Random Trees and General Branching Processes

Anna Rudas,[1] Bálint Tóth,[1] Benedek Valkó[1,2]

[1] Institute of Mathematics, Technical University Budapest, Egry József U. 1.,
   H-1111 Budapest, Hungary; e-mail: rudasa@math.bme.hu; balint@math.bme.hu
[2] Rényi Institute of Mathematics, Hungarian Academy of Sciences,
   Reáltanoda U. 13-15, H-1053 Budapest, Hungary; e-mail: valko@renyi.hu

ABSTRACT: We consider a tree that grows randomly in time. Each time a new vertex appears, it chooses exactly one of the existing vertices and attaches to it. The probability that the new vertex chooses vertex x is proportional to $w(\deg(x))$, a weight function of the actual degree of x. The weight function $w : \mathbb{N} \to \mathbb{R}_+$ is the parameter of the model.

In [4] and [11] the authors derive the asymptotic degree distribution for a model that is equivalent to the special case, when the weight function is linear. The proof therein strongly relies on the linear choice of w.

Using well-established results from the theory of general branching processes we give the asymptotical degree distribution for a wide range of weight functions. Moreover, we provide the asymptotic distribution of the tree itself as seen from a randomly selected vertex. The latter approach gives greater insight to the limiting structure of the tree.

Our proof is robust and we believe that the method may be used to answer several other questions related to the model. It relies on the fact that considering the evolution of the random tree in *continuous time*, the process may be viewed as a general branching process, this way classical results can be applied. © 2006 Wiley Periodicals, Inc.   Random Struct. Alg., 30, 000–000, 2007

Keywords: branching processes; preferential attachment; random trees

---

WILEY
InterScience®
DISCOVER SOMETHING GREAT

1. INTRODUCTION

In models of randomly growing networks, the concept of preferential attachment means that when a new vertex appears, it is more likely that it attaches to an already existing vertex if the latter already has more neighbors. One of the realizations of this idea is the Barabási–Albert graph. In this model, a tree grows in discrete time steps: at each step we add a new vertex and connect it to an already existing vertex. This choice is made randomly, using probabilities exactly proportional to the degree of the existing vertices. This model reproduces phenomena observed in certain real-world networks (see [2]), the power-law decay of the degree sequence, for example. This was proved in a mathematically precise way in [4] and, independently, in [11].

The probability with which a newly appearing vertex chooses its neighbor could depend on the degrees of the existing vertices in a more general way. We define a weight function, $w : \mathbb{N} \to \mathbb{R}_+$, and let the probability be proportional to this function of the degree. In [4] and [11], w is linear, their techniques strongly depend on martingales that are apparent in the system only in this special case.

General weight functions are considered in the model of Krapivsky and Redner [9]. There $w(k) \sim k^\gamma$, and non-rigorous results are obtained, showing the different behavior for $\gamma > 1$ and $\gamma \leq 1$. In the first region the limiting object does not have a non-trivial degree sequence: a single dominant vertex appears that is linked to almost every other vertex, the others having only finite degree. This statement is made precise and proved rigorously in [13]. (See also [5] for a related model.) The weight functions we consider will be such that this does not happen.

Our main results are the following. We determine the asymptotic distribution of the degree sequence, which equivalently gives the limiting distribution of the degree of a (uniformly) randomly selected vertex. We also look deeper into the structure of the tree: we give the asymptotic distribution of the subtree under a randomly selected vertex. Moreover, we present the asymptotic distribution of the whole tree, seen from a randomly selected vertex. For a general approach for asymptotic distribution of random subtrees of random trees, see [1]. These results give greater insight to the limiting structure of the random tree.

The key of our method is to place the process into continuous time in such a way that it fits in to the well-developed theory of general branching processes (see Section 5). The asymptotic behavior of the continuous time model gives that of the discrete time model, as pointed out in Section 3. A special case of our model (when w is linear) is equivalent to those studied in [4] and [11]. The robustness of our method is apparent in the fact that it gives a.s. results for a wide class of weight functions, it generalizes the results mentioned above.

The class of weight functions for which our theorems hold are not fully identified by the condition we make: it is sufficient, but we do not claim that it is also necessary. In the present paper we do not intend to analyze all the ways we believe this new approach can be used. We are planning to apply this technique to answer many other questions related to the randomly growing tree model.

The paper is organized as follows. We introduce the terminology and notation in the next section, then in Section 3 we define the model. We state our results in Section 4 and show the simplifications that arise in the linear case. After that, in Section 5 we give a brief introduction to the theory of general branching processes and quote the theorems that we rely on. We present our proofs in Section 6. We comment on the asymptotic growth of the tree in the last section.

2. ROOTED ORDERED TREES: TERMINOLOGY AND NOTATION

Throughout the paper it will be convenient to use genealogical phrasing. We will consider our tree evolving as a genealogical tree, where the individuals are the vertices and the parent child relations are the edges of the tree. It will be also convenient to follow the "birth orders" among children of the same parent. For this purpose we will consider our trees always as *rooted ordered trees* (sometimes called family trees or rooted planar trees). In the following paragraphs we introduce the (commonly known) terminology for these trees and also some additional notations needed for our theorems.

2.1. Vertices, Individuals, Trees

We will label the vertices of a rooted ordered tree using a subset of

$$\mathcal{N} := \bigcup_{n=0}^{\infty} \mathbb{Z}_+^n, \qquad \text{where} \quad \mathbb{Z}_+ := \{1, 2, \ldots\}, \quad \mathbb{Z}_+^0 := \{\emptyset\}.$$

$\emptyset$ denotes the root of the tree, its children are labeled with $\{1, 2, \ldots\}$, and in general the children of $x = (x_1, x_2, \ldots, x_k) \in \mathcal{N}$ are labeled by $(x_1, x_2, \ldots, x_k, 1), (x_1, x_2, \ldots, x_k, 2), \ldots$. Thus, if a vertex has the label $x = (x_1, x_2, \ldots, x_k) \in \mathcal{N}$ then this means that it is the x_k^{th} child of its parent, which is the x_{k-1}^{th} child of its own parent and so on. If $x = (x_1, x_2, \ldots, x_k)$ and $y = (y_1, y_2, \ldots, y_l)$ we will use the shorthanded notation xy for the concatenation $(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_l)$, and with a slight abuse of notation for $n \in \mathbb{Z}_+$ we use xn for $(x_1, x_2, \ldots, x_k, n)$.

We will identify a rooted ordered tree with the set of labels of its vertices, since this already contains the necessary information about the edges. It is clear that a $G \subset \mathcal{N}$ may represent a rooted ordered tree if and only if $\emptyset \in G$ and for each $(x_1, x_2, \ldots, x_k) \in G$ we have $(x_1, x_2, \ldots, x_{k-1}) \in G$ as well as $(x_1, x_2, \ldots, x_k - 1) \in G$, if $x_k > 1$.

The set of finite rooted ordered trees will be denoted by $\mathcal{G}$. We think about $G \in \mathcal{G}$ as an oriented tree with edges pointing from parents to children. The *degree* of a vertex $x \in G$ is just the number of its children in G:

$$\deg(x, G) := \max \{n \in \mathbb{Z}_+ : xn \in G\}.$$

The *nth generation* of $G \in \mathcal{G}$ is

$$G_{[n]} := \{x \in G : |x| = n\}, \qquad n \geq 0,$$

where $|x| = n$ iff $x \in \mathbb{Z}_+^n$.

The *n*th ancestor of $x = (x_1, x_2, \ldots, x_k) \in \mathcal{N}$ with $k \geq n$ is $x^n = (x_1, x_2, \ldots, x_{k-n})$ if $k > n$ and $x^n = \emptyset$ if $k = n$.

The *subtree* rooted at a vertex $x \in G$ is

$$G_{\downarrow x} := \{y : xy \in G\},$$

this is just the progeny of x viewed as a rooted ordered tree. Also (again with a slight abuse of notations), for an $x = (x_1, x_2, \ldots, x_n) \in \mathcal{N}$ with $|x| = n \geq k$ we use the notation $x_{\downarrow k} = (x_{n-k+1}, x_{n-k+2}, \ldots, x_n)$. This would be the new label given to $x \in G$ in the subtree $G_{\downarrow x^k}$.

Consider a $G \in \mathcal{G}$. An ordering $s = (s_0, s_1, \ldots, s_{|G|-1})$ of the elements of G is called *historical* if it gives a possible birth order of the vertices in G, formally if for each $0 \leq i \leq |G| - 1$ we have $\{s_0, s_1, \ldots, s_i\} \in \mathcal{G}$. The set of all historical orderings of $G \in \mathcal{G}$ will be denoted $\mathcal{S}(G)$. For a fixed $s \in \mathcal{S}(G)$ the rooted ordered trees

$$G(s, i) := \{s_0, s_1, \ldots, s_i\} \subset G \tag{1}$$

give the evolution of G in this historical ordering s.

2.2. Random Trees

Throughout the paper we will use Greek letters to denote random elements (of various distributions) selected from $\mathcal{N}$ and $\mathcal{G}$:

$$\xi, \zeta, \ldots \in \mathcal{N}, \qquad \Gamma, \Upsilon, \ldots \in \mathcal{G}.$$

Our results will deal with some asymptotic properties of a randomly chosen vertex in a certain random tree. We will investigate the asymptotic distribution of its degree, its progeny, and also the progeny, of its kth ancestor. In order to study the latter object, we introduce *rooted ordered trees with a marked vertex in generation k*:

$$\mathcal{G}^{(k)} := \{(G, u) \in \mathcal{G} \times \mathbb{Z}_+^k : u \in G_{[k]}\}.$$

$\mathcal{G}^{(0)}$ is identified with $\mathcal{G}$, since generation 0 consists of only the root, $\emptyset$. We can use the elements of $\mathcal{G}^{(k)}$ to describe the progeny of the kth ancestor of a random vertex: G is an ordered tree rooted in the kth ancestor of the selected point and $u \in G_{[k]}$ is the position of the random vertex in this tree. Clearly, if (G, u) describes the progeny of the kth ancestor, then for $0 \leq l \leq k$ the progeny of the lth ancestor is described by $(G_{\downarrow u^l}, u_{\downarrow k-l})$.

Thus, if $\pi^{(k)}$ is a distribution on $\mathcal{G}^{(k)}$, which describes the progeny of the kth ancestor of a chosen vertex, then, if $l < k$, the distribution of the progeny of the lth ancestor is

$$\pi^{(k,l)}(H, v) := \pi^{(k)}\big(\{(G, u) \in \mathcal{G}^{(k)} : G_{\downarrow u^l} = G, v = u_{\downarrow k-l}\}\big).$$

The sequence $\pi^{(k)}$ of probability measures on $\mathcal{G}^{(k)}$, $k = 0, 1, 2, \ldots$ is called *consistent* if for any $0 \leq l \leq k$, the identity $\pi^{(l)} = \pi^{(k,l)}$ holds.

Without presenting the precise formulation, it is clear that a consistent sequence $\pi^{(k)}$ of probability measures on $\mathcal{G}^{(k)}$ gives full insight to the limiting structure of the tree as seen from a random vertex, see Remark after Theorem 2.

We call a probability measure π on $\mathcal{G}$ *steady* if

$$\sum_{H \in \mathcal{G}} \pi(H) \sum_{x \in H_{[1]}} \mathbb{1}\{H_{\downarrow x} = G\} = \pi(G). \tag{2}$$

See the concept of fringe distributions in Aldous [1].

It is easy to check that in this case, for any $k = 1, 2, \ldots$, the similar identity

$$\sum_{H \in \mathcal{G}} \pi(H) \sum_{x \in H_{[k]}} \mathbb{1}\{H_{\downarrow x} = G\} = \pi(G)$$

follows. Equivalently, for any bounded function $\varphi : \mathcal{G} \to \mathbb{R}$ and any $k = 0, 1, 2, \ldots$,

$$\mathbf{E}\big(\big|\Gamma_{[k]}\big| \varphi(\Gamma_{\downarrow \xi})\big) = \mathbf{E}\big(\varphi(\Gamma)\big),$$

where Γ is a random element of $\mathcal{G}$ with distribution $\mathbf{P}(\Gamma = G) = \pi(G)$, and on the left-hand side ξ is a random vertex selected uniformly from the kth generation of Γ. (We don't have to worry about the fact that $\Gamma_{[k]}$ may be empty, since in that case the expression $\left|\Gamma_{[k]}\right|\varphi(\Gamma_{\downarrow\xi})$ is automatically 0.) Immediate consequences of this property are that the expected size of the kth generation is 1 for any $k \in \mathbb{N}$ (choose φ identically 1), and therefore the expected size of the whole tree is infinite.

2.2.1. Backward Extensions. Given a steady probability measure π on $\mathcal{G}$, define the probability measures $\pi^{(k)}$ on $\mathcal{G}^{(k)}$, $k = 0, 1, 2, \ldots$, by

$$\pi^{(k)}(G, u) := \pi(G).$$

One can easily check that, due to the steadiness of the distribution π, the sequence of probability measures $\pi^{(k)}$ on $\mathcal{G}^{(k)}$, $k = 0, 1, 2, \ldots$, is consistent.

3. THE RANDOM TREE MODEL

3.1. Discrete Time Model

Given a weight function $w : \mathbb{N} \to \mathbb{R}_+$, let us define the following discrete time Markov chain Υ_d on the countable state space $\mathcal{G}$, with initial state $\Upsilon_d(0) = \{\emptyset\}$. If for $n \geq 0$ we have $\Upsilon_d(n) = G$, then for a vertex $x \in G$ let $k := \deg(x, G) + 1$. Using this notation, let the transition probabilities be

$$\mathbf{P}(\Upsilon_d(n + 1) = G \cup \{xk\}) = \frac{w(\deg(x, G))}{\sum_{y \in G} w(\deg(y, G))}.$$

In other words, at each time step a new vertex appears and attaches to exactly one already existing vertex. If the tree at the appropriate time is G, then the probability of choosing vertex x in tree G is proportional to $w(\deg(x, G))$.

3.2. Continuous Time Model

Given a weight function $w : \mathbb{N} \to \mathbb{R}_+$, let $X(t)$ be a Markovian pure birth process with $X(0) = 0$ and birth rates

$$\mathbf{P}\big(X(t + \mathrm{d}t) = k + 1 \,\big|\, X(t) = k\big) = w(k)\,\mathrm{d}t + o(\mathrm{d}t).$$

Let $\rho : [0, \infty) \mapsto (0, \infty]$ be the density of the point process corresponding to the pure birth process $X(t)$ and $\hat{\rho} : (0, \infty) \to (0, \infty]$ the (formal) Laplace transform of ρ:

$$\hat{\rho}(\lambda) := \int_0^\infty e^{-\lambda t} \rho(t)\,\mathrm{d}t = \sum_{n=0}^\infty \prod_{i=0}^{n-1} \frac{w(i)}{\lambda + w(i)}. \tag{3}$$

The rightmost expression of $\hat{\rho}(\lambda)$ is easily computed, given the fact that the intervals between successive jumps of $X(t)$ are independent exponentially distributed random variables of parameters $w(0), w(1), w(2), \ldots$, respectively. Let

$$\underline{\lambda} := \inf\{\lambda > 0 : \hat{\rho}(\lambda) < \infty\}. \tag{4}$$

Throughout this paper we impose the following condition on the weight function w:

$$\lim_{\lambda \searrow \underline{\lambda}} \hat{\rho}(\lambda) > 1. \tag{M}$$

We are now ready to define our randomly growing tree $\Upsilon(t)$, which will be a continuous time, time-homogeneous Markov chain on the countable state space $\mathcal{G}$, with initial state $\Upsilon(0) = \{\emptyset\}$.

The jump rates are the following: if for a $t \geq 0$ we have $\Upsilon(t) = G$ then the process may jump to $G \cup \{xk\}$ with rate $w(\deg(x, G))$, where $x \in G$ and $k = \deg(x, G) + 1$. This means that each existing vertex $x \in \Upsilon(t)$ 'gives birth to a child' with rate $w(\deg(x, \Upsilon(t)))$ independently of the others.

Note that condition (M) implies

$$\sum_{k=0}^{\infty} \frac{1}{w(k)} = \infty$$

and hence it follows that the Markov chain $\Upsilon(t)$ is well defined for $t \in [0, \infty)$, it does not blow up in finite time. A rigorous proof of this statement follows from the connection with general branching processes (see Section 5) for which the related statement is derived in [6].

We define the *total weight* of a tree $G \in \mathcal{G}$ as

$$W(G) := \sum_{x \in G} w(\deg(x, G)).$$

Described in other words, the Markov chain $\Upsilon(t)$ evolves as follows: assuming $\Upsilon(t-) = G$, at time t a new vertex is added to it with total rate $W(G)$, which is attached with an oriented edge (pointing toward the newly added vertex) to the already existing vertex $x \in G$ with probability

$$\frac{w(\deg(x, G))}{\sum_{y \in G} w(\deg(y, G))}.$$

Therefore, if we only look at our process at the stopping times when a new vertex is just added to the randomly growing tree,

$$T_n := \inf\{t : |\Upsilon(t)| = n + 1\}, \tag{5}$$

then we get the discrete time model: $\Upsilon(T_n)$ has the same distribution as $\Upsilon_d(n)$, the discrete time model at time n.

It will sometimes be convenient to refer to the vertices in the order of their birth, not their genealogical code: let $\{\eta_k\} = \Upsilon(T_k) \setminus \Upsilon(T_k-)$ denote the vertex that appeared at T_k. Of course we will always have $\eta_0 = \emptyset$ and $\eta_1 = 1$.

4. RESULTS

4.1. Statement of Results

From condition (M) it follows that the equation

$$\hat{\rho}(\lambda) = 1$$

has a unique root λ^*.

Now we are ready to state our first theorem.

Theorem 1. *Consider a weight function w satisfying condition* (M) *and let λ^* be defined as above. Consider a bounded function $\varphi : \mathcal{G} \to \mathbb{R}$. Then the following limit holds almost surely:*

$$\lim_{t\to\infty} \frac{1}{|\Upsilon(t)|} \sum_{x\in\Upsilon(t)} \varphi(\Upsilon(t)_{\downarrow x}) = \lambda^* \int_0^\infty e^{-\lambda^* t} \mathbf{E}\big(\varphi(\Upsilon(t))\big)\, dt.$$

From Theorem 1 several statements follow regarding the asymptotic behavior of our random tree *as seen from a randomly selected vertex* ζ, chosen uniformly from $\Upsilon(t)$. As typical examples we determine the asymptotic distribution of the number of children, respectively, that of the whole subtree under the randomly chosen vertex, its kth ancestor, respectively. That is, the asymptotic distribution of $\deg(\zeta, \Upsilon(t)) \in \mathbb{N}$, $\Upsilon(t)_{\downarrow\zeta} \in \mathcal{G}$ and $(\Upsilon(t)_{\downarrow\zeta^{(k)}}, \zeta_{\downarrow k}) \in \mathcal{G}^{(k)}$.

In order to formulate these consequences of Theorem 1 we need to introduce some more notation. Let $G \in \mathcal{G}$ and one of its historical orderings $s = (s_0, s_1, \ldots, s_{|G|-1}) \in \mathcal{S}(G)$ be fixed. The historical sequence of total weights are defined as

$$W(G, s, i) := W(G(s, i)) \tag{6}$$

for $0 \le i \le |G| - 1$ while the respective weights of the appearing vertices are defined as

$$w(G, s, i) := w\left(\deg\big((s_i)^1, G(s, i-1)\big)\right), \tag{7}$$

for $1 \le i \le |G| - 1$. Since $\deg\big((s_i)^1, G(s, i-1)\big)$ is the degree of s_i's parent just before s_i appeared, $w(G, s, i)$ is the rate with which our random tree process jumps from $G(s, i-1)$ to $G(s, i)$.

Given the weight function $w : \mathbb{N} \to \mathbb{R}_+$ satisfying condition (M) and λ^* defined as before, define

$$p_w(k) := \frac{\lambda^*}{\lambda^* + w(k)} \prod_{i=0}^{k-1} \frac{w(i)}{\lambda^* + w(i)}, \tag{8}$$

$$\pi_w(G) := \sum_{s\in\mathcal{S}(G)} \frac{\lambda^*}{\lambda^* + W(G)} \prod_{i=0}^{|G|-2} \frac{w(G, s, i+1)}{\lambda^* + W(G, s, i)}. \tag{9}$$

Theorem 2. *Consider a weight function w that satisfies condition (M) and let λ^* be defined as before. Then the following limits hold almost surely:*

(a) *For any fixed $k \in \mathbb{N}$*

$$\lim_{t\to\infty} \frac{|\{x \in \Upsilon(t) : \deg(x, \Upsilon(t)) = k\}|}{|\Upsilon(t)|} = p_w(k).$$

(b) *For any fixed $G \in \mathcal{G}$*

$$\lim_{t\to\infty} \frac{\left|\{x \in \Upsilon(t) : \Upsilon(t)_{\downarrow x} = G\}\right|}{|\Upsilon(t)|} = \pi_w(G).$$

(c) *For any fixed* $(G, u) \in \mathcal{G}^{(k)}$

$$\lim_{t \to \infty} \frac{\left|\{x \in \Upsilon(t) : (\Upsilon(t)_{\downarrow x^{(k)}}, x_{\downarrow k}) = (G, u)\}\right|}{|\Upsilon(t)|} = \pi_w(G).$$

Furthermore, the functions p_w, π_w *are probability distributions on* $\mathbb{N}$ *and* $\mathcal{G}$, *respectively, and* π_w *is steady (i.e., identity (2) holds).*

Remarks.   **1.** Parts (a), (b), and (c) of Theorem 2, in turn, give more and more information about the asymptotic shape of the randomly growing tree $\Upsilon(t)$, as seen from a random vertex ζ chosen with uniform distribution. Part (a) identifies the a.s. limit as $t \to \infty$, of the degree distribution of ζ. Part (b) identifies the a.s. limit as $t \to \infty$, of the distribution of the progeny of ζ. Finally, part (c) does the same for the distribution of the progeny of the kth ancestor of the randomly selected vertex with the position of this vertex marked.

2. From part (c) it is easy to derive the asymptotic distribution of the progeny of the kth ancestor of the randomly selected vertex (as a rooted ordered tree without any marked vertices):

$$\lim_{t \to \infty} \frac{\left|\{x \in \Upsilon(t) : \Upsilon(t)_{\downarrow x^{(k)}} = G\}\right|}{|\Upsilon(t)|} = \pi_w(G) \left|G_{[k]}\right|.$$

The limit is the size-biased version of $\pi_w(G)$, with the biasing done by the size of the kth generation.

3. Since the distribution π_w is steady, part (c) identifies the asymptotic distribution of the *whole family tree* of the randomly selected vertex ζ (relatives of arbitrary degree included). Hence, asymptotically, as $t \to \infty$, the tree $\Upsilon(t)$ viewed from a random vertex ζ will have the following structure (we omit the precise formulation):

- there exists an infinite path of ancestors $\zeta^1, \zeta^2, \zeta^3, \dots$ "going back in time,"
- we have finite ordered random trees rooted at each vertex of this path,
- the tree rooted at ζ^k with the position of ζ marked on it has distribution $\pi_w^{(k)}$ on $\mathcal{G}^{(k)}$ where $\pi_w^{(k)}(G, u) = \pi_w(G)$.

4.2. Linear Weight Function

In the linear case $w(k) = \alpha k + \beta$ $(\alpha, \beta > 0)$ all computations are rather explicit. In this case the asymptotic degree distribution p_w (computed in [4], [11]) is reproduced, of course. But even in this simplest case the asymptotic distribution π_w of the subtree under a randomly selected vertex seems to be new.

For sake of completeness, in the rest of this section we perform these (explicit and straightforward) computations for the linear case. Multiplying the rate function with a positive constant only means the rescaling of time in our model; thus, it is enough to consider $w(k) = k + \beta$ (with $\beta > 0$). In this case it is straightforward to compute that condition (M) holds, $\hat{\rho}(\lambda) = \frac{\beta}{\lambda - 1}$, $\underline{\lambda} = 1$ and $\lambda^* = 1 + \beta$. Thus, both Theorems 1 and 2 hold.

For the asymptotic degree distribution we get

$$p_w(k) = (1 + \beta) \frac{(k - 1 + \beta)_k}{(k + 1 + 2\beta)_{k+1}},$$

where we used the shorthanded notation

$$(x)_k := \prod_{i=0}^{k-1}(x-i) = \frac{\Gamma(x+1)}{\Gamma(x-k+1)}, \qquad k = 0, 1, 2, \ldots.$$

For the calculation of $\pi_w(G)$ first we show that the sum that defines it contains identical elements. In order to avoid heavy notation, during the following computations we will use $n := |G| - 1$ and $\deg(x)$ instead of $\deg(x, G)$.

Clearly, for any $s \in \mathcal{S}(G)$

$$\prod_{i=0}^{n-1} w(G, s, i+1) = \prod_{x \in G} \left(\prod_{j=0}^{\deg(x)-1} w(j) \right) = \prod_{x \in G}(\deg(x) - 1 + \beta)_{\deg(x)}.$$

(Actually, the first equality holds for every weight function w.) It is also easy to see that for any $G \in \mathcal{G}$

$$W(G) = \sum_{x \in G}(\deg(x) + \beta) = |G|(1 + \beta) - 1,$$

thus, for any $s \in \mathcal{S}(G)$

$$\frac{\lambda^*}{\lambda^* + W(G)} \prod_{i=0}^{n-1} \frac{1}{\lambda^* + W(G, s, i)} = \frac{1}{(1+\beta)^n (n + 2 - (1+\beta)^{-1})_{n+1}}.$$

Therefore,

$$\pi_w(G) = |\mathcal{S}(G)| \frac{\prod_{x \in G}(\deg(x) - 1 + \beta)_{\deg(x)}}{(1+\beta)^n (n + 2 - (1+\beta)^{-1})_{n+1}}.$$

In the $\beta = 1$ case (i.e., if we consider random tree proposed in [2]) the previous calculations give

$$p_w(k) = \frac{4}{(k+1)(k+2)(k+3)}$$

and

$$\pi_w(G) = \frac{2|\mathcal{S}(G)|}{(2|G|+1)!!} \prod_{x \in G} \deg(x)!.$$

The value of $|\mathcal{S}(G)|$ cannot be written as the function of degrees of G only, but one can compute it using the values $|G_{\downarrow x}|$ for $x \in G$. For a given G and $x = (x_1, x_2, \ldots, x_n) \in G$ let us introduce the following notations (these will not be used in the other parts of the paper):

$$B(x) := \{y \in G : y = (x_1, x_2, \ldots, x_{n-1}, k), k > x_n\},$$

$$a(x) := \max(|G_{\downarrow x}| - 1, 1), \qquad b(x) := \max\left(\sum_{y \in B(x)} |G_{\downarrow x}|, 1 \right).$$

It is a simple exercise to prove that for a $G \in \mathcal{G}$ with $|G| > 1$

$$|\mathcal{S}(G)| = (|G| - 2)! \prod_{x \in G, x \neq \emptyset} a(x)^{-1} b(x)^{-1}.$$

The proof is left to the reader.

5. BRANCHING PROCESSES

The random tree model, defined in continuous time, has the big advantage that it fits into the framework of the well-established theory of general branching processes. We give a brief introduction to the fundamentals and state the theorems that we rely on in our proofs. We do not give a broad survey on the most general types of branching processes here, we choose to focus on the results that may be applied to our process. For more details see the monograph [6] or the papers [7], [12], and [14] and the references therein. For a survey on branching processes, trees, and superprocesses, see [10].

In the case of a general branching process, there is a population in which each individual reproduces at ages according to i.i.d. copies of a random point process ξ on $[0, \infty)$. We denote by $\xi(t)$ the ξ-measure of $[0, t]$, i.e. the random number of children an individual has at the age of t.

The individuals in the population are labeled with the elements of $\mathcal{N}$, as described in Section 3 (see (1)). The basic probability space is

$$(\Omega, \mathcal{A}, P) = \prod_{x \in \mathcal{N}} (\Omega_x, \mathcal{A}_x, P_x),$$

where $(\Omega_x, \mathcal{A}_x, P_x)$ are identical spaces on which ξ_x are distributed like ξ.

For each $x \in \mathcal{N}$ there is a $._{\downarrow x}$ shift defined on Ω by

$$(\omega_{\downarrow x})_y = \omega_{xy},$$

in plain words, $\omega_{\downarrow x}$ is the life of the progeny of x, regarding x as the ancestor.

The birth times σ_x of the individuals are defined in the obvious way: $\sigma_0 = 0$ and if $x' = xn$ with $n \in \mathbb{Z}_+$ then

$$\sigma_{x'} = \sigma_x + \inf\{t : \xi_x(t) \geq n\}. \tag{10}$$

The branching process is often counted by a random characteristic; this can be any real-valued process $\{\Phi : \mathbb{R} \times \Omega \to \mathbb{R}\}$. For each individual x, Φ_x is defined by

$$\Phi_x(t, \omega) = \Phi(t, \omega_{\downarrow x}),$$

in plain words $\Phi_x(t)$ denotes the value of Φ evaluated on the progeny of x, regarding x as the ancestor, at the time when x is of age t. We can think about $\Phi_x(t)$ as a "score" given to x when its age is t. With this,

$$Z_t^{\Phi} = \sum_{x \in \mathcal{N}} \Phi_x(t - \sigma_x)$$

is the branching process counted by the random characteristic Φ (the "total score" of the population at time t).

For our applications we only consider random characteristics that are 0 for $t < 0$ and equal to a bounded <u>deterministic</u> function of the rooted tree for $t \geq 0$.

This means that only those individuals contribute to Z_t^{Φ} that are born up to time t and their contribution is a deterministic function of their progeny tree. (Random characteristics may be defined in a more general way; see, e.g., [6], [7].) One of the important examples is $\Phi(t) = \mathbb{1}\{t \geq 0\}$ when Z_t^{Φ} is just the total number of individuals born up to time t.

The Laplace transform of $d\xi(t)$ will be of great importance, we denote this random variable by

$$\hat{\xi}(\lambda) := \int_0^\infty e^{-\lambda t} \, d\xi(t). \tag{11}$$

We shall be interested in *supercritical, Malthusian* processes, meaning that there exists a finite $0 < \lambda^* < \infty$ (the so-called Malthusian parameter) for which

$$\mathbf{E}\,\hat{\xi}(\lambda^*) = 1 \tag{12}$$

and also

$$\kappa = -\partial_\lambda \left(\mathbf{E}\,\hat{\xi}(\lambda) \right)\Big|_{\lambda=\lambda^*} = \mathbf{E} \int_0^\infty t e^{-\lambda^* t} \, d\xi(t) < \infty. \tag{13}$$

(The last property means that the process is Malthusian and the first means that it is super-critical.) Also, we require the reproduction to be non-lattice, which means that the jumps of $\xi(t)$ cannot be supported by any lattice $\{0, d, 2d, \ldots\}, d > 0$ with probability 1.

We quote here a weaker form of Theorem 6.3 from [12], using its extension, which appears in Section 7 of the same paper. This way the conditions of the original theorem are fulfilled automatically.

Theorem A [12].   *Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*, counted by two random characteristics $\Phi(t)$ and $\Psi(t)$ that have the properties described above (i.e., they are 0 for $k < 0$ and a deterministic bounded function of the progeny tree for $t \geq 0$). Suppose that there exists a $\underline{\lambda} < \lambda^*$ for which*

$$\mathbf{E}\,\hat{\xi}(\underline{\lambda}) < \infty.$$

Then almost surely

$$\frac{Z_t^\Phi}{Z_t^\Psi} \to \frac{\hat{\Phi}(\lambda^*)}{\hat{\Psi}(\lambda^*)} \quad as\ t \to \infty, \tag{14}$$

where $\hat{\Phi}(\lambda) = \int\limits_0^\infty \exp(-\lambda t) \mathbf{E}(\Phi(t)) \, dt.$

For a reader interested in how the Malthusian parameter and $\hat{\Phi}(\lambda)$ play a role in the theory, we give a short indication. This part may be skipped without any confusion.

The key observation is the so-called *basic decomposition*, namely that

$$Z_t^\Phi = \Phi(t) + \sum_{j \in \mathbb{N}} Z_{t-\sigma_j}^{\Phi_j}. \tag{15}$$

The advantage of this formula is that if we know the sequence $(\sigma_j)_{j \in \mathbb{Z}_+}$ (all the birth times of the children of the root), then $Z_{t-\sigma_j}^{\Phi_j}$ has the same conditional distribution as $Z_{t-\sigma_j}^\Phi$.

Therefore, using the notation $m_t^\Phi = \mathbf{E}Z_t^\Phi$, taking expectation on both sides of (15) in two steps (first conditionally on $(\sigma_j)_{j \in \mathbb{Z}_+}$, then taking expectation regarding $(\sigma_j)_{j \in \mathbb{Z}_+}$), we get

$$m_t^\Phi = \mathbf{E}(\Phi(t)) + \int_0^t m_{t-s}^\Phi \, d\mu(s), \tag{16}$$

where we used the notation $\mu(t) = \mathbf{E}\xi(t)$. Taking the Laplace transform of both sides,

$$\hat{m}(\lambda) = \hat{\Phi}(\lambda) + \hat{m}(\lambda)\hat{\mu}(\lambda), \tag{17}$$

so formally

$$\hat{m}(\lambda) = \frac{\hat{\Phi}(\lambda)}{1 - \hat{\mu}(\lambda)}. \tag{18}$$

Note that $\hat{\mu}(\lambda) = \mathbf{E}\,\hat{\xi}(\lambda)$. This shows that if there is a positive interval below λ^* where the Laplace transform is finite, then $1/(1 - \hat{\mu}(\lambda))$ has a simple pole at λ^* (it is easy to check that $\hat{\mu}'(\lambda^*) < 0$ and $\hat{\mu}''(\lambda^*) > 0$). So taking series expansion and inverse Laplace transform results that

$$m_t^{\Phi} = \frac{1}{\kappa}\hat{\Phi}(\lambda^*)\,e^{\lambda^* t} + o(e^{\lambda^* t}), \tag{19}$$

where κ is the finite constant defined in (13). This means that the ratio of the expectations of Z_t^{Φ} and Z_t^{Ψ} indeed tends to $\frac{\hat{\Phi}(\lambda^*)}{\hat{\Psi}(\lambda^*)}$. To get the almost sure convergence of $\frac{Z_t^{\Phi}}{Z_t^{\Psi}}$ to the same limit needs a lot more work and of course its proof is more elaborate (see [12] for example).

6. PROOFS

Proof of Theorem 1. Consider the continuous time branching process where the reproduction process $\xi(t)$ is the Markovian pure birth process $X(t)$, with rate function w, described at the beginning of Section 3.

Clearly, the time evolution of the population has the same distribution as the evolution of the continuous time Random Tree Model corresponding to the weight function w. The vertices are the respective individuals and edges are the parent–child relations.

It is also not hard to see that the function $\mathbf{E}\,\hat{\xi}(\lambda)$ for the branching process is the same as $\hat{\rho}(\lambda)$, which means that by condition (M) we may apply Theorem A with appropriate random characteristics. Given any bounded function $\varphi : \mathcal{G} \to \mathbb{R}$, setting the characteristics Φ, Ψ as $\Phi(t) := \varphi(\Upsilon(t))\mathbb{1}\{t \geq 0\}$ and $\Psi(t) := \mathbb{1}\{t \geq 0\}$ we get exactly the statement of Theorem 1. ∎

Proof of Theorem 2. (a) Apply Theorem 1 with the function

$$\varphi(G) := \mathbb{1}\{\deg(\emptyset, G) = k\}.$$

This gives that

$$\lim_{t \to \infty} \frac{|\{x \in \Upsilon(t) : \deg(x, \Upsilon(t)) = k\}|}{|\Upsilon(t)|} = \lambda^* \int_0^{\infty} e^{-\lambda^* t}\,\mathbf{P}\big(\deg(\emptyset, \Upsilon(t)) = k\big)\,\mathrm{d}t$$

almost surely. By the definition of σ_k (see (10)),

$$\mathbf{P}\big(\deg(\emptyset, \Upsilon(t)) = k\big) = \mathbf{P}\big(\sigma_k < t\big) - \mathbf{P}\big(\sigma_{k+1} < t\big).$$

Since

$$\lambda^* \int_0^{\infty} e^{-\lambda^* t}\,\mathbf{P}\big(\sigma_k < t\big)\,\mathrm{d}t = \mathbf{E}\big(e^{-\lambda^* \sigma_k}\big),$$

and σ_k is the sum of independent exponentially distributed random variables with parameters $w(0), w(1), \ldots, w(k-1)$, we get

$$\lambda^* \int_0^\infty e^{-\lambda^* t} \mathbf{P}\big(\deg(\emptyset, \Upsilon(t)) = k\big) \, dt = \frac{\lambda^*}{\lambda^* + w(k)} \prod_{i=0}^{k-1} \frac{w(i)}{\lambda^* + w(i)}.$$

This completes the proof of part (a) of the Theorem.

Using the identity

$$p_w(k) = \lambda^* \int_0^\infty e^{-\lambda^* t} \, \mathbf{P}\big(\deg(\emptyset, \Upsilon(t)) = k\big) \, dt$$

and the fact that $|\Upsilon(t)|$ is finite for every t with probability 1 it is straightforward to prove that p_w is indeed a probability distribution on $\mathbb{N}$.

(b) Let $G \in \mathcal{G}$ be fixed and denote $n = |G| - 1$. We apply Theorem 1 with $\varphi(H) = \mathbb{1}(H = G)$. We need to compute

$$\lambda^* \int_0^\infty e^{-\lambda^* t} \, \mathbf{P}\big(\Upsilon(t) = G\big) \, dt.$$

Consider the following random stopping times:

$$\tau_G := \sup\{t \geq 0 : \Upsilon(t) \subseteq G\},$$
$$\tau_G' := \sup\{t \geq 0 : \Upsilon(t) \subsetneq G\}.$$

That is, τ_G is the birth time of the first vertex not in G, while τ_G' is the minimum of τ_G and the time when we first have $\Upsilon(t) = G$, if the latter ever happens. Since

$$\mathbf{P}\big(\Upsilon(t) = G\big) = \mathbf{P}\big(\Upsilon(t) \subseteq G\big) - \mathbf{P}\big(\Upsilon(t) \subsetneq G\big) = \mathbf{P}\big(t < \tau_G\big) - \mathbf{P}\big(t < \tau_G'\big),$$

we get that

$$\lambda^* \int_0^\infty e^{-\lambda^* t} \, \mathbf{P}\big(\Upsilon(t) = G\big) \, dt = \mathbf{E}\big(e^{-\lambda^* \tau_G'} - e^{-\lambda^* \tau_G}\big)$$
$$= \mathbf{E}\big((e^{-\lambda^* \tau_G'} - e^{-\lambda^* \tau_G}) \mathbb{1}\{\tau_G' < \tau_G\}\big).$$

Note that by the definition we always have $\tau_G' \leq \tau_G$. The event $\{\tau_G' < \tau_G\}$ means that there is a t when $\Upsilon(t) = G$. On this event τ_G' gives the time when we first have $\Upsilon(t) = G$ and τ_G gives the appearance of the next vertex. Given the event $\{\tau_G' < \tau_G\}$, the conditional distribution of $\tau_G - \tau_G'$ is exponential with parameter $W(G)$ and it is (conditionally) independent of τ_G'. This leads to

$$\mathbf{E}\big((e^{-\lambda^* \tau_G'} - e^{-\lambda^* \tau_G}) \mathbb{1}\{\tau_G' < \tau_G\}\big) = \frac{\lambda^*}{\lambda^* + W(G)} \, \mathbf{E}\big(e^{-\lambda^* \tau_G'} \mathbb{1}\{\tau_G' < \tau_G\}\big).$$

Now, it is clear that the following two events are actually the same

$$\{\tau_G' < \tau_G\} = \big\{(\eta_0, \ldots, \eta_n) = (s_0, \ldots, s_n) \text{ for some } s \in \mathcal{S}(G)\big\}.$$

This implies that

$$\mathbf{E}\big(e^{-\lambda^* \tau'_G} \mathbb{1}\{\tau'_G < \tau_G\}\big)$$
$$= \sum_{s \in \mathcal{S}(G)} \mathbf{E}\big(e^{-\lambda^* T_n} \mathbb{1}\{(\eta_0, \ldots, \eta_n) = (s_0, \ldots, s_n)\}\big).$$

For T_n see (5) and the definition below it. Given $s \in \mathcal{S}(G)$ fixed,

$$\mathbf{P}\big((\eta_0, \ldots, \eta_n) = (s_0, \ldots, s_n)\big) = \prod_{i=0}^{n-1} \frac{w(G, s, i+1)}{W(G, s, i)}.$$

(See (3.2), (6), and (7) for the definitions.) Also, if $s \in \mathcal{S}(G)$ is fixed then *conditionally on the event* $\{(\eta_0, \ldots, \eta_n) = (s_0, \ldots, s_n)\}$ the random variables $T_{k+1} - T_k, k = 0, 1, \ldots, n-1$, are independent and exponentially distributed with parameters $W(G, s, k), k = 0, 1, \ldots, n-1$, respectively. This is an easy exercise: it may be proved by using the "lack of memory" of the exponential distribution and the fact that the minimum of independent exponentially distributed random variables with parameters $v_1, v_2, \ldots, v_l$ is also exponentially distributed with parameter $\sum_{i=1}^{l} v_i$. Hence, it is straightforward to get

$$\mathbf{E}\big(e^{-\lambda^* T_n} \mathbb{1}\{(\eta_0, \ldots, \eta_n) = (s_0, \ldots, s_n)\}\big) = \prod_{i=0}^{n-1} \frac{w(G, s, i+1)}{\lambda^* + W(G, s, i)}.$$

Collecting our previous calculations part (b) of Theorem 2 follows.

Using similar considerations as in the end of the proof of part (a) it is apparent that π_w is a probability distribution on $\mathcal{G}$.

(c) This is straightforward since for any $H \in \mathcal{G}$ and $(G, u) \in \mathcal{G}^{(k)}$ we have

$$\big|\{x \in H : (H_{\downarrow x^{(k)}}, x_{\downarrow k}) = (G, u)\}\big| = \big|\{x \in H : H_{\downarrow x} = G\}\big|.$$

The statement now follows from part (b).

The only thing left to prove is that π_w satisfies (2), i.e., it is steady. First observe that if $G_0 \in \mathcal{G}$ is fixed and ζ is a uniformly chosen random vertex in G_0 then the distribution of $\Gamma := (G_0)_{\downarrow \zeta}$ (which is a probability distribution on $\mathcal{G}$) is steady. (This follows by simple counting.) Equation (2) is linear in π; therefore, mixtures of steady distributions are also steady. Thus, if ζ is a uniformly chosen random vertex in $\Upsilon(t)$ then the distribution of $\Upsilon(t)_{\downarrow \zeta}$ (which is a *random* probability distribution on $\mathcal{G}$) is also steady. By part (b) with probability one these distributions converge (in distribution) to π_w and from this an easy consideration shows that π_w must satisfy (2).   ∎

7. ASYMPTOTIC GROWTH

In our theorems we determined the asymptotic ratio of vertices in $\Upsilon(t)$ satisfying certain properties. It is also natural to ask whether one can prove results about the asymptotic number of the respective vertices. As we have seen, this essentially requires to study the asymptotic behavior of Z_t^Φ for a suitable random characteristic Φ. This has been done in the framework of general branching processes; we shall give a short overview of the relevant results.

We have already seen that

$$\mathbf{E}(e^{-\lambda^* t} Z_t^\Phi) \to \frac{1}{\kappa} \hat{\Phi}(\lambda^*), \tag{20}$$

where κ is the constant defined in (13). Thus, we need to divide Z_t^Φ by $e^{\lambda^* t}$ to get something non-trivial. Let us quote a weaker form of Theorem 5.4 of [12].

Theorem B [12]. *Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*. Suppose that condition (M) holds and Φ is a random characteristic with properties described in Section 4. Then almost surely*

$$e^{-\lambda^* t} Z_t^\Phi \to \frac{1}{\kappa} \hat{\Phi}(\lambda^*) \Theta, \quad as \; t \to \infty,$$

where Θ is a random variable not depending on Φ.

The necessary and sufficient condition for the random variable Θ to be a.s. positive is the so-called $x \log x$ property of the reproduction process ξ:

$$\mathbf{E}(\hat{\xi}(\lambda^*) \log^+ \hat{\xi}(\lambda^*)) < \infty. \tag{L}$$

We quote Theorem 5.3 of [7].

Theorem C [7]. *Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*. If condition (L) holds then $\Theta > 0$ a.s. and $\mathbf{E}(\Theta) = 1$; otherwise $\Theta = 0$ a.s.*

Remark. This theorem is the generalization of the Kesten–Stigum theorem, which states this fact for Galton–Watson processes (see [8]).

Theorem B applies for the random tree model if condition (M) is fulfilled. We do not intend to identify the necessary and sufficient condition on the weight function w, which would guarantee that the corresponding reproduction process possesses property (L). Still it is worth pointing out that if $w(k) \to \infty$ as $t \to \infty$, then this property holds; this, by Theorem C, Θ is a.s. positive.

Lemma 1. *If a weight function w satisfies condition (M) and $w(n) \to \infty$, as $n \to \infty$, then the corresponding branching process satisfies condition (L).*

Proof. We will prove the existence of the second moment of $\hat{\xi}(\lambda^*)$ from which condition (L) trivially follows. Since $\hat{\xi}(\lambda^*) = \sum_{i=1}^\infty e^{-\lambda^* \sigma_k}$, we need

$$\mathbf{E} \left(\sum_{k=1}^\infty e^{-\lambda^* \sigma_k} \right)^2 < \infty. \tag{21}$$

The random variables $\sigma_{k+1} - \sigma_k$ are independent exponentials for $k = 0, 1, 2, \dots$ with parameters $w(0), w(1), \dots$, respectively; thus a simple computation yields that the expression in (21) is equal to

$$-\hat{\rho}(2\lambda^*) + 2 \sum_{i=0}^\infty \sum_{j=0}^i \left(\prod_{l=0}^j \frac{w(l)}{2\lambda^* + w(l)} \prod_{l=j+1}^i \frac{w(l)}{\lambda^* + w(l)} \right).$$

Transforming the double sum on the right, we get

$$\sum_{i=0}^{\infty}\left(\prod_{l=0}^{i}\frac{w(l)}{\lambda^*+w(l)}\right)\sum_{j=0}^{i}\prod_{l=0}^{j}\frac{\lambda^*+w(l)}{2\lambda^*+w(l)} \leq \sup_{i}\sum_{j=0}^{i}\prod_{l=0}^{j}\frac{\lambda^*+w(l)}{2\lambda^*+w(l)},$$

where we also used $\hat{\rho}(\lambda^*) = 1$. On the other hand,

$$\sum_{j=0}^{i}\prod_{l=0}^{j}\frac{\lambda^*+w(l)}{2\lambda^*+w(l)} = \sum_{j=0}^{i}\prod_{l=0}^{j}\frac{w(l)}{\underline{\lambda}+w(l)}\prod_{l=0}^{j}\frac{(\underline{\lambda}+w(l))(\lambda^*+w(l))}{w(l)(2\lambda^*+w(l))}.$$

Since $w(l) \to \infty$ and $\underline{\lambda} < \lambda^*$, we have

$$\frac{(\underline{\lambda}+w(l))(\lambda^*+w(l))}{w(l)(2\lambda^*+w(l))} < 1$$

if l is large enough. This leads to

$$\sum_{j=0}^{i}\prod_{l=0}^{j}\frac{\lambda^*+w(l)}{2\lambda^*+w(l)} < K \sum_{j=0}^{i}\prod_{l=0}^{j}\frac{w(l)}{\underline{\lambda}+w(l)} < K\hat{\rho}(\underline{\lambda}) < \infty$$

by condition (M), which completes the proof of the lemma. ∎

The distribution of Θ is usually hard to determine from the weight function w; however, one can characterize its moment generating function $\varphi(u) = \mathbf{E}e^{-u\Theta}$. Using the idea of the basic decomposition (15) one can write the following equation for $f(u,t) := \mathbf{E}e^{-u|\Upsilon(t)|}$:

$$f(u,t) = e^{-u}\mathbf{E}\prod_{j\geq 1}f(u,t-\tau_j).$$

By Theorem 4

$$\varphi(u) = \lim_{t\to\infty}f(ue^{-\lambda^*t},t), \tag{22}$$

which gives

$$\varphi(u) = \mathbf{E}\prod_{j\geq 1}\varphi(ue^{-\lambda^*\tau_j}). \tag{23}$$

It can be proved that this equation characterizes φ as there is no other bounded function satisfying it with a right derivative -1 at 0. (See Theorem 6.8.3 in [6].)

In the linear case ($w(k) = \alpha k + \beta$) the distribution of Θ may be explicitly calculated. Using the fact that $|\Upsilon(t)|$ is now a Markov process (which is an easy exercise to prove) and applying standard martingale techniques one can derive a partial differential equation for $f(u,t)$. Solving this pde and using the limit (22) we get the momentum generating function of ω from which one can identify the distribution of Θ as a gamma distribution with parameters $(\frac{\beta}{\beta+\alpha}, \frac{\beta}{\beta+\alpha})$.

Acknowledgments

REFERENCES

[1] D. J. Aldous, Asymptotic fringe distributions for general families of random trees, Ann Appl Probab 1 (1991), 228–266.

[2] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science 286 (1999), 509–512.

[3] B. Bollobás, "Mathematical results on scale-free random graphs," S. Bornholdt and H.G. Schuster (Editors), Handbook of Graphs and Networks, Wiley, 2002, pp. 1–34.

[4] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, The degree sequence of a scale-free random graph process, Random Struct Algor 18 (2001), 279–290.

[5] F. Chung, S. Handjani, and D. Jungreis, Generalizations of Polya's urn problem, Ann Combinator 7 (2003), 141–154.

[6] P. Jagers, Branching Processes with Biological Applications. Wiley, 1975.

[7] P. Jagers and O. Nerman, The growth and composition of branching populations, Adv Appl Probab 16 (1984), 221–259.

[8] H. Kesten and B. Stigum, A limit theorem for multidimensional Galton–Watson processes, Ann Math Statist 37 (1966), 1211–1223.

[9] P. L. Krapivsky and S. Redner, Organization of growing random networks, Phys Rev E 63 (2001), 066123.

[10] J-F. Le Gall, "Processus de branchement, arbres et superprocessus," J-P. Pier (Editor), Development of Mathematics 1950–2000, Birkhäuser, 2000, pp. 763–793.

[11] T. Móri, On random trees, Stud Sci Math Hungar 39 (2002), 143–155.

[12] O. Nerman, On the convergence of supercritical (C-M-J) branching processes, Z Wahrsch verw, Geb 57 (1981), 365–395.

[13] R. Oliveira and J. Spencer, Connectivity transitions in networks with super-linear preferential attachment, Internet Math, to appear.

[14] P. Olofsson, The $x \log x$ condition for general branching processes, J Appl Probab 35 (1998), 537–544.