

# COMMUNITY MODULATED RECURSIVE TREES AND POPULATION DEPENDENT BRANCHING PROCESSES

SHANKAR BHAMIDI, RUITUO FAN, NICOLAS FRAIMAN, AND ANDREW NOBEL

ABSTRACT. We consider random recursive trees that are grown via community modulated schemes that involve random attachment or degree based attachment. The aim of this paper is to derive general techniques based on continuous time embedding to study such models. The associated continuous time embeddings are **not** branching processes: individual reproductive rates at each time  $t$  depend on the composition of the entire population at that time. Using stochastic analytic techniques we show that various key macroscopic statistics of the continuous time embedding stabilize, allowing asymptotics for a host of functionals of the original models to be derived.

## 1. INTRODUCTION

The subject of this paper is the analysis of random recursive trees that are grown via community modulated schemes involving random attachment or degree based attachment. This topic lies at the intersection of two areas of modern probabilistic combinatorics: random recursive trees and community detection for networks. Before describing the models of primary interest, we give some preliminaries.

A rooted tree on  $n$  vertices labeled  $\{1, 2, \dots, n\}$  is called recursive if 1 is the root, and for each  $2 \leq i \leq n$ , the (unique) path from the root to vertex  $i$  has increasing labels. The name “recursive” suggests that these trees can be constructed recursively, by adding a new vertex at each time step, and the labels can be viewed as birth orders. Recursive trees have been studied for decades (*e.g.* see Smythe and Mahmoud [35] for an early survey), with applications to epidemics [28], pyramid schemes [17][18], convex hull algorithms [26], and modeling family trees of preserved copies of ancient or medieval texts [29], where vertices represents people or texts that arrive chronologically and are labeled by time.

In many studies of rooted trees, the uniform recursive tree (URT) model is used, *i.e.*, a tree is chosen uniformly at random from all recursive trees of a fixed size. From a statistical point of view, URTs serve as a natural null model for recursive trees. Recently, a wide array of extensions of URTs have been considered. For example, preferential attachment models [3], which give rise to power law degree distributions, are favored in the complex network community where real data exhibits heavy-tailed degrees. A recent multitype extension of preferential attachment based on the genealogy of a multitype branching process has been considered in [33]. There are also variants of URTs that introduce choices to the attachment rule [15][25].

---

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, 304 HANES HALL, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599

*E-mail addresses:* bhamidi@email.unc.edu, badhron@email.unc.edu, fraiman@email.unc.edu, nobel@email.unc.edu.

2010 *Mathematics Subject Classification.* Primary: 60C05, 05C80, 60F05, 60K35, 60H30, 60J70.

*Key words and phrases.* inhomogeneous random trees, random recursive trees, multi-type branching processes, continuous time branching processes, stochastic block model, community detection.

Instead of choosing one existing vertex, these models choose  $k$  previous vertices (with or without replacement) as candidates, and connect the new vertex to one of them based on certain optimization criterion. Related to these models is the so called scaled attachment random recursive tree (SARRT) [13], where at the  $n$ -th step the new vertex  $n$  is connected to vertex  $\lfloor nX_n \rfloor$ , with  $X_1, X_2, \dots$  being a sequence of i.i.d. random variables taking value in  $[0, 1)$ . When the  $X_i$  follow a uniform distribution, SARRT reduces to the usual URT.

However, all of these alternatives are still homogeneous in the sense that attachments are made based on the same rule, no matter how complicated. In this paper, we introduce an alternative to URT that models heterogeneity in the spirit of the well-studied *stochastic block model*. Recall that in the stochastic block model we start with  $n$  vertices that are partitioned into two or more classes (often referred to as “communities”), and that each pair of vertices are connected independently at random with probability depending only on their class membership. Just as the stochastic block model can be seen as a heterogeneous alternative to the Erdős-Rényi model (where each pair of vertices are connected independently with the same probability), our model, which we shall refer to as Community Modulated Recursive Tree (CMRT), generalizes URT in a similar way to allow for latent class labels.

**1.1. Model formulation.** Here we describe the CMRT canonical model. Extensions of this model to more than two types as well as other schemes of attachment including preferential attachment are explored in Section 1.3. First we fix some terminology for rooted trees. Since they are connected (undirected) graphs with no cycles, there exists a unique path from any given vertex  $v$  to the root. Any vertex  $u \neq v$  in this path is called an antecedent of  $v$ , and  $v$  its descendant; if  $u$  and  $v$  are adjacent (*i.e.* there exists an edge between them), we call  $u$  the parent of  $v$  and  $v$  its child. We also say that  $v$  is connected to  $u$  if they are adjacent. For recursive trees with multiple vertices identified as roots, these terms are defined with respect to the nearest root.

Recall that a uniform recursive tree (URT) can be constructed as follows: starting with a root vertex labeled 1, at each time  $n$  choose an existing vertex uniformly at random and add a new vertex labeled  $n$  connected to this chosen vertex. Note that this definition yields a growing tree-valued process, which we shall denote by  $\{\mathcal{U}_n\}_{n \geq 1}$ , where  $\mathcal{U}_n$  is the random recursive tree given by this process when it reaches size  $n$ . We shall refer to  $\{\mathcal{U}_n\}_{n \geq 1}$  as a URT and  $\mathcal{U}_n$  as a URT of size  $n$ .

Now we are ready to introduce Community Modulated Recursive Tree (CMRT). This is the “canonical” model in the sense that it is easiest to describe both in discrete time and through a continuous time embedding that allows for its analysis. Once the reader is familiar with this model, understanding extensions becomes easy. Throughout this paper we shall always refer to classes of vertices in random trees as “types”. For simplicity we first state results for a two-type model, and postpone the general case to Section 1.3. Specifically, assume that all vertices are of either type  $A$  or type  $B$ . To construct a Community Modulated Recursive Tree (CMRT) we begin with a single vertex 1 of type  $A$  and a single vertex 2 of type  $B$  that are connected by an edge. This initial tree is the CMRT of size 2. Vertices 1 and 2 as roots of type  $A$  and type  $B$ , respectively. At each subsequent stage  $n \geq 3$ , a new vertex labeled  $n$  is added to the given tree of size  $n - 1$  in a three-step process:

- (1) Vertex  $n$  flips a  $p$ -coin to determine if it is of type  $A$  (with probability  $p \in (0, 1]$ ) or of type  $B$  (with probability  $1 - p$ ).

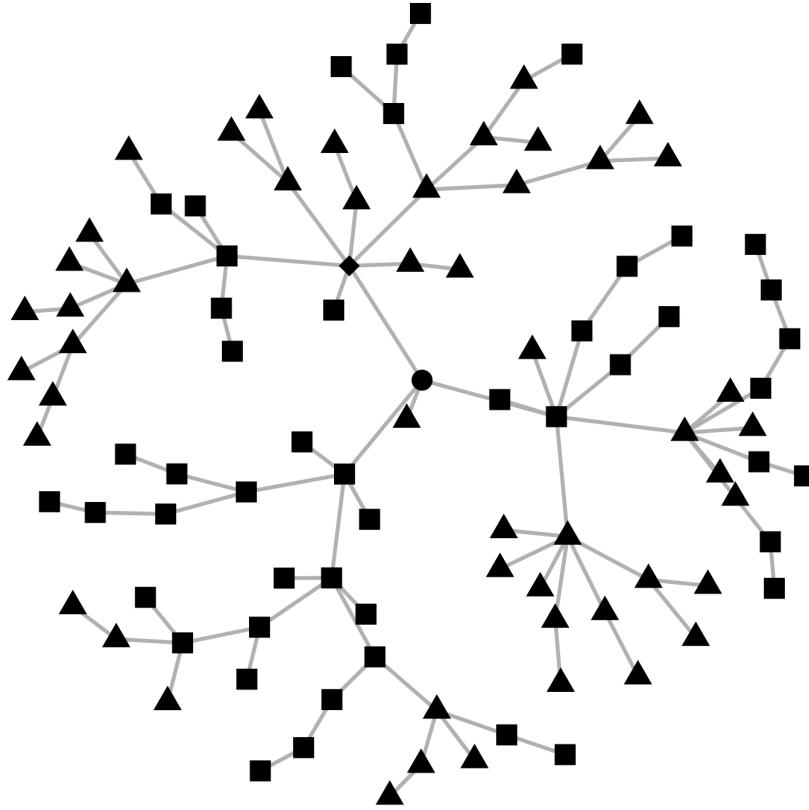


FIGURE 1.1. A (two-type) Community Modulated Recursive Tree of size 100, with parameters  $p = 0.5$  and  $q = 0.8$ . The type  $A$  and  $B$  roots are respectively plotted as circle and diamond, and other type  $A$  and  $B$  vertices as squares and triangles.

- (2) Vertex  $n$  flips a  $q$ -coin to decide if it will connect to vertices of its own type (with probability  $q \in [0, 1]$ ) or to vertices of the other type (with probability  $1 - q$ ).
- (3) Vertex  $n$  chooses an existing vertex of the type selected in step (2) uniformly at random and connects to it, forming a CMRT of size  $n$ .

Note that in (a) we do not lose any generality by excluding the case where  $p = 0$ , as one can switch the type labels if necessary. Adopting the terminology of stochastic block models, we call the CMRT assortative when  $q \geq 1/2$  and disassortative when  $q < 1/2$ . In general, the probability that a vertex chooses to connect to a vertex of the same type may vary across types. We assume here that they are the same for simplicity; the general case is discussed in Section 1.3.

We refer to the random tree  $\mathcal{T}_n$  produced at the  $n$ -th stage of the procedure above as a Community Modulated Recursive Tree (CMRT) of size  $n$ , and to the sequence  $\{\mathcal{T}_n\}_{n \geq 2}$  as a CMRT.

## 1.2. Special cases.

- (i) When  $p = 1$  and  $q = 1$ , the CMRT becomes a URT consisting of type  $A$  vertices plus an additional type  $B$  vertex (*i.e.* the type  $B$  root) connecting to the type  $A$  root. Thus one may consider this case to be degenerate.
- (ii) When  $p = 1$  and  $q \neq 1$ , all new vertices are of type  $A$  and have a fixed probability  $1 - q$  to connect to the unique type  $B$  vertex. This case too is degenerate in the sense that one can construct such a CMRT from a URT. To do so, take a URT consisting only of type  $A$  vertices, and add a single type  $B$  vertex connected to the root of the URT. Treat this type  $B$  vertex as the type  $B$  root. Then remove each edge between type  $A$  vertices independently with probability  $1 - q$ , and for each edge removed in this way, add an edge between the corresponding child and the type  $B$  root. It is easy to see that the recursive tree formed this way has the same distribution as a CMRT with  $p = 1$  and  $q \neq 1$ . This construction will be useful in some of the proofs.
- (iii) When  $q = 1$ , the CMRT can be partitioned into two disjoint, rooted subtrees associated with the type  $A$  and type  $B$  roots. Furthermore, each subtree, conditioned on its size, has the same distribution as a URT. However, in the absence of vertex labels, identifying these subtrees from an observed CMRT of finite size is non-trivial. Although not fully comparable, we note by way of comparison that in a stochastic block recovering the block assignment for each vertex is trivial if each block is connected and the probability to connect to vertices in other blocks is zero.
- (iv) When  $q = 0$ , each new vertex connects to a vertex of the opposite type. Thus for any path in the graph, adjacent vertices will have alternating types.

1.3. **Extensions.** Here we describe several natural extensions of the canonical model above.

1.3.1. *General CMRT.* One may readily extend CMRTs to  $K > 2$  types. As in the  $K = 2$  case we begin with one vertex of each type to which subsequent vertices may connect. The initial tree  $n = K$  is a URT of size  $K$ , the vertices of which are assigned to the  $K$  types based on uniform random permutation. We shall refer to these vertices as the roots of each type. At each subsequent stage  $n \geq 3$ , a new vertex labeled  $n$  is added to the given tree of size  $n - 1$  in a three-step process:

- (1) Vertex  $n$  is assigned to type  $i \in \{1, \dots, K\}$  with probability  $p_i$ , where we assume without loss of generality that  $p_1 = \max_{1 \leq i \leq K} p_i > 0$ .
- (2) Once its type has been determined, vertex  $n$  chooses to connect to a vertex of type  $j \in \{1, \dots, K\}$  with probability  $q_{ij}$ .
- (3) Vertex  $n$  chooses an existing vertex of the type selected in step (2) uniformly at random and connects to it, forming a CMRT of size  $n$ .

We refer to the sequence  $\{\mathcal{T}_n\}_{n \geq K}$  of stochastic trees produced in this way as a Community Modulated Recursive Tree (CMRT) with  $K$  types.

1.3.2. *Community Weighted Recursive Tree (CWRT).* There are other variants of multitype recursive trees that could be of potential interest. For example, one can generate CMRT with  $K$  types following steps (1) and (2) above, but let the attachment probabilities in step (3) depend on weights for transitions between types. In more detail, let  $\{\omega_{ij} : 1 \leq i, j \leq K\}$  be positive

weights, and let the incoming vertex  $n$  of type  $i$  attach to an existing vertex of type  $j$  with probability

$$\frac{\omega_{ij}}{\sum_{l=1}^K n_l \omega_{il}},$$

where  $n_l$  is the number of existing type  $l$  vertices in the tree. We refer to the sequence  $\{\mathcal{T}_n\}_{n \geq K}$  of stochastic trees produced in this way as a Community Weighted Recursive Tree (CWRT) with  $K$  types.

**1.3.3. Community Modulated Preferential Attachment (CMPA).** We can also define a “preferential attachment” version of CMRT, which we shall refer to as Community Modulated Preferential Attachment (CMPA) model. To construct a CMPA tree with  $K$  types, we follow the same rule as that of a CMRT with  $K$  types, except for the attachment step. Suppose that a vertex  $n$  of type  $i$  that has chosen to connect to a vertex of type  $j$ . Rather than choosing uniformly among vertices of type  $j$ , vertex  $n$  connects to a type  $j$  vertex  $v$  with probability proportional to  $D_v + \alpha_{ij}$ , where  $D_v$  is the out-degree of vertex  $v$ , and  $\alpha_{ij}$  are positive numbers. We refer to the sequence  $\{\mathcal{T}_n^*\}_{n \geq K}$  as a Community Modulated Preferential Attachment (CMPA) tree with  $K$  types.

**1.4. Contributions of the paper.** In addition to the formulation of the models described above, the main aim of this paper is to develop a common set of tools based on continuous time embedding of the associated discrete time processes that allows for asymptotics in the large network  $n \rightarrow \infty$  limit to be derived. The starting point, described in more detail in Section 2, is the embedding of recursive trees into a continuous time process. We emphasize that the continuous time process is **not** a branching process, in the sense that the instantaneous offspring growth rates of each vertex are intricately tied to the frequency of vertex types at that time. Using techniques from stochastic analysis, one can show that key macroscopic properties of the continuous time embedding (properly normalized) stabilize, thus allowing one to read off asymptotics for the discrete time processes. In particular, we shall study the limiting degree distribution, depth and height of these recursive trees, and establish the corresponding asymptotics.

**1.5. Organization of the paper.** The next section describes the population dependent branching process relevant for the canonical model in Section 2. Analogous extensions of the embedding for other models including the community modulated preferential attachment are described in the proofs (Section 5). Section 3 contains the statement of our main theoretical results. In Section 4 we provide a brief discussion on the relevance of this paper as well as related work. Section 5 contains proofs of all the results.

## 2. CONTINUOUS TIME EMBEDDING

In this section we describe the continuous time embedding of recursive trees, which is the major tool used in our theoretical analyses. To fix ideas we describe an embedding for the canonical two-type Community Modulated Recursive Tree introduced above. In order to state the basic embedding result, we introduce a (two-type) continuous time process that we refer to as a **population dependent** branching process (**pdBP**). We refer to the vertices in the continuous time process as “individuals” to differentiate them from vertices in the corresponding discrete time random tree. This process is not a (homogeneous or inhomogeneous) branching process in the usual sense as the reproduction processes of different individuals are not

independent. However, lengths of the time intervals between consecutive births are still exponentially distributed, and the parameters (expected number of births in a unit time interval) of these exponential variables are referred to as “rates”, similar to that of the usual branching process.

Each individual in the population dependent branching process has type  $A$  or  $B$  (not both) and lives forever. The process is initialized at time  $t = 0$  with two individuals, one of each type, that we will refer to as “ancestors”. At times  $t > 0$  each existing individual in the process gives birth to new individuals (offspring) of type  $A$  or  $B$  at rates that are specified below.

For each  $t \geq 0$  let  $n_A(t)$  and  $n_B(t)$  denote the number of individuals of type  $A$  and  $B$ , respectively, in the process, initialized with  $n_A(0) = n_B(0) = 1$ . Reproduction rates for individuals are as follows. For each  $t \geq 0$

- (i) Each type  $A$  individual gives birth to type  $A$  individuals at rate  $r_{AA}(t) = q$ , and to type  $B$  individuals at rate  $r_{AB}(t) = (1 - p)(1 - q)/p$ .
- (ii) Each type  $B$  individual gives birth to type  $A$  individuals at rate

$$r_{BA}(t) = \frac{n_A(t)}{n_B(t)} \cdot (1 - q)$$

and to type  $B$  individuals at rate

$$r_{BB}(t) = \frac{n_A(t)}{n_B(t)} \cdot \frac{q(1 - p)}{p}.$$

We introduce some notation that will be useful in what follows. For  $t \geq 0$ , let  $n(t) := n_A(t) + n_B(t)$  be the total number of individuals alive at time  $t$  and  $\mathcal{F}(t)$  be the  $\sigma$ -field generated by the process until time  $t$ . Also, denote by  $\{\mathcal{F}(t) : t \geq 0\}$  the natural filtration of the process. For each  $t \geq 0$  the genealogical structure of the population dependent branching process can be described by a recursive tree in which each individual corresponds to a vertex. In detail, the  $A$  and  $B$  ancestors are labeled 1 and 2, respectively, and are connected by an edge. The vertex for an individual is connected by edges to the vertices of its offspring, and is labeled by the absolute birth order of the individual in the overall process. Let  $\text{pdBP}(t)$  denote the recursive tree capturing the genealogical structure of the process until and including time  $t$ . The next lemma establishes a close connection between the continuous time process  $\{\text{pdBP}(t) : t \geq 0\}$  and the CMRT. The proof of the lemma is given in Section 5.

**Lemma 2.1.** *Let  $\{\text{pdBP}(t) : t \geq 0\}$  be the continuous time process described above. For  $n \geq 2$  define the stopping time  $T_n = \inf\{t \geq 0 : n(t) = n\}$ . Then  $\text{pdBP}(T_n) \stackrel{d}{=} \mathcal{T}_n$  where  $\{\mathcal{T}_n\}_{n \geq 2}$  is a CMRT. In fact,  $\{\text{pdBP}(T_n)\}_{n \geq 2} \stackrel{d}{=} \{\mathcal{T}_n\}_{n \geq 2}$  as processes.*

### 3. RESULTS

**3.1. Asymptotics for CMRT.** Our first result concerns the limiting distribution of the out-degrees, *i.e.*, the number of children, of vertices in the canonical CMRT. If one regards each edge in a rooted tree as directed from parent to child, then this definition is consistent with that of out-degree in directed graph.

**Theorem 3.1.** *Let  $\{\mathcal{T}_n : n \geq 1\}$  be the canonical CMRT, and for each  $k \geq 0$ , let  $N_k(n)$  denote the number of vertices with out-degree  $k$  in  $\mathcal{T}_n$ . Then for each fixed  $k \geq 0$  the ratio  $N_k(n)/n$  converges*

in probability to  $p_k$  where

$$p_k := \frac{p}{1+r_A^*} \left( \frac{r_A^*}{1+r_A^*} \right)^k + \frac{1-p}{1+r_B^*} \left( \frac{r_B^*}{1+r_B^*} \right)^k \quad (3.1)$$

and

$$r_A^* := q + \frac{(1-p)(1-q)}{p} \quad \text{and} \quad r_B^* := \frac{p(1-q)}{1-p} + q.$$

When  $p = 1$  the second term in  $p_k$  should be interpreted as 0.

**Remark 1.** The limiting degree distribution is a mixture of two geometric distribution shifted to the left by 1, with parameters  $1/(1+r_A^*)$  and  $1/(1+r_B^*)$ . This distribution is identical to the limiting degree distribution of a URT [28] if and only if  $p = 1/2$  or  $q = 1$ . Moreover, when  $p \neq 1/2$  and  $q \neq 1$ , the limiting proportion of leaves  $p_0$  satisfies

$$p_0 - \frac{1}{2} = \frac{p}{1+r_A^*} + \frac{1-p}{1+r_B^*} - \frac{1}{2} = \frac{(2p-1)^2(1-q)^2}{2(1-(1-2p)^2q^2)} > 0.$$

Thus in the case where the limiting degree distribution of the CMRT differs from that of the URT, the former has a larger proportion of leaves. In fact,  $p_0 \rightarrow 1$  as  $p \rightarrow 1$  and  $q \rightarrow 0$ . In this extreme case,  $\mathcal{T}_n$  looks like a “star” in which  $n-1$  type  $A$  vertices are connect to the type  $B$  root.

To better understand Theorem 3.1, we outline the following heuristic calculations based on the continuous time embedding. Consider the case where  $p < 1$ . Note that the proportions of type  $A$  and  $B$  vertices in the CMRT  $\{\mathcal{T}_n\}_{n \geq 2}$  converge almost surely to  $p$  and  $1-p$  respectively as  $n \rightarrow \infty$  by the strong law of large numbers. If one regards  $n_A(t)$  and  $n_B(t)$  as deterministic differentiable functions, formal solution of the rate equations (i) and (ii) in the definition of the process pdBP( $t$ ) yields  $n_A(t) = e^t$  and  $n_B(t) = (1-p)e^t/p$ . This suggests that, in analogy with the CMRT,

$$\frac{n_A(t)}{n_B(t)} \approx \frac{p}{1-p}$$

when  $t$  is large. One may make this rigorous using stopping time arguments. Plugging this approximation into (ii) yields corresponding approximations for the transition rates of type  $B$  individuals

$$r_{BA}(t) \approx \frac{p}{1-p} \cdot (1-q), \quad r_{BB}(t) \approx \frac{p}{1-p} \cdot \frac{q(1-p)}{p} = q.$$

Note that these approximations, and the transition rates (i) for type  $A$  individuals are independent of  $t$ . Combining the transition rates for type  $A$  we get  $r_{AA}(t) + r_{AB}(t) = r_A^*$ . Similarly, combining the approximate rates for type  $B$  individuals yields  $r_{BA}(t) + r_{BB}(t) \approx r_B^*$ .

These calculations suggest that one can approximate the population-dependent branching process by a time homogeneous multi-type branching process with types  $A, B$ , and rates  $r_A^*, r_B^*$ . Results of Jagers and Nerman [19], applied to the approximation, show that the type of a randomly chosen individual  $v$  and its age have joint limiting distribution  $\pi \times \text{Exponential}(1)$ , where  $\pi$  is a discrete measure on  $\{A, B\}$  with  $\pi(A) = p$  and  $\pi(B) = 1-p$  and  $\times$  denotes product measure. As a type  $A$  individual of age  $s$  in the homogeneous process has  $\text{Poisson}(sr_A^*)$  offspring we find

$$\mathbb{P}(v \text{ is of type } A \text{ and has } k \text{ offspring}) \approx p \int_0^\infty \mathbb{P}(\text{Poisson}(sr_A^*) = k) e^{-s} ds = \frac{p}{1+r_A^*} \left( \frac{r_A^*}{1+r_A^*} \right)^k.$$

A similar heuristic calculation can be used to motivate the limiting degree distribution for type  $B$  individuals. Note here however, that in our formal proofs we will not apply results from Jagers and Nerman [19] since we have **pdBP** instead of time homogeneous multi-type branching process.

It is worth noting that one may also carry out an approximation like that above for the URT. In this case the corresponding process is a (unitype) continuous time branching process with Poisson offspring rate 1. Results from Jagers and Nerman [19] show that, for the branching process, the limiting age distribution for a randomly chosen individual  $\nu$  is Exponential(1) and we find

$$\mathbb{P}(\nu \text{ has } k \text{ offspring}) \approx \int_0^\infty \mathbb{P}(\text{Poisson}(s) = k) e^{-s} ds = 2^{-k-1},$$

which is exactly the limiting degree distribution of URT.

Now suppose  $p = 1$ , using the construction in special case ii of Section 1.2 we have that for a randomly chosen individual  $\nu$  of type  $A$  in pdBP( $t$ )

$$\mathbb{P}(\nu \text{ has } k \text{ offsprings}) \approx \sum_{i=0}^{\infty} 2^{-k-i-1} \binom{k+i}{k} q^k (1-q)^i = \frac{q^k}{2^{k+1}} \sum_{i=0}^{\infty} \binom{k+i}{k} \left(\frac{1-q}{2}\right)^i = \frac{1}{q+1} \left(\frac{q}{q+1}\right)^k.$$

Here the infinite sum is calculated using binomial series, which is equivalent to summing up the probability mass function of a negative binomial distribution in this case.

Using the limiting degree distribution, one may derive consistent estimators of the model parameters  $p$  and  $q$  using the observed degrees of the tree. It is sufficient to consider the statistics  $N_0(n)$ , the number of leaves of  $\mathcal{T}_n$ , and  $N_1(n)$ , the number of vertices with out-degree 1 in  $\mathcal{T}_n$ .

**Corollary 3.2.** *If  $p \neq 1/2$  and  $q \neq 1$  then there exist consistent estimators  $\hat{p}_n$  and  $\hat{q}_n$  for  $p$  and  $q$  that can be computed by solving a quadratic equation involving only  $N_0(n)$  and  $N_1(n)$  (as given by (5.9) and (5.10)).*

**Remark 2.** In practice, if the number of vertices  $n$  is large, one may use a sub-sample to estimate  $N_0(n)$  and  $N_1(n)$ . If these estimates are consistent, the estimates of  $p$  and  $q$  obtained from them will be consistent as well.

**Remark 3.** As noted above, when  $p = 1/2$  or  $q = 1$  the limiting degree distribution if the CMRT matches that of the URT. In this setting it follows from results of [1] and [19] that the *local weak limits* of both the URT and the CMRT exist and are equal to the same infinite sin-tree. As such, distinguishing these two models using the densities of local statistics appears to be a difficult problem. Still, global statistics might remain informative. For example, when  $q = 1$ , the CMRT is just two disjoint URTs with the two roots connected. Using similar algorithm as introduced by Bubeck, Devroye and Lugosi [8] one can construct a confidence set for the roots and estimate  $p$  from that.

The next two results concern the limiting behavior of two global statistics for the CMRT  $\mathcal{T}_n$ , specifically, the maximal degree  $M_n$  of any vertex, and the height  $H_n$  of the tree, which we define to be the maximum distance from any vertex to the nearest root.

**Theorem 3.3.** *Let  $M_n$  denote the maximal degree of any vertex in the CMR  $\mathcal{T}_n$ . When  $p, q \neq 1$ , there exist constants  $C_1, C_2 > 0$  depending only on  $p$  and  $q$  such that*

$$\liminf_{n \rightarrow \infty} \frac{M_n}{\log n} \geq C_1 \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{M_n}{\log n} \leq C_2 \quad \text{a.s.}$$



Moreover, when  $q = 1$ ,

$$\frac{M_n}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{\log 2},$$

and when  $p = 1$  but  $q \neq 1$ ,

$$\frac{M_n}{n} \xrightarrow{\text{a.s.}} 1 - q.$$

**Theorem 3.4.** Let  $H_n$  denote the height of the CMR  $\mathcal{T}_n$ . When  $p \neq 1$  or  $q = 1$ ,

$$\frac{H_n}{\log n} \xrightarrow{\text{a.s.}} e.$$

When  $p = 1$  and  $q \neq 1$ ,

$$\liminf_{n \rightarrow \infty} \frac{H_n}{\log n} \geq qe \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{H_n}{\log n} \leq e \quad \text{a.s.}$$

**Remark 4.** When  $p = 1$  and  $q = 0$ , the tree  $\mathcal{T}_n$  looks like a “star” in which  $n-1$  type  $A$  vertices are connect to the type  $B$  root. In this case the height of the tree is  $H_n = 1$ , and  $\lim_{n \rightarrow \infty} H_n / \log n = 0$ .

**3.2. Results for general CMRT.** We now present results for  $K$  type CMRTs with arbitrary attachment probabilities.

**Theorem 3.5.** For each fixed  $k$ , let  $N_k(n)$  denote the number of vertices with out-degree  $k$  in a  $K$  type CMRT  $\mathcal{T}_n$ . Then

$$\frac{N_k(n)}{n} \xrightarrow{P} c_k$$

where

$$c_k := \sum_{i=1}^K \frac{p_i}{1+r_i} \left( \frac{r_i}{1+r_i} \right)^k.$$

Here

$$r_i = \frac{1}{p_i} \sum_{j=1}^K p_j q_{ji}.$$

When  $p_i = 0$  the  $i$ -th term in  $c_k$  should be interpreted as 0.

**Remark 5.** Analogously to the CMRT with two types, the limiting degree distribution of a  $K$ -type CMRT is a mixture of  $K$  shifted geometric distributions, which coincides with that of a URT if and only if the balance equation  $\sum_{j=1}^K p_j q_{ji} = p_i$  holds for all  $1 \leq i \leq K$ . Note that  $\sum_{j=1}^K p_j q_{ji}$  is the probability that the parent of a new vertex is of type  $i$ . Thus the balance equation essentially states that the type distribution for parents is the same as that of their children. In the continuous time embedding, the condition implies that all individuals will reproduce at approximately the same rate once the population stabilizes.

**Theorem 3.6.** *Let  $M_n$  denote the maximal degree in the  $K$ -type CMRT  $\mathcal{T}_n$ . When  $p_1 < 1$  and  $q_{ii} < 1$  for some  $1 \leq i \leq K$  there exists constants  $C_1$  and  $C_2$  depending only on  $p_i$  and  $q_{ij}$  such that*

$$\liminf_{n \rightarrow \infty} \frac{M_n}{\log n} \geq C_1 \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{M_n}{\log n} \leq C_2 \quad \text{a.s.}$$

When  $q_{ii} = 1$  for all  $1 \leq i \leq K$ ,

$$\frac{M_n}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{\log 2}$$

and when  $p_1 = 1$  but  $q_{11} < 1$ ,

$$\frac{M_n}{n} \xrightarrow{\text{a.s.}} \max_{2 \leq i \leq K} q_{1i}.$$

**Theorem 3.7.** *Let  $H_n$  denote the height of the  $K$ -type CMRT  $\mathcal{T}_n$ . When  $p_1 < 1$  or  $q_{11} = 1$ ,*

$$\frac{H_n}{\log n} \xrightarrow{\text{a.s.}} e.$$

When  $p_1 = 1$  and  $q_{11} < 1$ ,

$$\liminf_{n \rightarrow \infty} \frac{H_n}{\log n} \geq q_{11} e \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{H_n}{\log n} \leq e \quad \text{a.s.}$$

**3.3. Limiting degree distribution of CWRT.** Following the same arguments as in the case of community modulated recursive trees we can derive the limiting degree distribution of  $K$ -type community weighted recursive trees.

**Theorem 3.8.** *For each fixed  $k$ , let  $N_k(n)$  denote the number of vertices with out-degree  $k$  in a  $K$ -type CWRT  $\tilde{\mathcal{T}}_n$ . Then*

$$\frac{N_k(n)}{n} \xrightarrow{P} \tilde{c}_k$$

where

$$\tilde{c}_k := \sum_{i=1}^K \frac{p_i}{1 + \tilde{r}_i} \left( \frac{\tilde{r}_i}{1 + \tilde{r}_i} \right)^k$$

with

$$\tilde{r}_i = \sum_{j=1}^K \frac{p_j \omega_{ji}}{\sum_{l=1}^K p_l \omega_{jl}}.$$

**Remark 6.** Analogously to CMRTs, the limiting degree distribution of a  $K$ -type CWRT is a mixture of  $K$  geometric distributions with different parameters. In the special case  $\omega_{ij} \equiv 1$  the limiting distribution is identical to that of URT.

**3.4. Limiting degree distribution of CMPA.** We now describe results for the community modulated preferential attachment. Let  $p_i^* = \sum_{j=1}^K p_j q_{ji}$  be the probability for the parent of a new vertex to be of type  $i$  in CMPA (same as in CMRT). We also derive the limiting degree distribution of CMPA:

**Theorem 3.9.** For each fixed  $k$ , let  $N_k(n)$  denote the number of vertices with out-degree  $k$  in  $\mathcal{T}_n^*$ . Then

$$\frac{N_k(n)}{n} \xrightarrow{P} c_k^*$$

where

$$c_k^* := \sum_{i=1}^K \frac{p_i}{v_i} \frac{\Gamma(k + \alpha_i) \Gamma(1/v_i + \alpha_i)}{\Gamma(\alpha_i) \Gamma(k + 1 + 1/v_i + \alpha_i)}. \quad (3.2)$$

Here

$$v_i = \sum_{j=1}^K v_{ji}$$

with

$$v_{ji} = \frac{p_j q_{ji}}{\alpha_{ji} p_i + p_i^*},$$

and

$$\alpha_i = \frac{\sum_{j=1}^K v_{ji} \alpha_{ji}}{\sum_{j=1}^K v_{ji}} = \frac{\sum_{j=1}^K v_{ji} \alpha_{ji}}{v_i}.$$

And  $\Gamma(\cdot)$  is the gamma function. When  $p_i^* = 0$ , the  $i$ -th term in  $c_k^*$  should be interpreted as  $p_i$  when  $k = 0$ , and 0 otherwise.

**Remark 7.** Similar to CMRT, the limiting degree distribution of CMPA tree is a mixture of  $K$  distributions, and coincides with that of preferential attachment model [6] when  $\alpha_{ji} \equiv \alpha$  and the balance equation  $p_i^* = \sum_{j=1}^K p_j q_{ji} = p_i$  holds for all  $1 \leq i \leq K$ . Moreover, for large  $k$ , using Stirling's approximation we get  $c_k^* = O(k^{-1-1/v})$  where  $v = \max_i v_i$ . Consider the special case where  $\alpha_{ji} \equiv \alpha$ . we have  $\sum_{i=1}^K p_i^* = \sum_{j=1}^K p_j \sum_{i=1}^K q_{ji} = \sum_{j=1}^K p_j = 1$ , and  $\sum_{i=1}^K p_i^* (1/v_i - 1) = \alpha \sum_{i=1}^K p_i = \alpha$ . It follows that  $-1 - 1/v \geq -2 - \alpha$  and equality holds if and only if the balance equation holds. Therefore, the limiting degree distribution of CMPA tree typically has a heavier tail than that of preferential attachment tree with the same parameter  $\alpha$ . The tail is heavier when  $p_i^*$  is larger compared to  $p_i$ , which agrees with intuition.

#### 4. RELATED WORK AND DISCUSSION

- (a) **Branching processes and growing trees:** Branching processes have been used to study a variety of tree based stochastic structures. See Devroye [12] for a nice summary of classical results that can be proved via branching processes, both discrete and continuous. In this paper we make use of continuous time branching processes. Since Athreya and Karlin [2] introduced the continuous time embedding of urn processes, embeddings of this sort has been used to study a variety of tree structures. The well-studied Crump-Mode-Jagers branching process [9] is known to give rise to a rich family of stochastic trees, of which uniform recursive trees are a special case. Indeed, some of the results for the URT described below were proved via continuous time embeddings [14][32]. Although the standard techniques for analyzing asymptotic degree distributions were developed under the discrete setting, it could be fruitful to work with continuous time embeddings for more complicated models [4]. In general, this technique is more mathematically involved, but when tractable allows access to derive asymptotics for a host of functionals including degree distribution, maximal degree height and so on.

- (b) **Related work on the URT:** Our results on the asymptotics of CMRT are closely related to that of URT, which have been studied extensively. Let  $\{\mathcal{U}_n\}_{n \geq 1}$  be a URT, and  $N_k(n)$  ( $k \geq 0$ ) denote the number of vertices with out-degree  $k$  in  $\mathcal{U}_n$ . Moon[28] showed that

$$\frac{N_k(n)}{n} \xrightarrow{P} 2^{-k-1},$$

and Janson[20] further showed that

$$n^{-\frac{1}{2}}(N_k(n) - 2^{-k-1}n) \xrightarrow{d} V_k$$

jointly for all  $k \geq 0$ , where  $\{V_k\}_{k \geq 0}$  is a Gaussian process.

Let  $M_n$  and  $H_n$  denote respectively the maximal degree and height in  $\mathcal{U}_n$ . Here height is defined as length of the longest upward path from a leaf to the root. Devroye and Lu[14] showed that

$$\frac{M_n}{\log_2 n} \xrightarrow{\text{a.s.}} 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{EM_n}{\log_2 n} = 1.$$

Also, Devroye[11] and Pittel[32] showed that

$$\frac{H_n}{\log n} \xrightarrow{\text{a.s.}} e.$$

- (c) **Related work on the preferential attachment tree:** The preferential attachment tree and its variants have also been extensively studied in the literature. Recall that in linear preferential attachment trees, instead of uniform attachment, a new vertex connects to an existing vertex with probability proportional to  $\alpha$  plus that vertex's out-degree. The linear preferential attachment tree is known to have a limiting power law distribution with exponent depending on the model parameter  $\alpha$  [6]. In terms of models with multiple types, which are our focus here, to the best of our knowledge the only result we are aware of is Deijfen and Fitzner [10], who computed the limiting degree distribution for a special case of CMPA with two types heuristically and conducted a simulation study.
- (d) **Community detection:** As described in the introduction, one motivation for this work was to draw connections between the research area of community detection and that of stochastic trees. It is well nigh impossible to give a representative set of references to community detection, but the [16, 22, 30] and the references therein provide a good overview of applications, while [7, 23] provide an introduction to the burgeoning literature in the probability community. Loosely speaking, community detection is an unsupervised learning task (closely related to clustering) for obtaining insight into networks, in particular understanding subsets within a network, vertices within which seem more densely connected within the subset as opposed to outside the subset. A host of techniques have been proposed to extract such subsets within networks and there has also been an enormous effort in the probability community to evaluate the performance of proposed techniques on tractable network models with known community structure. The aim of this work was to derive math results for the asymptotics of various functionals related to community modulated growth schemes in networks. In work in progress we propose and study the performance of various techniques for extracting the underlying latent community structure if we only have access to the graph without any information on the corresponding types of vertices.

## 5. PROOFS

This section is organized as follows. First in Section 5.1 we prove the equivalence of the continuous time embedding, together with some of its basic properties. Then in Section 5.2 we derive the limiting degree distribution for Community Modulated Recursive Tree (CMRT) with two types and use it in Section 5.3 to give consistent estimators of model parameters. In Section 5.4 and 5.5 we prove results for maximal degree and height. Finally, Section 5.6, 5.7 and 5.8 extend the results to general CMRT and the other two variants: Community Weighted Recursive Tree (CWRT) and Community Modulated Preferential Attachment (CMPA).

**5.1. Proof for continuous-time embedding and some basic properties.** In this section we shall first give a proof of the continuous-time embedding (Lemma 2.1).

**Proof:** Assume that  $\{\text{pdBP}(T_n)\}_{2 \leq n \leq k} \stackrel{d}{=} \{\mathcal{T}_n\}_{2 \leq n \leq k}$  for a fixed integer  $k \geq 2$  (which holds for  $k = 2$  by definition). Conditioning on  $\{\text{pdBP}(T_n)\}_{2 \leq n \leq k}$ , it can be checked using properties of exponential distribution that the probability for the next individual born to be of type  $A$  is

$$\frac{n_A(T_n)r_{AA}(T_n) + n_B(T_n)r_{BA}(T_n)}{n_A(T_n)(r_{AA}(T_n) + r_{AB}(T_n)) + n_B(T_n)(r_{BA}(T_n) + r_{BB}(T_n))} = p$$

and the probability for the next type  $A$  individual born to have a parent of type  $A$  is

$$\frac{n_A(T_n)r_{AA}(T_n)}{n_A(T_n)r_{AA}(T_n) + n_B(T_n)r_{BA}(T_n)} = q.$$

Also, given a type  $\epsilon \in \{A, B\}$ , the probability for each individual of this type to give birth to the next type  $A$  individual is equal. Similarly one can check the corresponding probability for type  $B$  individuals. Thus the dynamics of  $\text{pdBP}(T_{k+1})$  conditioning on  $\{\text{pdBP}(T_n)\}_{2 \leq n \leq k}$  is the same as that of  $\mathcal{T}_{k+1}$  conditioning on  $\{\mathcal{T}_n\}_{2 \leq n \leq k}$ . Therefore by induction we have the desired result.  $\blacksquare$

Now that we have the continuous-time embedding, we shall proceed to derive some of its basic properties that will come in handy. In what follows, we shall assume that natural filtration  $\{\mathcal{F}(t) : t \geq 0\}$  is used throughout.

**Lemma 5.1.** *The process  $\{e^{-t}n_A(t)\}_{t \geq 0}$  is an  $\mathbb{L}^2$ -bounded positive martingale. In particular there exists a strictly positive finite random variable  $W$  such that*

$$e^{-t}n_A(t) \xrightarrow{\text{a.s.}} W, \quad \text{as } t \rightarrow \infty.$$

**Remark 8.** As will be evident from the calculation below, the marginal distribution of  $n_A(\cdot)$  is identical to that of a rate one Yule process starting with a single individual. In particular the limit random variable  $W \stackrel{d}{=} \exp(1)$ . Recall that a Yule process with rate  $\lambda$  is a time-inhomogeneous Poisson process with birth rate  $\lambda i$ , where  $i$  is the current population size.

**Proof:** First we introduce some preliminary notations that will be used extensively throughout the proofs. Recall that for a jump diffusion  $\{X(t) \in \mathbb{R}^n\}_{t \geq 0}$ , its infinitesimal generator  $\mathcal{A}$  is defined for functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$\mathcal{A}f(x) = \lim_{t \rightarrow 0^+} \frac{1}{t} (\mathbb{E}(f(X(t)) | X(0) = x) - f(x))$$

if the limit exists. Then by Dynkin's formula (see Øksendal and Sulem [31, Chapter 1.3] for a formulation) and Markov property of jump diffusion we have that  $\{X(t) - \int_0^t \mathcal{A}X(s)ds\}_{t \geq 0}$  is a

martingale. Note that if the diffusion term is not present and the jump part is an inhomogeneous Poisson process, we have that  $\mathcal{A}X(t) = \delta(X(t))\lambda(X(t))$  where  $\delta(x)$  and  $\lambda(x)$  are size and intensity of jump when the process is at  $x \in \mathbb{R}^n$ .

Now we are ready to introduce some martingales. Denote  $\exp(-t)n_A(t)$  by  $\tilde{n}_A(t)$ . Using the rates in Section 2 we have

$$\mathcal{A}\tilde{n}_A(t) = e^{-t}n_A(t) - e^{-t}n_A(t) = 0$$

and

$$\mathcal{A}\tilde{n}_A^2(t) = e^{-2t}(2n_A(t) + 1)n_A(t) - 2e^{-2t}n_A^2(t) = e^{-2t}n_A(t).$$

Thus by Dynkin's formula we have that both  $\{\tilde{n}_A(t) : t \geq 0\}$  and

$$\tilde{n}_A^2(t) - \int_0^t e^{-2s}n_A(s)ds, \quad t \geq 0$$

are martingales. Taking expectation of both martingales we get  $\mathbb{E}(n_A(t)) = \exp(t)$  and  $\mathbb{E}(\tilde{n}_A^2(t)) = 2 - \exp(-t)$ . Therefore  $\{\tilde{n}_A(t) : t \geq 0\}$  is  $\mathbb{L}^2$  bounded and the second statement follows from standard martingale convergence theorem. ■

**Lemma 5.2.** *Define  $Z(t) := pn_B(t) - (1-p)n_A(t)$ . Then  $\{Z(t) : t \geq 0\}$  is a martingale and further  $e^{-t}Z(t) \xrightarrow{\text{a.s.}} 0$ . This implies  $e^{-t}n_B(t) \xrightarrow{\text{a.s.}} (1-p)W/p$  where  $W$  is as in Lemma 5.1.*

**Proof:** Again using the rates in Section 2 we have

$$\mathcal{A}Z(t) = p\frac{1-p}{p}n_A(t) - (1-p)n_A(t) = 0$$

and

$$\mathcal{A}Z^2(t) = p^2\mathcal{A}n_B^2(t) + (1-p)^2\mathcal{A}n_A^2(t) - 2p(1-p)\mathcal{A}n_An_B(t),$$

with

$$\mathcal{A}n_B^2(t) = (2n_B(t) + 1)\frac{1-p}{p}n_A(t), \quad \mathcal{A}n_A^2(t) = (2n_A(t) + 1)n_A(t)$$

and

$$\mathcal{A}n_An_B(t) = n_A(t)\frac{1-p}{p}n_A(t) + n_B(t)n_A(t).$$

Thus by some elementary algebra and Dynkin's formula we have that both  $\{Z(t) : t \geq 0\}$  and

$$M(t) = Z^2(t) - \int_0^t (1-p)n_A(s)ds, \quad t \geq 0$$

are martingales. From Lemma 5.1 we get that  $\mathbb{E}(n_A(s)) = \exp(s)$ . Taking expectation of  $\mathbb{E}(M(t))$  shows

$$\mathbb{E}(Z^2(t)) = (1-p)(e^t - 1) + (2p-1)^2.$$

Now apply Markov's inequality to  $\exp(-4\log n)Z^2(2\log n)$  we have for any  $\epsilon > 0$ :

$$\mathbb{P}\left(e^{-4\log n}Z^2(2\log n) > \epsilon\right) \leq \frac{e^{-4\log n}((1-p)(e^{2\log n} - 1) + (2p-1)^2)}{\epsilon} = \frac{1-p}{n^2\epsilon} + \frac{4p^2-3p}{n^4\epsilon}.$$

Thus by the Borel–Cantelli lemma  $\exp(-2\log n)Z(2\log n) \xrightarrow{\text{a.s.}} 0$ . Since we know by Lemma 5.1 that  $\exp(-2\log n)n_A(2\log n) \xrightarrow{\text{a.s.}} W$ , we get  $\exp(-2\log n)n_B(2\log n) \xrightarrow{\text{a.s.}} (1-p)W/p$ .

Finally for any  $t > 0$  we can find a positive integer  $n$  such that  $2\log n \leq t < 2\log(n+1)$ . By monotonicity of  $n_B(\cdot)$  we have  $n_B(2\log n) \leq n_B(t) < n_B(2\log(n+1))$  and further

$$e^{-2\log(n+1)}n_B(2\log n) \leq e^{-t}n_B(t) < e^{-2\log n}n_B(2\log(n+1)).$$

Since the left hand side

$$e^{-2\log(n+1)} n_B(2\log n) = e^{-2\log n} n_B(2\log n) \cdot \frac{n^2}{(n+1)^2} \xrightarrow{\text{a.s.}} \frac{(1-p)W}{p}$$

and similarly the right hand side converges a.s. to the same limit, we have  $\exp(-t)n_B(t) \xrightarrow{\text{a.s.}} (1-p)W/p$ . This immediately implies  $\exp(-t)Z(t) \xrightarrow{\text{a.s.}} 0$ . ■

**Remark 9.** This lemma essentially proves that  $n_B(t)/n_A(t) \xrightarrow{\text{a.s.}} (1-p)/p$ , which is used in Section 3.1 for heuristic calculations.

**Lemma 5.3.** *The population size process  $n(t)$  satisfies  $e^{-t}n(t) \xrightarrow{\text{a.s.}} W + (1-p)W/p := W_\infty$ . In particular, the sequence of stopping times  $T_n$  satisfy*

$$T_n - \log n \xrightarrow{\text{a.s.}} -\log(W_\infty),$$

where  $W_\infty$  is a strictly positive finite random variable.

The first statement here is a direct corollary of Lemma 5.1 and 5.2, and the second statement follows by replacing  $t$  with  $T_n$ . Note that the first statement essentially says that the Malthusian parameter for process  $\{n(t) : t \geq 0\}$  is 1.

**5.2. Proof for limiting degree distribution.** In this section we shall prove Theorem 3.1.

To work in continuous time, we need to reformulate Theorem 3.1 via the embedding. Let  $N_{k,A}(n)$  and  $N_{k,B}(n)$  denote the number of type  $A$  and  $B$  vertices with out-degree  $k$  in  $\mathcal{T}_n$ . The plan is as follows: first we focus on type  $A$  vertices and prove the result below, and as similar result holds for type  $B$  vertices, using  $N_k(n) = N_{k,A}(n) + N_{k,B}(n)$  we completes the proof. Note that when  $p = 1$ , there is no need to consider type  $B$  vertices.

**Theorem 5.4.** *For a fixed integer  $k > 0$ , let  $n_{k,A}(t)$  denote the number of type  $A$  individuals with  $k$  offsprings in  $\text{pdBP}(t)$ . Then*

$$\frac{n_{k,A}(t)}{n(t)} \xrightarrow{P} p_{k,A}, \quad \text{as } t \rightarrow \infty.$$

Here

$$p_{k,A} := p \int_0^\infty \mathbb{P}(\text{Poisson}(r_A^* s) = k) e^{-s} ds = \frac{p}{1+r_A^*} \left( \frac{r_A^*}{1+r_A^*} \right)^k$$

where  $r_A^*$  is the total reproduction rate of type  $A$  individuals:

$$r_A^* = q + \frac{(1-p)(1-q)}{p}. \tag{5.1}$$

Thus by the embedding in Lemma 2.1 for  $\{\mathcal{T}_n\}_{n \geq 2}$  we have

$$\frac{N_{k,A}(n)}{n} \xrightarrow{P} p_{k,A}, \quad \text{as } n \rightarrow \infty.$$

**Proof:** Throughout the proof we work with the continuous time embedding. For any fixed constant  $0 < a < t$ , let  $n_{k,A}[t-a, t]$  be the number of type  $A$  individuals born in the interval  $[t-a, t]$  that have exactly  $k$  offsprings by time  $t$ . Given Lemma 5.3, it is enough to show the following two propositions.

**Proposition 5.5.**

$$\limsup_{a \rightarrow \infty} \limsup_{t \rightarrow \infty} e^{-t} (n_{k,A}(t) - n_{k,A}[t-a, t]) = 0, \quad a.s.$$

**Proof:** Since the population size  $n_A(t)$  grows exponentially, most type  $A$  individuals are born after time  $t - a$ . Indeed,  $n_{k,A}(t) - n_{k,A}[t-a, t] = n_{k,A}(t-a) \leq n_A(t-a)$ , and by Lemma 5.1 we have

$$\limsup_{t \rightarrow \infty} e^{-t} (n_{k,A}(t) - n_{k,A}[t-a, t]) \leq e^{-a} \lim_{t \rightarrow \infty} e^{-(t-a)} n_A(t-a) = e^{-a} W, \quad a.s.$$

Letting  $a \rightarrow \infty$  proves the proposition. ■

**Proposition 5.6.** *Recall the random variable  $W$  in Lemma 5.2. Then for each fixed  $a > 0$ , we have*

$$e^{-t} n_{k,A}[t-a, t] \xrightarrow{P} W \int_0^a \mathbb{P}(\text{Poisson}(r_A^* s) = k) e^{-s} ds$$

as  $t \rightarrow \infty$ .

This assertion needs some work and the proof follows a similar procedure as that in Bhamidi et al. [4, Section 4.2.2]. First, recall from Lemma 5.1 that  $n_A(t) \approx We^t$  for large  $t$ . For our proof we will need a finer concentration result that goes as follows:

**Lemma 5.7.**

$$\mathbb{P} \left( \sup_{t-a \leq s \leq t} |n_A(s) - We^s| < \sqrt{te^t} \right) \rightarrow 1$$

as  $t \rightarrow \infty$  where  $W$  is as in Lemma 5.1. Equivalently, we shall say that w.h.p. as  $t \rightarrow \infty$ ,  $\sup_{t-a \leq s \leq t} |n_A(s) - W \exp(s)| < \sqrt{t \exp(t)}$ .

**Proof:** First note that

$$e^{-s} n_A(s) - e^{-(t-a)} n_A(t-a), \quad s \geq t-a$$

is a martingale by Lemma 5.1 and recall that  $\mathbb{E}(\exp(-2s) n_A^2(s)) = 2 - \exp(-s)$  from the proof.

Now fix any  $T > t - a$ , Doob's  $L^2$ -maximal inequality (here we use the version stated in Øksendal [31, Theorem 3.2.4], see also Karatzas and Shreve [21, Theorem 3.8] for a proof) applied to the above martingale gives, for any  $C > 0$ :

$$\mathbb{P} \left( \sup_{t-a \leq s \leq T} |e^{-s} n_A(s) - e^{-(t-a)} n_A(t-a)| > C \right) \leq \frac{\mathbb{E}(e^{-T} n_A(T) - e^{-(t-a)} n_A(t-a))^2}{C^2} \quad (5.2)$$

Recall that  $\{\tilde{n}_A(t) := \exp(-t) n_A(t)\}_{t \geq 0}$  is a martingale so we have

$$\mathbb{E}(\tilde{n}_A(T) \tilde{n}_A(t-a)) = \mathbb{E}(\tilde{n}_A(t-a) \mathbb{E}(\tilde{n}_A(T) | \tilde{n}_A(t-a))) = \mathbb{E}(\tilde{n}_A^2(t-a)) = \mathbb{E}(e^{-2(t-a)} n_A^2(t-a)).$$

It follows that

$$\mathbb{E}(e^{-T} n_A(T) - e^{-(t-a)} n_A(t-a))^2 = \mathbb{E}(e^{-2(t-a)} n_A^2(t-a)) - \mathbb{E}(e^{-2T} n_A^2(T)) = e^{-(t-a)} - e^{-T}.$$

Plug this in (5.2) we get

$$\mathbb{P} \left( \sup_{t-a \leq s \leq T} |e^{-s} n_A(s) - e^{-(t-a)} n_A(t-a)| > C \right) \leq \frac{e^{-(t-a)} - e^{-T}}{C^2} \quad (5.3)$$

Now, let  $T \rightarrow \infty$  and use the a.s. convergence result from Lemma 5.1 to yield



$$\mathbb{P}(|W - e^{-(t-a)} n_A(t-a)| > C) \leq \frac{e^{-(t-a)}}{C^2}.$$

On the other hand, let  $T = t$  in (5.3) we get

$$\mathbb{P}\left(\sup_{t-a \leq s \leq t} |e^{-s} n_A(s) - e^{-(t-a)} n_A(t-a)| > C\right) \leq \frac{e^{-(t-a)} - e^{-t}}{C^2} \leq \frac{e^{-(t-a)}}{C^2}.$$

Finally, combine the above two inequalities we have

$$\mathbb{P}\left(\sup_{t-a \leq s \leq t} |e^{-t} n_A(t) - W| > 2C\right) \leq \frac{2e^{-(t-a)}}{C^2}.$$

Let  $C = \sqrt{t \exp(-t)}/2$  and after some simple algebraic manipulation we get the wanted result. ■

Now that we have Lemma 5.7, we will proceed to approximate the integral in Proposition 5.6 by  $e^{-t} n_{k,A}[t-a, t]$ . To do so, we divide the interval  $[t-a, t]$  into intervals of length  $\delta := e^{-t/3}$  and denote by  $\mathcal{I}_i$  ( $1 \leq i \leq a/\delta$ ) the  $i$ -th interval  $[t-a + (i-1)\delta, t-a + i\delta]$  (for simplicity we treat  $a/\delta$  as if it is an integer but this should not matter). Let  $t_i = t-a + (i-1)\delta$  and write  $n_A(\mathcal{I}_i)$  for the number of type  $A$  individuals born in  $\mathcal{I}_i$ .

On the event given in Lemma 5.7, we have

$$|n_A(t_i + \delta) - n_A(t_i) - W e^{t_i} (e^\delta - 1)| < 2\sqrt{t e^t}.$$

Since  $W \exp(t_i)(\exp(\delta) - 1) = \delta W \exp(t_i) + o_p(\sqrt{\exp(t)})$ , we can further get

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{|n_A(\mathcal{I}_i) - \delta W e^{t_i}| < 3\sqrt{t e^t}\right\}\right) \rightarrow 1 \quad (5.4)$$

as  $t \rightarrow \infty$ . This shows that the number of type  $A$  individuals born in  $\mathcal{I}_i$  is approximately  $\delta W e^{t_i}$ .

Unfortunately, even individuals born in the same interval could start reproducing at different times (*i.e.* have distinct birth times), and we need to at least align these individuals. To facilitate the analysis, for each type  $A$  individual born in  $\mathcal{I}_i$ , we call it “good” if it gives birth to no offspring (type  $A$  or  $B$ ) in  $\mathcal{I}_i$  and “bad” otherwise. Since the interval is small, we expect most individuals born to be good:

**Lemma 5.8.** *There exists a constant  $M > 0$  such that*

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{n_A^{bad}(\mathcal{I}_i) < M W t e^{t/3}\right\}\right) \rightarrow 1$$

as  $t \rightarrow \infty$ , where  $n_A^{bad}(\mathcal{I}_i)$  is the number of bad individuals born in  $\mathcal{I}_i$ .

**Proof:** Call a bad individual in  $\mathcal{I}_i$  a “direct” bad individual if it is an offspring of individuals born before  $\mathcal{I}_i$  and write  $n_A^{dir}(\mathcal{I}_i)$  for the number of direct bad individuals born in  $\mathcal{I}_i$ . As non-direct bad individuals in  $\mathcal{I}_i$  must have an antecedent that is direct bad in  $\mathcal{I}_i$ , we have  $n_A^{bad}(\mathcal{I}_i) \leq n_A^{desc}(\mathcal{I}_i)$ , where  $n_A^{desc}(\mathcal{I}_i)$  is the number of descendants of direct bad individuals in  $\mathcal{I}_i$ .

Note that a direct bad individual in  $\mathcal{I}_i$  has to satisfy the following two conditions:

- First it has to be an offspring of individuals born before  $\mathcal{I}_i$  and the number of such offsprings is at most  $n_A(\mathcal{I}_i)$ .

- Secondly it has to give birth to at least one offspring in  $\mathcal{I}_i$  and the corresponding (conditional) probability is at most

$$p_{bad} := \mathbb{P}(\text{Exp}(r_A^*) \leq \delta) \sim r_A^* \delta.$$

Combining the above conditions we have

$$n_A^{dir}(\mathcal{I}_i) \leq \text{Bin}(n_A(\mathcal{I}_i), p_{bad}).$$

where  $\leq$  denotes stochastic dominance. By (5.4) we have that w.h.p. as  $t \rightarrow \infty$ ,  $W \exp(2t/3)/2 \leq n_A(\mathcal{I}_i) \leq 2W \exp(2t/3)$  for all  $i$ . Condition on this event and use Hoeffding inequality for binomial distribution we get

$$\mathbb{P}\left(n_A^{dir}(\mathcal{I}_i) > n_\delta p_{bad} + \sqrt{n_\delta \log(n_\delta)}\right) \leq \frac{1}{n_A^2(\mathcal{I}_i)} \leq \frac{4}{W^2 \exp(4t/3)}$$

where  $n_\delta := 2W \exp(2t/3)$ . Since there exists  $C_1 > 0$  such that  $n_\delta p_{bad} + \sqrt{n_\delta \log(n_\delta)} < C_1 W t e^{t/3}$  for large enough  $t$ , and

$$\frac{a}{\delta} \cdot \frac{4}{W^2 \exp(4t/3)} \rightarrow 0$$

as  $t \rightarrow \infty$ , using a union bound we have

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{n_A^{dir}(\mathcal{I}_i) \leq C_1 W t e^{t/3}\right\}\right) \rightarrow 1 \quad (5.5)$$

as  $t \rightarrow \infty$ .

Conditioning on the event in (5.5) we have

$$n_A^{desc}(\mathcal{I}_i) \leq \sum_{j=1}^{n_A^{dir}(\mathcal{I}_i)} Y_j(\delta) \leq \sum_{j=1}^{C_1 W t \exp(t/3)} Y_j(\delta)$$

where  $\{Y_j(\cdot) : j \geq 1\}$  is an sequence of *i.i.d.* Yule process with rate  $r_A^*$ . Since  $Y_j(t)$  follows a geometric distribution with parameter  $\exp(-r_A^* t)$ , we have that for any positive integer  $C_2$ :

$$\mathbb{P}(Y_j(\delta) \geq C_2) = (1 - \exp(-r_A^* \delta))^{C_2-1} \leq (r_A^* \delta)^{C_2-1}$$

Using a union bound it follows (for simplicity the conditional is suppressed) that

$$\mathbb{P}(n_A^{desc}(\mathcal{I}_i) \geq C_1 C_2 W t e^{t/3}) \leq C_1 W t e^{t/3} \mathbb{P}(Y_j(\delta) \geq C_2) \leq C_1 W (r_A^*)^{C_2-1} t (e^{t/3})^{2-C_2}.$$

Apply union bound once again we get

$$\mathbb{P}\left(\bigcup_{i=1}^{a/\delta} \left\{n_A^{desc}(\mathcal{I}_i) \geq C_1 C_2 W t e^{t/3}\right\}\right) \leq \frac{a}{\delta} \cdot C_1 W (r_A^*)^{C_2-1} t (e^{t/3})^{2-C_2} \rightarrow 0$$

as  $t \rightarrow \infty$  for  $C_2 \geq 4$ . Recall that  $n_A^{bad}(\mathcal{I}_i) \leq n_A^{desc}(\mathcal{I}_i)$  and we can get rid of the conditional that was suppressed with the aid of (5.5). Then letting  $M = 4C_1$  completes the proof. ■

Now that we get the bad individuals under control, we can turn our attention to those good individuals. To start, combine Lemma 5.8 with (5.4) we have

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{|n_A^{good}(\mathcal{I}_i) - \delta W e^{t/3}| < 4\sqrt{t e^t}\right\}\right) \rightarrow 1 \quad (5.6)$$

as  $t \rightarrow \infty$ , where  $n_A^{good}(\mathcal{I}_i)$  is the number of good individuals in  $\mathcal{I}_i$ .

Next, note that good individuals in  $\mathcal{I}_i$  reproduce independently at rate  $r_A^*$  starting from time  $t_i + \delta = t_{i+1}$ . In particular the probability that a good individual has  $k$  offsprings by time  $t$  is

$$g_k(t - t_{i+1}) := \mathbb{P}(\text{Poisson}(r_A^*(t - t_{i+1})) = k).$$

Here  $t - t_{i+1}$  refers to the time left until time  $t$ . Since reproductions are all independent, we have (conditioning on  $\text{pdBP}(t_{i+1})$ )

$$n_{k,A}^{\text{good}}(\mathcal{I}_i) \stackrel{d}{=} \text{Bin}(n_A^{\text{good}}(\mathcal{I}_i), g_k(t - t_{i+1})) \quad (5.7)$$

where  $n_{k,A}^{\text{good}}(\mathcal{I}_i)$  is the number of good individuals in  $\mathcal{I}_i$  that have  $k$  offsprings by time  $t$ . Similar to the proof of (5.5), by (5.6), (5.7) and Hoeffding inequality for binomial distribution together with a union bound, there exists  $C > 0$  such that

$$\mathbb{P}\left(\bigcap_{i=1}^{a/\delta} \left\{ |n_{k,A}^{\text{good}}(\mathcal{I}_i) - \delta W e^{t_i} g_k(t - t_{i+1})| < CW \log(W) t e^{t/3} \right\}\right) \rightarrow 1 \quad (5.8)$$

as  $t \rightarrow \infty$ .

With both good and bad individuals under control, we are ready to prove what we started out for. Note that

$$\sum_{i=1}^{a/\delta} n_{k,A}^{\text{good}}(\mathcal{I}_i) \leq n_{k,A}[t - a, t] \leq \sum_{i=1}^{a/\delta} [n_{k,A}^{\text{good}}(\mathcal{I}_i) + n_A^{\text{bad}}(\mathcal{I}_i)].$$

For the bad individuals we know from Lemma 5.8 that

$$e^{-t} \sum_{i=1}^{a/\delta} n_A^{\text{bad}}(\mathcal{I}_i) \xrightarrow{P} 0 \quad \text{as} \quad t \rightarrow \infty.$$

On the other hand, for the good individuals we have from (5.8) that

$$e^{-t} \sum_{i=1}^{a/\delta} n_{k,A}^{\text{good}}(\mathcal{I}_i) - \sum_{i=1}^{a/\delta} \delta W e^{-(t-t_i)} g_k(t - t_{i+1}) \xrightarrow{P} 0 \quad \text{as} \quad t \rightarrow \infty.$$

Finally, note that by definition of Riemann integral we know

$$\sum_{i=1}^{a/\delta} \delta W e^{-(t-t_i)} g_k(t - t_{i+1}) \rightarrow W \int_0^a \mathbb{P}(\text{Poisson}(r_A^* s) = k) e^{-s} ds \quad \text{as} \quad t \rightarrow \infty.$$

This completes the proof for Proposition 5.6. ■

**5.3. Estimation of model parameters.** In this section we derive consistent estimators for model parameters  $p$  and  $q$  based on the limiting degree distribution (i.e. Corollary 3.2).

Specifically, from Theorem 3.1 we know

$$\frac{N_0(n)}{n} \xrightarrow{P} \frac{p}{1 + r_A^*} + \frac{1 - p}{1 + r_B^*}$$

and

$$\frac{N_1(n)}{n} \xrightarrow{P} \frac{p r_A^*}{(1 + r_A^*)^2} + \frac{(1 - p) r_B^*}{(1 + r_B^*)^2}.$$

Let  $\theta_A := 1/(1+r_A^*)$  and  $\theta_B := 1/(1+r_B^*)$ . If we can solve for unique  $\hat{p}$ ,  $\hat{\theta}_A$  and  $\hat{\theta}_B$  (up to a switch between two types) that satisfy the following equations

$$m_1 := \frac{N_0(n)}{n} = \hat{p}\hat{\theta}_A + (1-\hat{p})\hat{\theta}_B \quad (5.9)$$

$$m_2 := \frac{N_0(n) - N_1(n)}{n} = \hat{p}\hat{\theta}_A^2 + (1-\hat{p})\hat{\theta}_B^2, \quad (5.10)$$

then  $\hat{p}$ ,  $\hat{\theta}_A$  and  $\hat{\theta}_B$  are consistent estimators of  $p$ ,  $\theta_A$  and  $\theta_B$ . We can further solve for a consistent estimator  $\hat{q}$  for  $q$  from either  $\hat{\theta}_A$  or  $\hat{\theta}_B$  together with  $\hat{p}$ . Note that all estimators are consistent because (as we shall see) they are continuous functions of  $m_1$  and  $m_2$ .

To solve three unknowns from two equations (5.9) and (5.10), we need a third equation

$$\frac{\hat{p}}{\hat{\theta}_A} + \frac{1-\hat{p}}{\hat{\theta}_B} = 2 = p(1+r_A^*) + (1-p)(1+r_B^*) \quad (5.11)$$

where the second equality can be verified from the definitions of  $r_A^*$  and  $r_B^*$ .

From (5.9) we have

$$\hat{p} = \frac{m_1 - \hat{\theta}_B}{\hat{\theta}_A - \hat{\theta}_B}. \quad (5.12)$$

Plug it back to (5.10) and (5.11) we get  $m_1(\hat{\theta}_A + \hat{\theta}_B) - \hat{\theta}_A\hat{\theta}_B = m_2$  and  $\hat{\theta}_A + \hat{\theta}_B - m_1 = 2\hat{\theta}_A\hat{\theta}_B$ . Then we can solve for

$$\hat{\theta}_A\hat{\theta}_B = \frac{m_2 - m_1^2}{2m_1 - 1} \quad \text{and} \quad \hat{\theta}_A + \hat{\theta}_B = \frac{2m_2 - m_1}{2m_1 - 1},$$

so  $\hat{\theta}_A$  and  $\hat{\theta}_B$  are roots of the following quadratic equation:

$$x^2 - \frac{2m_2 - m_1}{2m_1 - 1}x + \frac{m_2 - m_1^2}{2m_1 - 1} = 0.$$

After we solve for  $\hat{\theta}_A$  and  $\hat{\theta}_B$  we can compute  $\hat{p}$  from (5.12) and

$$\hat{q} = \frac{\hat{p} - \hat{\theta}_A}{(2\hat{p} - 1)\hat{\theta}_A} = \frac{1 - \hat{p} - \hat{\theta}_B}{(1 - 2\hat{p})\hat{\theta}_B}$$

from definitions of  $\theta_A$  and  $\theta_B$ .

Note that these estimators fail when  $m_1 - 1/2$ ,  $m_2 - m_1^2$  or  $2m_2 - m_1$  is negative. However, these cases are unlikely to occur for CMRT as all three quantities have positive limits as  $n \rightarrow \infty$  (recall that  $m_1$  and  $m_2$  both depend on  $n$ ). When  $m_1 = 1/2$ , these estimators also fail: from the remark under Theorem 3.1 we know  $m_1$  has limit  $1/2$  if and only if  $p = 1/2$  or  $q = 1$ , and parameters  $p, q$  are not identifiable from the limiting degree distribution in these special cases.

**5.4. Proof for maximal degree.** In this section we shall prove Theorem 3.3. Throughout the proof we work with the continuous time embedding unless otherwise noted.

First we consider cases where  $p, q \neq 1$ . By Lemma 5.3 and Egoroff's Theorem, given any  $\epsilon > 0$ , we can choose  $K > 0$  such that (the dependence of  $K$  on  $\epsilon$  is suppressed throughout):

$$\mathbb{P}\left(\sup_n |T_n - \log n| < K\right) > 1 - \epsilon. \quad (5.13)$$

To ease notations, write  $T_n^+ = \log n + K$  and  $T_n^- = \log n - K$ .

**Lower bound:** Here it is enough to consider just the type  $A$  root. Recall from Section 2 that this vertex reproduces at constant rate

$$r_A^* = q + \frac{(1-p)(1-q)}{p}. \quad (5.14)$$

Denote by  $D(t)$  the out-degree of type  $A$  root at time  $t$ . Then  $D(t)$  follows a  $\text{Poisson}(r_A^* t)$  distribution. So for any  $0 < \gamma < 1$ , by standard tail bound for Poisson distribution we have

$$\mathbb{P}(D(T_n) \leq \gamma r_A^* \log n) \leq \mathbb{P}(D(T_n^-) \leq \gamma r_A^* \log n) \leq \exp\{-M \log n\} \quad (5.15)$$

for large enough  $n$ , conditioning on the event in (5.13). Here  $M > 0$  is a constant that depends on both  $\gamma$  and  $r_A^*$ .

Denote by  $M(t)$  the maximal number of offsprings an individual has by time  $t$ . Since  $M(t) \geq D(t)$ , it follows from (5.15) that

$$\mathbb{P}(M(T_n) \leq \gamma r_A^* \log n) \leq n^{-M} \quad (5.16)$$

conditioning on the event in (5.13).

Let  $\{n_k\}_{k \geq 1}$  be an increasing sequence of positive integers such that  $n_k \geq k^{2/M}$ . Then from (5.16) we know

$$\mathbb{P}(M(T_{n_k}) \leq \gamma r_A^* \log n_k) \leq \frac{1}{k^2} \quad (5.17)$$

for large  $k$ , conditioning on the event in (5.13). By Borel-Cantelli lemma and (5.13) it follows that

$$\mathbb{P}\left(\liminf_{k \rightarrow \infty} \frac{M(T_{n_k})}{\log n_k} \leq \gamma r_A^*\right) \leq \varepsilon.$$

From that we can also get

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \frac{M(T_n)}{\log n} \leq \gamma r_A^*\right) \leq \varepsilon.$$

Recall that  $M_n \stackrel{d}{=} M(T_n)$  from the embedding, let  $\gamma \rightarrow 1$  and  $\varepsilon \rightarrow 0$  we have

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \frac{M_n}{\log n} \geq r_A^*\right) = 1.$$

This completes the proof for the lower bound part, with  $C_1 = r_A^*$ .

Alternatively, we provide here another proof for the lower bound part that works when  $q \neq 0$ . Ideas used in the proof shall become useful later on.

Consider type  $A$  individuals alone. We call a type  $A$  individual “pure-blooded” if all of its antecedents are type  $A$  individuals. Define  $\text{pdBP}_A(t)$  to be the branching process consisting of all pure-blooded type  $A$  individuals in  $\text{pdBP}(t)$ . Let  $T_{n,A} := \inf\{t \geq 0 : |\text{pdBP}_A(t)| = n\}$ . Note that pure-blooded type  $A$  individuals give birth to new pure-blooded type  $A$  individuals at constant rate  $q$ , mimicking the results for CMRT (*i.e.* Lemma 2.1 and Lemma 5.3) it is not hard to see that:

- $\{\text{pdBP}_A(T_{n,A})\}_{n \geq 1} \stackrel{d}{=} \{\mathcal{U}_n\}_{n \geq 1}$  as processes, where  $\{\mathcal{U}_n\}_{n \geq 1}$  is a URT.
- The sequence of stopping times  $T_{n,A}$  satisfy

$$T_{n,A} - \frac{1}{q} \log n \xrightarrow{\text{a.s.}} -\log(W_A) \quad (5.18)$$

for a finite positive random variable  $W_A \stackrel{d}{=} \exp(1)$ .

Since we can transform results for URT into that of the branching process  $\text{pdBP}_A(t)$  through embedding, using Devroye and Lu [14] we get

$$\frac{M_A^{\text{pure}}(T_{n,A})}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{\log 2}$$

where  $M_A^{\text{pure}}(t)$  is the maximal number of offsprings a pure-blooded type  $A$  individual has by time  $t$ . Similar to (5.13) we have that by (5.18) and Egoroff's Theorem, given any  $\epsilon > 0$ , we can choose  $K_A > 0$  such that (again for simplicity the dependence of  $K_A$  on  $\epsilon$  is suppressed throughout):

$$\mathbb{P}\left(\sup_n |T_{n,A} - \frac{1}{q} \log n| < K_A\right) > 1 - \epsilon. \quad (5.19)$$

For each  $t > 0$ , let  $n^*(t)$  be the positive integer satisfying  $\frac{1}{q} \log(n^*(t)) + K_A \leq t < \frac{1}{q} \log(n^*(t) + 1) + K_A$ . Conditioning on the event in (5.19) we have

$$\frac{M_A^{\text{pure}}(t)}{t} \geq \frac{M_A^{\text{pure}}(\frac{1}{q} \log(n^*(t)) + K_A)}{\frac{1}{q} \log(n^*(t) + 1) + K_A} \geq \frac{M_A^{\text{pure}}(T_{n^*(t),A})}{\frac{1}{q} \log(n^*(t) + 1) + K_A} \xrightarrow{\text{a.s.}} \frac{q}{\log 2}$$

as  $t \rightarrow \infty$ , and it follows that

$$\mathbb{P}\left(\liminf_{t \rightarrow \infty} \frac{M_A^{\text{pure}}(t)}{t} \geq \frac{q}{\log 2}\right) > 1 - \epsilon.$$

Denote by  $M(t)$  the maximal number of offsprings an individual has by time  $t$ . Let  $\epsilon \rightarrow 0$  in the above inequality and note  $M(t) \geq M_A^{\text{pure}}(t)$ , we have

$$\mathbb{P}\left(\liminf_{t \rightarrow \infty} \frac{M(t)}{t} \geq \frac{q}{\log 2}\right) = 1.$$

Finally, since

$$\frac{T_n}{\log n} \xrightarrow{\text{a.s.}} 1, \quad (5.20)$$

by Lemma 5.3, and  $M_n \stackrel{d}{=} M(T_n)$  from the embedding, we get

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \frac{M_n}{\log n} \geq \frac{q}{\log 2}\right) = 1.$$

This completes the proof for the lower bound part when  $q \neq 0$ , with  $C_1 = \frac{q}{\log 2}$ .

**Upper bound:** First we consider the simpler type  $A$  individuals. Given  $\{\text{pdBP}(t)\}_{t \geq 0}$ , we couple it with another process where: whenever a type  $A$  individual is born to a type  $B$  individual in  $\{\text{pdBP}(t)\}_{t \geq 0}$ , chose uniformly at random a living type  $A$  individual, and treat the newborn individual as an offspring of that chosen individual. In this new process, if we look at type  $A$  individuals alone, it is not hard to see that they give birth to new type  $A$  individuals at constant rate 1. Denote by  $M_A(t)$  and  $\widetilde{M}_A(t)$  respectively the maximal number of offsprings a type  $A$  individual has by time  $t$  in  $\{\text{pdBP}(t)\}_{t \geq 0}$  and new process. Using the same argument as in the proof for lower bound we see that

$$\mathbb{P}\left(\limsup_{t \rightarrow \infty} \frac{\widetilde{M}_A(t)}{t} \leq \frac{1}{\log 2}\right) = 1.$$

Since  $\widetilde{M}_A(t) \geq M_A(t)$ , this immediately implies

$$\mathbb{P}\left(\limsup_{t \rightarrow \infty} \frac{M_A(t)}{t} \leq \frac{1}{\log 2}\right) = 1.$$

Denote by  $M_{n,A}$  the maximal degree of type  $A$  vertices in  $\mathcal{T}_n$ . Then it follows from (5.20) and  $M_{n,A} \stackrel{d}{=} M_A(T_n)$  that

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{M_{n,A}}{\log n} \leq \frac{1}{\log 2}\right) = 1.$$

Next we consider the more complicated type  $B$  individuals. For a given type  $B$  individual  $\nu$  and time  $T \in [0, t]$ , write  $n_\nu[T, t]$  for the number of offsprings this individual produced in time interval  $[T, t]$ . This is a pure birth process with rate

$$c \frac{n_A(t)}{n_B(t)} := \left(1 - q + \frac{(1-p)q}{p}\right) \frac{n_A(t)}{n_B(t)}.$$

Therefore for a fixed  $T$ , the following process

$$X(t) = n_\nu[T, t] - \int_T^t c \frac{n_A(s)}{n_B(s)} ds, \quad t \geq T, \quad (5.21)$$

is a martingale (here the infinitesimal generator is exactly the rate).

To handle the variability of  $X(t)$ , we will need its predictable quadratic variation process  $\langle X \rangle(t)$ . Note that

$$\mathcal{A} X^2(t) = \mathcal{A} n_\nu^2[T, t] + \mathcal{A} \left( \int_T^t c \frac{n_A(s)}{n_B(s)} ds \right)^2 - 2\mathcal{A} \left( n_\nu[T, t] \int_T^t c \frac{n_A(s)}{n_B(s)} ds \right)$$

with

$$\mathcal{A} n_\nu^2[T, t] = (2n_\nu[T, t] + 1) c \frac{n_A(t)}{n_B(t)},$$

$$\mathcal{A} \left( \int_T^t c \frac{n_A(s)}{n_B(s)} ds \right)^2 = 2 \left( \int_T^t c \frac{n_A(s)}{n_B(s)} ds \right) c \frac{n_A(t)}{n_B(t)}$$

and

$$\mathcal{A} \left( n_\nu[T, t] \int_T^t c \frac{n_A(s)}{n_B(s)} ds \right) = n_\nu[T, t] c \frac{n_A(t)}{n_B(t)} + \left( \int_T^t c \frac{n_A(s)}{n_B(s)} ds \right) c \frac{n_A(t)}{n_B(t)}.$$

It then follows from some elementary algebra that  $\mathcal{A} X^2(t) = c n_A(t)/n_B(t)$  and

$$\langle X \rangle(t) = \int_T^t c \frac{n_A(s)}{n_B(s)} ds, \quad t \geq T. \quad (5.22)$$

Now use Lemma 5.2 to choose  $T$  such that

$$\mathbb{P}\left(\sup_{t \geq T} \left| \frac{n_A(t)}{n_B(t)} - \frac{p}{1-p} \right| > \varepsilon\right) \leq \varepsilon. \quad (5.23)$$

Also, define the stopping time

$$S = \inf \left\{ t \geq T : \left| \frac{n_A(t)}{n_B(t)} - \frac{p}{1-p} \right| > \varepsilon \right\}.$$

Observe that by our choice of  $T$  we have  $\mathbb{P}(S < \infty) \leq \varepsilon$ . The idea here is to bound  $n_A(t)/n_B(t)$  around  $p/(1-p)$  after some finite time  $T$ , and show that what happened before time  $T$  does not have a noticeable effect in the long run.

Recall that we write  $T_n^+ = \log n + K$  and  $T_n^- = \log n - K$ . Consider the process  $\{X(t \wedge S) : t \geq T\}$  and note that for  $n$  large enough we have  $T_n^- \geq T$ . By the exponential martingale inequality from Liptser and Shiriyayev [24, Section 4.13, Theorem 5] with choices

$$K = 2 \quad \text{and} \quad \varphi(t) = c \left( \frac{p}{1-p} + \varepsilon \right) t,$$

we have for any  $\delta > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \sup_{t \leq T_n^+ \wedge S} X(t) \geq \delta c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right) &\leq \exp \left( -\kappa c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right) \\ &\quad + \mathbb{P}(\langle X \rangle(T_n^+ \wedge S) \geq 2\varphi(T_n^+)), \end{aligned} \quad (5.24)$$

where  $\kappa = (\delta + 2) \log \frac{\delta+2}{2} - \delta$ . By definition of  $S$  and the expression of  $\langle X \rangle(\cdot)$  in (5.22), we have that with probability one

$$\langle X \rangle(T_n^+ \wedge S) \leq \varphi(T_n^+),$$

so the second term on the right hand side of (5.24) vanishes. Further, using the expression of  $X(\cdot)$  in (5.21) we get

$$\mathbb{P} \left( n_{\nu}[T, T_n^+ \wedge S] \geq (\delta + 1)c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right) \leq \exp \left( -\kappa c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right). \quad (5.25)$$

For  $a < b$ , denote by  $M_B[a, b]$  the maximal number of offsprings produced by a type  $B$  individual in time interval  $[a, b]$  and let  $M_B(t) = M_B[0, t]$ . Then using a union bound and note that  $M_B[T, T_n \wedge S] \leq M_B[T, T_n^+ \wedge S]$  with probability at least  $1 - \varepsilon$ , we have

$$\mathbb{P} \left( M_B[T, T_n \wedge S] \geq (\delta + 1)c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right) \leq \varepsilon + n \exp \left( -\kappa c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right).$$

Further, from our choice of  $T$  in (5.23) it follows that

$$\mathbb{P} \left( M_B[T, T_n] \geq (\delta + 1)c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right) \leq 2\varepsilon + n \exp \left( -\kappa c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right).$$

Next, note that there exists  $L > 0$  such that

$$\mathbb{P}(M_B(T) \geq L) < \varepsilon.$$

As  $M_B(T_n) \leq M_B(T) + M_B[T, T_n]$ , combining the above results readily yields

$$\mathbb{P} \left( M_B(T_n) \geq L + (\delta + 1)c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right) \leq 3\varepsilon + n \exp \left( -\kappa c \left( \frac{p}{1-p} + \varepsilon \right) T_n^+ \right). \quad (5.26)$$

Denote by  $M_{n,B}$  the maximal degree of type  $B$  vertices in  $\mathcal{T}_n$ . Letting  $\varepsilon \rightarrow 0$  in (5.26) and note  $M_{n,B} \stackrel{d}{=} M_B(T_n)$ , we have

$$\mathbb{P} \left( M_{n,B} \geq L + (\delta + 1) \frac{cp}{1-p} T_n^+ \right) \leq n \exp \left( -\frac{\kappa cp}{1-p} T_n^+ \right). \quad (5.27)$$

Recall that  $\kappa = (\delta + 2) \log \frac{\delta+2}{2} - \delta$ . So given any  $\tilde{\varepsilon} > 0$ , for  $\delta > \max\{2e^2 - 2, \frac{(2+\tilde{\varepsilon})(1-p)}{cp} - 4\}$  we have  $\frac{\kappa cp}{1-p} T_n^+ > (2 + \tilde{\varepsilon}) \log n$  for large enough  $n$ , which makes the right hand side of (5.27) summable. By Borel-Cantelli lemma this implies

$$\mathbb{P} \left( \limsup_{n \rightarrow \infty} \frac{M_{n,B}}{\log n} \leq (\delta + 1) \frac{cp}{1-p} + \tilde{\varepsilon} \right) = 1.$$



Letting  $\tilde{\varepsilon} \rightarrow 0$  we get

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{M_{n,B}}{\log n} \leq C_2\right) = 1$$

where  $C_2 = \max\{\frac{(2e^2-1)cp}{1-p}, 2 - \frac{3cp}{1-p}\}$ . This completes the proof for the upper bound part.

This left us with the cases where  $p = 1$  or  $q = 1$ . When  $q = 1$ , as noted in Section 1.2,  $\{\mathcal{T}_n\}_{n \geq 2}$  looks like two disjoint URTs connected by a single edge between roots. By strong law of large numbers, with probability one these two subtrees have sizes proportional to  $p$  and  $1-p$  asymptotically. Therefore, appealing to existing results on maximal degree of URT [14] we have

$$\frac{M_n}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{\log 2}.$$

Last we consider the special case where  $p = 1$  but  $q \neq 1$ . We shall use the same notations as defined before. First, note that what we proved for type  $A$  individuals still holds. Specifically, recall that

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{M_{n,A}}{\log n} \leq \frac{1}{\log 2}\right) = 1.$$

However, we can no longer define  $T$  by (5.23). In fact, consider the unique type  $B$  vertex (*i.e.* the type  $B$  root) in  $\mathcal{T}_n$  and note that all type  $A$  vertices other than the root have a fixed probability  $1-q$  to connect to the type  $B$  root independently. By strong law of large numbers applied to binomial distribution we have

$$\frac{M_{n,B}}{n} \xrightarrow{\text{a.s.}} 1-q, \quad \text{and therefore} \quad \frac{M_n}{n} \xrightarrow{\text{a.s.}} 1-q.$$

This completes the proof for Theorem 3.3. ■

**5.5. Proof for height.** In this section we shall prove Theorem 3.4. Instead of proving the result from scratch using continuous time martingales as what we did for maximal degree, we present here a proof that appeals to existing results on first birth problem of branching processes.

Once again, we consider first cases where  $p, q \neq 1$ . The basic idea here is still the same, *i.e.* to bound  $n_A(t)/n_B(t)$  around  $p/(1-p)$  after some finite time and prove that what happened in the beginning does not really matter in the long term.

To obtain strong convergence, it is enough to prove that for any  $\delta, \varepsilon > 0$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{k \geq n} \left| \frac{H_k}{\log k} - e \right| > \delta\right) < \varepsilon \quad (5.28)$$

Define the event

$$E_1 = \left\{ \sup_{t \geq T} \left| \frac{n_A(t)}{n_B(t)} - \frac{p}{1-p} \right| < \eta \frac{p}{1-p} \right\}. \quad (5.29)$$

where  $0 < \eta < 1$  is any given constant. By Lemma 5.2 there exists  $T > 0$  such that  $\mathbb{P}(E_1) > 1 - \varepsilon/3$ . Also, choose  $N \in \mathbb{Z}^+$  such that  $\mathbb{P}(n(T) > N) < \varepsilon/3$ , and define another event  $E_2 = \{n(T) \leq N\}$ . Moreover, choose  $T_n^+$  and  $T_n^-$  in the same way as in Section 5.4, with  $\varepsilon$  replaced by  $\varepsilon/3$ , and define our last “good” event  $E_3 = \{T_n^- < T_n < T_n^+ \text{ for all } n \in \mathbb{Z}^+\}$ . In what follows, we condition on the event  $E_1 \cap E_2 \cap E_3$  and note that  $\mathbb{P}(E_1 \cap E_2 \cap E_3) > 1 - \varepsilon$  (for simplicity the conditional is suppressed throughout).

On this event, we have at most  $N$  individuals alive at time  $T$  and the ratio of  $n_A(t)$  to  $n_B(t)$  is bounded around  $p/(1-p)$  after that time. For a fixed individual  $v$  alive at time  $T$  and  $t > 0$ , denote by  $H_v(t)$  the height of the subtree root at  $v$  in  $\text{pdBP}(T+t)$ . To bound  $H_v(t)$ , we

now construct two processes. Recall our continuous time process as defined in Section 2, and consider a process where each type  $B$  individual gives birth to type  $A$  individuals at rate  $(1 - \eta)p(1 - q)/(1 - p)$  and type  $B$  individuals at rate  $(1 - \eta)q$ , while everything else stays the same. Denote by  $H_{min}(t)$  the height of this tree at time  $t$ ,  $B_{min}(n)$  the time when the first individual in the  $n$ -th generation is born, and define  $H_{max}(t)$  similarly using  $1 + \eta$  instead of  $1 - \eta$  in the rates. Since the reproduction rates of type  $A$  individuals are constants, and those of type  $B$  individuals only depend on  $n_A(t)/n_B(t)$ , we have  $H_{min}(t) \leq H_v(t) \leq H_{max}(t)$  where  $\leq$  denotes stochastic dominance as processes (*i.e.* for any monotone increasing functional  $f$  we have  $f(H_{min}(t)) \leq f(H_v(t)) \leq f(H_{max}(t))$  where  $\leq$  denotes the usual stochastic dominance).

From Biggins [5, Theorem 2] we know that

$$\lim_{n \rightarrow \infty} \frac{B_{min}(n)}{n} = \gamma_{min},$$

where  $\gamma_{min}$  can be calculated following the procedure given in the paper. First, compute the matrix  $\Phi(\theta)$  with entries

$$\Phi_{ij}(\theta) = \theta \int_0^\infty e^{-\theta t} \mathbb{E}(Z_{ij}(t)) dt = \frac{r_{ij}}{\theta}.$$

Here  $Z_{ij}(t)$  denotes the number of type  $j$  individuals born to a type  $i$  individual by time  $t$ , and  $r_{ij}$  denotes the rate at which a type  $i$  individual gives birth to type  $j$  individuals. Then take the largest eigenvalue  $\phi(\theta) = \lambda_{min}/\theta$  of  $\Phi(\theta)$ . In our case we have

$$\lambda_{min} = \frac{(2 - \eta)q + \sqrt{\eta^2 q^2 + 4(1 - \eta)(1 - q)^2}}{2}$$

by calculation. Finally, define

$$\mu(a) = \inf \left\{ e^{\theta a} \phi(\theta) : \theta > 0 \right\},$$

and compute  $\gamma_{min} = \inf \left\{ a : \mu(a) \geq 1 \right\} = 1/(\lambda_{min} e)$ .

Since

$$B_{min}(H_{min}(t)) \leq t \leq B_{min}(H_{min}(t) + 1),$$

dividing by  $H_{min}(t)$  and letting  $t \rightarrow \infty$  we get

$$\frac{H_{min}(t)}{t} \xrightarrow{\text{a.s.}} \lambda_{min} e. \quad (5.30)$$

Similarly we have

$$\frac{H_{max}(t)}{t} \xrightarrow{\text{a.s.}} \lambda_{max} e, \quad (5.31)$$

where

$$\lambda_{max} = \frac{(2 + \eta)q + \sqrt{\eta^2 q^2 + 4(1 + \eta)(1 - q)^2}}{2}.$$

As the eigenvalues are continuous with respect to  $\eta$ , and  $\lambda_{max} = \lambda_{min} = 1$  when  $\eta = 0$ , we can choose  $\eta$  in (5.29) small enough such that both  $\lambda_{max} - 1$  and  $1 - \lambda_{min}$  are smaller than  $\delta/3$ .

**Remark 10.** Since we will need to generalize the result to CMRT with more types, we include here an alternative argument using Perron-Frobenius theory of positive matrices [27, Chapter

8]. Note that when  $\eta = 0$ , the rate matrix consisting of  $r_{ij}$ 's is a positive matrix (*i.e.* a matrix where all entries are strictly positive)

$$\begin{pmatrix} q & \frac{(1-p)(1-q)}{p} \\ \frac{p(1-q)}{1-p} & q \end{pmatrix}$$

with left eigenvector  $(p, 1-p)$  corresponding to eigenvalue  $\lambda = 1$ . Since this eigenvector has strictly positive coordinates, we have that  $\lambda = 1$  is the unique largest eigenvalue of that rate matrix. Therefore the largest eigenvalue of rate matrix is continuous with respect to  $\eta$  around  $\eta = 0$ , and the previous result we established on  $\lambda_{max}$  and  $\lambda_{min}$  follows.

By (5.30), (5.31) and stochastic dominance (as processes) we see that

$$\lim_{s \rightarrow \infty} \mathbb{P} \left( \inf_{t \geq s} \frac{H_v(t)}{t} < \lambda_{min} e - \frac{\delta}{3} \right) = 0 \quad \text{and} \quad \lim_{s \rightarrow \infty} \mathbb{P} \left( \sup_{t \geq s} \frac{H_v(t)}{t} > \lambda_{max} e + \frac{\delta}{3} \right) = 0$$

hold for all  $v \in V$ , where  $V$  denotes the set of individuals alive at time  $T$ .

Since  $H_n \geq H_v(T_n^- - T)$ , we have

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \inf_{k \geq n} \frac{H_k}{T_k^- - T} < \lambda_{min} e - \frac{\delta}{3} \right) \leq \lim_{s \rightarrow \infty} \mathbb{P} \left( \inf_{t \geq s} \frac{H_v(t)}{t} < \lambda_{min} e - \frac{\delta}{3} \right) = 0.$$

On the other hand, note that  $H_n \leq \max_{u \in V} H_u(T_n^+ - T) + N$ . Using union bound with  $|V| \leq N$  we get

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{k \geq n} \frac{H_k - N}{T_k^+ - T} > \lambda_{max} e + \frac{\delta}{3} \right) \leq \sum_{u \in V} \lim_{s \rightarrow \infty} \mathbb{P} \left( \sup_{t \geq s} \frac{H_u(t)}{t} > \lambda_{max} e + \frac{\delta}{3} \right) = 0.$$

With

$$\left| \frac{H_k}{T_k^- - T} - \frac{H_k}{\log k} \right| < \frac{\delta}{3} \quad \text{and} \quad \left| \frac{H_k - N}{T_k^+ - T} - \frac{H_k}{\log k} \right| < \frac{\delta}{3}$$

for large enough  $k$ , by our choice of  $\eta$  and triangle inequality it follows that

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{k \geq n} \left| \frac{H_k}{\log k} - e \right| > \delta \right) = 0. \quad (5.32)$$

Finally, do not forget that we are conditioning on the event  $E_1 \cap E_2 \cap E_3$ , which occurs with probability at least  $1 - \varepsilon$ . Therefore we have (5.28) as desired.

Once again we are left with cases where  $p = 1$  or  $q = 1$ . When  $q = 1$ , recall that  $\{\mathcal{T}_n\}_{n \geq 2}$  looks like two disjoint URTs in this case. Using results for URT [32] we have

$$\frac{H_n}{\log n} \xrightarrow{\text{a.s.}} e.$$

Last we turn to cases where  $p = 1$  but  $q \neq 1$ . Consider only pure-blooded type  $A$  individuals as in Section 5.4 and use the same argument there we have

$$\mathbb{P} \left( \liminf_{n \rightarrow \infty} \frac{H_n}{\log n} \geq qe \right) = 1.$$

For upper bound we construct  $\{\mathcal{T}_n\}_{n \geq 2}$  from a URT  $\{\mathcal{U}_n\}_{n \geq 1}$  as described in special cases of Section 1.2 and note that by construction the height of  $\mathcal{T}_n$  is at most equal to that of  $\mathcal{U}_n$ . From results for URT [32] we know that

$$\mathbb{P} \left( \limsup_{n \rightarrow \infty} \frac{H_n}{\log n} \leq e \right) = 1.$$

This completes the proof for Theorem 3.4. ■

**5.6. Extension to general CMRT.** For CMRT with  $K$  types and arbitrary attachment probabilities, we will need a population-dependent branching process defined as follows:

- (a) **Initialization:** start with  $K$  individuals at  $t = 0$ , with one of each type. For any time  $t \geq 0$  and  $1 \leq i \leq K$ , let  $n_i(t)$  be the number of type  $i$  individuals. We have  $n_i(0) = 1$ . Denote by  $\mathcal{F}(t)$  the  $\sigma$ -field generated by the process until time  $t$ .
- (b) **Types:** Each individual in the system has a type  $\in \{1, 2, \dots, K\}$  and lives forever, while giving birth to individuals of all types.
- (c) **Reproduction:** At any time  $t$ , a living type  $i$  individual gives birth to type  $j$  ( $1 \leq j \leq K$ ) individuals at rate:

$$r_{ij}(t) = \frac{n_1(t)}{n_i(t)} \cdot \frac{p_j q_{ji}}{p_1}.$$

Then using the same arguments as in the proofs for CMRT with two types we can prove similar results in the general case as stated in Section 3.2.

**5.7. Extension to CWRT.** Similar to CMRT, we can define a population-dependent branching process with  $K$  types for CWRT, with birth rates replaced by

$$r_{ij}(t) = \frac{n_1(t)}{p_1} \cdot \frac{p_j \omega_{ji}}{\sum_{l=1}^K \omega_{jl} n_l(t)}.$$

Then following the same procedure as in the proof for CMRT we can derive the limiting degree distribution of CWRT, *i.e.* Theorem 3.8.

**5.8. Extension to CMPA.** For CMPA the birth rates have a more complex form and we need to introduce some notations.

For any time  $t \geq 0$ ,  $1 \leq i, j \leq K$  and individual  $v$ , let  $n_{ij}(t)$  be the number of type  $j$  individuals born to type  $i$  individuals, and  $d_v(t)$  be the number of offsprings born to individual  $v$ . Then at any time  $t$ , a living type  $i$  individual  $v$  gives birth to type  $j$  ( $1 \leq j \leq K$ ) individuals at rate:

$$r_v(t, j) = \frac{n_1(t)}{p_1} \cdot \frac{p_j q_{ji} (d_v(t) + \alpha_{ji})}{\alpha_{ji} n_i(t) + \sum_{l=1}^K n_{il}(t)}. \quad (5.33)$$

Note here that  $\sum_{l=1}^K n_{il}(t)$  is exactly the total number of offsprings born to type  $i$  individuals.

To derive the limiting degree distribution of CMPA, *i.e.* Theorem 3.9, we can still mimic the proof for CMRT. However, computation for the integral in the final step will be different. Recall that for CMRT we have the integral

$$\int_0^\infty \mathbb{P}(\text{Poisson}(r_A^* s) = k) e^{-s} ds = \frac{1}{1 + r_A^*} \left( \frac{r_A^*}{1 + r_A^*} \right)^k.$$

For CMPA, the (time-homogeneous) Poisson process is replaced by a time-inhomogeneous Poisson process. To get the birth rates, observe that similar to the proof of CMRT we have  $n_i(t)/n(t) \xrightarrow{\text{a.s.}} p_i$  and  $n_{ij}(t)/n(t) \xrightarrow{\text{a.s.}} p_j q_{ji}$  as  $t \rightarrow \infty$ . Therefore from (5.33) the reproduction of a type  $i$  individual  $v$  is approximately

$$\sum_{j=1}^K \frac{p_j q_{ji} (d_v(t) + \alpha_{ji})}{\alpha_{ji} p_i + \sum_{l=1}^K p_l q_{li}} = v_i (d_v(t) + \alpha_i).$$

This leads to a pure birth process with rate  $\nu_i(m + \alpha_i - 1)$ , where  $m$  is the current population size. We shall denote this process by  $\{\tilde{Y}(\nu_i, t) : t \geq 0\}$  (with  $\tilde{Y}(\nu_i, 0) = 1$ ). Then similar to CMRT we need to compute the integral  $\int_0^\infty \mathbb{P}(\tilde{Y}(\nu_i, s) = k + 1) e^{-s} ds$ .

Since transition probability function of  $\{\tilde{Y}(\nu_i, t) : t \geq 0\}$  can be computed explicitly (see *e.g.* [34, Proposition 6.1]), we have

$$\mathbb{P}(\tilde{Y}(\nu_i, s) = k + 1) = \frac{\Gamma(k + \alpha_i)}{\Gamma(\alpha_i)\Gamma(k + 1)} e^{-\alpha_i \nu_i s} (1 - e^{-\nu_i s})^k.$$

Plugging it into the integral we get

$$\int_0^\infty \mathbb{P}(Y(\nu_i, s) = k + 1) e^{-s} ds = \frac{\Gamma(k + \alpha_i)}{\Gamma(\alpha_i)\Gamma(k + 1)} \int_0^\infty e^{-\alpha_i \nu_i s} (1 - e^{-\nu_i s})^k e^{-s} ds.$$

Let  $x = e^{-\nu_i s}$  and by change of variable we have

$$\int_0^\infty e^{-\alpha_i \nu_i s} (1 - e^{-\nu_i s})^k e^{-s} ds = \frac{1}{\nu_i} \int_0^1 x^{1/\nu_i + \alpha_i - 1} (1 - x)^k dx = \frac{1}{\nu_i} \frac{\Gamma(k + 1)\Gamma(1/\nu_i + \alpha_i)}{\Gamma(k + 1 + 1/\nu_i + \alpha_i)}.$$

This leads to the constant in (3.2).

## 6. ACKNOWLEDGMENTS

The research of SB and RF was supported in part by the NSF (DMS-1606839, DMS-1613072) and the Army Research Office (W911NF-17-1-0010). The research of RF was supported in part by the NSF under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute.

## REFERENCES

- [1] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, The Annals of Applied Probability (1991), 228–266.
- [2] K. B Athreya and S. Karlin, *Embedding of urn schemes into continuous time markov branching processes and related limit theorems*, The Annals of Mathematical Statistics **39** (1968), no. 6, 1801–1817.
- [3] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, science **286** (1999), no. 5439, 509–512.
- [4] 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).
- [5] J. Biggins, *The first-and last-birth problems for a multitype age-dependent branching process*, Advances in Applied Probability **8** (1976), no. 3, 446–459.
- [6] B. e. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, *The degree sequence of a scale-free random graph process*, Random Structures & Algorithms **18** (2001), no. 3, 279–290.
- [7] C. Bordenave, M. Lelarge, and L. Massoulié, *Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs*, 2015 ieee 56th annual symposium on foundations of computer science, 2015, pp. 1347–1357.
- [8] S. Bubeck, L. Devroye, and G. Lugosi, *Finding adam in random growing trees*, Random Structures & Algorithms **50** (2017), no. 2, 158–172.
- [9] K. S Crump and C. J Mode, *A general age-dependent branching process. i*, Journal of Mathematical Analysis and Applications **24** (1968), no. 3, 494–508.
- [10] M. Deijfen and R. Fitzner, *Birds of a feather or opposites attract-effects in network modelling*, arXiv preprint arXiv:1612.03127 (2016).
- [11] L. Devroye, *Branching processes in the analysis of the heights of trees*, Acta Informatica **24** (1987), no. 3, 277–298.
- [12] 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.
- [13] L. Devroye, O. Fawzi, and N. Fraiman, *Depth properties of scaled attachment random recursive trees*, Random Structures & Algorithms **41** (2012), no. 1, 66–98.

- [14] L. Devroye and J. Lu, *The strong convergence of maximal degrees in uniform random recursive trees and dags*, *Random Structures & Algorithms* **7** (1995), no. 1, 1–14.
- [15] R. M D’Souza, P. L Krapivsky, and C. Moore, *The power of choice in growing trees*, *The European Physical Journal B* **59** (2007), no. 4, 535–543.
- [16] S. Fortunato, *Community detection in graphs*, *Physics reports* **486** (2010), no. 3-5, 75–174.
- [17] J. L Gastwirth, *A probability model of a pyramid scheme*, *The American Statistician* **31** (1977), no. 2, 79–82.
- [18] J. L Gastwirth and P. Bhattacharya, *Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable*, *Operations Research* **32** (1984), no. 3, 527–536.
- [19] P. Jagers and O. Nerman, *The asymptotic composition of supercritical, multitype branching populations*, *Lecture Notes in Mathematics* (1996), 40–54.
- [20] S. Janson, *Asymptotic degree distribution in random recursive trees*, *Random Structures & Algorithms* **26** (2005), no. 1-2, 69–83.
- [21] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, Vol. 113, Springer Science & Business Media, 2012.
- [22] B. Karrer and M. E. Newman, *Stochastic blockmodels and community structure in networks*, *Physical review E* **83** (2011), no. 1, 016107.
- [23] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, *Spectral redemption in clustering sparse networks*, *Proceedings of the National Academy of Sciences* **110** (2013), no. 52, 20935–20940.
- [24] R. Liptser and A. N. Shiryaev, *Theory of Martingales*, Vol. 49, Springer Science & Business Media, 2012.
- [25] H. M Mahmoud, *The power of choice in the construction of recursive trees*, *Methodology and Computing in Applied Probability* **12** (2010), no. 4, 763–773.
- [26] H. M Mahmoud and R. T. Smythe, *Asymptotic joint normality of outdegrees of nodes in random recursive trees*, *Random Structures & Algorithms* **3** (1992), no. 3, 255–266.
- [27] C. D Meyer, *Matrix analysis and applied linear algebra*, Vol. 71, Siam, 2000.
- [28] J. Moon, *The distance between nodes in recursive trees*, *London Mathematical Society Lecture Notes* **13** (1974), 125–132.
- [29] D. Najock and C. Heyde, *On the number of terminal vertices in certain random trees with an application to stemma construction in philology*, *Journal of Applied Probability* **19** (1982), no. 3, 675–680.
- [30] M. E. Newman, *Communities, modules and large-scale structure in networks*, *Nature physics* **8** (2012), no. 1, 25–31.
- [31] B. Øksendal, *Stochastic differential equations*, *Stochastic differential equations*, 2003, pp. 65–84.
- [32] B. Pittel, *Note on the heights of random recursive trees and random  $m$ -ary search trees*, *Random Structures & Algorithms* **5** (1994), no. 2, 337–347.
- [33] S. Rosengren, *A multi-type preferential attachment tree*, arXiv preprint arXiv:1704.03256 (2017).
- [34] S. M Ross, *Introduction to probability models*, Academic press, 2014.
- [35] R. T Smythe and H. M Mahmoud, *A survey of recursive trees*, *Theory of Probability and Mathematical Statistics* **51** (1995), 1–28.