed conditional on the history of the process till time $T_n^\theta$. Using Proposition 5.7,
\[
M_1(t) := \left( e^{-(2+\alpha)t} |\text{BP}_n(t + T_n^\theta) - n^\gamma | + \frac{1 - e^{-(2+\alpha)t}}{(2 + \alpha)} \right) \quad t \geq 0,
\]
and

\[ M_2(t) := e^{-2(2+\alpha)t}BP_n(t + T_{n'})^2 - \int_0^t \alpha e^{-2(2+\alpha)s}BP_n(s + T_{n'})ds - \frac{e^{-2(2+\alpha)t}}{2(2+\alpha)}, \quad t \geq 0, \]

are martingales. Using these expressions, it can be deduced that

\[ \sup_{t \geq 0} \mathbb{E}[M_1^2(t)] \leq C n^T \]

for some constant \( C > 0 \). An appeal to Doob's \( L^2 \)-maximal inequality then proves the first assertion of Proposition 8.14(b) which then results in the second assertion.

\[ \Box \]

Fix constant \( B \) and a sequence \( \omega_n = o(n^{\gamma/2}) \uparrow \infty \) and consider the following construction \( \tilde{\mathcal{F}}_n^+(B, \omega_n) \) related to the above continuous time construction of \( \mathcal{F}_n^\theta \):

(a) Run a rate one Yule process for time \( \gamma \log n + B \).

(b) Now every vertex in the Yule process switches dynamics so that it reproduces at rate equal to the number of children \( +1 + \alpha \). Grow this process for an additional time \( t_n^\omega := \frac{1-\gamma n}{2+\alpha} \log n + \frac{\omega_n}{n^{\gamma/2}} \).

Analogously define \( \tilde{\mathcal{F}}_n^-(B, \omega_n) \) where in the above construction we wait till time \( \log n - B \) before switching dynamics and run the new dynamics for time \( t_n^- := \frac{1-\gamma n}{2+\alpha} \log n - \frac{\omega_n}{n^{\gamma/2}} \). By Proposition 8.14, given any \( \epsilon > 0 \) we can choose a constant \( B = B(\epsilon) \) such that for any \( \omega_n \uparrow \infty \), we can produce a coupling between \( \mathcal{F}_n^\theta \) and \( \tilde{\mathcal{F}}_n^+(B, \omega_n) \) such that for all large \( n \), with probability at least \( 1 - \epsilon \), \( \mathcal{F}_n^\theta \subseteq \tilde{\mathcal{F}}_n^+(B, \omega_n) \) where we see the object on the left as a subtree of the object on the right with the same root. A similar assertion holds with \( \tilde{\mathcal{F}}_n^-(B, \omega_n) \subseteq \mathcal{F}_n^\theta \). Using these couplings, the following Proposition completes the proof of Theorem 3.16 with part(a) of the Proposition proving the lower bound while part(b) proving the upper bound.

**Proposition 8.15.** Fix \( B > 0 \) and \( \omega_n = o(\log n) \uparrow \infty \).

(a) Consider the degree of the root \( D_n^-(\rho) \) in \( \tilde{\mathcal{F}}_n^- (B, \omega_n) \). Then \( D_n^-(\rho) \gg n^{(1-\gamma)/(2+\alpha)} \log n / \omega_n \) whp.

(b) Consider the maximal degree \( M_n^+(1) \) in \( \tilde{\mathcal{F}}_n^+ (B, \omega_n) \). Then \( \exists A > 0 \) such that whp as \( n \to \infty \), \( M_n^+(1) \ll A n^{(1-\gamma)/(2+\alpha)} (\log n)^2 \).

**Proof:** We start with (a). Note that each individual in the original Yule process reproduces according to a rate one Poisson process. In particular standard bounds for a Poisson random variable implies that the degree of the root \( D_n^- \) is a rate one Poisson process. In particular standard bounds for a Poisson random variable implies that

\[ |\deg_n(\rho, \gamma \log n - B) - \gamma \log n| = O_P(\sqrt{\log n}). \]  

(8.32)

Now let \( \{Y_i(\cdot) : i \geq 1\} \) be a collection of independent rate one Yule processes. Comparing rates, the degree of the root after \( \gamma \log n - B \) we get that

\[ \deg_n(\gamma \log n - B + \cdot) \geq \sum_{i=1}^{\deg_n(\rho, \gamma \log n - B)} Y_i(\cdot), \]  

(8.33)

Using (8.32), Lemma 5.4 and standard tail bounds for the Geometric distribution now completes the proof.

Let us now prove (b). Recall that after the change point, dynamics are modulated by \( f_1(\cdot) = \cdot + 1 + \alpha \). Let \( A \) denote the smallest integer \( \geq \alpha + 1 \). Let \( \xi_{f_1} \) be the corresponding continuous time offspring process. Comparing rates we see that

\[ \xi_{f_1}(\cdot) \leq \sum_{i=1}^{A+2} Y_i(\cdot), \]  

(8.34)

where as before \( \{Y_i(\cdot) : i \geq 1\} \) is a collection of independent rate one Yule processes. For every vertex \( v \) write \( \deg_n(v) \) for the degree of the vertex at time \( \log n + B + t_n^\omega \) when we have finished constructing the process \( \tilde{\mathcal{F}}_n^+ (B, \omega_n) \). Abusing notation, write \( T_v \) for the time of birth of vertex \( v \). We will break up the proof of (b) into two cases:
(b1) Maximal degree for vertices born after \( \log n + B \): Define

\[ \mathcal{A}_n = \left\{ v \in \mathcal{T}_n^+(B, \omega_n) : T_v \in [\log n + B, \log n + B + t_n^+] , \text{ } \deg_n(v) > C n^{1 - \gamma} (\log n)^2 \right\}, \]

where \( C \) is an appropriate large constant that will be chosen later. The aim is to show that we can choose \( C \) such that \( \mathbb{P}(|\mathcal{A}_n|) \to 0 \) as \( n \to \infty \). This would then imply

\[ \mathbb{P}(\exists v \in \mathcal{T}_n^+(B, \omega_n), T_v \geq \log n + B \text{ } \deg_n(v) > C n^{1 - \gamma} (\log n)^2) \to 0. \]  

(8.35)

Let \( k_n := C n^{1 - \gamma} (\log n)^2 \) and let \( \mathcal{T}_n^+(t) \) denote the tree at time \( t \). Since the offspring distribution of each new vertex born at \( t > \log n + B \) is a Yule process then, by Lemma 5.4 the probably a new vertex has degree greater than \( k_n \) by time \( t_n^+ \) is given by

\[ P(\text{Geom}(e^{t - t_n^+}) \geq k_n) \leq e^{k_n e^{t - t_n^+}} \]

Note that new vertices are produced at rate \( (2 + \alpha) |\mathcal{T}_n^+(t)| - 1 \). As in the proof of Proposition 8.14 \( M(t) := e^{-(2 + \alpha) t} |\mathcal{T}_n^+(t)| + \frac{1}{(2 + \alpha)} e^{-(2 + \alpha) t}, t \geq \log n + B \) is a martingale. Noting \( \mathbb{E}[\mathcal{T}_n^+(\log n + B)] = \mathbb{E}B ) \) we get that

\[ \mathbb{E}[\mathcal{T}_n^+(t)] = C' n^y e^{(2 + \alpha) t} \text{ for } t \geq \log n + B \]

where \( C' \) is a constant depending only on \( B, \alpha \). Thus

\[ \mathbb{E}(|\mathcal{A}_n|) \leq C'' n^y \int_0^{t_n^+} e^{k_n e^{t - t_n^+}} e^{(2 + \alpha) t} \, dt \]

where \( C'' \) depends only on \( B, \alpha \) and it is sufficient to check the following lemma.

**Lemma 8.16.** Let

\[ I_n := n^y \int_0^{t_n^+} e^{-C(\log n)^2 n^{1 - \gamma} e^{t - t_n^+}} e^{(2 + \alpha) t} \, dt \]  

(8.36)

**Proof.** Writing \( a := \frac{1 - \gamma}{2 + \alpha} \) and \( b := 2 + \alpha \), algebraic manipulations result in the form:

\[ I_n \leq n^y (\log n)^{-2b} e^{b \frac{m_n}{n^{1/2}}} \Gamma b, C(\log n)^2 e^{-\frac{m_n}{n^{1/2}}} := \mathcal{E}_n. \]  

(8.37)

where \( \Gamma(b, z) = \int_0^\infty e^{-t} t^{b-1} \, dt \) is the upper incomplete Gamma function. Known asymptotics for the incomplete Gamma function \( \Gamma(b, z) = \Omega(z^{b-1} e^{-z}) \) as \( z \to \infty \) imply

\[ \mathcal{E}_n \sim n^y \log ne^{ -\frac{m_n}{n^{1/2}}} (\log n)^{-2} e^{ -\frac{m_n}{n^{1/2}}} \to 0. \]

\( \blacksquare \)

(b2) Maximal degree for vertices born before \( \log n + B \): We prove that vertices born before \( \gamma \log n + B \) cannot have too large of a maximal degree in \( \mathcal{T}_n^+(B, \omega_n) \). To simplify notation, write the following for the two times:

\[ \Delta_n := \gamma \log n + B, \quad Y_n := \gamma \log n + B + t_n^+. \]  

(8.38)

Further write \( \deg(v, t) \) for the degree of a vertex \( v \) at time \( t \) with the convention that \( \deg(v, t) := 0 \) for \( t < T_v \). Write \( \deg_n(v) := \deg(v, Y_n) \) for the final degree of \( v \) in \( \mathcal{T}_n^+(B, \omega_n) \). Finally in the construction of the tree \( \mathcal{T}_n^+(B, \omega_n) \), for any \( 0 \leq t \leq Y_n \), write \( \mathcal{T}_n^+(t) \) for the tree at time \( t \).

Fix \( C > 0 \) and let \( \mathcal{B}_n \) be the set of vertices born before \( \log n + B \) whose final degree is too large i.e.

\[ \mathcal{B}_n := \{ v \in \mathcal{B}_n : T_v \leq \log n + B, \deg_n(v) > C n^{1 - \gamma} (\log n)^2 \}. \]

where \( \deg_n(v) \) is the degree of vertex \( v \) in the final tree \( \mathcal{T}_n^+(B, \omega_n) \).

**Proposition 8.17.** We can choose \( C < \infty \) such that \( \mathbb{P}(\mathcal{B}_n \geq 1) \to 0 \) as \( n \to \infty \).
The plan is as follows: we control the maximal degree of vertices born in the early (pre $\Delta_n$) tree then show that none of these early vertices have time to accumulate too many edges in the remaining $Y_n - \Delta_n$ time period.

**Proof.** Consider the tree $\tilde{T}^+(\Delta_n)$. Let $M_n(\Delta_n) := \max_{v \in \tilde{T}^+(\Delta_n)} \deg(v, \Delta_n)$ be the maximal degree of vertices in $\tilde{T}^+(\Delta_n)$ at time $\Delta_n$. Let $\ell_n := 10 \log n$ and fix a sequence $\omega_n \rightarrow \infty$. By the union bound,

$$\mathbb{P}(B_n \geq 1) \leq \mathbb{P}(B_n \geq 1, |\tilde{T}^+(\Delta_n)| < \omega_n n^{\gamma}, M_n \leq \ell_n) + \mathbb{P}(|\tilde{T}^+(\Delta_n)| \geq \omega_n n^{\gamma}) + \mathbb{P}(M_n > \ell_n)$$

Lemmas 8.18 and 8.19 which bound the three terms on the right complete the proof of the Proposition.

**Lemma 8.18.** For $C$ large enough $\mathbb{P}(B_n \geq 1, |\tilde{T}^+(\Delta_n)| < \omega_n n^{\gamma}, M_n \leq \ell_n) \rightarrow 0$ as $n \rightarrow \infty$.

**Proof.** Let $G_n = |\tilde{T}^+(\Delta_n)| < \omega_n n^{\gamma}, M_n \leq \ell_n)$. It is sufficient to show $\mathbb{P}(B_n \geq 1|G_n) \rightarrow 0$. Conditional on $G_n$, we will construct a random variable that stochastically bounds the growth of degrees in the process $\tilde{T}^+(t)$ for $t \geq \Delta_n$. Let $\{X_i(\cdot) : 1 \leq i \leq n^{\gamma} \omega_n\}$ be a collection of independent rate one Yule processes each starting with $\ell_n + \lceil a \rceil$ individuals at time 0 and run each for time $t^*_i = \frac{1 - \gamma}{2 + \alpha} \log n + \frac{\omega_n}{n^{\gamma \alpha}}$. Consider $\mathcal{M} = \max_{1 \leq i \leq n^{\gamma} \omega_n} X_i(t^*_i)$.

On the event $G_n$, the degree evolution of $\tilde{T}^+(\Delta_n)$ after time $\Delta_n$ is as follows: Sample $\tilde{T}^+(\Delta_n)$ conditional on $G_n$ i.e. the event that there are fewer than $\omega_n n^{\gamma}$ vertices and the maximal degree is less than $\ell_n$. For each vertex, $v$, in $\tilde{T}^+(\Delta_n)$ we run an independent, rate 1 Yule process starting with $\deg(v, \Delta_n) + \alpha$ individuals for time $t^*_i$. Our new process starts each Yule process as if each individual has maximal degree at time $\gamma \log n + B$. In particular on the event $G_n$, the maximal degree $M_n(Y_n)$ at time $Y_n$ satisfies $M_n(Y_n) \leq \mathcal{M}$. The rest of the proof analyzes $\mathcal{M}$. Using the union bound gives,

$$\mathbb{P}(B_n \geq 1|G_n) \leq \mathbb{P}(\mathcal{M} \geq C n^{\frac{1+\gamma}{2+\alpha}} (\log n)^2) \leq \omega_n n^{\gamma} \mathbb{P}(X_i(t^*_i) \geq C n^{\frac{1+\gamma}{2+\alpha}} (\log n)^2)$$

Now for a rate one Yule process started with $m$ individuals at time zero say $Y^m(\cdot)$ for fixed $t$, $Y^m(t)$ is distributed as the sum of $m$ iid geometric random variables with $p = e^{-t}$. Thus

$$\mathbb{P}(Y^m(t) > \lambda) \leq m \mathbb{P}(\text{geom}(e^{-t}) > \lambda/m) \leq m \exp \left[ -\frac{\lambda}{m} e^{-t} \right].$$

Plugging in $m = \ell_n + \lceil a \rceil$, $t = t^*_i$, $\lambda = C n^{\frac{1+\gamma}{2+\alpha}} (\log n)^2$ we get,

$$\omega_n n^{\gamma} \mathbb{P}(X_i(t^*_i) \geq C n^{\frac{1+\gamma}{2+\alpha}} (\log n)^2) \leq K \omega_n n^{\gamma} \log n n^{-C}$$

which goes to zero for sufficiently large $C$.

**Lemma 8.19.** For $C$ large enough as $n \rightarrow \infty$,

$$\mathbb{P}(|\tilde{T}^+(\Delta_n)| \geq \omega_n n^{\gamma}) \rightarrow 0, \quad \mathbb{P}(M_n(\Delta_n) > \ell_n) \rightarrow 0.$$

**Proof.** We first prove the assertion on $|\tilde{T}^+(\Delta_n)|$. Note the size of the tree grows according to a rate one Yule process. Thus by Lemma 5.4 $|\tilde{T}^+(\Delta_n)| \sim \text{Geom}(e^{-\gamma \log n - B})$. Thus

$$\mathbb{P}(|\tilde{T}^+(\Delta_n)| \geq \omega_n n^{\gamma}) \leq \exp \left[ -\omega_n n^{\gamma} e^{-\gamma \log n - B} \right] \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

For the second assertion, note that for any $0 \leq t \leq \Delta_n$, the rate at which a new vertex is born is $|\tilde{T}^+(t)|$. Since the offspring distribution of each new vertex (before time $\Delta_n$) is a Poisson process, the probability that this new vertex has degree greater than $\ell_n$ conditional on $\tilde{T}^+(t)$ is

$$\mathbb{P}(\text{Poisson}(\Delta_n - t) \geq \ell_n) \leq \mathbb{P}(\text{Poisson}(\Delta_n) \geq \ell_n).$$
Thus writing \( N_n(\Delta_n) \) for the number of vertices with degree at least \( \ell_n \) by time \( \Delta_n \) and recalling that for \( t \leq \Delta_n \), \( \mathbb{E}(\mathcal{F}_n^+(t)) = e^t \) we have,

\[
\mathbb{E}(N_n(\Delta_n)) = \int_0^{\Delta_n} \mathbb{P}(\text{Poisson}(\Delta_n - t) \geq \ell_n) e^t \, dt \leq \frac{e^B}{n^\gamma} \mathbb{P}(\text{Poisson}(\Delta_n) \geq \ell_n).
\]

Since \( \Delta_n = \gamma \log n + B \) with \( \gamma < 1 \), exponential tail bounds for the Poisson distribution completes the proof.

\[ \blacksquare \]

9. Proofs: Convergence rates for model without change point

This section is dedicated to proving Theorem 3.3 and Theorem 3.4.

**Lemma 9.1.** Consider a continuous time branching process with attachment function \( f \) that satisfies Assumption 2.1. Fix \( \beta \in (0, \lambda^\star) \). There exist positive constants \( C_1, C_2 \) such that if \( h \) solves the renewal equation

\[
h(t) = e^{-\lambda^\star t} \phi(t) + \int_0^t h(t-s)e^{-\lambda^\star s} \mu_f(ds)
\]

with any \( \phi \) satisfying \( |\phi(s)| \leq C_\phi e^{\beta s} \) for all \( s \geq 0 \), for some \( C > 0 \), denoting \( h(\infty) = \lim_{t \to \infty} h(t) \), we have for all \( t \geq 0 \),

\[
|h(\infty) - h(t)| \leq C_1 C_\phi e^{-C_2 t}.
\]

**Proof.** We will use estimates about quantitative rates of convergence for renewal measures derived in [9] in the setting of the point process with i.i.d. inter-arrival times having distribution \( e^{-\lambda^\star s} \mu_f(ds) \). By Assumption 2.1 (iii), it is clear that the measure \( e^{-\lambda^\star s} \mu_f(ds) \) satisfies \( \int_0^\infty e^{\beta s} e^{-\lambda^\star s} \mu_f(ds) < \infty \) for some \( \beta' > 0 \). We thus, Assumption 1 of [9] is satisfied. Moreover, for any Borel set \( A \) in \([0,1]\), denoting by \( E \) the first time the root reproduces (which has an exponential distribution with rate \( f(0) \)), note that

\[
\mu_f(A) \geq \mathbb{E}(1 \{ E \in A \}) = \int_A f(0)e^{-f(0)x} \, dx \geq f(0)e^{-f(0)} \int_A d x
\]

and consequently, the distribution of the inter-arrival time is spread out in the sense of Assumption 2 of [9] taking \( c = 1/2, L = 1/2 \) and \( \overline{\mu} = f(0)e^{-(\lambda^\star + f(0))} \). Thus, Corollary 1 of [9] holds for the point process under consideration. For any \( x \geq 0 \), denote by \( U^x \) the renewal measure corresponding to the associated point process with time started at \( x \). The stationary version of this point process corresponds to a random starting time whose law is \( \mu^\star(ds) = m^{\star-1} e^{-\lambda^\star s} \mu_f(ds) \) (called the stationary delay distribution), where \( m^\star = \int_0^\infty u e^{-\lambda^\star u} \mu_f(du) \). From translation invariance, it follows that the renewal measure associated to this stationary version is given by \( U^\star(ds) = m^\star ds \). By Corollary 1 of [9], there exist constants \( C, C' > 0 \) and \( \beta'' < \beta' \) such that for any Borel set \( D \subset (0,\infty) \) and any \( x, t \geq 0 \),

\[
|U^x(D + t) - U^0(D + t)| \leq C e^{\beta'' x} e^{-C't}(U^0((0, \sup D)) + 1).
\]

Integration both sides of the above relation over \( x \) with respect to the stationary delay distribution \( \mu^\star(dx) \) and using Fubini’s theorem and the fact that \( \int_0^\infty e^{\beta'' s} e^{-\lambda^\star s} \mu_f(ds) < \infty \), we obtain

\[
|U^\star(D + t) - U^0(D + t)| \leq C e^{-C't}(U^0((0, \sup D)) + 1).
\]

This, in turn, implies that for any \( t \geq 0 \), if \( U^\star_{M,t} \) and \( U^0_{M,t} \) denote the measures defined by \( U^\star_{M,t}(D) = U^\star(D + t) \) and \( U^0_{M,t}(D) = U^0(D + t) \) for any Borel set \( D \subset [0, M] \), then using the fact that \( \lim_{t \to \infty} t^{-1} U^0((0, t)) = \frac{1}{m^\star} \) (which follows from the elementary renewal theorem),

\[
||U^\star_{M,t} - U^0_{M,t}||_{TV} \leq C M e^{-C't}.
\]
From standard results in renewal theory, observe that \( h(\infty) = \int_0^\infty e^{-\lambda^* s} \phi(s)U^*(ds) \) and \( h(t) = \int_0^t e^{-\lambda^*(t-s)} \phi(t-s)U^0(ds) \). Thus, for \( t \geq 0 \),
\[
|h(\infty) - h(t)| = \left| \int_0^\infty e^{-\lambda^* s} \phi(s)U^*(ds) - \int_0^t e^{-\lambda^*(t-s)} \phi(t-s)U^0(ds) \right| \\
\leq \left| \int_0^t e^{-\lambda^*(t-s)} \phi(s)U^*(ds) - \int_0^t e^{-\lambda^*(t-s)} \phi(t-s)U^0(ds) \right| + \int_t^\infty e^{-\lambda^* s} \phi(s)U^*(ds). \tag{9.2}
\]
As \( |\phi(s)| \leq C_\phi e^{bs} \) for all \( s \),
\[
\int_t^\infty e^{-\lambda^* s} \phi(s)U^*(ds) \leq C_\phi m^* e^{-|\lambda^*-\beta|s} ds = \frac{C_\phi}{m^* (\lambda^* - \beta)} e^{-(\lambda^* - \beta)t}. \tag{9.3}
\]
To estimate the first term in the bound (9.2), note that for \( t \geq 0 \),
\[
\begin{align*}
&\left| \int_0^t e^{-\lambda^* s} \phi(s)U^*(ds) - \int_0^t e^{-\lambda^*(t-s)} \phi(t-s)U^0(ds) \right| \\
&= \left| \int_0^t e^{-\lambda^*(t-s)} \phi(t-s)U^*(ds) - \int_0^t e^{-\lambda^*(t-s)} \phi(t-s)U^0(ds) \right| \\
&\leq \int_0^{t/2} e^{-\lambda^*(t-s)} \phi(t-s)U^*(ds) + \int_0^{t/2} e^{-\lambda^*(t-s)} \phi(t-s)U^0(ds) \\
&\quad + \int_0^{t/2} e^{-\lambda^*(t-s)} \phi(t-s)U^*(ds) - \int_0^{t/2} e^{-\lambda^*(t-s)} \phi(t-s)U^0(ds) \\
&\leq C_\phi e^{-(\lambda^* - \beta)t/2} U^*([0, t/2]) + C_\phi e^{-(\lambda^* - \beta)t/2} U^0([0, t/2]) + C_\phi |U^*_{t/2,t/2} - U^0_{t/2,t/2}| TV \leq C_1' C_\phi e^{C_2' t} \tag{9.4}
\end{align*}
\]
for constants \( C_1', C_2' > 0 \) not depending on \( \phi \), where we have used (9.1) along with the observations that \( U^*(][0, t/2] = \frac{t}{2m^*} \) and \( \lim_{t \to \infty} t^{-1} U^0([0, t/2]) = \frac{1}{2m^*} \). The lemma follows by using (9.3) and (9.4) in (9.2).

**Proof of Theorem 3.4** In the proof, \( C, C', C'' \), \( \sigma_X, \sigma_x, \beta' \), \( \beta \) will denote generic positive constants not depending on \( b_\phi \) and the specific choice of \( \phi \). Following [34], we write \( x = (x', i) \) to denote that \( x \) is the \( i \)-th child of \( x' \) and define for all \( c \geq 0 \),
\[
\mathcal{J}(t) = \{ x = (x', i) : \sigma_{x'} \leq t \} \quad \text{and} \quad \mathcal{J}_t = \mathcal{J}(t) \setminus \mathcal{J}_c \quad \text{and} \quad \mathcal{J}_c = \{ x = (x', i) : \sigma_{x'} \leq t \} \quad \text{and} \quad t + c < \sigma_x < \infty).
\]
Let \( T_t \) denote the number of vertices born by time \( t \) and let \( \mathcal{A}_n \) be the filtration generated by the entire life histories of the first \( n \) vertices (see [34] for detailed definitions). Define \( \mathcal{F}_t = \mathcal{A}_t \). For any \( s > 0 \), write \( \phi = \phi_s + \phi'_s \) where \( \phi_s(u) = \phi(u) \mathbf{1} \{ u < s \} \) and \( \phi'_s(u) = \phi(u) \mathbf{1} \{ u \geq s \} \). Note that
\[
\mathbb{E} \left[ e^{-\lambda^* t} Z_f^\phi(t) - W_\infty M_f^\phi(\infty) \right] \leq \mathbb{E} \left[ e^{-\lambda^* t} \left( Z_f^\phi(t) - Z_f^\phi(t) \right) \right] + \mathbb{E} \left[ e^{-\lambda^* t} Z_f^\phi(t) - W_\infty M_f^\phi(\infty) \right] \\
+ \mathbb{E} \left[ M_f^\phi(\infty) - M_f^\phi(\infty) \right] \mathbb{E} \left[ W_\infty \right]. \tag{9.5}
\]
The third term in the bound (9.5) can be bounded as
\[
\mathbb{E} \left[ M_f^\phi(\infty) - M_f^\phi(\infty) \right] W_\infty = M_f^\phi(\infty) = \frac{1}{m^*} \int_s^\infty e^{-\lambda^* u} \mathbb{E} (\phi(u)) du \\
\leq \frac{b_\phi}{m^*} \int_s^\infty e^{-\lambda^* u} \mathbb{E} (\xi_f(u) + 1) du \leq Cb_\phi e^{-(\lambda^* - \beta)s} \tag{9.6}
\]
for some \( \beta' < \lambda^* \) by virtue of Assumption 2.1(iii). The first term in the bound (9.5) can be bounded as
\[
\mathbb{E} \left[ e^{-\lambda^* t} \left( Z_f^\phi(t) - Z_f^\phi(t) \right) \right] = \mathbb{E} \left[ e^{-\lambda^* t} Z_f^\phi(t) \right] \leq M_f^\phi(\infty) + M_f^\phi(\infty). \tag{9.7}
\]
By the fact that $M_f^{\phi^t}(t)$ satisfies the renewal equation (3.3) (with $\phi^t$ in place of $\phi$) and Lemma 9.1 for $t \geq 0$,

$$\left|M_f^{\phi^t}(t) - M_f^{\phi^t}(\infty)\right| \leq C_1 b_\phi e^{-C_2 t}.$$  
Using this estimate and (9.6) in (9.7), we obtain

$$E \left|e^{-\lambda^* t} \left(Z_f^{\phi^t}(t) - Z_f^{\phi^t}(t)\right)\right| \leq C_1 b_\phi e^{-C_2 t} + C b_\phi e^{-\lambda^* t}.$$  
Using (9.6) and (9.8) in (9.5), for any $t, s \geq 0$,

$$E \left|e^{-\lambda^* t} Z_f^{\phi^t}(t) - W_{\infty} M_f^{\phi^t}(\infty)\right| \leq E \left|e^{-\lambda^* t} Z_f^{\phi^t}(t) - W_{\infty} M_f^{\phi^t}(\infty)\right| + C_1 b_\phi e^{-C_2 t} + 2C b_\phi e^{-\lambda^* t}.$$  
Now, we estimate the first term in the above bound. Observe that as $\phi(0) = 0$ for all $u \geq s$, every individual that contributes to $Z_f^{\phi^t}(t + s)$ must be born after time $t$. Therefore,

$$Z_f^{\phi^t}(t + s) = \sum_{x \in S(t)} Z_{f,x}^{\phi^t}(t + s - \sigma_x)$$  
where for any vertex $x$ and any $u \geq 0$, $Z_{f,x}^{\phi^t}(u)$ denotes the aggregate $\phi$-score at time $\sigma_x + u$ treating the vertex $x$ as the root.

For $t, c \geq 0$ such that $s \geq c$, write

$$X(t, s, c) = \sum_{x \in S(t) \setminus S(t,c)} e^{-\lambda^* s_x} \left(e^{-\lambda^* (t + s - \sigma_x)} Z_{f,x}^{\phi^t}(t + s - \sigma_x) - M_f^{\phi^t}(t + s - \sigma_x)\right).$$  
and write $W_t = \sum_{x \in S(t)} e^{-\lambda^* s_x}$, $W_{t,c} = \sum_{x \in S(t) \setminus S(t,c)} e^{-\lambda^* s_x}$. Following equation (3.36) in [34], we obtain

$$\left|e^{-\lambda^* (t + s)} Z_f^{\phi^t}(t + s) - W_{\infty} M_f^{\phi^t}(\infty)\right| \leq |X(t, s, c)| + \sum_{x \in S(t) \setminus S(t,c)} e^{-\lambda^* s_x} \left|M_f^{\phi^t}(t + s - \sigma_x) - M_f^{\phi^t}(\infty)\right|$$

$$+ \left|\sum_{x \in S(t,c)} e^{-\lambda^* s_x} \left(e^{-\lambda^* (t + s - \sigma_x)} Z_{f,x}^{\phi^t}(t + s - \sigma_x) - M_f^{\phi^t}(\infty)\right)\right| + M_f^{\phi^t}(\infty) |W_t - W_{\infty}|.$$  
Note that

$$\text{Var}(X(t, s, c) | S_t) = \sum_{x \in S(t) \setminus S(t,c)} e^{-2\lambda^* s_x} V_f^{\phi^t}(t + s - \sigma_x)$$  
where $V_f^{\phi^t}(t) = \text{Var}(e^{-\lambda^* t} Z_f^{\phi^t}(t))$. Recall $M_f^{\phi^t}(t) = E\left(Z_f^{\phi^t}(t)\right)$ and $V_f^{\phi^t}(t) = \text{Var}\left(Z_f^{\phi^t}(t)\right)$. From Theorem 3.2 of [29], $V_f^{\phi^t}(t) = h \ast U(t)$, where

$$h(t) = \text{Var}(\phi(t) + \int_0^t m_f^{\phi^t}(t-u) \xi_f(u) du)$$
and $U(\cdot) = \sum_{t=0}^{\infty} \mu^{\phi^t}(\cdot)$ denotes the renewal measure.

As $\phi(t) \leq b_\phi (\xi_f(t) + 1)$ for all $t$ and Assumption 3.2 holds,

$$e^{-2\lambda^* t} E(\phi(t))^2 \leq (b_\phi)^2 E\left(e^{-\lambda^* t} (1 + \xi_f(t))\right)^2 \leq 2(b_\phi)^2 E\left(e^{-2\lambda^* t} + \lambda^* 2\int_0^\infty e^{-\lambda^* u} \xi_f(u) du\right)^2 \leq C(b_\phi)^2.$$  
As $\text{E}(\xi_f(t) + 1) \leq C e^{\beta t}$ by Assumption 2.1 (iii), therefore $E(\phi(t)) \leq b_\phi E(\xi_f(t) + 1) \leq b_\phi C e^{\beta t}$. Hence, by the fact that $M_f^{\phi^t}(t)$ satisfies the renewal equation (3.3) and Lemma 9.1 for $t \geq 0$,

$$\left|M_f^{\phi^t}(t) - M_f^{\phi^t}(\infty)\right| \leq C_1 b_\phi e^{-C_2 t}.$$  
Moreover,

$$M_f^{\phi^t}(\infty) = \int_0^\infty e^{-\lambda^* u} E(\phi(u)) du \leq \frac{b_\phi}{m^*} \int_0^\infty E\left(e^{-\lambda^* u} (1 + \xi_f(u))\right) du \leq C b_\phi.$$
Using (9.13) and (9.14), we obtain for all \( t \geq 0 \),
\[
M_f^{\phi}(t) \leq C' b_\phi. \tag{9.15}
\]

From (9.12) and (9.15), we conclude for all \( t \geq 0 \),
\[
e^{-2\lambda^* t} h(t) = \text{Var}\left(e^{-\lambda^* t} \phi_f(t) + \int_0^t e^{-\lambda^* (t-u)} m_f^{\phi_f}(t-u)e^{-\lambda^* u} \xi_f(du)\right)
\leq 2e^{-2\lambda^* t} \mathbb{E}(\phi_f(t))^2 + 2\mathbb{E}\left(\int_0^t M_f^{\phi_f}(t-u)e^{-\lambda^* u} \xi_f(du)\right)^2
\leq 2C(b_\phi)^2 + 2(Cb_\phi)^2 \mathbb{E}\left(\int_0^\infty e^{-\lambda^* u} \xi_f(du)\right)^2 \leq C'(b_\phi)^2.
\]

Thus, for all \( t \geq 0 \),
\[
V_f^{\phi_f}(t) = \int_0^\infty e^{-2\lambda^* (t-u)} h(t-u)e^{-2\lambda^* u} U(du) \leq C'(b_\phi)^2 \int_0^\infty e^{-2\lambda^* u} U(du) = C'(b_\phi)^2 \sum_{\ell=0}^\infty \hat{\mu}_f(2\lambda^*)^\ell
= \frac{C'(b_\phi)^2}{1 - \hat{\mu}_f(2\lambda^*)} = C'(b_\phi)^2. \tag{9.16}
\]

Using this bound in (9.11), we obtain
\[
\mathbb{E}\left(\text{Var}(X(t, s, c)|\mathcal{F}_t)\right) \leq C''(b_\phi)^2 \mathbb{E}\left(\sum_{x \in \mathcal{F}(t) \setminus \mathcal{F}(t, c)} e^{-2\lambda^* x} \right) \leq C''(b_\phi)^2 e^{-\lambda^* t} \mathbb{E}(W_t) = C''(b_\phi)^2 e^{-\lambda^* t}.
\]

Moreover, \( \mathbb{E}(X(t, s, c)|\mathcal{F}_t) = 0 \). Thus, we obtain
\[
\mathbb{E}|X(t, s, c)| \leq \sqrt{\mathbb{E}(X(t, s, c))^2} = \sqrt{\text{Var}(X(t, s, c))} \leq \sqrt{C''(b_\phi)^2} e^{-\lambda^* t/2}.
\tag{9.17}
\]

Using (9.13),
\[
\mathbb{E}\left(\sum_{x \in \mathcal{F}(t, c)} e^{-\lambda^* x} \left| M_f^{\phi_f}(t+s-\sigma_x) - M_f^{\phi_f}(\infty)\right|\right) \leq C_1 b_\phi e^{-C_2(s-c)} \mathbb{E}(W_t) = C_1 b_\phi e^{-C_2(s-c)}. \tag{9.18}
\]

To estimate the third term in the bound (9.10), observe that upon conditioning on \( \mathcal{F}_t \) and noting that \( \sup_{t<\infty} M_f^{\phi_f}(t) \leq C' b_\phi \),
\[
\mathbb{E}\left(\sum_{x \in \mathcal{F}(t, c)} e^{-\lambda^* x} \left( e^{-\lambda^* (t+s-\sigma_x)} Z_f^{\phi_f}(t+s-\sigma_x) - M_f^{\phi_f}(\infty)\right)\right) \leq \mathbb{E}\left(\sum_{x \in \mathcal{F}(t, c)} e^{-\lambda^* x} \left( M_f^{\phi_f}(t+s-\sigma_x) + M_f^{\phi_f}(\infty)\right)\right) \leq C' b_\phi \mathbb{E}(W_{t,c}). \tag{9.19}
\]

Consider the characteristic \( \phi^\beta(v) = e^{\lambda^* v} \int_{v+c}^\infty e^{-\lambda^* u} \xi_f(du) \), \( v \geq 0 \). Then \( W_{t,c} = e^{\lambda^* t} Z_f^{\phi_f}(t) \). Note that
\[
\mathbb{E}(\phi^\beta(t)) = e^{\lambda^* t} \mathbb{E}\left(\int_{t+c}^\infty e^{-\lambda^* u} \xi_f(du)\right) = e^{\lambda^* t} \mathbb{E}\left(\int_{t+c}^\infty \lambda^* e^{-\lambda^* v} (\xi_f(v) - \xi_f(t+c))dv\right)
\leq e^{\lambda^* t} \mathbb{E}\left(\int_{t+c}^\infty \lambda^* e^{-\lambda^* v} \xi_f(v)dv\right) \leq Ce^{\lambda^* t} \int_{t+c}^\infty \lambda^* e^{-\lambda^* v} e^{\beta v} dv \leq \frac{CA^e}{\lambda^* - \beta^e} e^{\lambda^* t} = \frac{CA^e}{\lambda^* - \beta^e}.
\]

Hence, by Lemma 9.1
\[
\left| M_f^{\phi^\beta}(t) - M_f^{\phi^\beta}(\infty)\right| \leq C_1 e^{-C_2 t}. \tag{9.20}
\]

Moreover, by Lemma 3.5 of [34],
\[
M_f^{\phi^\beta}(\infty) = \int_{t}^{\infty} \frac{(1 - \mu_f, \lambda^* (u))du}{1 - \mu_f, \lambda^* (u)}du
\]
where $\mu_{f, \lambda^*}(u) = \int_0^u e^{-\lambda^* v} \mu_f(dv)$. Now, for any $u \geq 0$,

$$1 - \mu_{f, \lambda^*}(u) = \int_u^\infty e^{-\lambda^* v} \mu_f(dv) \leq \int_u^\infty \lambda^* e^{-\lambda^* v} \mu_f(v) dv \leq C \int_u^\infty \lambda^* e^{-\lambda^* v} e^{-\beta v} dv = \frac{C \lambda^*}{\lambda^* - \beta} e^{-(\lambda^* - \beta)u}$$

and hence,

$$\int_c^\infty (1 - \mu_{f, \lambda^*}(u)) du \leq \int_c^\infty \frac{C \lambda^*}{\lambda^* - \beta} e^{-(\lambda^* - \beta)u} du = \frac{C \lambda^*}{(\lambda^* - \beta)^2} e^{-(\lambda^* - \beta)c}.$$ 

This bound implies that there exists $C > 0$ such that for all $c > 0$,

$$M_{f}^{\phi_t}(\infty) \leq C e^{-(\lambda^* - \beta)c}. \quad (9.21)$$

Combining (9.20) and (9.21),

$$\mathbb{E}(W_t, c) = M_{f}^{\phi_t}(t) \leq C_1 e^{-c_2 t} + C e^{-(\lambda^* - \beta)c}.$$ 

Using this in (9.19), we get

$$\mathbb{E} \left( \sum_{x \in \mathcal{F}(t, c)} e^{-\lambda^* \sigma_x} (e^{-\lambda^* (t+s-\sigma_x)} Z_{f_x}^{\phi_t}(t+s-\sigma_x) - M_{f}^{\phi_t}(\infty)) \right) \leq C' b_\phi \left( e^{-c_2 t} + e^{-(\lambda^* - \beta)c} \right). \quad (9.22)$$

To estimate the last term in the bound (9.10), observe that for any $t \geq 0$, $W_{\infty} = \sum_{x \in \mathcal{F}(t)} e^{-\lambda^* \sigma_x} W_{\infty}^x$, where $W_{\infty}^x$ corresponds to $W_{\infty}$ treating vertex $x$ as the root (and hence are i.i.d and have the same distribution as $W_{\infty}$). Moreover, by Theorem 4.1 of [29], Var($W_{\infty}$) $< \infty$. Using these observations,

$$\mathbb{E}(W_t - W_\infty)^2 = \mathbb{E} \left( \sum_{x \in \mathcal{F}(t)} e^{-\lambda^* \sigma_x} (1 - W_{\infty}) \right)^2 = \text{Var}(W_{\infty}) \mathbb{E} \left( \sum_{x \in \mathcal{F}(t)} e^{-2\lambda^* \sigma_x} \right) \leq \text{Var}(W_{\infty}) e^{-\lambda^* t} \mathbb{E}(W_t) = \text{Var}(W_{\infty}) e^{-\lambda^* t}.$$ 

Together with the fact that $\sup_{t \leq \infty} M_{f}^{\phi_t}(t) \leq C b_\phi$, this implies that for $t \geq 0$,

$$\mathbb{E} \left( M_{f}^{\phi_t}(\infty) \mid W_t - W_\infty \right) \leq \sqrt{\mathbb{E} \left( M_{f}^{\phi_t}(\infty) \mid W_t - W_\infty \right)^2} \leq C' b_\phi e^{-\lambda^* t/2}. \quad (9.23)$$

Using (9.17), (9.18), (9.22) and (9.23) and the bound (9.10), we obtain $D, D_1, D_2, D_3 > 0$ not dependin on $b_\phi, t, s, c$ such that

$$\mathbb{E} \left( e^{-\lambda^* (t+s)} Z_{f}^{\phi_t}(t+s) - W_\infty M_{f}^{\phi_t}(\infty) \right) \leq D b_\phi \left( e^{-D_{1}t} + e^{-D_{2}s} + e^{-D_{3}(s-c)} \right). \quad (9.24)$$

Using (9.24) in (9.9), we obtain

$$\mathbb{E} \left| e^{-\lambda^* t} Z_{f}^{\phi_t}(t) - W_\infty M_{f}^{\phi_t}(\infty) \right| \leq D b_\phi \left( e^{-D_{1}t} + e^{-D_{2}s} + e^{-D_{3}(s-c)} \right) + C_1 b_\phi e^{-c_2 t} + 2 C b_\phi e^{-(\lambda^* - \beta)c}.$$ 

The lemma now follows by taking $s = t$ and $c = t/2$.

Recall $\lambda_{f_{k}}$, $\lambda_{f_{l}}$ for $k, l \geq 0$ from (3.4), with $f_{1}$ replaced by $f$ (as this section considers the model without change point).

**Lemma 9.2.** Consider a continuous time branching process with attachment function $f$ that satisfies Assumptions 2.1, 5.1 and 3.2. There exist $\omega_{1}, \epsilon^* \in (0, 1)$ and positive constants $C, \omega_{2}$ such that for all $e \leq \epsilon^*$ and all $T \in \left[ \frac{1-e^*}{\lambda^*} \log n, \frac{1-e^*}{\lambda^*} \log n \right]$,

$$\mathbb{E} \left( \sup_{t \in [0, 2 \log n/\lambda^*]} \left| e^{-\lambda^* T} \sum_{\ell=0}^\infty \lambda_{f}(\ell) D(\ell, T) - \sum_{\ell=0}^\infty \lambda_{f}(\ell) p_{f} W_{\infty} \right| \right) \leq C n^{-\omega_{2}},$$

where $\lambda_{f}(\ell) = \sum_{\ell=0}^\infty \lambda_{f}(\ell)$. Now, for any $u \geq 0$,
and for any $k \geq 0$,
\[
\mathbb{E} \left( n^{\omega_1} \sup_{t \in [0, 2c \log n / \lambda^*)} \left| e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) D(\ell, T) - \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) p_\ell W_\infty \right| \right) \leq C(k + 1) n^{-\omega_2}.
\]

**Proof.** For any $t$, consider the characteristic $\phi(s) = \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) 1 \{ \xi_f(s) = \ell \}$. Then $Z^{(k)}_f(s) = \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) D(\ell, s)$, and as $\phi$ satisfies the hypothesis of Theorem 3.4 with $b_0 = C e^{\lambda^* t}$ for some $C > 0$ (which is a consequence of $\lim_{t \to \infty} e^{\lambda^* t} = \lim_{t \to \infty} e^{\lambda^* t}$ for any $\epsilon \in (0, 1)$, any $t \in [0, 2c \log n / \lambda^*)$ and any $T \in [\frac{1}{\lambda^*} \log n, \frac{1+c}{\lambda^*} \log n]$),
\[
\mathbb{E} \left( e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) D(\ell, T) - \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) p_\ell W_\infty \right) \leq C_1 C e^{\lambda^* t} e^{-\frac{C_1}{1+c} \log n} \leq C_1 C e^{2c \log n} e^{-\frac{C_1}{1+c} \log n}.
\]
Therefore, choosing $e^n$ small enough, there exists $\theta_1 > 0$ such that for any $\epsilon \leq e^n$, any $t \in [0, 2c \log n / \lambda^*)$ and any $T \in [\frac{1}{\lambda^*} \log n, \frac{1+c}{\lambda^*} \log n]$,
\[
\mathbb{E} \left( e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) D(\ell, T) - \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) p_\ell W_\infty \right) \leq \frac{1}{n^{\theta_1}}.
\]
(9.25)

Take any $\theta_2 \in (0, \theta_1)$ and a partition of $[0, 2c \log n / \lambda^*)$ into $t_0 < t_1 < \cdots < t_{(2c \log n / \lambda^*)+1}$ of mesh $n^{-\theta_2}$. By Lemma 7.4, for any $j$ and any $t \in [t_j, t_{j+1}]$, there exist constants $C, C' > 0$ independent of $e, n$ such that
\[
\begin{align*}
&\left| e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) D(\ell, T) - \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t_0) p_\ell W_\infty \right| \\
&\leq e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \left| \lambda^{(k)}_\ell (t) - \lambda^{(k)}_\ell (t_0) \right| D(\ell, T) + \sum_{\ell=0}^{\infty} \left| \lambda^{(k)}_\ell (t) - \lambda^{(k)}_\ell (t_0) \right| p_\ell W_\infty \\
&\leq C n^{C' e} \sum_{\ell=0}^{\infty} (\ell + 1) D(\ell, T) + \frac{C}{n^{\theta_2}} \sum_{\ell=0}^{\infty} (\ell + 1) p_\ell W_\infty \\
&\leq \frac{2C}{n^{1-(1+C')e+\theta_2}} Z(T) + \frac{2C}{n^{\theta_2}} W_\infty. \quad (9.26)
\end{align*}
\]

Using (9.25), (9.26) and the union bound, we obtain for any $\omega' > 0$,
\[
\begin{align*}
\mathbb{E} \left( n^{\omega'} \sup_{t \in [0, 2c \log n / \lambda^*)} \left| e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) D(\ell, T) - \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t) p_\ell W_\infty \right| \right) &\leq \mathbb{E} \left( n^{\omega'} \sup_{1 \leq j \leq [(2c \log n / \lambda^*)+1]} \left| e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t_j) D(\ell, T) - \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t_0) p_\ell W_\infty \right| \right) \\
&\quad + \mathbb{E} \left( \frac{2C n^{\omega'}}{n^{1-(1+C')e+\theta_2}} Z(T) + \frac{2C n^{\omega'}}{n^{\theta_2}} W_\infty \right) \\
&\leq n^{\omega'} \sum_{j=0}^{[(2c \log n / \lambda^*)]} 1 \left| e^{-\lambda^* T} \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t_j) D(\ell, T) - \sum_{\ell=0}^{\infty} \lambda^{(k)}_\ell (t_0) p_\ell W_\infty \right| \\
&\quad + n^{\omega'} \mathbb{E} \left( \frac{2C}{n^{1-(1+C')e+\theta_2}} Z(T) + \frac{2C}{n^{\theta_2}} W_\infty \right) \leq \frac{C'' n^{\omega} \log n}{n^{\theta_1-\theta_2-\omega'}} + \frac{C'' n^{\theta_2-2(1+C')e+\omega'}}{n^{\theta_2-\omega'}}
\end{align*}
\]
for some constant $C'' > 0$. Taking $\epsilon^n < \theta_2 / (2 + C')$ and any $\omega' < \min(\theta_1 - \theta_2, \theta_2 - (2 + C') \epsilon^n, 1)$, this proves the first assertion in the lemma. The second assertion follows similarly upon noting that $\lambda^{(k)}_\ell (t) \leq \lambda^{(k)}_\ell$ for each $k \geq 0$ (and thus the constant $C$ in the expectation bound can be chosen uniformly over $k$) and using Corollary 7.6 in place of Lemma 7.4 (which accounts for the $(k+1)$ in the bound).
Proof of Theorem 3.3 Take $\epsilon^* \leq \epsilon^*$ (where $\epsilon^*$ is as in Lemma 9.2) and any $\epsilon \leq \epsilon^*$. We will abbreviate

$$\mathcal{I}_n := \sup_{t \in [0, 2 \log n / \lambda^*]} \left| \sum_{\ell = 0}^{n} \lambda_\ell(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) - n^{1 - \epsilon} \sum_{\ell = 0}^{\infty} \lambda_\ell(t) p_\ell W_\infty \right|,$$

$$\mathcal{I}^{(k)}_n := \sup_{t \in [0, 2 \log n / \lambda^*]} \left| \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) - n^{1 - \epsilon} \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) p_\ell W_\infty \right|.$$ 

Observe that for any $k \geq 0$, using the fact that $\lambda_\ell(\cdot)$ is an increasing function and $\lambda_\ell(0) = 1$ for each $\ell \geq 0$,

\[
\sup_{t \in [0, 2 \log n / \lambda^*]} \left| \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) - n^{1 - \epsilon} \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) p_\ell \right| 
\leq \frac{\mathcal{I}^{(k)}_n}{\sum_{\ell = 0}^{\infty} \lambda_\ell(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right)} + \frac{\mathcal{I}^{(k)}_n}{\left[ \sum_{\ell = 0}^{\infty} \lambda_\ell(t) p_\ell W_\infty \right]} \frac{\mathcal{I}^{(k)}_n}{\sum_{\ell = 0}^{\infty} \lambda_\ell(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right)} \mathcal{I}^{(k)}_n + \frac{\mathcal{I}_n}{\sum_{\ell = 0}^{\infty} \lambda_\ell(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right)} \mathcal{I}_n = \frac{\mathcal{I}^{(k)}_n}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)} + \frac{\mathcal{I}_n}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)}.
\]

Recalling $\omega_1$ from Lemma 9.2

\[
n^{\omega_1}_{\epsilon} \sum_{k = 0}^{\infty} 2^{-k} \left( \sup_{t \in [0, 2 \log n / \lambda^*]} \left| \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) - n^{1 - \epsilon} \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) p_\ell \right| \right) \leq \frac{n^{1 - \epsilon}}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)} \sum_{k = 0}^{\infty} 2^{-k} \left( \frac{\mathcal{I}^{(k)}_n}{n^{1 - \epsilon - \omega_1}} + \frac{\mathcal{I}_n}{n^{1 - \epsilon - \omega_1}} \right).
\]

Using Lemma 9.2, for any $\eta > 0$,

\[
\mathbb{P} \left( \sum_{k = 0}^{\infty} 2^{-k} \left( \frac{\mathcal{I}^{(k)}_n}{n^{1 - \epsilon - \omega_1}} + \frac{\mathcal{I}_n}{n^{1 - \epsilon - \omega_1}} \right) > \eta \right) \leq \eta^{-1} \sum_{k = 0}^{\infty} 2^{-k} \frac{1}{n^{1 - \epsilon - \omega_1}} \mathbb{E} \left[ \mathcal{I}^{(k)}_n + \mathcal{I}_n \right] \leq \eta^{-1} \sum_{k = 0}^{\infty} 2^{-k} (k + 2) C n^{-\omega_2} \leq C' \eta^{-1} n^{-\omega_2}
\]

for positive constants $C, C'$. Moreover, $\frac{n^{1 - \epsilon}}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)} \to_{\lambda^* \rightarrow \infty} \frac{\lambda^* m^*}{W_\infty}$ as $n \to \infty$. By Lemma 5.8 Combining these observations,

\[
n^{\omega_1}_{\epsilon} \sum_{k = 0}^{\infty} 2^{-k} \left( \sup_{t \in [0, 2 \log n / \lambda^*]} \left| \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) - n^{1 - \epsilon} \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) p_\ell \right| \right) \to_{n \to \infty} 0.
\]

Moreover, it is straightforward to check that

\[
\sup_{t \in [0, 2 \log n / \lambda^*]} \left| D \left( k, \frac{1 - \epsilon}{\lambda^*} \log n + t \right) - \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) \right| \leq \frac{1}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)} \sup_{t \in [0, 2 \log n / \lambda^*]} \left| D \left( k, \frac{1 - \epsilon}{\lambda^*} \log n + t \right) - \sum_{\ell = 0}^{\infty} \lambda_\ell^{(k)}(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) \right| + \frac{1}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)} \sup_{t \in [0, 2 \log n / \lambda^*]} \left| Z \left( \frac{1 - \epsilon}{\lambda^*} \log n + t \right) - \sum_{\ell = 0}^{\infty} \lambda_\ell(t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) \right|.
\]
Abbreviate
\[
\hat{\mathcal{F}}_n^{(k)} := \sup_{t \in [0, 2e \log n/\lambda^*]} \left| D \left( k, \frac{1 - \epsilon}{\lambda^*} \log n + t \right) - \sum_{\ell = 0}^{\infty} \lambda^{(k)}_{\ell} (t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) \right|
\]
\[
\hat{\mathcal{F}}_n := \sup_{t \in [0, 2e \log n/\lambda^*]} \left| Z \left( \frac{1 - \epsilon}{\lambda^*} \log n + t \right) - \sum_{\ell = 0}^{\infty} \lambda_{\ell} (t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) \right|
\]

By conditioning on \( \mathcal{F}_n \left( \frac{1 - \epsilon}{\lambda^*} \log n \right) \) and applying Lemma 7.11, we obtain \( \omega'_1 \in (0, 1), \omega'_2 > 0 \) not depending on \( \epsilon \) such that for any \( \eta > 0 \),
\[
P \left( \sum_{k=0}^{\infty} 2^{-k} \left( \frac{\hat{\mathcal{F}}_n^{(k)}}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)^{1-\omega'_i}} \right) > \eta \right| \mathcal{F}_n \left( \frac{1 - \epsilon}{\lambda^*} \log n \right) \)
\[
= P \left( \sum_{k=0}^{\infty} 2^{-k} \left( \frac{\hat{\mathcal{F}}_n^{(k)}}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)^{1-\omega'_i}} \right) > \sum_{k=0}^{\infty} \left( \frac{3}{2} \right)^{-k} \frac{\eta}{3} \left| \mathcal{F}_n \left( \frac{1 - \epsilon}{\lambda^*} \log n \right) \right) \right.
\]
\[
\leq \sum_{k=0}^{\infty} P \left( \frac{\hat{\mathcal{F}}_n^{(k)}}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)^{1-\omega'_i}} > \left( \frac{4}{3} \right)^{k} \frac{\eta}{3} \left| \mathcal{F}_n \left( \frac{1 - \epsilon}{\lambda^*} \log n \right) \right) \right)
\]
\[
\leq C e^{C \epsilon^2 r \log n/\lambda^*} \eta^{-2} Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)^{-\omega'_i} \sum_{k=0}^{\infty} (k + 1)^2 \left( \frac{3}{4} \right)^{2k} = C' n^{2C \epsilon^2/\lambda^*} \eta^{-2} Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)^{-\omega'_2} \quad (9.29)
\]

for positive constants \( C, C' \). As \( \frac{n^{1-\epsilon}}{Z(\lambda^*)^{1-\omega'_i}} \lambda^{m^*} \frac{P}{W^*} \), the bound above converges to zero almost surely if \( \epsilon^{**} \) is chosen sufficiently small and \( \epsilon \leq \epsilon^{**} \). Similarly,
\[
P \left( \sum_{k=0}^{\infty} 2^{-k} \left( \frac{\hat{\mathcal{F}}_n}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)^{1-\omega'_i}} \right) > \epsilon \right| \mathcal{F}_n \left( \frac{1 - \epsilon}{\lambda^*} \log n \right) \)
\[
\leq C' n^{2C \epsilon^2/\lambda^*} \epsilon^{-2} Z \left( \frac{1 - \epsilon}{\lambda^*} \log n \right)^{-\omega'_2} \quad (9.30)
\]

Using (9.28), (9.29), (9.30) and recalling that \( \frac{n^{1-\epsilon}}{Z(\lambda^*)^{1-\omega'_i}} \lambda^{m^*} \frac{P}{W^*} \) as \( n \to \infty \), we conclude
\[
n^{(1-\epsilon)\omega'_i} \sum_{k=0}^{\infty} 2^{-k} \left( \sup_{t \in [0, 2e \log n/\lambda^*]} \left| D \left( k, \frac{1 - \epsilon}{\lambda^*} \log n + t \right) - \sum_{t=0}^{\infty} \lambda^{(k)}_{\ell} (t) D \left( \ell, \frac{1 - \epsilon}{\lambda^*} \log n \right) \right| \right) \to 0. \quad (9.31)
\]

Choosing \( \omega^* = \min(\omega_1, (1 - \epsilon)\omega'_i) \), we conclude from (9.29) and (9.31) that
\[
n^{\omega^*} \sum_{k=0}^{\infty} 2^{-k} \left( \sup_{t \in [0, 2e \log n/\lambda^*]} \left| \frac{D \left( k, \frac{1 - \epsilon}{\lambda^*} \log n + t \right)}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n + t \right)} - \frac{\sum_{t=0}^{\infty} \lambda^{(k)}_{\ell} (t) p_{\ell}}{\sum_{t=0}^{\infty} \lambda_{\ell} (t) p_{t}} \right| \right) \to 0. \quad (9.32)
\]

Finally, we claim that for each \( k \geq 0, t \geq 0, \)
\[
\sum_{t=0}^{\infty} \lambda^{(k)}_{\ell} (t) p_{\ell} = p_{k}. \quad (9.33)
\]

To see this, observe that the following limits hold as \( n \to \infty \):
\[
\frac{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n + t \right)}{n^{1-\epsilon}} \overset{P}{\longrightarrow} e^{\lambda^* t W_{\infty}}, \quad \frac{D \left( k, \frac{1 - \epsilon}{\lambda^*} \log n + t \right)}{n^{1-\epsilon}} \overset{P}{\longrightarrow} p_{k} e^{\lambda^* t W_{\infty}}
\]

and thus,
\[
\frac{D \left( k, \frac{1 - \epsilon}{\lambda^*} \log n + t \right)}{Z \left( \frac{1 - \epsilon}{\lambda^*} \log n + t \right)} \overset{P}{\longrightarrow} p_{k}.
But from (9.32),
\[
\frac{D(k, \frac{1}{\lambda^*} \log n + t)}{Z(\frac{1}{\lambda^*} \log n + t)} \rightarrow p \frac{\sum_{\ell=0}^\infty \lambda^{(k)}_\ell(t) p_\ell}{\sum_{\ell=0}^\infty \lambda_\ell(t) p_\ell}.
\]
(9.33) follows from the above two observations. The lemma now follows from (9.32) and (9.33).

10. PROOFS: CHANGE POINT DETECTION

Recall \(\lambda_\ell, \lambda^{(k)}_\ell\) for \(k, \ell \geq 0\) defined in (3.4) and the functional \(\Phi_a : \mathcal{P} \rightarrow \mathcal{P}\) defined for each \(a > 0\) in (3.5).

**Lemma 10.1.** \(\lim_{a \rightarrow \infty} \Phi_a(p) = p^k\) (where the limit is taken in the coordinate-wise sense).

**Proof.** For each \(k \geq 0\), by Lemma 5.8 (ii), \(\lim_{t \rightarrow \infty} e^{-\lambda_\ell^* t} m_{f_1}(t) = \frac{1}{\lambda_1^* m^*}\) and \(\lim_{t \rightarrow \infty} e^{-\lambda_\ell^* t} m_{f_1}(t) = \frac{1}{\lambda_1^* m^*}\) and consequently,
\[
\lim_{t \rightarrow \infty} e^{-\lambda_\ell^* t} \lambda^*_\ell(t) = \frac{w_\ell}{\lambda_1^* m^*}, \quad \lim_{t \rightarrow \infty} e^{-\lambda_\ell^* t} \lambda^{(k)}_\ell(t) = \frac{p_1^k w_\ell}{\lambda_1^* m^*}.
\]
(10.1)
Moreover, it is easy to see from (3.4) that for any \(\ell, k \geq 0\), \(e^{-\lambda_\ell^* t} \lambda_\ell(t) \leq \left(\sup_{u \geq 0} e^{-\lambda_1^* u} m_{f_1}(u)\right) w_\ell\) and \(e^{-\lambda_\ell^* t} \lambda^{(k)}_\ell(t) \leq \left(\sup_{u \geq 0} e^{-\lambda_1^* u} m_{f_1}(u)\right) w_\ell\) for all \(t \geq 0\) and this bound is finite. By this observation, we can apply the dominated convergence theorem and (10.1) in the formula of \(\Phi_a(p)\) to obtain the lemma.

**Lemma 10.2.** For any \(s, t \geq 0\) and any \(j, k \geq 0\),
\[
\sum_{\ell=0}^\infty \lambda^{(j)}_\ell(t) \lambda^*_\ell(s) = \lambda_j(s + t), \quad \sum_{\ell=0}^\infty \lambda^{(j)}_\ell(t) \lambda^{(k)}_\ell(s) = \lambda_j^{(k)}(s + t).
\]
Consequently, for any \(p \in \mathcal{P}\),
\[
\Phi_s(\Phi_t(p)) = \Phi_{s+t}(p).
\]
**Proof.** We will only prove the first assertion. The second one follows similarly. Denote by \(PA^{(j)}(\cdot)\) the continuous time branching process with attachment function \(i \mapsto f_1(i + j)\) and denote by \(D_n^{(j)}(t, \ell)\) the number of vertices of degree \(\ell\) at time \(t\) (excluding the root). Then
\[
\mathbb{E}\left[PA^{(j)}(t + s) \mid \mathcal{F}_n(t)\right] = \sum_{\ell=j}^\infty 1 \left\{ \xi_{f_1}^{(j)}(t) = \ell - j \right\} \left( 1 + \int_0^s m_{f_1}(s - v) \mu_1^{(j)}(dv) \right) \left( 1 + \int_0^t m_{f_1}(t - u) \mu_1^{(j)}(du) \right)
\]
where the first term denotes the expected number of vertices born to the root in the process in the time interval \([t, t + s]\) and the second term denotes the expected number of vertices born to the root in the time interval \([t, t + s]\) to those vertices born in the time interval \([0, t]\). Taking expectation on both sides of the above expression and noting that \(\lambda_j(t + s) = \mathbb{E}\left[PA^{(j)}(t + s)\right]\) and \(\mathbb{E}\left[D_n^{(j)}(t, \ell)\right] = \int_0^t m_{f_1}(t - u) \mu_1^{(j)}(du)\), we obtain
\[
\lambda_j(t + s) = \sum_{\ell=0}^\infty \left( p_{j} \xi_{f_1}^{(j)}(t) + \int_0^t m_{f_1}(t - u) \mu_1^{(j)}(du) \right) \left( 1 + \int_0^s m_{f_1}(s - v) \mu_1^{(j)}(dv) \right) = \sum_{\ell=0}^\infty \lambda^{(j)}_\ell(t) \lambda^*_\ell(s).
\]
To prove the semigroup property, note that for each \(k \geq 0\),
\[
(\Phi_s(\Phi_t(p)))_k = \left( \sum_{\ell=0}^\infty (\Phi_t(p))_\ell \lambda^{(k)}_\ell(s) \right)_k = \sum_{\ell=0}^\infty (\Phi_t(p))_\ell \lambda^{(k)}_\ell(s) = \sum_{\ell=0}^\infty \frac{\sum_{j=0}^\infty p_j \lambda^{(j)}_\ell(t) \lambda^{(k)}_\ell(s)}{\sum_{j=0}^\infty p_j \lambda^{(j)}_\ell(s)} = \sum_{j=0}^\infty p_j \lambda^{(k)}_\ell(s + t) = (\Phi_{s+t}(p))_k.
\]
Lemma 10.3. For any $a > 0$ and any $p \in \mathcal{P}$ such that $p \neq p^1$, we have $\Phi_a(p) \neq p$.

Proof. Suppose there exists $a > 0$ and $p \neq p^1$, such that $\Phi_a(p) = p$. Then by Lemma 10.2, for any $n \geq 1$, $\Phi_{na}(p) = p$. Letting $n \to \infty$ and using Lemma 10.1 we obtain $p^1 = p$ which gives a contradiction.

Now we are ready to prove Theorem 3.17.

Proof of Theorem 3.17. Recall $\omega^*$, $\epsilon^{**}$ from Theorem 3.3 applied to the branching process with attachment function $f_0$ and fix any $\epsilon \leq \epsilon^{**}$. Let $\lambda_0^*$ denote the associated Malthusian rate. Take any $n_0 \geq 1$ such that $h_n \geq 1/\gamma$ for all $n \geq n_0$. Observe that for any $\eta > 0$ and any $n \geq n_0$,

$$
\mathbb{P}\left(n^{\omega^*} \sum_{k=0}^{\infty} 2^{-k} \sup_{1/h_n \leq t \leq \gamma} \left| D(k, T_{[nt]}) - p_k^0 \right| > \eta \right) \\
\leq \mathbb{P}\left(n^{\omega^*} \sum_{k=0}^{\infty} 2^{-k} \left( \sup_{t \in \{0,2,\ldots,\log n/\lambda_0^*\}} \left| D\left(\log(1+\epsilon\lambda_0^*/n) + t\right) - D\left(\log(1+\epsilon\lambda_0^*/n) + t\right) - p_k^0\right| \right) > \eta \right) \\
+ \mathbb{P}\left(T_{[n/h_n]} < \frac{1-\epsilon}{\lambda_0^*} \log n \right) \mathbb{P}\left(T_{[n\gamma]} > \frac{1+\epsilon}{\lambda_0^*} \log n \right) \right)
$$

The first term in the above bound converges to zero by Theorem 3.3. Further,

$$
\mathbb{P}\left(T_{[n/h_n]} < \frac{1-\epsilon}{\lambda_0^*} \log n \right) \to 0 \quad (10.2)
$$

because $T_{[n/h_n]} \to 1$ as $n \to \infty$ by Lemma 5.8 (ii) and by assumption, $\frac{\log h_n}{\log n} \to 0$. Similarly,

$$
\mathbb{P}\left(T_{[n\gamma]} > \frac{1+\epsilon}{\lambda_0^*} \log n \right) \to 0 \quad (10.3)
$$

because $\frac{T_{[n\gamma]}}{\log(n/\gamma)} \to 1$ as $n \to \infty$. Thus, we conclude

$$
n^{\omega^*} \sum_{k=0}^{\infty} 2^{-k} \sup_{1/h_n \leq t \leq \gamma} \left| D(k, T_{[nt]}) - p_k^0 \right| \mathbb{P} \to 0 \quad (10.4)
$$

as $n \to \infty$ which, along with the fact that $\omega^* \in (0,1)$, implies

$$
n^{\omega^*} \sum_{k=0}^{\infty} 2^{-k} \sup_{1/h_n \leq t \leq \gamma} \left| D(k, T_{[nt]}) \right| - D(k, T_{[n/h_n]}) \mathbb{P} \to 0.
$$

As $\frac{\log h_n}{\log n} \to 0$ as $n \to \infty$, the above implies

$$
b_n \sum_{k=0}^{\infty} 2^{-k} \sup_{1/h_n \leq t \leq \gamma} \left| D(k, T_{[nt]}) \right| - D(k, T_{[n/h_n]}) \mathbb{P} \to 0.
$$

From this observation and the definition of $\hat{T}_n$, we conclude that

$$
\mathbb{P}\left(\hat{T}_n \geq \gamma \right) \to 1 \quad \text{as} \quad n \to \infty.
$$

Moreover, by Theorem 3.6 for any $t > \gamma$ and any $k \geq 0$, $\left| \frac{D(k, T_{[nt]})}{nt} - \left(\Phi_{a_k(p^0)}\right)_k \right| \mathbb{P} \to 0$ and hence, by (10.4) and the dominated convergence theorem, as $n \to \infty$,

$$
\sum_{k=0}^{\infty} 2^{-k} \left| \frac{D(k, T_{[nt]})}{nt} - \frac{D(k, T_{[n/h_n]})}{n/h_n} \right| \mathbb{P} \to \sum_{k=0}^{\infty} 2^{-k} \left| \left(\Phi_{a_k(p^0)}\right)_k - p_k^0 \right|.
$$
As \( a_t > 0 \) for each \( t > \gamma \) and \( \mathbf{p}^0 \neq \mathbf{p}^1 \), by Lemma 10.3, \( \Phi_{a_t}(\mathbf{p}^0) \neq \mathbf{p}^0 \) and hence, the limit above is strictly positive. From the definition of \( \hat{T}_n \) and the above, we conclude that for each \( t > \gamma \),

\[
P(\hat{T}_n \leq t) \to 1 \quad \text{as } n \to \infty.
\]

The theorem follows from (10.5) and (10.6).

Acknowledgements

SB and IC were partially supported by NSF grants DMS-1613072, DMS-1606839 and ARO grant W911NF-17-1-0010.

References

[1] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Reviews of modern physics 74 (2002), no. 1, 47.
[2] D. Aldous, Asymptotic fringe distributions for general families of random trees, The Annals of Applied Probability (1991), 228–266.
[3] K. B. Athreya and S. Karlin, Embedding of urn schemes into continuous time markov branching processes and related limit theorems, Ann. Math. Statist. 39 (196812), no. 6, 1801–1817.
[4] K. B. Athreya and P. E. Ney, Branching processes, Springer-Verlag, New York-Heidelberg, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196. MR0373040 (51 #9242)
[5] J. Bai, Estimating multiple breaks one at a time, Econometric theory 13 (1997), no. 03, 315–352.
[6] J. Bai and P. Perron, Estimating and testing linear models with multiple structural changes, Econometrica (1998), 47–78.
[7] J. Bai and P. Perron, Computation and analysis of multiple structural change models, Journal of applied econometrics 18 (2003), no. 1, 1–22.
[8] A. L. Barabási and R. Albert, Emergence of scaling in random networks, science 286 (1999), no. 5439, 509–512.
[9] J-B Bardet, A Christen, and J Fontbona, Quantitative exponential bounds for the renewal theorem with spread-out distributions, arXiv preprint arXiv:1504.06184 (2015).
[10] F. Bergeron, P. Flajolet, and B. Salvy, Varieties of increasing trees, Colloquium on trees in algebra and programming, 1992, pp. 24–48.
[11] S. Bhamidi, Universal techniques to analyze preferential attachment trees: Global and local analysis, preparation. Version August 19 (2007).
[12] S. Bhamidi, S. N Evans, and A. Sen, Spectra of large random trees, Journal of Theoretical Probability 25 (2012), no. 3, 613–654.
[13] S. Bhamidi, J. Jin, and A. Nobel, Change point detection in network models: Preferential attachment and long range dependence, arXiv preprint arXiv:1508.02043 (2015).
[14] B. Bollobás, Random graphs, Cambridge University Press, 2001.
[15] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, The degree sequence of a scale-free random graph process, Random Struct. Algorithms 18 (May 2001), no. 3, 279–290.
[16] E Brodsky and B. S Darkhovsky, Nonparametric methods in change point problems, Vol. 243, Springer Science & Business Media, 2013.
[17] S. Bubeck, L. Devroye, and G. Lugosi, Finding adam in random growing trees, Random Structures & Algorithms 50 (2017), no. 2, 158–172.
[18] S. Bubeck, E. Mossel, and M. Z Rácz, On the influence of the seed graph in the preferential attachment model, IEEE Transactions on Network Science and Engineering 2 (2015), no. 1, 30–39.
[19] M. Csörgő and L. Horváth, Limit theorems in change-point analysis, Vol. 18, John Wiley & Sons Inc, 1997.
[20] N. Curien, T. Duquesne, I. Kortchemski, and I. Manolescu, Scaling limits and influence of the seed graph in preferential attachment trees, arXiv preprint arXiv:1406.1758 (2014).
[21] L. Devroye, Branching processes and their applications in the analysis of tree structures and tree algorithms, Probabilistic methods for algorithmic discrete mathematics, 1998, pp. 249–314.
[22] L. Devroye and J. Lu, The strong convergence of maximal degrees in uniform random recursive trees and dags, Random Structures & Algorithms 7, no. 1, 1–14, available at https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.3240070102.
[23] M. Drmota, Random trees: an interplay between combinatorics and probability, Springer Science & Business Media, 2009.
[24] R. Durrett, Random graph dynamics, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2007. MR2271734 (2008c:05167)
[25] P. Flajolet and R. Sedgewick, Analytic combinatorics, cambridge University press, 2009.
[26] C. Goldschmidt and J. B Martin, Random recursive trees and the bolthausen-sznitman coalescent, Electron. J. Probab 10 (2005), no. 21, 718–745.
[27] C. Holmgren, S. Janson, et al., *Fringe trees, crump–mode–jagers branching processes and m-ary search trees*, Probability Surveys 14 (2017), 53–154.
[28] P. Jagers, *Branching processes with biological applications*, Wiley-Interscience [John Wiley & Sons], London-New York-Sydney, 1975. Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics. MR0488341
[29] P. Jagers and O. Nerman, *The growth and composition of branching populations*, Advances in Applied Probability 16 (1984), no. 2, 221–259.
[30] P. Jagers and O. Nerman, *The growth and composition of branching populations*, Adv. in Appl. Probab. 16 (1984), no. 2, 221–259. MR742953 (86j:60193)
[31] S. Janson, *Functional limit theorems for multitype branching processes and generalized pólya urns*, Stochastic Processes and their Applications 110 (2004), no. 2, 177–245.
[32] H. Mahmoud, *Pólya urn models*, Chapman and Hall/CRC, 2008.
[33] T. F. Móri, *Degree distribution nearby the origin of a preferential attachment graph*, Electron. Comm. Probab 12 (2007), 276–282.
[34] O. Nerman, *On the convergence of supercritical general (cmj) branching processes*, Probability Theory and Related Fields 57 (1981), no. 3, 365–395.
[35] M. Newman, *Networks: an introduction*, Oxford University Press, 2010.
[36] M. E. Newman, *The structure and function of complex networks*, SIAM review 45 (2003), no. 2, 167–256.
[37] J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998. Reprint of 1997 original. MR1600720 (99c:60414)
[38] A. B Olshen, E. Venkatraman, R. Lucito, and M. Wigler, *Circular binary segmentation for the analysis of array-based dna copy number data*, Biostatistics 5 (2004), no. 4, 557–572.
[39] S. Resnick and G. Samorodnitsky, *Asymptotic normality of degree counts in a preferential attachment model*, arXiv preprint arXiv:1504.07328 (2015).
[40] A. Rudas, B. Tóth, and B. Valkó, *Random trees and general branching processes*, Random Structures & Algorithms 31 (2007), no. 2, 186–202.
[41] R. T Smythe and H. M Mahmoud, *A survey of recursive trees*, Theory of Probability and Mathematical Statistics 51 (1995), 1–28.
[42] J. Szymański, *On a nonuniform random recursive tree*, North-Holland Mathematics Studies 144 (1987), 297–306.
[43] J. Szymanski, *On the maximum degree and the height of a random recursive tree*, Random graphs, 1990, pp. 313–324.
[44] R. Van Der Hofstad, *Random graphs and complex networks*, Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf (2009).
[45] Y.-C. Yao, *Estimating the number of change-points via schwarz’criterion*, Statistics & Probability Letters 6 (1988), no. 3, 181–189.
[46] N. R Zhang and D. O Siegmund, *A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data*, Biometrics 63 (2007), no. 1, 22–32.