Random Tree Growth with Branching Processes – a Survey

Anna Rudas and Bálint Tóth

Abstract We investigate the asymptotic properties of a random tree growth model which generalizes the basic concept of preferential attachment. The Barabási-Albert random graph model is based on the idea that the popularity of a vertex in the graph (the probability that a new vertex will be attached to it) is proportional to its current degree. The dependency on the degree, the so-called weight function, is linear in this model. We give results which are valid for a much wider class of weight functions. This generalized model has been introduced by Krapivsky and Redner in the physics literature. The method of re-phrasing the model in a continuous-time setting makes it possible to connect the problem to the well-developed theory of branching processes. We give local results, concerning the neighborhood of a "typical" vertex in the tree, and also global ones, about the repartition of mass between subtrees under fixed vertices.

1 Random Tree Model

A natural concept of randomly growing trees is the class of models where at each discrete time step, a new vertex appears, and is attached to an already existing vertex randomly, the distribution of the choice depending on the actual degrees of the vertices currently apparent in the tree. This dependence on the degree structure is characterised by a *weight function* $w : \mathbb{N} \to \mathbb{R}_+$.

The concept of preferential attachment generally means that w is an increasing function. One of the realizations of this idea is the Barabási - Albert graph [2], which model reproduces certain phenomena observed in real-world networks, the power-

Anna Rudas
Institute of Mathematics, Budapest University of Technology, e-mail: rudasa@math.bme.hu

Bálint Tóth
Institute of Mathematics, Budapest University of Technology, e-mail: balint@math.bme.hu

law decay of the degree sequence, for example. This was proved in a mathematically precise way in [3] and, independently, in [15]. The Barabási - Albert model, for trees, corresponds to the special case of the model considered in this survey, when w is chosen to be linear. In [3] and [15], the techniques used strongly depend on martingales that are apparent in the system only in the linear case. For a survey on random graph models that produce scale-free behavior, see Bollobás and Riordan [4] and Chapter 4 of Durrett [7].

General weight functions are considered in the model of Krapivsky and Redner [12], [13]. 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 regime the limiting object does not have a non-trivial degree sequence: a single dominant vertex appears which is linked to almost every other vertex, the others having only finite degree. This statement is made precise and proved rigorously in [17]. See also [5] for a related model. In this survey, following [20], we consider weight functions for which this does not happen, our class includes the second regime $\gamma \leq 1$ mentioned above.

Certain similar random recursive trees and random plane-oriented trees have been studied before, see, for example, [21].

1.1 Notation

We consider rooted ordered trees, which are also called family trees or rooted planar trees in the literature.

In order to refer to these trees it is convenient to use genealogical phrasing. The tree is thus regarded as the coding of the evolution of a population stemming from one individual (the root of the tree), whose "children" form the "first generation" (these are the vertices connected directly to the root). In general, the edges of the tree represent parent-child relations, the parent always being the one closer to the root. The birth order between brothers is also taken into account, this is represented by the tree being an ordered tree (planar tree). The vertices are labelled by the set

$$\mathscr{N} = \bigcup_{n=0}^{\infty} \mathbb{Z}_+^n, \quad \text{where} \quad \mathbb{Z}_+ := \{1, 2, \dots\}, \quad \mathbb{Z}_+^0 := \{\emptyset\} ,$$

as follows. $\emptyset$ denotes the root of the tree, its firstborn child is labeled by 1, the second one by 2, etc., all the vertices in the first generation are thus labeled with the elements of $\mathbb{Z}$. Similarly, in general, the children of $x = (i_1, i_2, \dots, i_k)$ are labeled by $(i_1, i_2, \dots, i_k, 1)$, $(i_1, i_2, \dots, i_k, 2)$, etc. Thus, if a vertex has label $x = (i_1, i_2, \dots, i_k) \in \mathscr{N}$, then it is the i_k^{th} child of its parent, which is the i_{k-1}^{th} child of its own parent and so on. If $x = (i_1, i_2, \dots, i_k)$ and $y = (j_1, j_2, \dots, j_l)$ then we will use the shorthand notation xy for the concatenation $(i_1, i_2, \dots, i_k, j_1, j_2, \dots, j_l)$, and with a slight abuse of notation for $n \in \mathbb{Z}$, we use xn for $(i_1, i_2, \dots, i_k, n)$.

There is a natural partial ordering $\preceq$ on $\mathcal{N}$, namely, $x \preceq z$ if x is ancestor of z, so if $\exists y \in \mathcal{N}$ such that $z = xy$. If we want to exclude equality, we use $x \prec z$, which thus means that $x \preceq z$ but $x \neq z$.

We can identify a rooted ordered tree with the set of labels of the vertices, since this set already identifies the set of edges in the tree. It is clear that a subset $G \subset \mathcal{N}$ may represent a rooted ordered tree iff $\emptyset \in G$, and for each $(i_1, i_2, \ldots, i_k) \in G$ we have $(i_1, i_2, \ldots, i_k - 1) \in G$ if $i_k > 1$, and $(i_1, i_2, \ldots, i_{k-1}) \in G$ if $i_k = 1$.

$\mathcal{G}$ will denote the set of all finite, rooted ordered trees. The *degree* of vertex $x \in G$ will denote the number of its children in G:

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

Thus *degree* in this paper is one less than the usual graph theoretic degree, except for the root, where it is the same. 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.

1.2 The Model

First we describe the growth model in discrete time. Then we present a continuous-time model which, at certain random stopping times, is equivalent to the previous one. The method of investigating the discrete time growth model by introducing the continuous-time setting described below appears in [17] and in [20], independently of each other.

1.2.1 Discrete Time Model

Given the 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

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

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 the tree G is proportional to $w(\deg(x, G))$.

1.2.2 Continuous Time Model

Given the 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 $X(t)$, namely let

$$\rho(t) = \lim_{\varepsilon \to 0} \varepsilon^{-1} \mathbf{P}((t, t+\varepsilon) \text{ contains a point from } X) . \tag{1}$$

We denote the (formal) Laplace transform of ρ by $\widehat{\rho} : (0,\infty) \to (0,\infty]$:

$$\widehat{\rho}(\lambda) := \int_0^\infty \mathrm{e}^{-\lambda t} \rho(t) \mathrm{d}t = \sum_{n=1}^\infty \prod_{i=0}^{n-1} \frac{w(i)}{\lambda + w(i)} . \tag{2}$$

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

$$\underline{\lambda} := \inf\{\lambda : \widehat{\rho}(\lambda) < \infty\} . \tag{3}$$

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

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

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 $\mathscr{G}$, 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

$$\sum_{k=0}^\infty \frac{1}{w(k)} = \infty \tag{4}$$

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. If, on the other hand, (4) does not hold, then $\underline{\lambda}$ is not finite. A rigorous proof of this statement follows from the connection with general branching processes (see Sect. 3) for which the related statement is derived in [9].

We define the *total weight* of a tree $G \in \mathscr{G}$ 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 $W(G)$, and it is attached with an edge to exactly one already existing vertex, which is $x \in G$ with probability

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

Therefore, if we only look at our continuous time 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.

1.3 Some Additional Notation

We are ready to introduce the notations τ_x and σ_x, as follows. Let τ_x be the birth time of vertex x,

$$\tau_x := \inf\{t > 0 \ : x \in \Upsilon(t)\} \ . \tag{6}$$

Let σ_x be the time we have to wait for the appearance of vertex x, starting from the moment that its birth is actually possible (e.g. when no other vertex is obliged to be born before it). Namely, let

(a) $\sigma_\emptyset := 0$,
(b) $\sigma_{y1} := \tau_{y1} - \tau_y$, for any $y \in \mathscr{N}$,
(c) and $\sigma_{yk} := \tau_{yk} - \tau_{y(k-1)}$, for each $y \in \mathscr{N}$ and $k \geq 2$.

Also, we will need the concept of historical orderings of the vertices in a finite tree, as follows.

Consider a $G \in \mathscr{G}$. 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 $\{s_0, s_1, \ldots, s_i\} \in \mathscr{G}$. The set of all historical orderings of $G \in \mathscr{G}$ will be denoted $\mathscr{S}(G)$. For a fixed $s \in \mathscr{S}(G)$ the rooted ordered trees

$$G(s,i) := \{s_0, s_1, \ldots, s_i\} \subset G \tag{7}$$

give the evolution of G in this historical ordering s.

Let $G \in \mathscr{G}$ and one of its historical orderings $s = (s_0, s_1, \ldots, s_{|G|-1}) \in \mathscr{S}(G)$ be fixed. The historical sequence of total weights are defined as

$$W(G,s,i) := W(G(s,i)) \tag{8}$$

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

$$w(G,s,i) := w\left(\deg\left((s_i)^1, G(s,i-1)\right)\right) , \tag{9}$$

for $1 \le i \le |G|-1$, here $(s_i)^1$ denotes the parent of s_i. Since $\deg\left((s_i)^1, G(s,i-1)\right)$ 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 $G(s,i)$.

2 Local Properties

We ask questions about the neighborhood of the "typical" vertex (e.g. sampled uniformly randomly from the tree), after a long time. We state theorems regarding the limit distribution of the number of children, and of the subtree under the uniformly selected vertex. We then comment on the case of linear weight functions, where the formulae gain a simpler explicit form. After that we state a more general version of these theorems, which turns out to be the analogue of an important theorem in the theory of general branching processes (see Sect. 3). At the end of the section we give an argument on the convergence in Theorem 2, different from that of the analoguous theorem discussed in [16]. Although this approach does not give almost sure convergence, it is elementary and instructive, and gives convergence in probability.

2.1 Statement of Results

Note that from condition (M) it follows that the equation

$$\widehat{\rho}(\lambda) = \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{w(i)}{\lambda + w(i)} = 1 \tag{10}$$

has a unique root $\lambda^* > 0$, called the Malthusian parameter. Now we are ready to state our theorems, quoted from [20].

Theorem 1. *Consider a weight function w which satisfies condition* (M) *and let λ^* be defined by* (10)*. For any $t \ge 0$ let ζ_t denote a random vertex which is, once $\Upsilon(t)$ is given, selected uniformly from $\Upsilon(t)$. Then the following limits hold almost surely:*

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

$$\lim_{t\to\infty} \mathbf{P}\left(\deg(\zeta_t, \Upsilon(t)) = k\right) = \lim_{t\to\infty} \frac{|\{x \in \Upsilon(t) : \deg(x, \Upsilon(t)) = k\}|}{|\Upsilon(t)|} = p_w(k) ,$$

where the limit degree distribution p_w on $\mathbb{N}$ is given by

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

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

$$\lim_{t\to\infty} \mathbf{P}\Big(\Upsilon(t)_{\downarrow\zeta_t} = G\Big) = \lim_{t\to\infty} \frac{\Big|\{x \in \Upsilon(t) : \Upsilon(t)_{\downarrow x} = G\}\Big|}{|\Upsilon(t)|} = \pi_w(G) ,$$

where the limit subtree distribution π_w on $\mathscr{G}$ is given by

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

These results are direct consequences of the following theorem, as shown in detail in [20], with the choices of $\varphi : \mathscr{G} \to \mathrm{I\!R}$ as

$$\varphi(H) = 1\!\!1_{\{\deg(\emptyset,H)=k\}} ,$$

or, respectively,

$$\varphi(H) = 1\!\!1_G(H) .$$

Theorem 2. *Consider a weight function w satisfying condition* (M) *and let λ^* be defined by equation* (10)*. Consider a bounded function $\varphi : \mathscr{G} \to \mathrm{I\!R}$. Then the following limit holds almost surely:*

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

Remark. Based on Theorem 2 it is also possible to “look back” from the uniform random vertex, and ask local questions regarding its parent, grandparent, etc. This Theorem is an application of a result of Nerman, see the remark after Theorem 4, and for the detailed proof see [20]. The related general concept of fringe distributions can be found in Aldous [1].

2.2 Linear Weight Function

In the linear case $w(k) = \alpha k + \beta$ $(\alpha, \beta > 0)$, all computations regarding the distributions p_w and π_w in Theorem 1 are rather explicit. In [3] and [15], the degree distribution is given, for any fixed finite number of vertices. Our computations in the linear case (see below), reproduce the asymptotic degree distribution p, as the size of the tree tends to infinity.

For sake of completeness, 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, $\widehat{\rho}(\lambda) = \frac{\beta}{\lambda-1}$, $\underline{\lambda} = 1$ and $\lambda^* = 1+\beta$. Thus both Theorems 2 and 1 hold.

For the asymptotic degree distribution we get

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

where we used the shorthand notation

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

For the calculation of $\pi(G)$ first we show that the sum which 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 $s \in \mathscr{S}(G)$

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

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

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

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

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

Therefore

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

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

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

and

$$\pi(G) = \frac{2\,|\mathscr{S}(G)|}{(2\,|G|+1)!!}\prod_{x\in G}\deg(x)!$$

Although the value of $|\mathscr{S}(G)|$ cannot be written as the function of degrees of G only, one can compute it using the values $\left|G_{\downarrow x}\right|$ for $x \in G$, for the detalis see [20].

2.3 Proof of Convergence in Probability

In this section we present a proof of the convergence in probability of certain ratios of variables, see Theorem 3. The result formulated here exists in a stronger form, namely, the convergence holds in the almost sure sense, as stated by Theorem 4 in Sect. 3. The proof presented here, though, uses more elementary methods and it is probabilistically instructive.

In order to simplify technicalities, we restrict the class of weight functions from those satisfying condition (M) to a somewhat smaller, but still very wide class. We demand in this section that $w(k) \to \infty$ as $k \to \infty$, with the weight function varying regularly.

$$w(k) = k^{\gamma} + v(k) \tag{11}$$

at some $0 < \gamma \le 1$ and $v(k) = o(k^{\gamma})$ as $k \to \infty$. (We do not need monotonicity for w.)

Let us fix $w(0) = 1$, which can be done without loss of generality, since multiplying all $w(k)$ by a constant just corresponds to rescaling time in the continuous time model.

Let λ^* be the constant defined by (10), and with the letters φ and ψ we denote positive bounded functions $\varphi, \psi : \mathscr{G} \to \mathbb{R}^+$.

Define

$$Z_t^{\varphi} := \sum_{x \in \Upsilon(t)} \varphi(\Upsilon(t)_{\downarrow x})$$

(the analogous definition in Sect. 3 is (24)). We use the notation

$$\kappa := -\partial_{\lambda} \widehat{\rho}(\lambda)\Big|_{\lambda=\lambda^*} = \int_0^{\infty} t \mathrm{e}^{-\lambda^* t} \rho(t) \mathrm{d}t < \infty \,. \tag{12}$$

We also introduce the notation

$$\widehat{\varphi}(\lambda) := \int_0^{\infty} \mathrm{e}^{-\lambda s} \mathbf{E}\left(\varphi(\Upsilon(s))\right) \mathrm{d}s \,. \tag{13}$$

Theorem 3. *Let w satisfy the conditions described in the beginning of Sect. 2.3. Then*

$$\frac{Z_t^{\varphi}}{Z_t^{\psi}} \to \frac{\widehat{\varphi}(\lambda^*)}{\widehat{\psi}(\lambda^*)}$$

in probability, as $t \to \infty$.

To prove Theorem 3 we need Lemmas 1, 2 and 3 below.

Lemma 1.

$$\mathbf{E}(\mathrm{e}^{-\lambda^* t} Z_t^{\varphi}) \to \frac{1}{\kappa} \widehat{\varphi}(\lambda^*) =: d_{\varphi}, \quad \text{as } t \to \infty. \tag{14}$$

Remark. See Sect. 3 and Theorem 5 therein for the analogous, more general result. There we see that $\mathrm{e}^{-\lambda^* t} Z_t^{\varphi}$ itself converges almost surely to a random variable with the appropriate expectation.

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

$$Z_t^{\varphi} = \varphi(\Upsilon(t)) + \sum_{j \in \mathbb{N}} Z_t^{\varphi_j} , \tag{15}$$

where j runs over the children of the root, $\varphi_j(G) := \varphi(G_{\downarrow j})$, and recall that τ_j is the birth time of vertex j.

The advantage of this formula is due to the fact that given the sequence $(\tau_j)_{j \in \mathbb{Z}_+}$, $Z_t^{\varphi_j}$ has the same conditional distribution as $Z_{t-\tau_j}^{\varphi}$.

At this point observe that if Z_t^{φ} is of some exponential order $\mathrm{e}^{\lambda t}$, then λ must be the one defined by equation (10). This can be seen if we take expectation of both sides in equation (15), supposing that $\lim_{t\to\infty} \mathrm{e}^{-\lambda t} Z_t^{\varphi}$ exists almost surely, we can write

$$\begin{aligned} \mathbf{E}(\lim_{t\to\infty} \mathrm{e}^{-\lambda t} Z_t^{\varphi}) &= \sum_{j \in \mathbb{N}} \mathbf{E}\left(\mathrm{e}^{-\lambda \tau_j} \lim_{t\to\infty} \mathrm{e}^{-\lambda (t-\tau_j)} Z_t^{\varphi_j}\right) \\ &= \sum_{j \in \mathbb{N}} \mathbf{E}\left(\mathrm{e}^{-\lambda \tau_j}\right) \mathbf{E}\left(\lim_{t\to\infty} \mathrm{e}^{-\lambda (t-\tau_j)} Z_t^{\varphi_j}\right) = \mathbf{E}\left(\sum_{j \in \mathbb{N}} \mathrm{e}^{-\lambda \tau_j}\right) \mathbf{E}\left(\lim_{t\to\infty} \mathrm{e}^{-\lambda t} Z_t^{\varphi}\right) , \end{aligned}$$

since $\lim_{t\to\infty} \mathrm{e}^{-\lambda (t-\tau_j)} Z_t^{\varphi_j} \overset{\mathrm{d}}{=} \lim_{t\to\infty} \mathrm{e}^{-\lambda t} Z_t^{\varphi}$.

So if the limit exists almost surely, and is non-zero and finite, then

$$\mathbf{E}\left(\sum_{j \in \mathbb{N}} \mathrm{e}^{-\lambda \tau_j}\right) = \widehat{\rho}(\lambda) = 1$$

must hold (compare with (10)).

For the convergence itself, using the notation

$$m_t^{\varphi} := \mathbf{E}\left(Z_t^{\varphi}\right) , \tag{16}$$

taking expectation on both sides of (15) in two steps (first conditionally on $(\tau_j)_{j \in \mathbb{Z}_+}$, then taking expectation regarding $(\tau_j)_{j \in \mathbb{Z}_+}$), we get

$$m_t^{\varphi} = \mathbf{E}(\varphi(\Upsilon(t))) + \int_0^t m_{t-s}^{\varphi} \rho(s) \mathrm{d}s . \tag{17}$$

Taking the Laplace transform of both sides, we have

$$\widehat{m}(\lambda) = \widehat{\varphi}(\lambda) + \widehat{m}(\lambda)\widehat{\rho}(\lambda) ,$$

so formally

$$\widehat{m}(\lambda) = \frac{\widehat{\varphi}(\lambda)}{1 - \widehat{\rho}(\lambda)} .$$

From condition (M) it follows that there is an interval of positive length below λ^* where the Laplace transform is finite, so $1/(1-\widehat{\rho}(\lambda))$ has a simple pole at λ^* (it is easy to check that $\widehat{\rho}'(\lambda^*) < 0$ and $\widehat{\rho}''(\lambda^*) > 0$). Taking series expansion and inverse Laplace transform results that

$$m_t^{\varphi} = \frac{1}{\kappa}\widehat{\varphi}(\lambda^*)\,\mathrm{e}^{\lambda^* t} + o(\mathrm{e}^{\lambda^* t}) ,$$

so, indeed, the statement of the lemma holds.

□

Recall the notation in Sect. 1.2.2, the birth times of the vertices in the first generation of the tree, $(\tau_j)_{j>0}$, constitute the point process X. The density function ρ has already been introduced, see (1). Similarly, let us denote the second correlation function by ρ_2, namely, for $u \neq s$, let

$$\rho_2(u,s) := \lim_{\varepsilon,\delta\to 0} (\varepsilon\delta)^{-1}\mathbf{P}\left((u,u+\varepsilon) \text{ and } (s,s+\delta) \text{ both contain a point from } X\right) ,$$

and we define it to be 0 if $u = s$.

The following estimates are needed.

Lemma 2. *Suppose that w satisfies the conditions described in the beginning of Sect. 2.3. Then*

(a)

$$C_1 := \int_0^\infty \mathrm{e}^{-2\lambda^* s}\rho(s)\mathrm{d}s < 1 , \tag{18}$$

(b)

$$C_2 := \int_0^\infty \int_0^\infty \mathrm{e}^{-\lambda^*(u+s)}\rho_2(u,s)\mathrm{d}u\mathrm{d}s < \infty . \tag{19}$$

Proof. The first statement is obvious, considering that

$$\int_0^\infty \mathrm{e}^{-2\lambda^* s}\rho(s)\mathrm{d}s = \widehat{\rho}(2\lambda^*) < \widehat{\rho}(\lambda^*) < 1 ,$$

since $\widehat{\rho}$ strictly decreases and $\lambda^* > 0$.

As for statement (19), write C_2 as follows:

$$C_2 = 2\sum_{1\le i<j} \mathbf{E}\left(\mathrm{e}^{-\lambda^*(\tau_i+\tau_j)}\right) .$$

Since for any $i < j$, τ_j can be decomposed as the sum of the two independent variables $\tau_j = \tau_i + (\tau_j - \tau_i)$, it can be seen that

$$\mathbf{E}\left(\mathrm{e}^{-\lambda^*\tau_j}\right) = \mathbf{E}\left(\mathrm{e}^{-\lambda^*\tau_i}\mathrm{e}^{-\lambda^*(\tau_j-\tau_i)}\right) = \mathbf{E}\left(\mathrm{e}^{-\lambda^*\tau_i}\right)\mathbf{E}\left(\mathrm{e}^{-\lambda^*(\tau_j-\tau_i)}\right) .$$

It now follows that

$$C_2 = 2\sum_{1\le i<j} \mathbf{E}(\mathrm{e}^{-2\lambda^*\tau_i})\mathbf{E}(\mathrm{e}^{-\lambda^*(\tau_j-\tau_i)}) = 2\sum_{1\le i<j} \mathbf{E}(\mathrm{e}^{-2\lambda^*\tau_i})\frac{\mathbf{E}(\mathrm{e}^{-\lambda^*\tau_j})}{\mathbf{E}(\mathrm{e}^{-\lambda^*\tau_i})} .$$

From here we get the estimate

$$C_2 = 2\sum_{1\le i<j} \frac{\mathbf{E}(\mathrm{e}^{-2\lambda^*\tau_i})}{\mathbf{E}(\mathrm{e}^{-\lambda^*\tau_i})}\mathbf{E}(\mathrm{e}^{-\lambda^*\tau_j}) \le 2\left(\sum_i \frac{\mathbf{E}(\mathrm{e}^{-2\lambda^*\tau_i})}{\mathbf{E}(\mathrm{e}^{-\lambda^*\tau_i})}\right)\left(\sum_j \mathbf{E}(\mathrm{e}^{-\lambda^*\tau_j})\right) ,$$

where the second sum is just $\widehat{\rho}(\lambda^*) = 1$, while the first is

$$\sum_{n=1}^{\infty}\prod_{k=0}^{n-1}\frac{\lambda^*+w(k)}{2\lambda^*+w(k)} = \sum_{n=1}^{\infty}\prod_{k=0}^{n-1}\left(1-\frac{\lambda^*}{2\lambda^*+w(k)}\right).$$

So far this was all general, but under the specific assumptions (see (11)) on w, the final expression is finite. If $\gamma = 1$ then the logarithm of the product is $(-1+o(1))\lambda^*\log n + O(1)$, so for any $\varepsilon > 0$, the sum can be bounded by $O(1)+\sum n^{-(\lambda^*+\varepsilon)}$. This is finite, since $\lambda^* > \underline{\lambda} = 1$. If $0<\gamma<1$, then the sum can be bounded by $\sum \exp(-cn^{1-\gamma})$ for some $c > 0$.

This completes the proof of Lemma 2.

□

Lemma 3.

$$\mathbf{E}(\mathrm{e}^{-2\lambda^* t}Z_t^{\varphi}Z_t^{\psi}) \to C\widehat{\varphi}(\lambda^*)\,\widehat{\psi}(\lambda^*)$$

for some constant $C > 0$. *(This constant depends on the weight function* w.*)*

Proof. According to the basic decomposition, we can write

$$\begin{aligned} Z_t^{\varphi}Z_t^{\psi} &= \varphi(\Upsilon(t))\,\psi(\Upsilon(t)) \\ &+\varphi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\psi_j} + \psi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\varphi_j} \\ &+\sum_{j=1}^{\infty} Z_t^{\varphi_j}Z_t^{\psi_j} + \sum_{i\ne j}^{\infty} Z_t^{\varphi_i}Z_t^{\psi_j} . \end{aligned}$$

Taking expectation yields

$$\begin{aligned} m_t^{\varphi,\psi} := \mathbf{E}(Z_t^{\varphi}Z_t^{\psi}) &= \mathbf{E}\left(\varphi(\Upsilon(t))\,\psi(\Upsilon(t))\right) \\ &+\mathbf{E}\left(\varphi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\psi_j} + \psi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\varphi_j}\right) \\ &+\int_0^t m_{t-s}^{\varphi,\psi}\rho(s)\mathrm{d}s + \int_0^t\int_0^t m_{t-u}^{\varphi}m_{t-s}^{\psi}\rho_2(u,s)\mathrm{d}u\mathrm{d}s , \end{aligned} \qquad (20)$$

recall the notation $m_t^{\varphi} = \mathbf{E}(Z_t^{\varphi})$ from (16).

After multiplying the equation by $\mathrm{e}^{-2\lambda^* t}$, we can easily identify the limit (as $t\to\infty$) of the first, second and fourth terms in (20), as follows.

First term: since φ and ψ are bounded, $\lim_{t\to\infty} \mathrm{e}^{-2\lambda^* t}\mathbf{E}\left(\varphi(\Upsilon(t))\psi(\Upsilon(t))\right)$ is trivially 0.

Second term: let φ be bounded by the constant $D < \infty$, then

$$\mathrm{e}^{-2\lambda^* t}\mathbf{E}\left(\varphi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\psi_j}\right) \le D\mathrm{e}^{-2\lambda^* t}\int_0^t m_{t-s}^{\psi}\rho(s)\mathrm{d}s$$
$$= D\mathrm{e}^{-2\lambda^* t}\left(m_t^{\psi} - \mathbf{E}(\psi(\Upsilon(t)))\right) ,$$

the limit of which is 0 since m_t^{ψ} is of order $\mathrm{e}^{\lambda^* t}$ (see (17)), and since ψ is bounded.

Fourth term: Let us introduce $\tilde{m}_t^{\varphi} := \mathrm{e}^{-\lambda^* t}m_t^{\varphi}$, and $\tilde{m}_t^{\varphi,\psi} := \mathrm{e}^{-2\lambda^* t}m_t^{\varphi,\psi}$. By Lemma 1, $\lim_{t\to\infty}\tilde{m}_t^{\varphi} = d_{\varphi}$ and $\lim_{t\to\infty}\tilde{m}_t^{\psi} = d_{\psi}$. With these, and using Lemma 2,

$$\lim_{t\to\infty}\int_0^t\int_0^t \tilde{m}_{t-s}^{\varphi}\tilde{m}_{t-u}^{\psi}\mathrm{e}^{-\lambda^*(u+s)}\rho_2(u,s)\mathrm{d}u\mathrm{d}s = C_2 d_{\varphi}d_{\psi},$$

by dominated convergence.

This way we see

$$\tilde{m}_t^{\varphi,\psi} = \int_0^t \tilde{m}_{t-s}^{\varphi,\psi}\mathrm{e}^{-2\lambda^* s}\rho(s)\mathrm{d}s + C_2 d_{\varphi}d_{\psi} + \varepsilon_t , \tag{21}$$

where $\varepsilon_t \to 0$ as $t\to\infty$.

Now let us assume for a moment that the limit $d_{\varphi,\psi} := \lim_{t\to\infty}\tilde{m}_t^{\varphi,\psi}$ does exist. In this case dominated convergence could also be used in the third (normalized) term of (20), and the following would be true:

$$\lim_{t\to\infty}\int_0^t \mathrm{e}^{-2\lambda^* t}m_{t-s}^{\varphi,\psi}\rho(s)\mathrm{d}s = \lim_{t\to\infty}\int_0^t \tilde{m}_{t-s}^{\varphi,\psi}\mathrm{e}^{-2\lambda^* s}\rho(s)\mathrm{d}s = C_1 d_{\varphi,\psi} ,$$

recall the notation and result in Lemma 2.

This way *if* $\tilde{m}_t^{\varphi,\psi}$ was convergent, then its limit could only be

$$d_{\varphi,\psi} = \frac{C_2}{1-C_1}d_{\varphi}d_{\psi} ,$$

recall that $C_1 < 1$, by Lemma 2.

To show that the limit really exists, first note that $\tilde{m}_t^{\varphi,\psi}$ is bounded. This is true since with $M_t^{\varphi,\psi} := \sup_{s<t}\tilde{m}_s^{\varphi,\psi}$ and $M^{\varphi} := \sup_{s>0}\tilde{m}_s^{\varphi}$, we get

$$M_t^{\varphi,\psi} \le E + M_t^{\varphi,\psi}C_1 + M^{\varphi}M^{\psi}C_2 ,$$

where E is an upper bound for ε_t. This way $M_t^{\varphi,\psi}$ is bounded by a constant independent of t (again, $C_1 < 1$), thus $\tilde{m}_t^{\varphi,\psi}$ is bounded.

Let us introduce the difference of $\tilde{m}_t^{\varphi,\psi}$ and its supposed limit,

$$n_t := \tilde{m}_t^{\varphi,\psi} - \frac{C_2}{1-C_1}d_{\varphi}d_{\psi} \tag{22}$$

and rearrange equation (21),

$$n_t = \int_0^t n_{t-s} \mathrm{e}^{-2\lambda^* s} \rho(s) \mathrm{d}s + \bar{\varepsilon}_t \ ,$$

where $\bar{\varepsilon}_t \to 0$ as $t \to \infty$.

Since we have shown that $\tilde{m}_t^{\varphi,\psi}$ is bounded, so is n_t. Let $N_t := \sup_{s \geq t} |n_s|$, $\bar{E}_t := \sup_{s \geq t} |\bar{\varepsilon}_s|$, and fix arbitrarily $0 < u < t_0$. For these and for all $t > t_0$

$$\begin{aligned} |n_t| \leq |\bar{\varepsilon}_t| + & \left| \int_0^u n_s \mathrm{e}^{-2\lambda^*(t-s)} \rho(t-s) \mathrm{d}s \right| \\ & + \left| \int_u^t n_s \mathrm{e}^{-2\lambda^*(t-s)} \rho(t-s) \mathrm{d}s \right| \ . \end{aligned}$$

Recall that $\int_0^\infty \mathrm{e}^{-\lambda^* t} \rho(t) \mathrm{d}t = \widehat{\rho}(\lambda^*) = 1$ and $\int_0^\infty \mathrm{e}^{-2\lambda^* t} \rho(t) \mathrm{d}t = \widehat{\rho}(2\lambda^*) = C_1$, and thus

$$|n_t| \leq \bar{E}_{t_0} + \mathrm{e}^{-\lambda^*(t-u)} N_0 + N_u C_1 \ .$$

This way

$$N_{t_0} \leq \bar{E}_{t_0} + \mathrm{e}^{-\lambda^*(t_0-u)} N_0 + N_u C_1 \ .$$

Letting $t_0 \to \infty$ with u remaining fixed,

$$N_\infty \leq N_u C_1 \ ,$$

and now letting $u \to \infty$

$$N_\infty \leq N_\infty C_1 \ .$$

Since $C_1 < 1$ this means that $N_\infty = 0$, so $\tilde{m}_t^{\varphi,\psi}$ is convergent and its limit is

$$\lim_{t\to\infty} \tilde{m}_t^{\varphi,\psi} = \lim_{t\to\infty} \mathrm{e}^{-2\lambda^* t} \mathbf{E}(Z_t^\varphi Z_t^\psi) = \frac{C_2}{1 - C_1} d_\varphi d_\psi \ ,$$

as stated by the lemma. □

Now we are ready to prove Theorem 3.

Proof (Proof of Theorem 3).

Let $A_t := \mathrm{e}^{-\lambda^* t} Z_t^\varphi$ and $B_t := \mathrm{e}^{-\lambda^* t} Z_t^\psi$. Denote the limits of their expectations $a := \lim_{t\to\infty} \mathbf{E}(A_t)$ and $b := \lim_{t\to\infty} \mathbf{E}(B_t)$. From Lemma 3 we see that $\mathbf{E}(A_t B_t) \to Cab$, and also $\mathbf{E}(A_t^2) \to Ca^2$ and $\mathbf{E}(B_t^2) \to Cb^2$. This implies that

$$\mathbf{E}\left((bA_t - aB_t)^2\right) \to 0$$

so $(bA_t - aB_t) \to 0$ in L^2 and thus in probability, too.

Now fix any positive $\delta, \eta > 0$, then

$$\begin{aligned}
&\mathbf{P}\left(\left|\frac{A_t}{B_t}-\frac{a}{b}\right|>\delta\right)= \\
&\quad\mathbf{P}\left(\left\{\left|\frac{A_t}{B_t}-\frac{a}{b}\right|>\delta\right\}\cap\{B_t\geq\eta\}\right)+\mathbf{P}\left(\left\{\left|\frac{A_t}{B_t}-\frac{a}{b}\right|>\delta\right\}\cap\{B_t<\eta\}\right) \\
&\quad\leq\mathbf{P}\left(\left|bA_t-aB_t\right|>b\delta\eta\right)+\mathbf{P}(B_t<\eta).
\end{aligned}$$

Since the first term tends to 0 by the previous observation, it remains to show that in the limit, B_t does not have a positive mass at 0, and then the statement of the theorem is true.

But since $(B_t)_{t>0}$ is tight, in every subsequence there is a sub-subsequence $(t_n)_{n>0}$ along which B_{t_n} converges weakly to some random variable Y. By (15) for this variable, in distribution,

$$Y=\sum_{j=1}^{\infty}\mathrm{e}^{-\lambda^*\tau_j}Y_j\ ,$$

where the Y_j are iid with the same distribution as Y.

This means that

$$\mathbf{P}(Y=0)=\mathbf{P}(Y_j=0\text{ for all }\ j)=\lim_{k\to\infty}(\mathbf{P}(Y=0))^k\ .$$

It follows that if Y had a positive mass at 0, then Y would be a random variable that is almost surely 0. Since we know that its expectation tends to a positive limit, this could only happen if $\mathbf{E}(B_t^2)$ converged to ∞, but in fact it converges to a finite positive limit, according to Lemma 3. Thus, Y does not have a positive mass at 0, so the statement of Theorem 3 holds.

□

3 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 branching processes. We give a brief introduction to the fundamentals and state the theorems that we rely on. We do not give a broad survey on the most general types of branching processes here, we choose to focus on the results which may be applied to our process. For more details see the monograph [9] or the papers [10], [16], [18] and the references therein. For a survey on branching processes, trees and superprocesses, see [14].

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 $[0,t]$, this the random number of children an individual has up to time t. (In the case of our model, ξ is the Markovian pure birth process X.)

The individuals in the population are labelled with the elements of $\mathscr{N}$, the same way as described in Sect. 1.1. The basic probability space is

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

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

For each $x \in \mathscr{N}$ there is a $\cdot_{\downarrow x}$ 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 τ_x of the individuals are defined in the obvious way: $\tau_0 = 0$ and if $x' = xn$ with $n \in \mathbb{Z}_+$ then

$$\tau_{x'} = \tau_x + \inf\{t : \xi_x(t) \geq n\} \, , \tag{23}$$

just like in the random tree model, see (6).

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

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

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,

$$Z_t^{\Phi} := \sum_{x \in \mathscr{N}} \Phi_x(t - \tau_x) \tag{24}$$

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 which are 0 for $t < 0$ and equal to a bounded deterministic function of the rooted tree for $t \geq 0$.

This means that only those individuals contribute to Z_t^{Φ} which 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. [9], [10].) One of the important examples is $\Phi(t) = 1\!\!1\{t \geq 0\}$ when Z_t^{Φ} is just the total number of individuals born up to time t.

The Laplace-transform of $\mathrm{d}\xi(t)$ is of great importance, we denote this random variable by:

$$\widehat{\xi}(\lambda) := \int_0^{\infty} \mathrm{e}^{-\lambda t} \mathrm{d}\xi(t) \, . \tag{25}$$

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

$$\mathbf{E}\,\widehat{\xi}(\lambda^*) = 1 \, , \tag{26}$$

and also

$$\kappa = -\partial_\lambda \left(\mathbf{E}\,\widehat{\xi}(\lambda) \right)\Big|_{\lambda=\lambda^*} = \mathbf{E}\int_0^\infty t e^{-\lambda^* t} \mathrm{d}\xi(t) < \infty\,. \tag{27}$$

(The last property means that the process is Malthusian and the first means that it is supercritical.)

Also, we require the reproduction to be non-lattice, which means that the jumps of $\xi(t)$ cannot be supported by any lattice $\{0, d, 2d, \dots\}$, $d > 0$ with probability one.

We quote here a weaker form of Theorem 6.3 from [16], using its extension which appears in Section 7 of the same paper. This way the conditions of the original theorem are fulfilled automatically.

Theorem 4 (Nerman, [16]). *Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*, counted by two random characteristics $\Phi(t)$ and $\Psi(t)$ which have the properties described above (i.e. they are 0 for $t < 0$ and a deterministic bounded function of the progeny tree for $t \geq 0$). Suppose that there exists a $\underline{\lambda} < \lambda^*$ for which*

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

Then, almost surely

$$\frac{Z_t^\Phi}{Z_t^\Psi} \to \frac{\widehat{\Phi}(\lambda^*)}{\widehat{\Psi}(\lambda^*)} \quad \text{as } t \to \infty\,,$$

where $\widehat{\Phi}(\lambda) = \int_0^\infty \exp(-\lambda t)\mathbf{E}(\Phi(t))\mathrm{d}t$.

Remark. 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 $\mathbf{E}\,\widehat{\xi}(\lambda)$ for the branching process is the same as $\widehat{\rho}(\lambda)$, and the two definitions (12) and (27) for κ agree. This means that by condition (M) we may apply Theorem 4 with appropriate random characteristics. Given any bounded function $\varphi : \mathscr{G} \to \mathbb{R}$, 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 2.

We have already seen in Sect. 2.3, Lemma 1, that

$$\mathbf{E}(e^{-\lambda^* t} Z_t^\Phi) \to \frac{1}{\kappa}\hat{\Phi}(\lambda^*)\,. \tag{28}$$

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 [16].

Theorem 5 (Nerman, [16]).

Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*. Suppose that condition (M) holds and Φ is a random characteristic with properties described before. Then almost surely

$$\mathrm{e}^{-\lambda^* t} Z_t^{\Phi} \to \frac{1}{\kappa} \widehat{\Phi}(\lambda^*) W, \quad \text{as } t \to \infty \,, \tag{29}$$

where W is a random variable that does not depend on Φ.

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

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

We quote Theorem 5.3 of [10].

Theorem 6 (Jagers-Nerman, [10]). *Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*. If condition (L) holds then $W > 0$ a.s. and $\mathbf{E}(W) = 1$; otherwise $W = 0$ a.s.*

Remark. This theorem is the generalization of the Kesten-Stigum theorem, which states this fact for Galton-Watson processes (see [11]).

4 Global Properties

The questions discussed in Sect. 2 investigated local properties of the random tree: after a long time evolution, we asked about the neighborhood of a typical (uniformly selected) vertex.

When we want to look “too far away” from the random vertex though, the general theorem (Theorem 2) is of no use. What is the probability, for example, that the random vertex is descendant of the first, second, or k^{th} child of the root? These are the types of questions that we address.

The results in this section are from [19], we give sketches of the basic ideas and proofs.

4.1 Notation and Question

As we have seen in Sect. 3, the normalized size of the tree, $\exp(-\lambda^* t)|\Upsilon(t)|$, converges almost surely to a random variable. Throughout this section, we denote it by

$$\Theta := \lim_{t \to \infty} \mathrm{e}^{-\lambda^* t} |\Upsilon(t)| \,. \tag{30}$$

For every $x \in \mathscr{N}$, we introduce the variables Θ_x, corresponding to the growth of the subtree under x, analogously to Θ,

$$\Theta_x := \lim_{t \to \infty} \mathrm{e}^{-\lambda^* (t - \tau_x)} |\Upsilon_{\downarrow x}(t)| \,.$$

The letter Θ refers to the variable corresponding to the root, but when we want to emphasize this, we sometimes write $\Theta_\emptyset$ instead.

The most important relation between the different Θ_x variables in the tree is the so-called basic decomposition (according to the first generation in the tree):

$$\Theta_\emptyset = \lim_{t\to\infty} \mathrm{e}^{-\lambda^* t}|\Upsilon(t)| = \lim_{t\to\infty} \mathrm{e}^{-\lambda^* t}\left(1 + \sum_{k=1}^{\infty} |\Upsilon_{\downarrow k}(t)|\right) =$$
$$= \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda^*\tau_k}\left(\lim_{t\to\infty} \mathrm{e}^{-\lambda^*(t-\tau_k)}|\Upsilon_{\downarrow k}(t)|\right) = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda^*\tau_k}\Theta_k ,$$

or similarly for any x,

$$\Theta_x = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda^*(\tau_{xk}-\tau_x)}\Theta_{xk} . \tag{31}$$

Now we turn our attention to another random variable, which we denote by Δ_x. In the introduction of this section, we asked about the probability that for a fixed vertex x, after a long time, a randomly chosen vertex is descendant of x. This probability tends to an almost sure limit as $t \to \infty$,

$$\Delta_x := \lim_{t\to\infty} \frac{|\Upsilon_{\downarrow x}(t)|}{|\Upsilon(t)|} = \mathrm{e}^{-\lambda^*\tau_x} \lim_{t\to\infty} \frac{\mathrm{e}^{-\lambda^*(t-\tau_x)}|\Upsilon_{\downarrow x}(t)|}{\mathrm{e}^{-\lambda^* t}|\Upsilon(t)|} = \frac{\mathrm{e}^{-\lambda^*\tau_x}\Theta_x}{\Theta_\emptyset} . \tag{32}$$

Note that the value of Θ, obviously apparent in the continuous-time model, seems to disappear when we investigate the discrete-time evolution of the tree. The questions we naturally pose concern those properties which are observable also in the discrete-time setting. The variable Δ_x is obviously one of these observables since it can be written as the limit of a sequence of random variables from the discrete time model:

$$\Delta_x = \lim_{n\to\infty} \frac{|\Upsilon^d{}_{\downarrow x}(n)|}{|\Upsilon^d(n)|} .$$

The question is, does Θ really "disappear" in the problems concerning the discrete-time evolution? Recall that in Sect. 2, the main theorem discusses the behavior of ratios of the form Z_t^φ / Z_t^ψ, and since both the numerator and the denominator converges to constant times the common global variable, Θ itself cancels out. In this section we are interested whether Θ is merely an artificial side-product of the continuous-time embedding, or there exist properties observable in the discrete-time model, where Θ itself appears.

4.2 Markov Property

Definition 1. We say that a system of random variables $(Y_x)_{x\in\mathcal{N}}$ constitutes a *tree-indexed Markov field* if for any $x \in \mathcal{N}$, the distribution of the collection of variables $(Y_y : x \prec y)$, and that of $(Y_z : x \not\preceq z)$, are conditionally independent, given Y_x.

We state the following Lemma (recall that σ_x is the time we have to wait before vertex x is born, as counted from the birth time of his youngest brother, see Sect. 1.3).

Lemma 4. *For each $x \in \mathcal{N}$ let V_x denote the vector $V_x := (\sigma_x, \Theta_x)$. Then the collection of variables $\mathscr{A}_x := (V_y : x \prec y)$ and $\mathscr{B}_x := (V_z : x \not\preceq z;\, \sigma_x)$ are conditionally independent, given Θ_x.*

Proof. Recall (31), the decomposition of Θ_x according to the first generation of the subtree under x,

$$\Theta_x = \sum_{j=1}^{\infty} \mathrm{e}^{-\lambda^*(\tau_{xj}-\tau_x)}\Theta_{xj} = \sum_{j=1}^{\infty} \mathrm{e}^{-\lambda^*(\sigma_{x1}+\sigma_{x2}+\ldots+\sigma_{xj})}\Theta_{xj}\,. \tag{33}$$

Decompose the Θ_{xj} variables similarly, and so on, to arrive at the conclusion that Θ_x, as well as the whole collection $\mathscr{A}_x$, is in fact a function of the variables $(\sigma_y : x \prec y)$.

By the same argument, the collection $(V_z : x \not\preceq z;\, \sigma_x)$ is a function of the collection of variables $(\sigma_y : x \not\prec y;\, \Theta_x)$.

Now since for all $x \neq y$, σ_x and σ_y are independent, it is clear that $(\sigma_y : x \prec y)$ and $(\sigma_y : x \not\prec y)$ are independent sets of variables. Given Θ_x, the two collections $\mathscr{C}_x := (\sigma_y : x \prec y;\, \Theta_x)$, and $\mathscr{D}_x := (\sigma_y : x \not\prec y;\, \Theta_x)$ are conditionally independent. Since the variables in $\mathscr{A}$ are functions of $\mathscr{C}$ nad similarly, $\mathscr{B}$ is function of $\mathscr{D}$, the statement of the lemma follows.

□

Corollary 1. *The variables $(\Theta_x)_{x\in\mathcal{N}}$ constitute a tree-indexed Markov field.*

Proof. Direct consequence of Lemma 4, since $V_x = (\sigma_x, \Theta_x)$. □

It is clear that if $x \not\preceq x'$ and $x' \not\preceq x$, then Θ_x and $\Theta_{x'}$ are independent. Using Lemma 4, any moment can be computed in the non-independent case. We give formula for the covariance here for example:

Corollary 2. *Let $x' = xy$ for some $y \in \mathcal{N}$, then*

$$\mathbf{Cov}(\Theta_x, \Theta_{x'}) = \mathbf{E}(\mathrm{e}^{-\lambda^*\tau_y})\mathbf{Var}(\Theta).$$

Proof. With the notation $|y| = n$,

$$
\begin{aligned}
\mathbf{E}\left(\Theta_x\Theta_{xy}\right) &= \mathbf{E}\left(\Theta_{xy}\sum_{z:\,|z|=n} \mathrm{e}^{-\lambda^*(\tau_{xz}-\tau_x)}\Theta_{xz}\right)\\
&= \mathbf{E}\left(\mathrm{e}^{-\lambda^*(\tau_{xy}-\tau_x)}\right)\mathbf{E}\left(\Theta_{xy}^2\right)+\mathbf{E}\left(\Theta_{xy}\right)\sum_{z:\,|z|=n,z\neq y}\mathbf{E}\left(\mathrm{e}^{-\lambda^*(\tau_{xz}-\tau_x)}\right)\mathbf{E}\left(\Theta_{xz}\right)\\
&= \mathbf{E}\left(\mathrm{e}^{-\lambda^*\tau_y}\right)\mathbf{E}\left(\Theta^2\right)+\left(\mathbf{E}(\Theta)\right)^2\left(1-\mathbf{E}\left(\mathrm{e}^{-\lambda^*\tau_y}\right)\right)\\
&= \left(\mathbf{E}(\Theta)\right)^2+\mathbf{E}\left(\mathrm{e}^{-\lambda^*\tau_y}\right)\mathbf{Var}(\Theta)\ ,
\end{aligned}
$$

since by the results in Lemma 4, Θ_{xz} is independent of $(\tau_{xz}-\tau_x)$. Since $\mathbf{E}(\Theta_x)\mathbf{E}(\Theta_{x'}) = (\mathbf{E}(\Theta))^2$, the formula for the covariance follows. □

Let us introduce the following variables, indexed by $\mathscr{N}$. For the root let $R_\emptyset := 1$ and for any other vertex y' which has a parent y, so for any $y' = yk$ with $k \in \mathbb{N}_+$, let

$$
R_{yk} := \lim_{t\to\infty}\frac{|\Upsilon_{\downarrow yk}(t)|}{|\Upsilon_{\downarrow y}(t)|} = \frac{\mathrm{e}^{-\lambda^*(\tau_{yk}-\tau_y)}\Theta_{yk}}{\Theta_y} = \frac{\Delta_{yk}}{\Delta_y}\ . \tag{34}
$$

Notice that for $x = (i_1 i_2 \dots i_n)$, Δ_x is a telescopic product,

$$
\Delta_x = \Delta_{i_1}\frac{\Delta_{i_1 i_2}}{\Delta_{i_1}}\frac{\Delta_{i_1 i_2 i_3}}{\Delta_{i_1 i_2}}\cdots\frac{\Delta_{i_1\dots i_n}}{\Delta_{i_1\dots i_{n-1}}} = R_{i_1}R_{i_1 i_2}R_{i_1 i_2 i_3}\cdots R_{i_1\dots i_n}\ . \tag{35}
$$

This decomposition is of interest due to the following Theorem.

Theorem 7. *With the variables R_x defined as above, let $U_x := (R_x, \Theta_x)$. Given any sequence of positive integers $(i_n)_{n=1}^\infty$, the sequence of variables $U_\emptyset$, U_{i_1}, $U_{i_1 i_2}$, $U_{i_1 i_2 i_3}$, … constitutes a Markov chain, which is homogeneous: the transition probabilities from U_y to U_{yk} depend on k, but not on y.*

Proof. Let y, x, z be vertices in a progeny line, so let x be parent of z, and y be parent of x. Given Θ_x, then, from Corollary 1, Θ_y and Θ_z are conditionally independent. We show that so are R_y and R_z.

Consider that

$$
R_z = \frac{\mathrm{e}^{-\lambda^*(\tau_z-\tau_x)}\Theta_z}{\Theta_x}\ ,
$$

so R_z is a function of $\mathscr{A}_x$ (recall the notation in Lemma 4). At the same time,

$$
R_y = \frac{\mathrm{e}^{-\lambda^*(\tau_y-\tau_{y'})}\Theta_y}{\Theta_{y'}}\ ,
$$

where y' is the parent of y, so R_y is a function of Θ_y and the collection $(\sigma_v : x \not< v)$, which implies that R_y is a function of $\mathscr{B}_x$.

According to Lemma 4, $\mathscr{A}_x$ and $\mathscr{B}_x$ are conditionally independent, given Θ_x, thus the proof is complete.

□

4.3 Fragmentation of Mass

The Δ_x variables can be thought of as relative "masses" which the limiting tree "gives" to the individual vertices. For any fixed vertex x, Δ_x is the limit of the probability that a uniformly selected vertex, is descendant of x. If we thus look at all the x vertices in any fixed generation of the tree, the sum of the respective Δ_x values is obviously 1.

How can we describe this fragmentation of mass 1, from the root to generation one, and then to generation two, etc? We investigate this question in the following, in different cases of the choice of the weight function.

4.3.1 Linear Weight Function

Let us consider the case when the weight function is linear. That is, for some $\alpha \geq 0$ and $\beta > 0$, for all $k \in \mathbb{N}_+$, let the weight of a vertex of degree k be

$$w(k) = \alpha k + \beta \ .$$

The corresponding Malthusian parameter (solution of (10)), is now $\lambda^* = \alpha + \beta$. For the sake of computational simplicity it is convenient to re-scale w so that $\lambda^* = 1$, thus $\alpha = 1 - \beta$,

$$w(k) = (1-\beta)k + \beta \ , \tag{36}$$

where $0 < \beta \leq 1$. Note that $\beta = 1$ is allowed, and means that the weight function is constant, which corresponds to the Yule-tree model (see [22]). Also, for any integer $m > 1$, we allow the choice of $\beta = \frac{m}{m-1}$. In this case w linearly decreases, and it hits level zero at m, meaning that each vertex can have at most m children.

When a new vertex is added to the system, the sum of the weights in the tree increases by two terms, with $1-\beta$ because of the parent, and with β because of the new vertex. Thus, each time a new vertex is added, the total growth rate increases by 1, *independently* of the choice of the parent. This intuitively explains why the size of the tree grows exponentially in time, with parameter $\lambda^* = 1$.

The previous observation means that $N_t := |\Upsilon(t)|$ is a Markov process, which, at time t, increases by one with rate $N_t - 1 + \beta$. Thus it is straightforward to set up a partial differential equation for $f(u,t) := \mathbf{E}(\mathrm{e}^{-uN_t})$, which can be solved explicitly. By taking the limit $\lim_{t\to\infty} f(u\mathrm{e}^{-t}, t)$, one arrives at the conclusion that Θ has Gamma$(1,\beta)$ distribution.

The fact that the growth rate is independent of the structure of the tree, implies that anything that can be computed from the discrete time model, is independent of Θ. This is in accordance with the distribution of Θ being Gamma. To see this connection, consider for example that

$$\Delta_1 = \frac{e^{-\lambda^*\tau_1}\Theta_1}{\Theta_\emptyset} = \frac{e^{-\lambda^*\tau_1}\Theta_1}{\sum_{k=1}^{\infty} e^{-\lambda^*\tau_k}\Theta_k} = \frac{\Theta_1}{\Theta_1 + \sum_{k=2}^{\infty} e^{-\lambda^*(\tau_k-\tau_1)}\Theta_k}\,, \tag{37}$$

which shows that Δ_1 is a random variable of the form $\frac{X}{X+Y}$, where X and Y are independent. For the ratio to be independent of the denominator (thus Δ_1 to be independent of $\Theta_\emptyset$), X has to be of a Gamma distribution. This result is in accordance with the above considerations.

This all implies that in the linear case, Δ_x is the product of independent variables (see (35)), since now the Markov chain in Theorem 7 consists of independent elements.

From this observation it follows that Θ, according to the first generation, splits into the vector

$$\left(e^{-\lambda^*\tau_k}\Theta_k\right)_{k\in\mathbb{N}_+}\,,$$

which is of a Poisson-Dirichlet distribution. For a precise formulation of this fact, see [6].

4.3.2 Binary Tree

We now consider weight functions which ensure that each vertex can have two children at maximum. First let us investigate the distribution of Θ in this case, then we construct the system of the Δ_x variables.

Let $w(0) = a > 0$, $w(1) = 1$, and $w(k) = 0$ for $k \geq 2$. (We fix $w(1) = 1$ for the sake of computational simplicity, this is a different scaling from the one used in the linear case). The Malthusian parameter λ^* is now the positive solution of the equation

$$(\lambda^*)^2 + \lambda^* - a = 0\,. \tag{38}$$

Consider the basic decomposition of Θ,

$$\Theta = e^{-\lambda^*\sigma_1}\Theta_1 + e^{-\lambda^*(\sigma_1+\sigma_2)}\Theta_2\,. \tag{39}$$

Let the moment generating function be $\varphi(u) := \mathbf{E}(e^{-u\Theta})$. Using (39), computing $\varphi(u)$ in two steps, by first taking conditional expectation with σ_1 and σ_2 remaining fixed, then expectation with regards to Θ_1 and Θ_2, yields the integral equation

$$\varphi(u) = \int_0^\infty \int_0^\infty \varphi(ue^{-\lambda^* x})\varphi(ue^{-\lambda^* x}e^{-\lambda^* y})ae^{-ax}e^{-y}\,dy\,dx\,. \tag{40}$$

Now, with two changes of variables, and twice differentiating the equation, one arrives at the differential equation

$$\varphi''(u) = c\frac{\left(\varphi(u)^2 - \varphi(u)\right)}{u^2} - c\frac{\varphi'(u)}{u} + \frac{(\varphi'(u))^2}{\varphi(u)}\,, \tag{41}$$

where we introduced the shorthand notation

$$c := \frac{\lambda^* + 1}{\lambda^*}\,. \tag{42}$$

The boundary values are $\varphi(0) = 1$ and $\varphi'(0) = -\mathbf{E}(\Theta) = -\frac{1}{\lambda^*\kappa}$, where κ is easily computed (recall (12)),

$$\kappa = -\frac{\partial}{\partial\lambda}\left(\frac{a}{\lambda+a} + \frac{a}{\lambda+a}\frac{1}{\lambda+1}\right)\bigg|_{\lambda=\lambda^*}\,. \tag{43}$$

Introducing $g(u) = \log\varphi(u)$, equation (41) is equivalent to

$$g''(u) = cu^{-2}\left(\mathrm{e}^{g(u)} - 1\right) - cu^{-1}g'(u)\,. \tag{44}$$

We could have computed the moments of Θ already from (39), but (44) offers a simple method. From the series expansion of $\mathrm{e}^{g(u)}$,

$$g''(u) = c\frac{\mathrm{e}^{g(u)} - 1 - g'(u)u}{u^2} \to \frac{c}{2}(-g''(0) + (g'(0))^2), \quad \text{as } u \to 0\,.$$

This way $g''(0) = \frac{c}{2+c}(g'(0))^2$, and so $\mathbf{E}(\Theta^2) = (1 + \frac{c}{2+c})(\mathbf{E}(\Theta))^2$.

As for other moments of Θ, one can find a simple recursive formula for the derivatives of g. The derivation of this recursion is not illuminating, so we just state the result, namely,

$$g^{(k+2)} = c\frac{(\mathrm{e}^g)^{(k)}}{u^2} - (ck + k(k-1))\frac{g^{(k)}}{u^2} - (c+2k)\frac{g^{(k+1)}}{u}, \quad (\forall\, k \geq 1)\,,$$

from which all the values of $g^{(k)}(0)$ can be computed.

Now we turn our attention to the structure of the Δ_x system, meaning that we want to construct the Δ_x variables step by step, from generation to generation. For the rest of this section we treat the distribution of Θ as known.

Recall that σ_1, σ_2, Θ_1 and Θ_2 are independent. Thus, from

$$\Theta = \mathrm{e}^{-\lambda^*\sigma_1}(\Theta_1 + \mathrm{e}^{-\lambda^*\sigma_2}\Theta_2),$$

the conditional distribution of σ_1, given Θ, is straightforward to calculate. Then, given Θ and σ_1, the conditional distribution of Θ_1 follows. After this, given now Θ, σ_1 and Θ_1, the conditional distribution of σ_2 can be determined. Finally, if we know Θ, σ_1, Θ_1 and σ_2, then Θ_2 is a deterministic function of these.

Based on these considerations and on Theorem 7, now we can construct the system of the Δ_x variables in the following steps.

1. Pick $\Theta_\emptyset$ at random, according to its distribution.
2. *First generation*

   a. Pick (σ_1, Θ_1) according to their conditional distribution, given $\Theta_\emptyset$. These three numbers define $\Delta_1 = R_1 = \frac{\exp(-\lambda^* \sigma_1)\Theta_1}{\Theta}$.
   b. Pick (σ_2, Θ_2) similarly, according to their conditional distribution, given $\Theta_\emptyset$ and (σ_1, Θ_1). At this point we can compute $\Delta_2 = R_2 = \frac{\exp(-\lambda^*(\sigma_1+\sigma_2))\Theta_2}{\Theta_\emptyset}$.

3. *Second generation*

   a. Repeat the steps seen before for the progeny of vertex 1, to get R_{11} and R_{12}. Using the Markov property described in Theorem 7, this is done only using the information carried by Θ_1, conditionally independently of Θ and also of Θ_2. Since we already know R_1, we can now compute the values $\Delta_{11} = R_1 R_{11}$ and $\Delta_{12} = R_1 R_{12}$.
   b. Independently of the previous step, split Θ_2 to get R_{21} and R_{22}, thus also Δ_{21} and Δ_{22}.

4. *Subsequent generations* are constructed similarly.

Note that the special choice of $a = 2$ corresponds to the negative linear, binary case of Section 4.3.1 (take $m = 2$ there). Then $\lambda^* = 1$, $\kappa = \frac{1}{2}$, and $\mathbf{E}(\Theta) = 2$. It can be checked that in this case, equation (41) is solved by $\varphi(u) = (1+u)^{-2}$, so indeed, if $a = 2$, then Θ is of Gamma(1,2) distribution. Also, by checking that no other function of the form $(1+u)^{-\alpha}$ satisfies (41), it is verified that no other choice of a allows Θ to be of a Gamma distribution, so in all $a \neq 2$ cases the structure of the tree is indeed dependent on Θ.

These results are extendable, in a very similar way, to the case where the weight function allows the vertices to have at most m children for some finite $m > 2$. Then, an m^{th} order differential equation can be derived for the moment generating function, and again, any moment of Θ can be computed. The structure of variables Δ_x is also constructed analogously.

In [8], certain scaling exponents are derived explicitly, characterizing the decay of the distribution of Δ_x.

5 Acknowledgements

We thank the anonymous referee for the thorough work, and especially for his/her idea which shortened the proof of Lemma 2 considerably. We also thank Balázs Ráth and Imre Péter Tóth for stimulating discussions. The research work of the authors is partially supported by the following OTKA (Hungarian National Research Fund) grants: K 60708 (for A. R. and B. T.), TS 49835 (for A. R.).

References

1. David Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1(2):228–266, 1991.
2. Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
3. Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.
4. Béla Bollobás and Oliver M. Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34. Wiley-VCH, Weinheim, 2003.
5. Fan Chung, Shirin Handjani, and Doug Jungreis. Generalizations of Polya's urn problem. *Ann. Comb.*, 7(2):141–153, 2003.
6. Rui Dong, Christina Goldschmidt, and James B. Martin. Coagulation-fragmentation duality, Poisson-Dirichlet distributions and random recursive trees. *Ann. Appl. Probab.*, 16(4):1733–1750, 2006.
7. Richard Durrett. *Random Graph Dynamics*. Cambridge University Press, Cambridge, 2007. Cambridge Series in Statistical and Probabilistic Mathematics.
8. Emmanuel Guitter François David, Philippe Di Francesco and Thordur Jonsson. Mass distribution exponents for growing trees. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(02):P02011, 2007.
9. Peter Jagers. *Branching processes with biological applications*. Wiley-Interscience [John Wiley & Sons], London, 1975. Wiley Series in Probability and Mathematical Statistics — Applied Probability and Statistics.
10. Peter Jagers and Olle Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, 16(2):221–259, 1984.
11. H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Statist.*, 37:1211–1223, 1966.
12. P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63(6):066123, May 2001.
13. P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85(21):4629–4632, Nov 2000.
14. Jean-François Le Gall. Processus de branchement, arbres et superprocessus. In *Development of mathematics 1950–2000*, pages 763–793. Birkhäuser, Basel, 2000.
15. T. F. Móri. On random trees. *Studia Sci. Math. Hungar.*, 39(1-2):143–155, 2002.
16. Olle Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrsch. Verw. Gebiete*, 57(3):365–395, 1981.
17. Roberto Oliveira and Joel Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.*, 2(2):121–163, 2005.
18. Peter Olofsson. The $x \log x$ condition for general branching processes. *J. Appl. Probab.*, 35(3):537–544, 1998.
19. Anna Rudas. Global properties of a randomly growing tree. *Work in progress*.
20. Anna Rudas, Bálint Tóth, and Benedek Valkó. Random trees and general branching processes. *Random Struct. Algorithms*, 31(2):186–202, 2007.
21. Robert T. Smythe and Hosam M. Mahmoud. A survey of recursive trees. *Teor. Probab. Mat. Statist.*, (51):1–29, 1994.
22. G. Udny Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Royal Society of London Philosophical Transactions Series B*, 213:21–87, 1925.