Asymptotic behaviour of random growing trees

PhD thesis

ANNA RUDAS

Institute of Mathematics,
Budapest University of Technology and Economics

Supervisor:
Prof. Bálint Tóth

2012

Ezen értekezés bírálatai és a védésről készült jegyzőkönyv a későbbiekben a Budapesti Műszaki és Gazdaságtudományi Egyetem Természettudományi Karának Dékáni Hivatalában elérhető.

Alulírott Rudas Anna hozzájárulok a doktori értekezésem interneten történő korlátozás nélküli nyilvánosságra hozatalához.

Alulírott Rudas Anna kijelentem, hogy ezt a doktori értekezést magam készítettem, és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

.....
Rudas Anna

Contents

Acknowledgements

First of all I thank my supervisor, Prof. Bálint Tóth, for grabbing my attention with his lectures about probability theory and stochastic processes during my undergraduate studies, and for providing unwavering support and help later on, throughout the elongated period of my PhD years.

I also thank Benedek Valkó and Imre Péter Tóth, co-authors of the papers [41] and [42], for the enjoyable work together.

I gratefully acknowledge the comments and suggestions of Dr. Tamás F. Móri and Dr. András Telcs, who read through the whole thesis toroughly, pointed out mistakes and gave additional useful references.

Balázs Ráth is one of those friends without whom I would doubtedly have finished the present dissertation, his friendship helped me through much trouble since my early graduate years. My thanks go to many more of my friends who gave me continuous support.

I am thankful for the high quality of teaching with which my high school Mathematics teacher, Péterné Pósfai strengthened my interest and enthusiasm towards the field.

Finally I would like to express my gratefulness to my family, especially my parents and sister, for all three of them being with me all the way.

1 Introduction

1.1 The model class and context

The present PhD dissertation investigates the asymptotic properties of a random tree growth model which generalizes the basic concept of preferential attachment.

In this family of tree growth models, the tree stems from a root in the beginning, and vertices are added one at a time, the new vertex always attaching to exactly one already existing vertex. The rule by which the new vertex chooses its "parent", is dependent on the degree distribution apparent in the tree at the time the vertex is born. This dependence on the degree structure is characterised by a *weight function* $w : \mathbb{N} \to \mathbb{R}_+$, which function is the parameter of the model.

The models can be either in discrete time, when a vertex is born in every second, or in continuous time, then birth times are random. For the problems we discuss, these two versions are equivalent and can be translated into each other (details in Section 2.2). The classical models and results of the area use the discrete time setting. However, for the proofs we give, the continuous-time version is much more natural and convenient, so this is what we will use. The method of investigating the discrete time growth model by introducing the continuous-time setting described here, appears in [39], later in [41], and, independently, in Oliveira and Spencer [36]. An otherwise widely applied technique by the name of Poisson clumping, is based on the related, general idea of transforming a discrete time model to continuous time via the introduction of independent, exponentially distributed random times (see for example [1]).

One of the famous models, a realization of preferential attachment, is the Barabási - Albert graph [3], where the random choice of the parent for the new vertex is made using probabilities exactly proportional to the degree of the existing vertices. The tree case of this model corresponds to the the special case of the model considered in this dissertation, namely, when w is chosen to be linear. The Barabási - Albert graph reproduces certain phenomena observed in real-world networks, the power-law decay of the degree sequence, for example. This was proved in a mathematically precise way in Bollobás et al. [8] and, independently, in Móri [32]. Several more detailed results on the linear weight function case can be found in Móri [33], regarding the maximum degree, and in Móri [34], regarding the differences in the asymptotic degree distribution in the lower levels of the tree, compared to that in the whole tree. In these papers, the techniques 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 behaviour, see Bollobás and Riordan [9] and Chapter 4 of Durrett [17].

The concept of preferential attachment generally means that the weight function w is an increasing function. In the family of models that we are interested in, this is not necessarily true. General weight functions are considered in the papers of Krapivsky and Redner [25] and [26], where $w(k) \sim k^\gamma$, and non-rigorous results are obtained, showing the different behaviour 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 which is linked

to almost every other vertex, the others having only finite degree. This statement is made precise and proved rigorously in Oliveira and Spencer [36]. See also Chung et al. [10] for a related model. The weight functions we consider in the present dissertation are such that the model does not "blow up" this way, our class includes the second regime $\gamma \leq 1$ mentioned above.

In the work of Dereich and Mörters [11], the authors take a closer look at the temporal evolution of the degrees of individual vertices, in the same sublinear preferential attachment case as we do. This paper refers to our work [41]. Certain random recursive trees and random plane-oriented trees similar to our setting have also been studied before in Smythe and Mahmoud [44].

Population growth models, studied excessively in the theory of branching processes (see e.g. Jagers [22]), are intimately related to our model. This connection is the basis for many of the proofs in the present dissertation, as discussed in detail in Section 3.

Tree growth processes based on fragmentation processes are closely related to our investigation of the *global properties* of the model, as we will point out in Section 4 (see Remark 4.5 therein). Limiting objects called "random real trees" and "continuum random trees" were introduced, to which the evolving trees converge, after an appropriate rescaling of the *distances* on the tree. Much of the structure of these limiting objects is understood, see e.g. Haas, Miermont et al. [19, 20, 21].

Our concept of the limiting measure μ in Section 4 is different from these. It is a measure on the set of leaves of the infinite complete tree (with each vertex having exactly K children), which is a metric space, but the metric structure is trivial: it is not a result of any spatial scaling, and it carries no information about the tree growth process. On the other hand, the weights given by μ are a result of an appropriate rescaling of the tree size, where size means *cardinality*. In short, we are really interested in the asymptotic weight distribution, and not the asymptotic metric structure. This asymptotic weight distribution is also studied in the Physics literature, see e.g. Berestycki [5], where a quantity analogous to the local dimension is calculated for a continuous time fragmentation process.

Similarly, in the limiting continuous trees obtained in Haas, Miermont et al. [19, 20, 21] by a spatial rescaling of the evolving tree, the metric structure is of main interest, and the Hausdorff dimension and Hausdorff measure of *sets* are the natural questions to ask, see Duquesne and Le Gall [15, 16]. However, in our model it is not the set, but the measure which captures the long-term structure of the tree well, and of which the dimension is interesting.

The continuous time version of our tree growth process can also be translated into a branching random walk, with time turning into displacement. Then the asymptotic growth can be described analogously, see the Biggins theorem in [7] or Lyons [30]. However, with that point of view, the natural questions about the limiting structure are quite different.

In the last decades there has been much progress in describing the asymptotic structure of randomly evolving trees. As for a graph limit definition specifically related to our results, introduced by Benjamini and Schramm, see [4] and the remarks after the results

in Section 3.3. For a general, detailed survey of research on random graph dynamics, see the book of Durrett [17], the reference book of Drmota [14], and the yet unpublished work of van der Hofstad [46]. A very useful reference on the real world inspiraton for studying huge random graphs, on the way preferential attachment models fit into this research, and also on the limit objects for graph sequences, is Lovász [29].

1.2 Results, key methods, sources

The questions that we ask about the limiting behaviour of the random tree model are categorized into two classes. *Local properties* focus on the neighborhood of the *typical* vertex (e.g. sampled uniformly randomly after a long time) of the random tree, this is the subject of Section 3. *Global properties* capture phenomena observable by looking at the whole tree in the limit (e.g. asymptotic speed of tree growth in the continuous time setting, and the "limiting success level" of a fixed vertex in the limit), these are presented in Section 4.

Results.

Our main *local 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. These results are stated in Theorems 3.1 and 3.2. We also present Theorem 3.3, which is a weaker result than Theorem A in Nerman [35], but the proof uses more elementary methods and it is probabilistically instructive.

When turning to *global properties*, we first investigate the asymptotic speed of tree growth in the continuous time setting. We also ask questions about the limiting "success level" of some fixed vertex, and this leads us to the concept of a certain random measure μ on the leaves of the limiting tree, which captures a global property of the tree growth in a natural way.

We prove the following results.

1. The limiting entropies (as time tends to infinity) of the random measures on the different generations converge to a constant with probability one, as we let the generation level to infinity. This constant h is called the entropy of the limiting measure μ. This result is stated in Theorem 4.1.

2. The Hausdorff and the packing dimension of the random limiting measure μ are constant and equal with probability one. The entropy and the dimension satisfy the usual simple relation $dimension = \frac{entropy}{Ljapunov\,exponent}$. Moreover, the local dimension of μ equals the Hausdorff dimension at μ-almost every point. This result is stated in Theorem 4.2.

3. Given the so-called *weight function* w, which determines the rule of the tree growth, we provide an explicit formula for the entropy, and thus for the Hausdorff dimension, in terms of w. The computation is presented in Section 4.8, which constitutes the proof of Theorem 4.3.

Key methods.

The method regarding *local properties* is to embed the model into continuous time (details in Section 2.2). The greatest advantage of this setting is that it reveals the connection between the original, discrete time random tree model and the extensively studied framework of general branching processes (see Section 3.5). Our main *local results* gain their proofs through this relation, except for the proof of Theorem 3.3 which does not rely on the strong theorems in branching processes.

The proofs of our *global results* also contain the continuous time embedding, but the key element of those is a Markov process appearing naturally in the construction of a μ-typical leaf of the tree. After some discussion of the tree structure, the Markov property is easy to see. Some technical difficulties arise from the non-compactness of the state space.

Sources.

The numbered Theorems, Lemmas and Propositions, their proofs, together with most of the comments and introductory texts in the dissertation are from the following papers.

- [39] A. Rudas, Random tree growth with general weight function. Posted to `arxiv.org` in October 2004, unpublished. (The proof of Theorem 3.3 first appeared here, and was later published as Section 2.3 of [40].)
- [41] A. Rudas, B. Tóth, B. Valkó, Random Trees and General Branching Processes. Posted to `arxiv.org` and submitted to *Random Structures and Algorithms* in March 2005. Revised version accepted for publication and final version posted to `arxiv.org` in March 2006. Appeared in *Random Structures and Algorithms* online version in October 2006, printed version in 2007. (Theorems 3.1 and 3.2.)
- [40] A. Rudas, B. Tóth, Random tree growth with branching processes - a survey. Conference talk by A. Rudas at the workshop on Large-scale Random Graphs held in Budapest in August 2006. Final version submitted June 2008, published as Chapter 4 of *Handbook of Large-Scale Random Networks* in 2009. (Theorem 3.3 and the results in Section 4.3.)
- [42] A. Rudas, I. P. Tóth, Entropy and Hausdorff Dimension in Random Growing Trees. Posted to `arxiv.org` in April 2010, revised version in June 2011. Accepted for publication in *Stochastics and Dynamics* in December 2011. Proof reading received in June 2012, current status of final version is Online Ready on `www.worldscientific.com`. (Theorems 4.1, 4.2 and 4.3.)

2 Terminology, notation and the model

In the first subsection we introduce the (commonly known) terminology for rooted ordered trees, together with supplementary notation needed for our model, the statement of results and our proofs.

In the second subsection, we define two versions of the random tree model: embedded into discrete and into continuous time. We show in the end of Section 2.2.2 that the random tree evolving according to the continuous time setting, is equivalent to the Tree evolving in the discrete time setting, if we look at it at random stopping times.

Section 2.2 also contains some model specific notation, used extensively throughout the dissertation.

2.1 Vertices, individuals, trees

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, we will do so throughout the dissertation. 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).

Let us fix a subset of positive integers, $\mathbb{I}$, and let us label the vertices of a rooted ordered tree using the elements of

$$\mathcal{N} := \bigcup_{n=0}^{\infty} \mathbb{I}^n, \quad \text{where} \quad \mathbb{I}^0 := \{\emptyset\} \, .$$

We will consider slightly different cases of the model in Sections 3 and 4, and define $\mathbb{I}$ in the two sections accordingly, as follows.

- Throughout Section 3 we choose $\mathbb{I} = \mathbb{Z}^+$, this corresponds to the fact that any vertex can have any number of children.
- In Section 4, except for a short analysis of the linear weight function case in Section 4.3.1, we will fix a positive integer $K \in \mathbb{N}$, and choose $\mathbb{I} := \{1, 2, \ldots, K\}$. This means that starting from Section 4.4, we restrict the weight function in such a way that the vertices can have at most K number of children. We assume $K \geq 2$ to avoid the trivial case when only one child is born per parent. (In that case the tree growth is linear and the tree has no interesting structure.)

In our notation $\emptyset$ denotes the root of the tree, its children are labelled with the elements of $\mathbb{I}$, and in general the children of $x = (x_1, x_2, \ldots, x_k) \in \mathcal{N}$ are labelled 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 shorthand 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 $n \in \mathbb{I}$ we use xn for $(x_1, x_2, \ldots, x_k, n)$.

There is a natural partial ordering $\prec$ on $\mathcal{N}$, namely, $x \prec z$ if x is ancestor of z, so if $\exists y \in \mathcal{N}$, $y \neq \emptyset$ such that $z = xy$. We use $x \preceq z$ meaning $x \prec z$ or $x = z$.

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 the number of its children in G, so this terminology differs a little bit from the usual:

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

The n^{th} *generation* of $G \in \mathcal{G}$ is

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

where $|x| = n$ iff $x \in \mathbb{I}^n$.

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 $k = n$. In Section 4 we will also use the notation $p(x) = x^1$ for the parent of x.

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

$$G_{\downarrow x} := \{y : xy \in G\}, \tag{1}$$

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 $G_{\downarrow x^k}$.

2.2 The random tree model

As the parameter of the random tree model, we fix a weight function $w : \mathbb{N} \to \mathbb{R}_+$.

For the definition of the discrete time model, we do not need any further restrictions on w. In the continuous time case, we impose certain restrictions on w, see (M), these are needed for the model definition, and also for our results in Section 3. In Section 4, we will require $w(k) = 0, \quad k \geq K$, which will on one hand make sure that each vertex can have at most K children, and on the other hand, it automatically implies condition (M).

2.2.1 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

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

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 $w(\deg(x, G))$.

2.2.2 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

$$\rho(t) = \lim_{\varepsilon \to 0} \varepsilon^{-1} \mathbf{P}((t, t + \varepsilon) \text{ contains a point from } X)\ . \tag{2}$$

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

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

The rightmost expression of $\widehat{\rho}(\lambda)$ 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

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

Throughout the dissertation we impose the following condition on the weight function w:

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

Remark. At certain Sections of the dissertation, we will restrict ourselves to certain smaller subclasses of weight functions, to be specified later. The condition described above is nevertheless always fulfilled.

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 3.5) for which the related statement is derived in [22].

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 towards the newly added vertex) to the already existing vertex $x \in G$ with probability

$$\frac{w(\deg(x, G))}{\sum_{y \in G} w(\deg(y, G))}. \tag{4}$$

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.

We have already introduced the notation $G_{\downarrow x}$ for the *subtree* rooted at x of the tree G in (1). In the random tree model, we will use the two forms

$$\Upsilon_{\downarrow x}(t) = (\Upsilon(t))_{\downarrow x} \tag{6}$$

exchangeably throughout the dissertation. This is the subtree of $\Upsilon(t)$ rooted at x, which is the set of descendants of x (including x) that are born up to time t. Note that t here is total time, and not the time since birth of x. In particular, $|\Upsilon_{\downarrow x}(0)| = 0$ if x is not the root.

Throughout the dissertation, we will use τ_x to denote the birth time of vertex x,

$$\tau_x := \inf\{t > 0 \ : \ x \in \Upsilon(t)\} \ . \tag{7}$$

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 him). Namely, let

(a) $\sigma_\emptyset := 0$,

(b) $\sigma_{y1} := \tau_{y1} - \tau_y$, for any $y \in \mathcal{N}$,

(c) and $\sigma_{yi} := \tau_{yi} - \tau_{y(i-1)}$, for each $y \in \mathcal{N}$ and $i \geq 2$, $i \in \mathbb{I}$.

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 $\eta_1 = 1$.

3 Local properties

3.1 Introduction

In the present Section we investigate the *local properties* of the random tree after a long time of its evolution. We ask questions about the neighborhood of the "typical" vertex (e.g. sampled uniformly randomly) of the random tree, after a long time.

This Section relies on the paper [41], joint work with Bálint Tóth and Benedek Valkó, and on the survey [40], joint work with Bálint Tóth.

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 [2]. 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, as already introduced in Section 2.2. Even without referring to the well-developed theory of general branching processes, this makes it possible to give an argument for the convergence of certain general ratios, see Theorem 3.3 in Section 3.4. This theorem has its stronger analogue, Theorem A in Section 3.5, but our proof presented here is different from that in [35], and although this approach does not give almost sure convergence, it is elementary and instructive, and gives convergence in probability.

The greatest advantage of the continuous time setting is definitely that it reveals the connection between the original, discrete time random tree model and the extensively studied framework of general branching processes (see Section 3.5). Our main *local* results gain their proofs through this relation. As an earlier application of a similar idea, see the paper [38] of B. Pittel, in which the author establishes the connection with a Crump–Mode branching process, and proves his results about the height of the uniform and general ordered recursive tree, and also for a random m-ary search tree.

The present Section is organized as follows. We introduce some additional notation about historical birth orderings, rooted ordered trees with a marked vertex at generation k, distributions on these sets, the steady property of probability measures on $\mathcal{G}$, and backward extensions of measures, needed specifically for understanding the "local properties" of the random tree, in Section 3.2. After this, we state the main results of this part of the thesis in Section 3.3. In the same Section we spend some time to present the excplicit calculations that can be done in the special case when the weight function is linear, in Section 3.3.1. Section 3.4 is devoted to the proof of convergence in probability, not referring to the methods used in the theory of general branching processes. Then we give a brief introduction to the field of general branching processes in Section 3.5, and state the relevant results. The last subsection, 3.6 contains the proofs of the main *local* results, Theorems 3.1 and 3.2.

3.2 Notation

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 $0 \le i \le |G| - 1$ 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

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

give the evolution of G in this historical ordering s.

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

$$\zeta, \cdots \in \mathcal{N}, \qquad \Gamma, \cdots \in \mathcal{G}$$

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 k^{th} 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 k^{th} ancestor of a random vertex: G is an ordered tree rooted in the k^{th} 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 k^{th} ancestor, then for $0 \le l \le k$ the progeny of the l^{th} ancestor is described by $(G_{\downarrow u^l}, u_{\downarrow k-l})$.

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

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

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

Without presenting the precise formulation, it is clear that a consistent sequence $\boldsymbol{\pi}^{(k)}$ of probability measures on $\mathcal{G}^{(k)}$ gives full insight to the limiting structure of the tree as seen from a random vertex, see Remarks 3.1, 3.2 and 3.3 after Theorem 3.2.

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

$$\sum_{H \in \mathcal{G}} \boldsymbol{\pi}(H) \sum_{x \in H_{[1]}} \mathbb{1}\{H_{\downarrow x} = G\} = \boldsymbol{\pi}(G). \tag{8}$$

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

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

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

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

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

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

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

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

3.3 Results

From condition (M) it follows that the equation

$$\widehat{\rho}(\lambda) = 1 \tag{9}$$

has a unique root λ^*.

Now we are ready to state our first theorem.

Theorem 3.1. *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:*

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

From Theorem 3.1 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 k^{th} 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 3.1 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

$$W(G, s, i) := W(G(s, i)) \tag{10}$$

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

$$w(G, s, i) := w\left(\deg\left((s_i)^1, G(s, i-1)\right)\right). \tag{11}$$

for $1 \leq i \leq |G| - 1$. 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 $G(s, i)$.

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

$$\begin{aligned} \boldsymbol{p}_w(k) &:= \frac{\lambda^*}{\lambda^* + w(k)} \prod_{i=0}^{k-1} \frac{w(i)}{\lambda^* + w(i)}, \\ \boldsymbol{\pi}_w(G) &:= \sum_{s \in \mathcal{S}(G)} \frac{\lambda^*}{\lambda^* + W(G)} \prod_{i=0}^{|G|-2} \frac{w(G, s, i+1)}{\lambda^* + W(G, s, i)}. \end{aligned}$$

Theorem 3.2. *Consider a weight function w which satisfies condition* (M) *and let λ^* be defined as before. Then the following limits hold almost surely:*

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

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

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

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

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

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

Furthermore, the functions $\boldsymbol{p}_w, \boldsymbol{\pi}_w$ are probability distributions on $\mathbb{N}$ and $\mathcal{G}$, respectively, and $\boldsymbol{\pi}_w$ is steady (i.e. identity (8) *holds).*

Remark 3.1. *Parts (a), (b) and (c) of Theorem 3.2, 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 k^{th} ancestor of the randomly selected vertex with the position of this vertex marked.*

Remark 3.2. *From part (c) it is easy to derive the asymptotic distribution of the progeny of the k^{th} ancestor of the randomly selected vertex (as a rooted ordered tree without any marked vertices):*

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

The limit is the size-biased version of $\boldsymbol{\pi}_w(G)$, with the biasing done by the size of the k^{th} generation.

Remark 3.3. *Since the distribution $\boldsymbol{\pi}_w$ 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 $\zeta^1, \zeta^2, \zeta^3, \ldots$ '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 $\boldsymbol{\pi}_w^{(k)}$ on $\mathcal{G}^{(k)}$ where $\boldsymbol{\pi}_w^{(k)}(G,u) = \boldsymbol{\pi}_w(G)$.

A related graph limit definition had been introduced by Benjamini and Schramm, see [4].

3.3.1 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 $\boldsymbol{p}_w$ (computed in [8], [32]) is reproduced, of course. But even in this simplest case the asymptotic distribution $\boldsymbol{\pi}_w$ 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, $\widehat{\rho}(\lambda) = \frac{\beta}{\lambda - 1}$, $\underline{\lambda} = 1$ and $\lambda^* = 1 + \beta$. Thus both Theorems 3.1 and 3.2 hold.

For the asymptotic degree distribution we get

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

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 $\boldsymbol{\pi}_w(G)$ 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 $s \in \mathcal{S}(G)$

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

(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

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

In the $\beta = 1$ case (i.e. if we consider random tree proposed in [3]) the previous calculations give

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

and

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

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 dissertation):

$$B(x) := \{y \in G : y = (x_1, x_2, \ldots, x_{n-1}, k), k > x_n\},$$
$$a(x) := \max(|G_{\downarrow x}| - 1, 1), \qquad b(x) := \max\left(\textstyle\sum_{y\in B(x)} |G_{\downarrow x}|\,, 1\right).$$

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 a simple exercise and is left to the reader of the dissertation.

3.4 Convergence in probability

In this section we present a proof of the convergence in probability of certain ratios of variables, see Theorem 3.3 below. The result formulated here exists in a stronger form, namely, the convergence holds in the almost sure sense, as stated by Theorem A in Section 3.5. 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), see Section 2.2.2 to a somewhat smaller, but still very wide class. We demand in this section that

$$w(k) = k^\gamma + v(k) \tag{12}$$

with some $0 < \gamma \leq 1$ and $v(k) = o(k^\gamma)$ as $k \to \infty$, which implies that the weight function varies regularly. Note that this way $w(k) \to \infty$ as $k \to \infty$, but monotonicity for w is not required.

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 (9), and with the letters Φ and Ψ we denote positive bounded functions $\Phi, \Psi : \mathcal{G} \to \mathbb{R}$.

Define

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

(the analogous definition in Section 3.5 is (26)). We use the notation

$$\kappa := -\partial_\lambda \widehat{\rho}(\lambda)\Big|_{\lambda=\lambda^*} = \int_0^\infty t e^{-\lambda^* t} \rho(t)\, \mathrm{d}t < \infty\,. \tag{13}$$

(the analogous definition in Section 3.5 is (27)). We also introduce the notation

$$\widehat{\Phi}(\lambda) := \int_0^\infty e^{-\lambda s} \mathbf{E}\left(\Phi(\Upsilon(s))\right)\, \mathrm{d}s\,. \tag{14}$$

Theorem 3.3. *Let w satisfy condition* (12)*. Then*

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

Remark 3.4. *An analogous theorem is valid for any w satisfying* (M)*, and convergence holds in the almost sure sense, see Theorem 3.5 in Section 3.5.*

To prove Theorem 3.3 we need Lemmas 3.1, 3.2 and 3.3 below.

Lemma 3.1.

$$\mathbf{E}(e^{-\lambda^* t} Z_t^\Phi) \to \frac{1}{\kappa}\widehat{\Phi}(\lambda^*) =: d_\Phi, \quad \text{as } t \to \infty. \tag{15}$$

Remark 3.5. *See Section 3.5 and Theorem 3.5 therein for the analogous, more general result. There we see that $e^{-\lambda^* t} Z_t^\Phi$ itself converges almost surely to a random variable with the appropriate expectation.*

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

$$Z_t^\Phi = \Phi(\Upsilon(t)) + \sum_{j \in \mathbb{N}} Z_{t-\tau_j}^{\Phi_j} \,, \tag{16}$$

where j runs over the children of the root, $\Phi_j(G) := \Phi(G_{\downarrow j})$, 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 $(\tau_j)_{j \in \mathbb{Z}_+}$, $Z_t^{\Phi_j}$ has the same conditional distribution as $Z_{t-\tau_j}^\Phi$.

At this point observe that if Z_t^Φ is of some exponential order $e^{\lambda t}$, then λ must be the one defined by equation (9). This can be seen if we take expectation of both sides in equation (16), supposing that $\lim_{t\to\infty} e^{-\lambda t} Z_t^\Phi$ exists almost surely (and is a non-zero, finite random variable, with finite expectation), we can write

$$\mathbf{E}(\lim_{t\to\infty} e^{-\lambda t} Z_t^\Phi) = \sum_{j\in\mathbb{N}} \mathbf{E}\left(e^{-\lambda\tau_j} \lim_{t\to\infty} e^{-\lambda(t-\tau_j)} Z_{t-\tau_j}^{\Phi_j}\right)$$

$$= \sum_{j\in\mathbb{N}} \mathbf{E}\left(e^{-\lambda\tau_j}\right) \mathbf{E}\left(\lim_{t\to\infty} e^{-\lambda(t-\tau_j)} Z_{t-\tau_j}^{\Phi_j}\right) = \mathbf{E}\left(\sum_{j\in\mathbb{N}} e^{-\lambda\tau_j}\right) \mathbf{E}\left(\lim_{t\to\infty} e^{-\lambda t} Z_t^\Phi\right) \,,$$

since $\lim_{t\to\infty} e^{-\lambda(t-\tau_j)} Z_{t-\tau_j}^{\Phi_j} \stackrel{\mathrm{d}}{=} \lim_{t\to\infty} e^{-\lambda t} Z_t^\Phi$.

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

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

must hold (compare with (9)).

For the convergence itself, using the notation

$$m_t^\Phi := \mathbf{E}\left(Z_t^\Phi\right) \,, \tag{17}$$

taking expectation on both sides of (16) in two steps (first conditionally on $(\tau_j)_{j\in\mathbb{Z}_+}$, then taking expectation regarding $(\tau_j)_{j\in\mathbb{Z}_+}$), we get

$$m_t^\Phi = \mathbf{E}(\Phi(\Upsilon(t))) + \int_0^t m_{t-s}^\Phi \rho(s)\,\mathrm{d}s \,. \tag{18}$$

Taking the Laplace transform of both sides, we have

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

so formally

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

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

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

so, indeed, the statement of the lemma holds.

□

Recall the notation in Section 2.2.2, the birth times of the vertices in the first generation of the tree, $(\tau_j)_{j>0}$, constitute the point process X. Similarly to the notation for the density function ρ, let us denote the second correlation function by ρ_2, namely, for $u \neq s$, let

$$\rho_2(u,s) := \lim_{\varepsilon,\delta\to 0} (\varepsilon\delta)^{-1}\mathbf{P}\left((u,u+\varepsilon) \text{ and } (s,s+\delta) \text{ both contain a point from } X\right) \ , \quad (19)$$

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

The following estimates are needed.

Lemma 3.2. *Suppose that w satisfies the conditions described in the beginning of Section 3.4. Then*

(a)

$$C_1 := \int_0^\infty e^{-2\lambda^* s}\rho(s)\,\mathrm{d}s < 1 \ , \quad (20)$$

(b)

$$C_2 := \int_0^\infty \int_0^\infty e^{-\lambda^*(u+s)}\rho_2(u,s)\,\mathrm{d}u\,\mathrm{d}s < \infty \ . \quad (21)$$

Proof. The first statement is obvious, considering that

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

since $\widehat{\rho}$ strictly decreases and $\lambda^* > 0$. As for statement (21), write C_2 as follows:

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

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

$$\mathbf{E}(e^{-\lambda^*\tau_j}) = \mathbf{E}(e^{-\lambda^*\tau_i}e^{-\lambda^*(\tau_j-\tau_i)}) = \mathbf{E}(e^{-\lambda^*\tau_i})\mathbf{E}(e^{-\lambda^*(\tau_j-\tau_i)}),$$

It now follows that

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

From here we get the estimate

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

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 (12)) on w, the final expression is finite, as follows. If $\gamma = 1$ then the logarithm of the product is $-\lambda^*\log n + o(\log n)$. This means that the sum is $\sum n^{-(\lambda^* - g(n))}$ for some $g(n) \to 0$, thus the sum is finite, since λ^* is strictly greater than 1, by $\lambda^* > \underline{\lambda} = 1$. 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 3.2.

□

Lemma 3.3.

$$\mathbf{E}(e^{-2\lambda^* t} Z_t^\Phi Z_t^\Psi) \to C\widehat{\Phi}(\lambda^*)\widehat{\Psi}(\lambda^*)$$

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

Proof. According to the basic decomposition, we can write

$$\begin{aligned} Z_t^\Phi Z_t^\Psi &= \Phi(\Upsilon(t))\Psi(\Upsilon(t)) \\ &\quad + \Phi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\Psi_j} + \Psi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\Phi_j} \\ &\quad + \sum_{j=1}^{\infty} Z_t^{\Phi_j} Z_t^{\Psi_j} + \sum_{i\ne j}^{\infty} Z_t^{\Phi_i} Z_t^{\Psi_j}\ . \end{aligned}$$

Taking expectation yields

$$\begin{aligned} m_t^{\Phi,\Psi} := \mathbf{E}(Z_t^\Phi Z_t^\Psi) &= \mathbf{E}\left(\Phi(\Upsilon(t))\Psi(\Upsilon(t))\right) \\ &\quad + \mathbf{E}\left(\Phi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\Psi_j} + \Psi(\Upsilon(t))\sum_{j=1}^{\infty} Z_t^{\Phi_j}\right) \\ &\quad + \int_0^t m_{t-s}^{\Phi,\Psi}\rho(s)\,\mathrm{d}s + \int_0^t\int_0^t m_{t-u}^{\Phi} m_{t-s}^{\Psi}\rho_2(u,s)\,\mathrm{d}u\,\mathrm{d}s\ , \end{aligned} \tag{22}$$

recall the notation $m_t^\Phi = \mathbf{E}(Z_t^\Phi)$ from (17).

After multiplying the equation by $e^{-2\lambda^* t}$, we can easily identify the limit (as $t \to \infty$) of the first, second and fourth terms in (22), as follows.

First term: since Φ and Ψ are bounded, $\lim_{t\to\infty} e^{-2\lambda^* t}\mathbf{E}\left(\Phi(\Upsilon(t))\Psi(\Upsilon(t))\right)$ is trivially 0.

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

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

the limit of which is 0 since m_t^Ψ is of order $e^{\lambda^* t}$ (see (18)), and since Ψ is bounded.

Fourth term: Let us introduce $\tilde{m}_t^\Phi := e^{-\lambda^* t} m_t^\Phi$, and $\tilde{m}_t^{\Phi,\Psi} := e^{-2\lambda^* t} m_t^{\Phi,\Psi}$. By Lemma 3.1, $\lim_{t\to\infty} \tilde{m}_t^\Phi = d_\Phi$ and $\lim_{t\to\infty} \tilde{m}_t^\Psi = d_\Psi$. With these, and using Lemma 3.2,

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

by dominated convergence.

This way we see

$$\tilde{m}_t^{\Phi,\Psi} = \int_0^t \tilde{m}_{t-s}^{\Phi,\Psi} e^{-2\lambda^* s}\rho(s)\,\mathrm{d}s + C_2 d_\Phi d_\Psi + \varepsilon_t , \tag{23}$$

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

Now let us assume for a moment that the limit $d_{\Phi,\Psi} := \lim_{t\to\infty} \tilde{m}_t^{\Phi,\Psi}$ does exist, and is finite. In this case dominated convergence could also be used in the third (normalized) term of (22), and the following would be true:

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

recall the notation and result in Lemma 3.2.

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

$$d_{\Phi,\Psi} = \frac{C_2}{1-C_1} d_\Phi d_\Psi ,$$

recall that $C_1 < 1$, by Lemma 3.2.

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

$$M_t^{\Phi,\Psi} \le E + M_t^{\Phi,\Psi} C_1 + M^\Phi M^\Psi C_2 ,$$

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

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

$$n_t := \tilde{m}_t^{\Phi,\Psi} - \frac{C_2}{1-C_1} d_\Phi d_\Psi \tag{24}$$

and rearrange equation (23),

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

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

Since we have shown that $\tilde{m}_t^{\Phi,\Psi}$ 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 $t > t_0$

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

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

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

This way

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

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 $\tilde{m}_t^{\Phi,\Psi}$ is convergent and its limit is

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

as stated by the lemma. □

Now we are ready to prove Theorem 3.3.

Proof of Theorem 3.3. Let $A_t := e^{-\lambda^* t} Z_t^\Phi$ and $B_t := e^{-\lambda^* t} Z_t^\Psi$. 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 3.3 we see that $\mathbf{E}(A_t B_t) \to \tilde{C}ab$, and also $\mathbf{E}(A_t^2) \to \tilde{C}a^2$ and $\mathbf{E}(B_t^2) \to \tilde{C}b^2$, for some constant $\tilde{C} > 0$. 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, being bounded in L_1, 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 (16) for this variable, in distribution,

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

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 3.3. Thus, Y does not have a positive mass at 0, so the statement of Theorem 3.3 holds.

□

3.5 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 which may be applied to our process. For more details see the monograph [22] or the papers [23], [35], [37] and the references therein. For a survey on branching processes, trees and superprocesses, see [28].

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.

The individuals in the population are labelled with the elements of $\mathcal{N}$, as described in Section 2.1 (see (2.1)). 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 τ_x of the individuals are defined in the obvious way: $\tau_0 = 0$ and if $x' = xn$ with $n \in \mathbb{Z}_+$ then

$$\tau_{x'} = \tau_x + \inf\{t : \xi_x(t) \geq n\}. \tag{25}$$

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

$$\Phi_x(t, \omega) = \Phi(t, \omega_{\downarrow x}), \tag{26}$$

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,

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

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 which 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 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. [22], [23].) 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 $\mathrm{d}\xi(t)$ will be of great importance, we denote this random variable by:

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

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

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

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 [35], 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 (Nerman, [35])**.** *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 $\widehat{\Phi}(\lambda) = \int_0^\infty \exp(-\lambda t) \mathbf{E}(\Phi(t)) \, \mathrm{d}t$.

In Theorems 3.1 and 3.2, we determined the asymptotic ratio of vertices in $\Upsilon(t)$ satisfying certain properties. The proofs of these Theorems rely on Theorem A above.

It is also natural to ask questions about the asymptotic number of the respective vertices, as we will do so in Section 4. This essentially requires to study the asymptotic behavior of Z_t^Φ for a suitable random characteristic Φ. Here we give two of the results in the framework of general branching processes that relate to this subject, and we will refer to these Theorems in Section 4.

As we have seen in Section 3.4, Lemma 3.1,

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

thus we need to divide Z_t^Φ by $e^{\lambda^* t}$ to get something non-trivial. We quote here a weaker form of Theorem 5.4 of [35].

Theorem B (Nerman, [35])**.** *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*

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

where $\tilde{\Theta}$ is a random variable that does not depend on Φ.

The necessary and sufficient condition for the random variable $\tilde{\Theta}$ 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 [23].

Theorem C (Jagers-Nerman, [23])**.** *Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*. If condition (L) holds then $\tilde{\Theta} > 0$ a.s. and $\mathbf{E}(\tilde{\Theta}) = 1$; otherwise $\tilde{\Theta} = 0$ a.s.*

Remark 3.6. *This theorem is the generalization of the Kesten-Stigum theorem, which states this fact for Galton-Watson processes (see [24]).*

We do not intend to identify the necessary and sufficient condition on the weight function w in the random tree model which would guarantee that the corresponding reproduction process possesses property (L), still we will give certain sufficient conditions for Theorem B and Theorem C in Section 4. There, we will also derive results for the random variable $\tilde{\Theta}$, in the case of a special choice of Φ.

3.6 Proofs of results

Proof of Theorem 3.1. 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 2.2.

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 $\widehat{\rho}(\lambda)$ 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 3.1. □

Proof of Theorem 3.2. (a) Apply Theorem 3.1 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 τ_k (see (25)):

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

Since

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

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

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

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

Note that since

$$\boldsymbol{p}_w(k) = \prod_{i=0}^{k-1} \frac{w(i)}{\lambda^* + w(i)} - \prod_{i=0}^{k} \frac{w(i)}{\lambda^* + w(i)}$$

is a telescopic sum adding up to 1, $\boldsymbol{p}_w$ is indeed a probability distribution on $\mathbb{N}$.

(b) Let $G \in \mathcal{G}$ be fixed and denote $n = |G| - 1$. We apply Theorem 3.1 with $\varphi(H) = \mathbb{1}(H = G)$. We need to compute

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

Consider the following random stopping times:

$$\begin{aligned} \tau_G &:= \sup\{t \geq 0 : \Upsilon(t) \subseteq G\}, \\ \tau'_G &:= \sup\{t \geq 0 : \Upsilon(t) \subsetneq G\}. \end{aligned}$$

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

$$\begin{aligned} \lambda^* \int_0^\infty e^{-\lambda^* t}\, \mathbf{P}\big(\Upsilon(t) = G\big)\, \mathrm{d}t &= \mathbf{E}\big(e^{-\lambda^* \tau'_G} - e^{-\lambda^* \tau_G}\big) \\ &= \mathbf{E}\big((e^{-\lambda^* \tau'_G} - e^{-\lambda^* \tau_G}) \mathbb{1}\{\tau'_G < \tau_G\}\big). \end{aligned}$$

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

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

This implies that

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

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

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

(See (4), (10) and (11) for the definitions.) Also, if $s \in \mathcal{S}(G)$ is fixed then *conditionally on the event* $\{(\eta_0,\dots,\eta_n) = (s_0,\dots,s_n)\}$ the random variables $T_{k+1} - T_k$, $k = 0,1,\dots,n-1$, are independent and exponentially distributed with parameters $W(G,s,k)$, $k = 0,1,\dots,n-1$, 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 $\nu_1,\nu_2,\dots,\nu_l$ is also exponentially distributed with parameter $\sum_{i=1}^{l}\nu_i$. Hence it is straightforward to get

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

Collecting our previous calculations part (b) of Theorem 3.2 follows.

Using the identity

$$\boldsymbol{\pi}_w(G) = \lambda^* \int_0^\infty e^{-\lambda^* t}\,\mathbf{P}\big(\Upsilon(t) = G\big)\,\mathrm{d}t,$$

and the fact that $|\Upsilon(t)|$ is finite for every t with probability 1 it is straightforward to prove that $\boldsymbol{\pi}_w$ is indeed 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

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

The statement now follows from part (b).

The only thing left to prove is that $\boldsymbol{\pi}_w$ satisfies (8), 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 (8) is linear in $\boldsymbol{\pi}$, 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 $\boldsymbol{\pi}_w$ and from this an easy consideration shows that $\boldsymbol{\pi}_w$ must satisfy (8). □

4 Global properties

4.1 Introduction

The results in Section 3 focus on the *local properties* of the random tree, namely, they give results concerning the neighbourhood of a uniformly random vertex, which is chosen from the tree after a long time of tree evolution. In this Section we concentrate on *global properties* of the limiting tree.

One of the first *global* questions about the continuous time model is the speed of the tree growth. We have already seen results that show that the size of the tree at time t is of order $\exp(\lambda^* t)$. There are several results about the limit distribution, as $t \to \infty$, if we scale by this order. Moreover, in certain subclasses of the weight function, the momentums of the distribution can be calculated, as we will see in Section 4.3.

It is also natural to pose the following question. Let us fix a vertex, say the first vertex in the first generation (first child of the root). What is the "limiting success level" of this vertex, compared to the other vertices in the same generation? What we mean by this is the number of descendants of this vertex, after a long time of tree evolution, compared to the number of descendants of its brothers.

Another formulation of the same question is to fix a vertex, let the tree grow for a long time, then choose a vertex uniformly at random from the big tree, and ask the probability that this random vertex is descendant of the fixed vertex. Clearly, if we look at these limiting probabilities for let us say the first generation, we get a distribution, itself being random, that codes an important information of the evolution of the tree.

If one looks at the system of these limiting (as time evolution of the tree tends to infinity) random distributions on the different generations of the tree, it is tempting to ask something about the limiting measure of this system, when letting the generation level tend to infinity. We will define the above concepts properly, and will denote this overall limiting measure by μ.

Having a random measure in our hand, which describes a global property of the limiting infinite system, it is natural to ask about the *Hausdorff (and packing) dimension of this measure*. On the other hand, the dimension of the measure depends on a parameter of the underlying metric, which is arbitrary. To rule out this (trivial) dependence, it is usual to ask about the *entropy of the limiting measure*, which depends on the growth process only. This is the natural equivalent of the dimension from a dynamical point of view.

The key to our results is a Markov process appearing naturally in the construction of a μ-typical leaf of the tree. After some discussion of the tree structure, the Markov property will be easy to see. Some technical difficulties will arise from the non-compactness of the state space.

The model choice is special in the sense that we only allow a finite degree for each vertex, but it is general in the sense that after having fixed the maximum number of children K a vertex may have, the weight function w, which determines the rule of attachment, can be any positive-valued function on $\{0, 1, \ldots, K-1\}$.

This Section relies on the paper [42], joint work with Imre Péter Tóth, and on the survey [40], joint work with Bálint Tóth.

Structure of the present section

The present Section is structured as follows. We introduce some additional notation, and also repeat some facts from Section 3 that we need, in Section 4.2. After that, we focus on the asymptotic growth of the tree, with the appropriate scaling, in Section 4.3, first recalling some earlier results and giving general arguments, then going into detailed calculations about the momentums of the distribution in two special choices of the weight funtion. Namely, in Section 4.3.1 we investigate the linear case, and then in Section 4.3.2 we turn our attention to the class of models with Bounded Degree. After this, we will keep this model and stick to it throughout the whole section.

In Section 4.4 we identify random measures on the finite generation levels of the tree in the $t \to \infty$ limit, and then take the limit of these, as the generation level tends to infinity. This way, a random measure is defined on the *leaves of the infinite tree*, in Section 4.4.1, and this measure is the limiting object that we are interested in. For the sake of the dissertation to be self contained, we briefly review the definitions of local dimension, Hausdorff dimension and packing dimension of measures, in Section 4.4.2.

After all this, we are ready to state our results in Section 4.5. The proofs are structured into three subsections. In Section 4.6 we give the main line of the proofs of Theorems 4.1 and 4.2, but delay the proofs of certain auxiliary lemmas to Section 4.7, and then we conclude with the calculation of the entropy in Section 4.8, which constitutes the proof of Theorem 4.3.

4.2 Notation and model choice

Except for a short subsection (see Section 4.3.1), we will restrict our weight function to the following class of functions.

We fix a positive integer $K > 2$, and we require the weight function to be zero above K:

$$w(k) = 0, \quad k \geq K. \tag{30}$$

As is clear from the model definition, see Section 2.2, this restriction makes sure that any vertex can have at most K children. The vertices of the random tree are labeled with elements of

$$\mathcal{N} := \bigcup_{n=0}^{\infty} \mathbb{I}^n,$$

as in (2.1), but now with the choice of $\mathbb{I} = \{1, 2, \ldots, K\}$. Since we require (30), the weight function automatically fulfils (M), the condition stated in Section 2.2.2.

The same way as in Section 2.2.2, let τ_x be the birth time of vertex x, and σ_x 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 him). For the precise definition see (7).

Recall also the function $\widehat{\rho}\colon (0,\infty) \to (0,\infty)$ defined in (3) (see Section 2.2.2),

$$\widehat{\rho}(\lambda) := \mathbf{E}\sum_{j=1}^{K} e^{-\lambda\tau_j} = \sum_{j=1}^{K}\prod_{i=0}^{j-1}\frac{w(i)}{\lambda + w(i)}\ .$$

Recall that the reason for the notation $\widehat{\rho}$ is that this function is the Laplace transform of the density of the point process formed by birth times in the first generation of the tree, see (3) in Section 2.2.2.

In the present section we use some of the results in Section 3, we list them here.

1. The equation
$$\widehat{\rho}(\lambda) = 1$$
has a unique root $\lambda^* > 0$, the Malthusian parameter.

2. This λ^* gives the rate of exponential growth of the tree size almost surely. The normalized size of the tree converges almost surely to a random variable, which we denote by
$$\Theta := \lim_{t\to\infty} e^{-\lambda^* t}|\Upsilon(t)|\ . \tag{31}$$
Remark 4.1. *See Section 4.3 for results on the distribution of* Θ.

3. Θ is almost surely positive, and
$$0 < \mathbf{E}\Theta < \infty,$$
which implies (also) that almost surely the process $\Upsilon(t)$ does not blow up in finite time.

4. Moreover,
$$\mathbf{E}\Theta^2 < \infty. \tag{32}$$

The first statement is in our setting obvious from the definition, since we have assumed $2 \le K < \infty$. The second and third follow from Theorem B and Theorem C in Section 3.5, we will also refer to it in Section 4.3. The last statement is also implicit from Section 3.5, the calculation for the variance is presented in Section 4.3.2. Alternatively, the finiteness of the variance follows from Theorem 6.8.1 in [22], which states L^2 convergence of the normalized size under the condition $\mathbf{E}[(\sum_{i=1}^{K} e^{-\lambda^*\tau_i})^2] < \infty$, which is again obvious, since $K < \infty$.

Remark 4.2. *The process* $\Upsilon(t)$ *has an alternative construction, which we state here and refer to later. Define a countably infinite number of independent random variables* $\tilde{\sigma}_x$*, indexed with the elements of* $\mathcal{N}$*, as follows. Let* $\tilde{\sigma}_\emptyset = 0$*, and for* $x = i_1 i_2 \dots i_n$*, let* $\tilde{\sigma}_x$ *be exponentially distributed with parameter* $w(i_n - 1)$*. Denoting the parent of* x *by* $p(x)$*, we define* $\tilde{\tau}_\emptyset = 0$ *and*

$$\tilde{\tau}_x = \tilde{\tau}_{p(x)} + \tilde{\sigma}_{p(x)1} + \tilde{\sigma}_{p(x)2} + \ldots + \tilde{\sigma}_{p(x)i_n}.$$

It is straightforward that with $\tilde{\Upsilon}(t) := \{x \in \mathcal{N} : \tilde{\tau}_x \le t\}$*, the process* $\tilde{\Upsilon}$ *has the same distribution as* Υ.

4.3 Asymptotic growth

The most natural question, already posed in Section 4.2 (see (31)) is how fast the tree itself is growing, in the continuous time setting. In this subsection we derive results regarding the distribution of Θ, in various choices of the weight function w.

Let us repeat here the two conditions (M) from Section 2.2.2, and (L) from Section 3.5, as well as Theorem B and Theorem C from Section 3.5. These are needed to derive results on the limiting random variable Θ. Conditions M and L on the weight function w are

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

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

Theorem B (Nerman, [35])**.** *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*

$$e^{-\lambda^* t} Z_t^{\Phi} \to \frac{1}{\kappa} \widehat{\Phi}(\lambda^*) \tilde{\Theta} =: \Theta \,, \quad \text{as } t \to \infty \,, \tag{33}$$

where $\tilde{\Theta}$ is a random variable that does not depend on Φ.

Theorem C (Jagers-Nerman, [23])**.** *Consider a supercritical, Malthusian branching process with Malthusian parameter λ^*. If condition (L) holds then $\tilde{\Theta} > 0$ a.s. and $\mathbf{E}(\tilde{\Theta}) = 1$; otherwise $\tilde{\Theta} = 0$ a.s.*

If we choose the random characteristic as $\Phi_x(t) = \mathbb{1}\{\tau_x < t\}$, then $Z_t^{\Phi} = |\Upsilon(t)|$. Theorem B applies 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, thus by Theorem C, Θ is a.s. positive.

Lemma 4.1. *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 $\widehat{\xi}(\lambda^*)$ from which condition (L) trivially follows. Since $\widehat{\xi}(\lambda^*) = \sum_{i=1}^{\infty} e^{-\lambda^* \tau_k}$, we need

$$\mathbf{E}\left(\sum_{k=1}^{\infty} e^{-\lambda^* \tau_k} \right)^2 < \infty. \tag{34}$$

The random variables $\tau_{k+1} - \tau_k$ are independent exponentials for $k = 0, 1, 2, \ldots$ with parameters $w(0), w(1), \ldots$, respectively, thus a simple computation yields that the expression in (34) is equal to

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

Transforming the double sum on the right, we get

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

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

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

Let λ_1 be an arbitrary value for which $\underline{\lambda}<\lambda_1<\lambda^*$. Since $w(l)\to\infty$, we have

$$\frac{(\lambda_1+w(l))(\lambda^*+w(l))}{w(l)(2\lambda^*+w(l))}<1$$

if l is large enough. This leads to

$$\sum_{j=0}^{i}\prod_{l=0}^{j}\frac{\lambda^*+w(l)}{2\lambda^*+w(l)} < C\sum_{j=0}^{i}\prod_{l=0}^{j}\frac{w(l)}{\lambda_1+w(l)} < C\widehat{\rho}(\lambda_1)<\infty,$$

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 $\phi(u)=\mathbf{E}e^{-u\Theta}$. Using the idea of the basic decomposition (16) one can write the following equation for $f(u,t):=\mathbf{E}e^{-u|\Upsilon(t)|}$

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

By Theorem B,

$$\phi(u)=\lim_{t\to\infty}f(ue^{-\lambda^* t},t) \tag{35}$$

which gives

$$\phi(u)=\mathbf{E}\prod_{j\ge 1}\phi(ue^{-\lambda^*\tau_j}).$$

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 [22].)

4.3.1 Linear weight function

In this subsection we forget about the restriction that the degrees of the vertices are maximized by K. Let us consider instead the case when the weight function is linear. That is, for some $\alpha\ge 0$ 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 (9)), 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 [45]). 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 $f(u, t) := \mathbf{E}(e^{-uN_t})$, which can be solved explicitly. By taking the limit $\lim_{t\to\infty} f(ue^{-t}, t)$, one arrives at the conclusion that Θ has Gamma$(1, \beta)$ distribution.

For a further note on the linear weight function case, see Section 4.6.3. In order to state that remark though, we will first need some more notation, introduced in Section 4.4.

4.3.2 Trees with bounded degree

We now turn back to the case introduced in Section 4.2, we consider weight functions which ensure that each vertex can have a certain number of children at maximum, denoted by K. From now on, throughout Section 4, we will restrict our investigations to this subclass of models (the only, brief exception being the note on the linear case in Section 4.6.3).

In the present subsection, we derive the first and second moments of Θ in the case of binary trees ($K = 2$). The details of the computation are somewhat cumbersome, even in this simplest case, so we only sketch the method. By some extra work, following and expanding the main steps described here, the interested reader can, theoretically, compute any specific, finite moment of θ recursively. Also, the reader gets a quick insight to the method by which one can derive the moments of Θ in the case of greater, finite K values as well. The main motivation for this subsection is to provide the reader with the basic ideas of the steps of this computation.

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

$$(\lambda^*)^2 + \lambda^* - a = 0 \ . \tag{37}$$

Consider the basic decomposition of Θ,

$$\Theta = e^{-\lambda^*\sigma_1}\Theta_1 + e^{-\lambda^*(\sigma_1+\sigma_2)}\Theta_2 \; . \tag{38}$$

Let the moment generating function be $\varphi(u) := \mathbf{E}(e^{-u\Theta})$. Using (38), 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

$$\varphi(u) = \int_0^\infty \int_0^\infty \varphi(ue^{-\lambda^* x})\varphi(ue^{-\lambda^* x}e^{-\lambda^* y})ae^{-ax}e^{-y}\,\mathrm{d}y\,\mathrm{d}x \; . \tag{39}$$

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

$$\varphi''(u) = c\frac{(\varphi(u)^2 - \varphi(u))}{u^2} - c\frac{\varphi'(u)}{u} + \frac{(\varphi'(u))^2}{\varphi(u)} \; , \tag{40}$$

where we introduced the shorthand notation

$$c := \frac{\lambda^* + 1}{\lambda^*} \; . \tag{41}$$

The boundary values are $\varphi(0) = 1$ and $\varphi'(0) = -\mathbf{E}(\Theta) = -\frac{1}{\lambda^*\kappa}$, where κ is easily computed (recall (27)),

$$\kappa = -\frac{\partial}{\partial\lambda}\left(\frac{a}{\lambda + a} + \frac{a}{\lambda + a}\frac{1}{\lambda + 1}\right)\Bigg|_{\lambda=\lambda^*} \; . \tag{42}$$

Introducing $g(u) = \log\varphi(u)$, equation (40) is equivalent to

$$g''(u) = cu^{-2}\left(e^{g(u)} - 1\right) - cu^{-1}g'(u) \; . \tag{43}$$

We could have computed the moments of Θ already from (38), but (43) offers a simple method. From the series expansion of $e^{g(u)}$,

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

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,

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

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

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 (40) 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 (40), 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.

In the case where the weight function allows the vertices to have at most K children for some finite $K > 2$, these reasonings are very similar. Then, a K^{th} order differential equation can be derived for the moment generating function, and again, theoretically, any moment of Θ can be computed.

4.4 Limiting objects

For every $x \in \mathcal{N}$, we introduce the variables Θ_x, corresponding to the growth of the subtree under x. The definition is analogous to Θ in (31), recall also the notation (6) in Section 2.1 for subtrees of the random tree,

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

(Θ refers to $\Theta_\emptyset$, the variable corresponding to the root). Clearly, for every $x \in \mathcal{N}$, the random variables Θ_x are identically distributed. The basic relation between the different Θ_x variables in the tree is that for any $x \in \mathcal{N}$,

$$\Theta_x = \sum_{i=1}^{K} e^{-\lambda^*(\tau_{xi}-\tau_x)}\Theta_{xi} \ ,$$

which is straightforward from $\left|\Upsilon_{\downarrow x}(t)\right| = 1 + \sum_{i=1}^{K} \left|\Upsilon_{\downarrow xi}(t)\right|$.

Now let us ask the following question. Fix a vertex $x \in \mathcal{N}$, and at time t, draw a vertex ζ_t uniformly randomly from $\Upsilon(t)$. What is the probability that ζ_t is descendant of x, so $x \prec \zeta_t$? As shown in (44) below, this probability tends to an almost sure limit Δ_x as $t \to \infty$, which can be expressed using the τ and Θ random variables,

$$\Delta_x := \lim_{t\to\infty} \frac{\left|\Upsilon_{\downarrow x}(t)\right|}{\left|\Upsilon(t)\right|} = e^{-\lambda^*\tau_x} \lim_{t\to\infty} \frac{e^{-\lambda^*(t-\tau_x)}\left|\Upsilon_{\downarrow x}(t)\right|}{e^{-\lambda^* t}\left|\Upsilon(t)\right|} = \frac{e^{-\lambda^*\tau_x}\Theta_x}{\Theta_\emptyset} \ . \tag{44}$$

We can now, for any $n \in \mathbb{N}$, define a random measure μ_n on the finite set $\mathbb{I}^n = \{x \ : |x| = n\}$, the n^{th} generation of the full tree, by

$$\mu_n(\{x\}) := \Delta_x \ .$$

This is a probability measure almost surely, which follows from the facts $\Delta_\emptyset = 1$ and $\Delta_y = \sum_{i=1}^{K} \Delta_{yi}$.

Let H_n denote the entropy of μ_n, that is

$$H_n = -\sum_{|x|=n} \Delta_x \log \Delta_x \ .$$

4.4.1 A measure as the limiting object for the tree

Let $\partial\mathcal{N}$ denote the set of leaves of the complete tree: $\partial\mathcal{N} = \{1, 2, \ldots, K\}^\infty$. The concatenation xy makes sense for $x \in \mathcal{N}$ and $y \in \partial\mathcal{N}$, and then $xy \in \partial\mathcal{N}$. Also, for $x \in \mathcal{N}$ and $z \in \partial\mathcal{N}$, we write $x \prec z$ if $\exists y \in \partial\mathcal{N}$ such that $z = xy$. For $x \in \mathcal{N}$ we denote the set of leaves under x by $\partial\mathcal{N}(x) = \{z \in \partial\mathcal{N} \, : \, x \prec z\}$.

Let $x_{|l}$ denote the first l letters of the string x, or in different words, let it denote the ancestor of x on the l-th level of the tree, and let $\partial\mathcal{N}$ be equipped with the usual metric

$$d(x, y) = \Lambda^{\max\{n\in\mathbb{N} \, : \, x_{|n}=y_{|n}\}} \, , \tag{45}$$

where $0 < \Lambda < 1$ is an arbitrary constant. This constant is often chosen to be $1/e$, which makes certain formulae appear simpler. Yet we will not fix the value, so that our formulae express the dependence of the studied quantities on this arbitrary choice.

With the help of the μ_n random limiting measures, we define μ on the cylinder sets $\partial\mathcal{N}(x)$ of $\partial\mathcal{N}$ by

$$\mu(\partial\mathcal{N}(x)) := \mu_n(\{x\}) = \Delta_x \, , \text{ if } |x| = n \, ,$$

and then we extend μ from $\{\partial\mathcal{N}(x) \, : \, x \in \mathcal{N}\}$ to the sigma-algebra generated (on $\partial\mathcal{N}$). Our results concern the properties of this extended random measure μ.

Remark 4.3. *The limiting relative weights Δ_x defined in (44) make sense and are interesting in both the discrete and the continuous time setting of the random tree model, just like the measure μ and the entropy H_n. The limiting "absolute" weights Θ_x however, while playing a central role in our proofs, only make sense in the continuous time setting. This is the reason why we work with the continuous time setting in this Section of the thesis.*

4.4.2 Dimensions of measures: definitions

For the reader's convenience, let us review the definitions of local dimension, Hausdorff dimension and packing dimension of measures. The lower and upper local dimensions of μ at x are defined in [18] (2.15) and (2.16) as

$$\begin{aligned} \underline{\dim}_{\mathrm{loc}}\mu(x) &= \liminf_{r\to 0} \frac{\log \mu(B(x,r))}{\log r}, \\ \overline{\dim}_{\mathrm{loc}}\mu(x) &= \limsup_{r\to 0} \frac{\log \mu(B(x,r))}{\log r}, \end{aligned}$$

where $B(x, r)$ is the ball of radius r centred at x. If the lower and upper local dimensions coincide at some x, they are called the local dimension at x. The Hausdorff and packing dimensions of μ are defined in [18] (10.8) and (10.9) as

$$\begin{aligned} \dim_{\mathrm{H}} \mu &= \sup\{s : \underline{\dim}_{\mathrm{loc}}\mu(x) \geq s \text{ for } \mu\text{-almost all } x\}, \\ \dim_{\mathrm{P}} \mu &= \sup\{s : \overline{\dim}_{\mathrm{loc}}\mu(x) \geq s \text{ for } \mu\text{-almost all } x\}. \end{aligned}$$

The name of these dimensions come from the fact ([18] (10.10) and (10.11)) that

$$\begin{aligned}\dim_{\mathrm{H}}\mu &= \inf\{\dim_{\mathrm{H}} E:\ E \text{ is a Borel set with } \mu(E)>0\},\\ \dim_{\mathrm{P}}\mu &= \inf\{\dim_{\mathrm{P}} E:\ E \text{ is a Borel set with } \mu(E)>0\}.\end{aligned}$$

We are ready to state our results.

4.5 Results

Theorem 4.1. *The limiting entropy*

$$h := \lim_{n\to\infty} \frac{1}{n} H_n$$

exists and is constant with probability one.

Theorem 4.2. *The Hausdorff dimension* $\dim_H \mu$ *and the packing dimension* $\dim_P \mu$ *of the measure* μ *are constant and equal with probability one, and* h *and the dimensions satisfy the relation*

$$\dim_H \mu = \dim_P \mu = \frac{h}{-\log \Lambda},$$

where Λ *is from (45). Moreover, the local dimension of* μ *equals* $\dim_H \mu = \dim_P \mu$ *at* μ*-almost every point.*

Theorem 4.3. *Furthermore, an explicit formula for* h *is given:*

$$h = \mathbf{E}\left(\sum_{i=1}^{K} \lambda^* \tau_i e^{-\lambda^* \tau_i}\right).$$

This can be computed given the weight function w*.*

4.6 Proofs of Theorems 4.1 and 4.2

4.6.1 Idea of the proof

The random limiting measure μ depends on the random growth of the tree. The idea of the proof is the following: we define a random leaf in the limiting tree according to the measure μ. The way the random leaf is defined is based on a step-by-step construction of the subsequent generations of the limiting tree, together with a step-by-step construction of a path from the root to the random leaf. This is done in such a way that a Markov process appears naturally along this path, and the local dimension of the measure μ in this random point can be computed as an ergodic average. It follows that this average is constant with probability one, unconditionally. Thus, although the measure depends on the random tree growth, this ergodic average is constant, and it is the local dimension of the measure in all the μ-typical leafs of the limiting tree. This implies that this constant is the Hausdorff (and also the packing) dimension of μ with probability one. Some technical difficulty comes from the fact that the state space of the key Markov process is continuous and non-compact, so to apply ergodic theorems, one has to work for the existence of the invariant measure (while uniqueness is easy).

4.6.2 Markov structure of the tree

The content of this short section is mainly repetition of material from [40]. These concepts and statements allow for a good understanding of the tree structure, on which our main construction (in Section 4.6.4) relies. Lemma 4.2 will also be used formally in Section 4.6.4 to get an easy proof of the fact that our step-by-step construction of the limiting tree is equivalent to the original model (Proposition 4.2).

Definition 4.1. *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 4.2. *For each $x \in \mathcal{N}$ let V_x denote the vector $V_x := (\sigma_x, \Theta_x)$. Then the collections of variables $\mathcal{A}_x := (V_y : x \prec y)$ and $\mathcal{B}_x := (V_z : x \not\preceq z; \sigma_x)$ are conditionally independent, given Θ_x.*

Proof. Recall Remark 4.2, the alternative construction of $\Upsilon(t)$. From that, it is straightforward that the collection $\mathcal{A}_x$ is in fact constructed by the set of independent variables $A_x := (\sigma_y : x \prec y)$.

Similarly, recall (4.4), and decompose $\Theta_{p(x)}$, where $p(x)$ is the parent of vertex x,

$$\Theta_{p(x)} = \sum_{j=1}^{K} e^{-\lambda^*(\tau_{p(x)j}-\tau_{p(x)})}\Theta_{p(x)j} = \sum_{j=1}^{K} e^{-\lambda^*(\sigma_{p(x)1}+\sigma_{p(x)2}+\cdots+\sigma_{p(x)j})}\Theta_{p(x)j} \ .$$

This means that if we take the set of variables $B_x := (\sigma_y : x \not\prec y)$, then $\mathcal{B}_x$ is constructed by $B_x \cup \{\Theta_x\}$.

Given Θ_x, the two collections $A_x \cup \{\Theta_x\}$ and $B_x \cup \{\Theta_x\}$ are conditionally independent, this way the same is true for $\mathcal{A}_x$ and $\mathcal{B}_x$, so the statement of the lemma follows.

□

Corollary 4.1. *The variables $(\Theta_x)_{x\in\mathcal{N}}$ constitute a tree-indexed Markov field.*

Proof. Direct consequence of Lemma 4.2, 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 4.2, any moment can be computed in the non-independent case. We give formula for the covariance here for example:

Corollary 4.2. *Let $x' = xy$ for some $y \in \mathcal{N}$, then*

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

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

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

since by the results in Lemma 4.2, Θ_{xz} is independent of $(\tau_{xz} - \tau_x)$. Since $\mathbf{E}(\Theta_x)\mathbf{E}(\Theta_{x'}) = \left(\mathbf{E}(\Theta)\right)^2$, the formula for the covariance follows. □

Definition 4.2. *We introduce the variables R_x, indexed by $\mathcal{N}$. For the root we leave $R_\emptyset$ undefined. For any other vertex y' which has a parent y, so for any $y' = yi$ with $i \in \mathbb{I}$, let*

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

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

$$\Delta_x = \Delta_{i_1}\frac{\Delta_{i_1 i_2}}{\Delta_{i_1}}\frac{\Delta_{i_1 i_2 i_3}}{\Delta_{i_1 i_2}} \dots \frac{\Delta_{i_1 \dots i_n}}{\Delta_{i_1 \dots i_{n-1}}} = R_{i_1} R_{i_1 i_2} R_{i_1 i_2 i_3} \dots R_{i_1 \dots i_n}\,. \tag{46}$$

Equivalently, for $|x| = n$,

$$\log \Delta_x = \sum_{l=1}^{n} \log R_{x_{|l}}, \tag{47}$$

where recall that $x_{|l}$ denotes the first l letters of the string x.

The decomposition described above is of interest due to the following Proposition.

Proposition 4.1. *With the variables R_x defined as above, let $U_x := (R_x, \Theta_x)$. Given any sequence $(i_n)_{n=1}^\infty$, $i_n \in \mathbb{I}$, 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 4.1, Θ_y and Θ_z are conditionally independent. We show that so are the pairs (R_y, Θ_y) and (R_z, Θ_z).

Consider that

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

so the pair (R_z, Θ_z) is a function of $\mathcal{A}_x$ (recall the notation in Lemma 4.2). At the same time,

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

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

According to Lemma 4.2, $\mathcal{A}_x$ and $\mathcal{B}_x$ are conditionally independent, given Θ_x, thus the proof is complete.

□

4.6.3 A further note on the linear case

For the sake of a short note we briefly return to the case already discussed in Section 4.3.1, because it is worth pointing out that in the linear case, the Markov chain in Proposition 4.1 in fact consists of independent elements.

In the scope of the present short subsection, following the notation in (36), let the weight function be given by

$$w(k) = (1-\beta)k + \beta \quad , \; k \in \mathbb{N},$$

for some $0 < \beta \leq 1$. Recall that in this case, the distribution of Θ is Gamma$(1, \beta)$. Also, the growth rate is independent of the structure of the tree, which 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

$$\Delta_1 = \frac{e^{-\lambda^*\tau_1}\Theta_1}{\Theta_\emptyset} = \frac{e^{-\lambda^*\tau_1}\Theta_1}{\sum_{k=1}^{\infty} e^{-\lambda^*\tau_k}\Theta_k} = \frac{\Theta_1}{\Theta_1 + \sum_{k=2}^{\infty} e^{-\lambda^*(\tau_k - \tau_1)}\Theta_k} \; , \tag{48}$$

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 (46), so indeed, the Markov chain in Proposition 4.1 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 [12].

4.6.4 Construction of the random leaf

We will now give a different construction of the tree from the ones seen before. Namely, we construct the system of $V_x = (\sigma_x, \Theta_x)$ variables starting from the root, and going step-by-step, from generation to generation. Together with these, we compute the R_x and Δ_x variables, and use them to construct a random path $\{y_n\}$ starting from the root, stepping from generation to generation, towards the set of leaves of the infinite tree. The y_n will

be chosen from the children of y_{n-1} in a "size-biased" way. We will use this path in the proofs of our results. For the sake of simple notation, we suppose for a moment that the maximum number of children of any vertex is two, that is, $K = 2$. It is straightforward to construct the corresponding generations and the random path for any $K < \infty$. 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. Keeping that in mind, using

$$\Theta = e^{-\lambda^* \sigma_1}(\Theta_1 + e^{-\lambda^* \sigma_2}\Theta_2), \tag{49}$$

we will consider the conditional joint distribution of $(\sigma_2, \Theta_1, \Theta_2)$, given Θ. (Of course, σ_1 is – conditionally – a deterministic function of these, but we will not use the value.)

Now we can construct the generations, together with the random path y_n, in the following steps.

1. Pick $\Theta_\emptyset$ at random, according to its distribution, and fix $\sigma_\emptyset = 0$. Also, fix $y_0 = \emptyset$.
2. *First generation*
   (a) Pick $(\sigma_2, \Theta_1, \Theta_2)$ according to their conditional distribution, given $\Theta_\emptyset$
   (b) Define $\Delta_1 = R_1 = \frac{\Theta_1}{\Theta_1 + e^{-\lambda^* \sigma_2}\Theta_2}$ (which is equal to $\frac{e^{-\lambda^* \sigma_1}\Theta_1}{\Theta}$, and happens not to depend on σ_1). Also define $\Delta_2 = R_2 = \frac{e^{-\lambda^* \sigma_2}\Theta_2}{\Theta_1 + e^{-\lambda^* \sigma_2}\Theta_2}$.
   (c) Choose y_1 according to the conditional probabilities $\mathbf{P}(y_1 = 1 | \Theta, \sigma_2, \Theta_1, \Theta_2) = R_1$ and $\mathbf{P}(y_1 = 2 | \Theta, \sigma_2, \Theta_1, \Theta_2) = R_2$.
3. *Second generation*
   (a) Repeat the steps seen before for the progeny of vertex 1, to get $(\sigma_{12}, \Theta_{11}, \Theta_{12})$ and also R_{11} and R_{12}. This is done only using the information carried by Θ_1, conditionally independently of (Θ, Θ_2). This conditional independence is the consequence of Corollary 4.1. 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 steps, use Θ_2 to get $(\sigma_{22}, \Theta_{21}, \Theta_{22})$, R_{21} and R_{22}. We then also have Δ_{21} and Δ_{22}.
   (c) Choose y_2 from the children of y_1, according to the conditional distribution given by the R_x variables in the second generation. Namely, if $y_1 = 1$,

$$\mathbf{P}(y_2 = 11 | y_1 = 1, \sigma_{12}, \Theta_{11}, \Theta_{12}) = R_{11}$$
$$\mathbf{P}(y_2 = 12 | y_1 = 1, \sigma_{12}, \Theta_{11}, \Theta_{12}) = R_{12},$$

   and if $y_1 = 2$,

$$\mathbf{P}(y_2 = 21 | y_1 = 2, \sigma_{22}, \Theta_{21}, \Theta_{22}) = R_{21}$$
$$\mathbf{P}(y_2 = 22 | y_1 = 2, \sigma_{22}, \Theta_{21}, \Theta_{22}) = R_{22},$$

   conditionally independently of the entire past of the construction.

4. *n-th generation*

   (a) Having constructed all the Θ_x with $|x| = n-1$, split these all in the way above, conditionally independently of each other (and the entire past of the construction), to get the R_z and Δ_z variables in the $n-th$ generation. In particular,

$$R_{xi} = \frac{e^{-\lambda^*(\sigma_{x1}+\cdots+\sigma_{xi})}\Theta_{xi}}{\Theta_x}.$$

   (b) According to the value of y_{n-1}, choose y_n from its children, according to the corresponding R_z distribution (conditionally independently of the entire past).

Remark 4.4. *As mentioned before, our model is intimately related to a branching process, as discussed in Section 3. In branching processes, the idea of size biasing is not at all new, as its importance is emphasized e.g. in [31].*

Remark 4.5. *This step-by-step construction of the tree is similar to the fragmentation processes discussed e.g. in [6]. There the usage of "randomly tagged branches" based on size-biased choices is a standard technique, see [6], Section 1.2.3. Note however, that our step-by-step construction is* not *a fragmentation process in the classical sense. In particular, the sequence of measures μ_n is* not *Markov: the process also "remembers" the values Θ_x which influence how the weight $\mu_n(\{x\})$ at x is further "fragmented".*

Proposition 4.2. *With $V_x = (\sigma_x, \Theta_x)$ as before, the distribution of $\{V_x\}_{x\in\mathcal{N}}$ in the above construction is identical to the distribution in the randomly growing tree model.*

Proof. The statement we are proving is about the joint distribution of countably infinitely many (real-valued) random variables, so this joint distribution can be viewed as a measure on $\mathbb{R}^{\mathbb{N}}$, [1] with the σ-algebra of measurable sets being the σ-algebra generated by cylinder sets – defined in terms of finitely many of the σ_x and Θ_x. So to prove that the two measures on $\mathbb{R}^{\mathbb{N}}$ – given by the two constructions – coincide, it is enough to see that they coincide on such cylinder sets.

In terms of joint distributions: It is enough to see that the distributions of $\{V_x\}_{x\in\mathcal{N}}$ coming from the two constructions have identical finite-dimensional marginals. In particular, it is enough to show that for every n, the distribution of $\{V_x\}_{x\in\mathcal{N},|x|\le n}$ in the above construction is identical to the distribution in the randomly growing tree model.

This is easy to see by induction:

- For $n = 0$ we have chosen the law of $\Theta_\emptyset$ properly by construction, also $\sigma_\emptyset = 0$ as it should be.
- For $n = 1$, the $\{V_x\}_{x\in\mathcal{N},|x|=1}$ are constructed to have the right conditional joint distribution, given $\Theta_\emptyset$, so the $n = 0$ statement implies the $n = 1$ statement. In particular, the Θ_x for $|x| = 1$ are distributed as they should be.

[1] we could write $([0,\infty)\times[0,\infty))^{\mathcal{N}}$, but a measure on this can be viewed as a special case of a measure on $\mathbb{R}^{\mathbb{N}}$.

- For $n \geq 2$, the same argument (the construction) gives inductively that the joint distribution of the $\{V_x\}_{x \in W}$ is what it should be, for any family W of x-es which consists of a vertex and its children. However, the construction also ensures the conditional independence of $\{V_y\}_{x \prec y}$ and $\{V_z\}_{x \not\preceq z}$ given Θ_x, as in Lemma 4.2. This, together with the joint distributions of the $\{V_x\}_{x \in W}$ (with W as above) already characterizes the joint distribution of $\{V_x\}_{x \in \mathcal{N}, |x| \leq n}$.

□

From now on, we will use the alternative construction of the tree in our discussion, so Proposition 4.2 is used all the time in the proof, but this will not be formally mentioned.

Definition 4.3. *Denote by Υ the σ-algebra generated by $\{\sigma_x \mid x \in \mathcal{N}\}$, which contains the full tree evolution.*

Note that for any $x \in \mathcal{N}$, Θ_x is measurable with respect to Υ, so Υ is also the σ-algebra generated by $\{\sigma_x, \Theta_x \mid x \in \mathcal{N}\}$, namely all the data about the tree – but not about the random leaf – during the parallel construction of the tree and the random leaf just presented.

The usefulness of the random leaf we constructed is shown by the following:

Lemma 4.3. *Conditioned on Υ, the conditional distribution of the leaf $\lim_n y_n$ is exactly the measure μ. Similarly, the conditional distribution of y_n is exactly μ_n.*

Proof. The second statement can be seen by induction: μ_0 obviously gives weight 1 to the single point $\emptyset = y_0$. Later, by construction of y_{n+1}, for any $x \in \mathcal{N}$ with $|x| = n$ and any $i \in \mathbb{I}$ we have $\mathbf{P}(y_{n+1} = xi \,|\, y_n = x, \Upsilon) = R_{xi}$, so if we assume inductively that $\mathbf{P}(y_n = x \,|\, \Upsilon) = \mu_n(\{x\}) = \Delta_x$, then $\mathbf{P}(y_{n+1} = xi \,|\, \Upsilon) = \Delta_x R_{xi} = \Delta_{xi} = \mu_{n+1}(\{xi\})$ for any $|xi| = n + 1$, so y_{n+1} is indeed distributed according to μ_{n+1}.

The first statement is an immediate consequence of the second, since for any cylinder set $\partial\mathcal{N}(x)$, if $|x| = n$, we have $\mathbf{P}(y_\infty \in \partial\mathcal{N}(x) \,|\, \Upsilon) = \mathbf{P}(y_n = x \,|\, \Upsilon) = \mu_n(\{x\}) = \mu(\partial\mathcal{N}(x))$. □

Corollary 4.3. *Conditioned on the tree, the conditional expectation of $-\log \Delta_{y_n}$ is exactly H_n.*

Proof. Indeed, by the above lemma,

$$\mathbf{E}(-\log \Delta_{y_n} \,|\, \Upsilon) = -\sum_{|x|=n} \mathbf{P}(y_n = x \,|\, \Upsilon) \log \Delta_x = -\sum_{|x|=n} \mu_n(\{x\}) \log \Delta_x =$$
$$= -\sum_{|x|=n} \Delta_x \log \Delta_x = H_n.$$

□

4.6.5 Markov processes along the random path

The key to the proof is the following easy observation:

Proposition 4.3. *The stochastic process $X_n = \Theta_{y_n}$ ($n = 0, 1, 2, \dots$) is a homogeneous Markov process. By "homogeneous" we mean that the transition kernel does not depend on n.*

Proof. This is clear from the construction in Section 4.6.4. Indeed, when constructing Θ_{y_n}, only the value of $\Theta_{y_{n-1}}$ is used, and the construction is the same on every level. □

The reason to construct in Section 4.6.4 the entire tree of pairs (Θ_x, Δ_x) step by step – and not just the random path $\{y_n\}$ on an already existing tree – was exactly to make the Markov property of Θ_{y_n} obvious. A direct proof without the step-by-step construction would also not be hard, but according to our taste, the underlying phenomena are more transparent this way.

Based on this proposition and equation (47), the proof of our main results will be a reference to an appropriate ergodic theorem. However, there are two issues to deal with before. First, the state space of our Markov processes is continuous and even non-compact, so the unique existence of the invariant measure needs to be discussed. This is done in the next proposition. Second, the quantity $-\log R_{y_n}$, of which we want to calculate the ergodic average, is not an observable on the state space of X_n, so this state space needs to be extended. This obvious extension will be done in Corollary 4.4.

Before starting the main arguments, let us formulate, as a lemma, an easy observation about the distribution of Θ. We will use this in the arguments both for the uniqueness and the existence of the invariant measure of X_n. From now on, we will use the notation $\mathbb{R}^+$ for the set of *positive* real numbers:

$$\mathbb{R}^+ = (0, \infty).$$

It is important that 0 is not included, e.g. when we speak of functions being continuous or nonzero on $\mathbb{R}^+$.

Lemma 4.4. *Θ is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^+$, with a density function π which is continuous and strictly positive on $\mathbb{R}^+$.*

Proof. Start from the decomposition (4.4). It shows that Θ is of the form $\Theta = e^{-\lambda^* \sigma_1} \widehat{\Theta}$ where σ_1 is independent of $\widehat{\Theta}$, which immediately implies that Θ must be equivalent to Lebesgue measure on the interval from zero to its maximal value. On the other hand, $\Theta \geq e^{-\lambda^* \sigma_1} \Theta_1 + e^{-\lambda^* (\sigma_1 + \sigma_2)} \Theta_2$ implies that Θ is not bounded, since Θ_1 and Θ_2 are independent and distributed as Θ, and their prefactors can be arbitrarily close to 1. The same decomposition, applied once again, also implies that the density π is even a continuous function (more precisely, can be chosen to be continuous), since Θ being absolutely continuous w.r.t. Lebesgue measure implies that so is $\widehat{\Theta}$ (since $K < \infty$), the density of which is once again smoothened by $\Theta = e^{-\lambda^* \sigma_1} \widehat{\Theta}$. □

For the discussion of the invariant measures, let P denote the transition kernel of X_n – that is, $P(t)$ is the conditional distribution of X_{n+1} under the condition $X_n = t$ (for every $t \in \mathbb{R}^+$). We also use it as the operator acting on measures by $\eta P = \int_{\mathbb{R}^+} P(t)\,\mathrm{d}\eta(t)$.

Proposition 4.4. *The transition kernel P of the Markov process $X_n = \Theta_{y_n}$ has exactly one invariant measure.*

Proof. Recall that the decomposition (4.4) is the key relation between the Θ_x-es of the different generations, on which the construction of X_n – and thus every property of the transition kernel – is based.

The key observation is that $P(t)$ is equivalent to Lebesgue measure (on $\mathbb{R}^+$, of course) for every $t \in \mathbb{R}^+$. This (and more) is explicitly stated and proven in Lemma 4.12. However, since we feel that this statement is really intuitive, let us give a rough reasoning here as well.

First, Lemma 4.4 implies that the distribution of Θ is equivalent to Lebesgue measure on $\mathbb{R}^+$. Recall now the construction in Section 4.6.4, the essence of which is that $P(t)$ is the conditional distribution of Θ' under the condition $\Theta = t$, where Θ' is a random choice from the set $\{\Theta_1, \dots, \Theta_K\}$. Look again at the relation between Θ and $\{\Theta_1, \dots, \Theta_K\}$, which is the decomposition (4.4), or the simplified form for $K = 2$, which is (49). It shows that given any value of t, the condition $\Theta = t$ doesn't rule out any of the possible values of a Θ_i with $1 \le i \le K$. Also, the conditioning on $\Theta = t$ doesn't spoil the absolute continuity of Θ_i, and the method of randomly choosing Θ' from $\{\Theta_1, \dots, \Theta_K\}$ also preserves absolute continuity. With this, the key observation is shown. Again, see Lemma 4.12 for a detailed proof.

This observation about $P(t)$ implies that for any measure η on $\mathbb{R}^+$, the first iterate ηP is already equivalent to Lebesgue measure. This in turn implies that any invariant measure $\eta = \eta P$ is equivalent to Lebesgue measure, so any two invariant measures are equivalent.

Suppose now indirectly that there exist two different invariant probability measures. Then two different *extremal* invariant probability measures also have to exist. But two different extremal invariant probability measures must be mutually singular, which contradicts the previous argument. Thus there is at most one invariant probability measure.

The existence follows from Lemma 4.6 and Lemma 4.5. Indeed, the limiting measure ν of Lemma 4.5 has to be invariant by Lemma 4.6. □

Lemma 4.5. *The sequence of random variables $X_n = \Theta_{y_n}$ is weakly convergent to some measure ν on $\mathbb{R}^+$.*

To keep our arguments easy to follow, we delay the proof to Section 4.7.2.

Lemma 4.6. *P is continuous with respect to weak convergence of measures.*

The proof is delayed to Section 4.7.3.

Corollary 4.4. *The stochastic process $Y_n = (\Theta_{y_n}, R_{y_n})$ $(n = 1, 2, \dots)$ is a homogeneous Markov process, for which the transition kernel has exactly one invariant measure.*

Proof. Notice that during the construction of the tree in Section 4.6.4, R_{y_n} is constructed by using only the value of $\Theta_{y_{n-1}}$ (not even $R_{y_{n-1}}$), in a time-homogeneous way. Thus Y_n is really homogeneous Markov. Let $\tilde{P}$ denote the transition kernel. From the construction, $\tilde{\eta}\tilde{P}$ depends only on the first marginal of $\tilde{\eta}$, and on this marginal it acts exactly like P. So for any measure $\hat{\nu}$ with first marginal ν, $\tilde{\nu} := \hat{\nu}\tilde{P}$ is invariant by the invariance of ν under P. The uniqueness is obvious from the uniqueness of ν. □

Now we are ready to apply an ergodic theorem on the sequence $-\log R_{y_n}$ to get the central technical result, from which our first two theorems easily follow.

Corollary 4.5. *The limit* $h := -\lim_{n\to\infty} \frac{1}{n} \log \Delta_{y_n}$ *exists and is constant with probability one.*

Proof. $-\log R_{y_n}$ is an observable on the state space of Y_n, and h is exactly the ergodic average of this observable by (47). So it is guaranteed to be constant by the unique existence of the invariant measure and Theorem 1.1 in Chapter X of [13]. We give the details of the (standard) argument now.

Theorem 1.1 in Chapter X of [13] states that "If $\{x_n, n \geq 0\}$ is a stationary Markov process, and if z is an invariant random variable, then z is measurable on the sample space of x_0". To formally apply this theorem to our process, we first need to construct a stationary version of Y_n. Namely, let $\tilde{Y}_n$ be the Markov process with generator $\tilde{P}$ started from $\tilde{Y}_0$ which is distributed according to the unique invariant measure $\tilde{\nu}$. For this process, the ergodic average of an observable, being an invariant random variable (see [13], Chapter X for the definition), is by the above theorem measurable on the state space – that is, constant with probability one, conditioned on the initial value (more precisely, for $\tilde{\nu}$-a.e. initial value). But in our case, this constant is indeed independent of the initial value – actually, it is constant for *every* initial value, since $\tilde{P}$ brings any measure (e.g. a point measure concentrated on any point) into a measure equivalent with $\tilde{\nu}$. Now notice that the property that the ergodic average is the same constant with probability one, independently of the initial state, is a property of the transition kernel $\tilde{P}$ only (and not of $\tilde{Y}_n$ as a stochastic process), so it also holds for the process Y_n. □

Remember that $\frac{1}{n}H_n$ is a conditional expectation of $-\frac{1}{n}\log \Delta_{y_n}$ by Corollary 4.3. So since we have just shown the almost sure convergence of $-\frac{1}{n}\log \Delta_{y_n}$, the almost sure convergence of $\frac{1}{n}H_n$ follows, if we have e.g. dominated convergence. This will be guaranteed by the following lemma.

Lemma 4.7. *Let* $\tilde{\mu}$ *be any Borel probability measure on* $\partial\mathcal{N}$*, with* $K < \infty$*. Using the notation in Section 4.4.1, for every* $x \in \partial\mathcal{N}$ *let*

$$f_n(x) := -\frac{1}{n} \log \tilde{\mu}(\partial\mathcal{N}(x|_n)).$$

Then $\bar{f} := \sup_n f_n$ *is integrable with respect to the measure* $\tilde{\mu}$*.*

The proof is delayed to Section 4.7.1. Now we are ready to prove the main results of Section 4.

Proof of Theorem 4.1. For every $x \in \partial\mathcal{N}$ let $f_n(x) = -\frac{1}{n}\log\mu_n(\{x|_n\}) = -\frac{1}{n}\log\mu(\partial\mathcal{N}(x|_n))$. By Lemma 4.3, Corollary 4.5 states exactly that for almost every realization of the tree, $f_n(x)$ converges μ-almost surely to h.

Now divide the statement of Corollary 4.3 by n to get

$$\frac{1}{n}H_n = \mathbf{E}\left(-\frac{1}{n}\log\Delta_{y_n}|\Upsilon\right) = \int_{\{\bar{x}\in\mathcal{N}:|x|=n\}} -\frac{1}{n}\log(\mu_n(\{\bar{x}\}))\,\mathrm{d}\mu_n(\bar{x}) = \int_{\partial\mathcal{N}} f_n(x)\,\mathrm{d}\mu(x).$$

We can now apply the dominated convergence theorem to finish the proof, since we can use the supremum as an integrable dominating function, see Lemma 4.7. □

Proof of Theorem 4.2. We first show the second statement of the theorem by showing that the local dimension of μ at the leaf $\lim_n y_n$ is exactly $\frac{h}{-\log\Lambda}$ where h is from Corollary 4.5. Let $B(x,r)$ denote the r-neighbourhood of the point $x \in \partial\mathcal{N}$ w.r.t. the metric (45). For $r = \Lambda^n$, this neighbourhood is formed exactly by the descendants of $x|_n$, so $B(x,\Lambda^n) = \partial\mathcal{N}(x|_n)$. The μ-measure of this set is

$$\mu(B(x,\Lambda^n)) = \mu(\partial\mathcal{N}(x|_n)) = \mu_n(\{x|_n\}) = \log\Delta_{x|_n},$$

while the logarithm of the diameter of this set is $n\log\Lambda$. Thus the local dimension of μ at the leaf x is

$$\dim_{\mathrm{loc}}\mu(x) = \lim_{n\to\infty}\frac{\mu(B(x,\Lambda^n))}{n\log\Lambda} = \lim_{n\to\infty}\frac{-\frac{1}{n}\log\Delta_{x|_n}}{-\log\Lambda}$$

(if this limit exists), by the definition in (46) and (46).

Applying that to $x = \lim_n y_n$, Lemma 4.3 and Corollary 4.5 say that this limit indeed exists and is equal to $\frac{h}{-\log\Lambda}$ for μ-almost every x, which is what we wanted to show.

The first statement of the theorem in now an immediate consequence of the definitions of the Hausdorff and packing dimension of a measure in (46) and (46).

□

4.7 Proofs of auxiliary lemmas

4.7.1 The lemma for dominated convergence of the entropies

In this section we prove Lemma 4.7.

Proof of Lemma 4.7. For arbitrary $M < \infty$, let us define the set

$$F_M^{(n)} := \{x : f_n(x) \geq M\} = \{x : -\frac{1}{n}\log\tilde{\mu}(\partial\mathcal{N}(x|_n)) \geq M\} = \{x : \tilde{\mu}(\partial\mathcal{N}(x|_n)) \leq e^{-nM}\}.$$

Since f_n takes constant values on the K^n cylinder sets, we have

$$\tilde{\mu}(F_M^{(n)}) \leq K^n e^{-nM} = \left(Ke^{-M}\right)^n. \tag{50}$$

Now we define

$$F_M := \{x : \bar{f}(x) > M\} = \bigcup_n \{x : f_n(x) > M\} \subseteq \bigcup_n F_M^{(n)}.$$

By (50), for $M > \log(2K)$,

$$\tilde{\mu}(F_M) \leq \sum_{n=1}^{\infty} \left(Ke^{-M}\right)^n < 2Ke^{-M}.$$

Thus, since $\bar{f} \geq 0$,

$$\int \bar{f}(x)\, \mathrm{d}\tilde{\mu}(x) < \sum_{M=1}^{\infty} M\tilde{\mu}(\{x : M-1 \leq \bar{f}(x) < M\}) < \infty.$$

□

4.7.2 Limiting distribution of Θ_{y_n} along the random path

In this section we prove Lemma 4.5. We begin with three lemmas of elementary probability whose statements do not rely on the setting of the present dissertation.

The first one is a trivial generalization of the ordinary weak law of large numbers. We could call it "Weak law of large numbers with arbitrary weights". For this purpose, we will consider a sequence of probability vectors $\{\underline{p}^n\}_{n=1}^{\infty}$, where, again, each $\underline{p}^n$ is a probability vector $\underline{p}^n = (p_1^n, p_2^n, \ldots, p_{N_n}^n)$. We plan to calculate weighted averages of independent random variables with weight vectors $\underline{p}^n$. We expect such an average to be close to the expectation, if every term has a sufficiently small weight. So we will say that the sequence $\{\underline{p}^n\}_{n=1}^{\infty}$ is proper if

$$\lim_{n\to\infty} \max\{p_j^n : 1 \leq j \leq N_n\} = 0\,.$$

Lemma 4.8. *Let ν_0 be a probability distribution on $\mathbb{R}$ with finite expectation m. Let $\{\underline{p}^n\}_{n=1}^{\infty}$ be a proper sequence of weight vectors, and let ν_n be the distribution of*

$$\sum_{j=1}^{N_n} p_j^n Z_j$$

where $Z_1, Z_2, \ldots, Z_{N_n}$ are independent random variables with distribution ν_0. Then

$$\nu_n \Rightarrow m\,.$$

Note that this is the usual weak law if $p_j^n = \frac{1}{n}$ $(j = 1, \ldots, n)$.

Proof. The proof is trivial following the standard proof of the weak law with characteristic functions. □

Now we turn to a lemma which could be called "size-biased sampling with arbitrary extra weights". For this purpose, let $\underline{p} = (p_1, p_2, \dots, p_N)$ be a probability vector, and let $Z_1, Z_2, \dots, Z_N$ be random variables on $\mathbb{R}^+$ (meaning $\mathbf{P}(Z_j > 0) = 1$). We will say that the random variable V is the size-biased random choice from $Z_1, Z_2, \dots, Z_N$ with extra weights $p_1, p_2, \dots, p_N$, if it is constructed the following way:

1. Generate a realization of $(Z_1, Z_2, \dots, Z_N)$, and call it $(z_1, z_2, \dots, z_N)$.
2. Having that, choose a random integer J from the index set $\{1, 2, \dots, N\}$ with the weight
$$\frac{p_j z_j}{\sum_{j=1}^N p_j z_j}$$
given to each j.
3. Set $V = z_J$.

Note that this is the usual size-biased random choice if all the p_j are equal. Our lemma states that this size-biased random choice with extra weights behaves just like the ordinary one, provided that every weight is small.

To state the lemma, let ν_0 be a probability distribution on $\mathbb{R}^+$ with finite expectation m. We will say that the measure ν is the size-biased version of ν_0, if it is absolutely continuous with respect to ν_0, and the density is $\rho(t) = \frac{1}{m}t$. In other words, $\nu(A) = \frac{1}{m}\int_A t \,\mathrm{d}\nu_0(t)$.

Lemma 4.9. *Let ν_0 be a probability distribution on $\mathbb{R}^+$ with finite expectation m. Let $\{\underline{p}^n\}_{n=1}^\infty$ be a proper sequence of weight vectors, and (for each n) let $Z_1^n, Z_2^n, \dots, Z_{N_n}^n$ be independent random variables with distribution ν_0. Let V_n be the random choice from $Z_1^n, Z_2^n, \dots, Z_{N_n}^n$ with extra weights $p_1^n, p_2^n, \dots, p_{N_n}^n$. Let ν be the size-biased version of ν_0. Then*
$$V_n \Rightarrow \nu\,.$$

Proof. Let F denote the cumulative distribution function of ν, that is, $F(t) = \nu([0,t])$. Let F_n denote the cumulative distribution function of V_n. For some fixed t, we write it in the form
$$F_n(t) = \mathbf{E}(\mathbf{P}(V_n \le t \mid \{Z_j^n\}_{j=1}^{N_n})). \tag{51}$$
The conditional probability inside is just the weight of j-s with $Z_j \le t$, so
$$\mathbf{P}(V_n \le t \mid \{Z_j^n\}_{j=1}^{N_n}) = \frac{\sum_{j=1}^{N_n} p_j^n Z_j^n \mathbb{1}(Z_j^n \le t)}{\sum_{j=1}^{N_n} p_j^n Z_j^n}\,.$$
According to Lemma 4.8 the denominator converges weakly (and thus, in probability) to $\mathbf{E}(Z_1^n) = m > 0$ as $n \to \infty$. Similarly, the numerator converges in probability to
$$\mathbf{E}(Z_1^n \mathbb{1}(Z_1^n \le t)) = \int_{\mathbb{R}^+} \tilde{t}\mathbb{1}(\tilde{t} \le t)\,\mathrm{d}\nu_0(\tilde{t}) = m\nu([0,t])\,.$$

This implies that the quotient converges weakly to $\nu([0,t]) = F(t)$. Since this quotient is a conditional probability, it is obviously bounded by 1, so (51) implies that $F_n(t) \to F(t)$. □

The following lemma is just a re-statement of the previous one. This is the form that we will use.

Lemma 4.10. *Let ν_0 be a probability distribution on $\mathbb{R}^+$ with finite expectation, and let ν be its size-biased version. Let φ be a bounded continuous function on $\mathbb{R}^+$. Then for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any probability vector $(p_1, p_2, \ldots, p_N)$ which satisfies that*

$$\max\{p_j : 1 \le j \le N\} \le \delta,$$

if $Z_1, Z_2, \ldots, Z_N$ are independent with distribution ν_0, then the size-biased random choice (called V) from $Z_1, Z_2, \ldots, Z_N$ with extra weights $p_1, p_2, \ldots, p_N$ satisfies

$$\left| \mathbf{E}(\varphi(V)) - \int \varphi(t)\, \mathrm{d}\nu(t) \right| < \varepsilon\,.$$

Before proving Lemma 4.5, we need one more tiny statement about the structure of the growing tree.

Lemma 4.11. *For any vertex $x \in \mathcal{N}$, let*

$$T_x := e^{-\lambda^* \tau_x}, \tag{52}$$

and for every x with $|x| = n$ let

$$p_x := \frac{T_x}{\sum_{|y|=n} T_y} \text{ for } |x| = n\,. \tag{53}$$

Then the sequence $p^{n,max} := \max\{p_x : |x| = n\}$ converges to zero in probability.

Proof. We prove the stronger statement that $p^{n,\max}$ converges to zero with probability one. We use the form

$$p^{n,\max} = \frac{\max\{T_x : |x| = n\}}{\sum_{|y|=n} T_y}. \tag{54}$$

We show that the numerator converges to zero with probability one, while the denominator converges to a positive limit with probability one.

1. If the numerator does not converge to zero, then there is some $\varepsilon > 0$ and there are infinitely many vertices $x \in \mathcal{N}$ with $T_x > \varepsilon$. Then, for all these x we have $\tau_x < \tau^* := \frac{-\log \varepsilon}{\lambda^*}$, so infinitely many vertices are born within the finite time τ^*. This is known to have probability zero – see comment at (3).

2. Iterating the decomposition of Θ, we get

$$\Theta = \sum_{|x|=n} T_x \Theta_x. \tag{55}$$

Let Σ_n denote the σ-algebra generated by $\{\sigma_x : x \in \mathcal{N}, |x| \leq n\}$ – that is, the complete history of the tree growth up to the n-th level. Similarly, let Σ denote the σ-algebra generated by $\{\sigma_x : x \in \mathcal{N}\}$. Clearly $\Sigma_n \subset \Sigma_{n+1}$, Σ is generated by $\cup_n \Sigma_n$, and Θ is Σ-measurable. So Lévy's 'upward' theorem ensures that $\mathbf{E}(\Theta \mid \Sigma_n) \to \Theta$ with probability one. However, if $|x| = n$, then Θ_x is independent of Σ_n, while T_x is Σ_n-measurable, so (55) implies that

$$\mathbf{E}(\Theta \mid \Sigma_n) = \sum_{|x|=n} T_x \mathbf{E}\Theta_x = \mathbf{E}\Theta \sum_{|x|=n} T_x,$$

so with probability one the denominator of (54) converges to $\frac{\Theta}{\mathbf{E}\Theta} \neq 0$.

□

Now we can complete the goal of this subsection:

Proof of Lemma 4.5. Actually we give the limit explicitly. Let ν be the measure on $\mathbb{R}^+$ with density function $cx\pi(x)$, where $\pi(x)$ is the density of Θ, and $c = \frac{1}{\mathbf{E}\Theta}$ is a normalizing constant. We will show that

$$X_n \Rightarrow \nu. \tag{56}$$

Let us look directly at $X_n = \Theta_{y_n}$ for some fixed n. This can also be constructed in the following way:

1. Generate the birth times τ_x for all vertices x with $|x| = n$ (that is, on the n-th level of the tree). This defines the values $T_x = e^{-\lambda^* \tau_x}$, $|x| = n$. For better transparency, let us normalize these values to get a probability distribution on the n-th level of the tree, as before, see (53) for the definition of p_x.

2. Also generate the random variables Θ_x for $|x| = n$, which are independent of the p_x.

3. Now y_n is chosen from the points $|x| = n$ according to the distribution μ_n, so the weight given to some x is

$$\frac{\Delta_x}{\sum_{|z|=n} \Delta_z} = \frac{T_x \Theta_x}{\sum_{|z|=n} T_z \Theta_z} = \frac{p_x \Theta_x}{\sum_{|z|=n} p_z \Theta_z}.$$

So, having the values p_x fixed, the value $X_n = \Theta_{y_n}$ is the result of a size-biased sampling from the independent random variables Θ_x, $|x| = n$, with additional weights p_x – just like in the context of Lemma 4.9 and Lemma 4.10.

Now we can prove (56). Let φ be a fixed bounded continuous function on $\mathbb{R}^+$, let M_φ be an upper bound of $|\varphi|$, and let $m_\varphi = \int_{\mathbb{R}^+} \varphi \, \mathrm{d}\nu$ (which satisfies $|m_\varphi| \le M_\varphi$). Let $\varepsilon > 0$ be arbitrary.

Choose $\delta > 0$ according to Lemma 4.10 so that if all the p_x on some level $|x| = n$ are at most δ, then

$$\left|\mathbf{E}\left(\varphi(X_n) \mid \{p_x\}\right) - m_\varphi\right| < \varepsilon.$$

Lemma 4.11 implies that there exists an n_0 such that for all $n > n_0$,

$$\mathbf{P}(\max\{p_x : |x| = n\} > \delta) < \frac{\varepsilon}{2M_\varphi}\,.$$

Let $\Omega_{n,\delta}$ denote the event that $\max\{p_x : |x| = n\} \le \delta$. For $n > n_0$ we get

$$\left|\mathbf{E}\left(\varphi(X_n)\right) - m_\varphi\right| \le \int \left|\mathbf{E}\left(\varphi(X_n) - m_\varphi \mid \{p_x\}\right)\right| \mathrm{d}\mathbf{P} =$$

$$= \int_{\Omega_{n,\delta}^c} \left|\mathbf{E}\left(\varphi(X_n) - m_\varphi \mid \{p_x\}\right)\right| \mathrm{d}\mathbf{P} + \int_{\Omega_{n,\delta}} \left|\mathbf{E}\left(\varphi(X_n) - m_\varphi \mid \{p_x\}\right)\right| \mathrm{d}\mathbf{P} \le$$

$$\le 2M_\varphi \mathbf{P}\left(\Omega_{n,\delta}^c\right) + \int_{\Omega_{n,\delta}} \varepsilon \, \mathrm{d}\mathbf{P} \le \varepsilon + \varepsilon = 2\varepsilon.$$

□

4.7.3 Weak continuity of the transition kernel

This section is devoted to the proof of Lemma 4.6.

Proof of Lemma 4.6. We first show in Lemma 4.12 that the transition kernel P can be written as $(\eta P)(B) = \int_{\mathbb{R}^+} \int_B k(t,s) \, \mathrm{d}s \, \mathrm{d}\eta(t)$ where the kernel function $k(t,s)$ is continuous in the first variable (actually it is continuous in both variables). Lemma 4.13 – which is a pure probability statement – says that such a kernel is continuous with respect to weak convergence of measures. □

In the lemma, we show a little more than what is needed for the above proof. In particular, we also show that the kernel function $k(t,s)$ is nowhere zero on $\mathbb{R}^+ \times \mathbb{R}^+$, because this is used in the proof of Proposition 4.4.

Lemma 4.12. *The transition kernel P can be written as $(\eta P)(B) = \int_{\mathbb{R}^+} \int_B k(t,s) \, \mathrm{d}s \, \mathrm{d}t$ where the kernel function $k(t,s)$ is continuous in both variables (in its domain $(t,s) \in \mathbb{R}^+ \times \mathbb{R}^+$), and strictly positive.*

Proof. For the time of the proof, let Θ and Θ' denote two consecutive values of the process, say $\Theta := X_n = \Theta_{y_n}$, $\Theta' = X_{n+1} = \Theta_{y_{n+1}}$. So the kernel function $k(t,s)$ is just the conditional density of Θ' (as a function of s), under the condition $\Theta = t$. So

$$k(t,s) = \frac{\rho(t,s)}{\pi(t)},$$

where $\rho(t,s)$ is the joint density of the pair (Θ,Θ'), and $\pi(t)$ is its first marginal – that is, the density of Θ.

We know from Lemma 4.4 that Θ is indeed absolutely continuous w.r.t. Lebesgue measure, and the density π is continuous and nonzero on $\mathbb{R}^+$. Knowing this, we now show that $\rho(t,s)$ is also continuous in both variables and nonzero on $\mathbb{R}^+\times\mathbb{R}^+$, which completes the proof.

We restrict to the case $K=2$. The case of a general $K<\infty$ causes no additional difficulty other than messy notation. Following the construction of the tree in Section 4.6.4, we start with $\sigma_1,\sigma_2,\Theta_1,\Theta_2$ independent, with σ_i being exponentially distributed with parameter $w(i-1)/\lambda^*$ and Θ_i being distributed as Θ $(i=1,2)$. We introduce the temporary notation $S_i=e^{-\lambda^*\sigma_i}$ and denote its density by g_i. Explicit calculation gives that

$$g_i(u)=\frac{w(i-1)}{\lambda^*}u^{\frac{w(i-1)}{\lambda^*}-1}\mathbb{1}_{(0,1)}(u), \tag{57}$$

of which we will only use that $u\,g_1(u)$ is bounded.

Denote the joint density of $(S_1,S_2,\Theta_1,\Theta_2)$ by

$$f(u_1,u_2,t_1,t_2)=g_1(u_1)g_2(u_2)\pi(t_1)\pi(t_2)\,.$$

We define

$$\Theta=S_1\Theta_1+S_1S_2\Theta_2=S_1(\Theta_1+S_2\Theta_2)\,.$$

To get the appropriate joint distributions, in the probability vector $(S_1,S_2,\Theta_1,\Theta_2)$ we replace S_1 by Θ, so let us denote the joint density of $(\Theta,S_2,\Theta_1,\Theta_2)$ by $\tilde f$. The density transformation formula gives

$$\tilde f(t,u_2,t_1,t_2)=\frac{1}{t_1+u_2t_2}f(\frac{t}{t_1+u_2t_2},u_2,t_1,t_2)=\frac{1}{t}\frac{t}{t_1+u_2t_2}g_1(\frac{t}{t_1+u_2t_2})g_2(u_2)\pi(t_1)\pi(t_2)\,.$$

According to the construction, Θ' is chosen to be either Θ_1 or Θ_2, with conditional probabilities (given (S_2,Θ_1,Θ_2) and conditionally independently of Θ)

$$\begin{aligned}\mathbf{P}(\Theta'=\Theta_1|S_2,\Theta_1,\Theta_2)&=\frac{\Theta_1}{\Theta_1+S_2\Theta_2},\\ \mathbf{P}(\Theta'=\Theta_2|S_2,\Theta_1,\Theta_2)&=\frac{S_2\Theta_2}{\Theta_1+S_2\Theta_2}\,.\end{aligned}$$

So the joint density of (Θ,Θ') is

$$\begin{aligned}\rho(t,s)&=\iint_{\mathbb{R}^2}\frac{s}{s+u_2t_2}\tilde f(t,u_2,s,t_2)\,\mathrm{d}t_2\,\mathrm{d}u_2+\iint_{\mathbb{R}^2}\frac{u_2s}{t_1+u_2s}\tilde f(t,u_2,t_1,s)\,\mathrm{d}t_1\,\mathrm{d}u_2=\\ &=\iint_{\mathbb{R}^2}\bar f_1(t,s,u_2,t_2)\,\mathrm{d}t_2\,\mathrm{d}u_2+\iint_{\mathbb{R}^2}\bar f_2(t,s,u_2,t_1)\,\mathrm{d}t_1\,\mathrm{d}u_2\end{aligned}$$

All there is left is to show that both integrals on the right hand side are continuous and nonzero for $(t,s)\in\mathbb{R}^+\times\mathbb{R}^+$. Now the integrands $\bar f_1$ and $\bar f_2$ are not exactly continuous,

but they are continuous *on their supports.* [2] On the other hand, for every $(t,s) \in \mathbb{R}^+ \times \mathbb{R}^+$ the support of each integrand is a nice set (described in the footnote) with a boundary of Lebesgue measure zero. That is, for *every* $(t_0, s_0) \in \mathbb{R}^+ \times \mathbb{R}^+$,

$$\bar{f}_1(t,s,u_2,t_2) \overset{(t,s)\to(t_0,s_0)}{\longrightarrow} \bar{f}_1(t_0,s_0,u_2,t_2) \text{ for Lebesgue-a.e. } (u_2,t_2) \in \mathbb{R}^2 .$$

To get the desired continuity of the first integral by the Lebesgue dominated convergence theorem, we only need to find an integrable (in (u_2,t_2)) uniform (in (t,s) near (t_0,s_0)) upper bound for

$$\bar{f}_1(t,s,u_2,t_2) = \frac{s}{s+u_2t_2}\frac{1}{t}\frac{t}{s+u_2t_2} g_1\Big(\frac{t}{s+u_2t_2}\Big) g_2(u_2)\pi(s)\pi(t_2) .$$

The first factor is at most 1, and the product $\frac{t}{s+u_2t_2} g_1\big(\frac{t}{s+u_2t_2}\big)$ is bounded because $u\, g_1(u)$ is bounded due to (57). So we have

$$\bar{f}_1(t,s,u_2,t_2) \le C\frac{1}{t}\pi(s) g_2(u_2)\pi(t_2) \le C\Big(\frac{1}{t_0}+1\Big)(\pi(s)+1) g_2(u_2)\pi(t_2)$$

if (t,s) is close enough to (t_0,s_0), since $\frac{1}{t}\pi(s)$ is continuous in (t_0,s_0). This upper bound is clearly integrable in (u_2,t_2), so the dominated convergence theorem ensures that the integral is also continuous.

The second integral in (58) can be shown to be continuous in exactly the same way. Thus the continuity of $k(t,s)$ is proven.

To get that $\rho(t,s)$ (and thus $k(t,s)$) is strictly positive on $\mathbb{R}^+ \times \mathbb{R}^+$, we only need to note that the support of the integrand is nonempty for every $(t,s) \in \mathbb{R}^+ \times \mathbb{R}^+$ in both integrals on the right hand side of (58). This comes again from (49), which shows that any pair of positive values is possible for (Θ, Θ_1) (in case of the first integrand) or for (Θ, Θ_2) (in case of the second integrand). (See the footnote 2 for explicit formulae.) The integrands are of course also non-negative, so both integrals are positive. □

Lemma 4.13. *Let $k : \mathbb{R}^+ \times \mathbb{R}^+ \to [0,\infty)$ be a function continuous in the first variable, such that for every $t \in \mathbb{R}^+$ the function $k(t,.)$ is a probability density on $\mathbb{R}^+$ – that is, $\int_{\mathbb{R}^+} k(t,s)\,\mathrm{d}s = 1$. Let the operator P be defined on Borel probability measures of $\mathbb{R}^+$ by*

$$(\eta P)(B) := \int_{\mathbb{R}^+}\int_B k(t,s)\,\mathrm{d}s\,\mathrm{d}\eta(t)$$

for every Borel probability measure η on $\mathbb{R}^+$ and every Borel set $B \subset \mathbb{R}^+$. Then P is continuous with respect to weak convergence of measures.

This lemma is an easy consequence of the following:

[2] The supports of the two integrands are actually not the same. Both of them are characterized by the system of inequalities $\{0 < t_1, t_2;\ 0 < u_2 < 1;\ 0 < \frac{t}{t_1+u_2t_2} < 1\}$, but with the choice $s = t_1$ or $s = t_2$, respectively.

Lemma 4.14. *Let $k : \mathbb{R}^+ \times \mathbb{R}^+ \to [0, \infty)$ be a function as in Lemma 4.13, and for every $t \in \mathbb{R}^+$ let K_t denote the measure on $\mathbb{R}^+$ with density $k(t, .)$. Then if t_n is a sequence in $\mathbb{R}^+$ converging to t, then K_{t_n} converges to K_t weakly.*

Proof. By assumption, $\{k(t_n, .)\}_{n=1}^{\infty}$ is a sequence of density functions converging pointwise to the density function $k(t, .)$. This implies weak convergence of the corresponding measures through the Fatou lemma: for any Borel set $B \subset \mathbb{R}^+$

$$\liminf_{n\to\infty} K_{t_n}(B) = \liminf_{n\to\infty} \int_B k(t_n, s)\,\mathrm{d}s \overset{\text{Fatou}}{\geq} \int_B \liminf_{n\to\infty} k(t_n, s)\,\mathrm{d}s = \int_B k(t, s)\,\mathrm{d}s = K_t(B),$$

similarly

$$\liminf_{n\to\infty} K_{t_n}(B^c) \geq K_t(B^c),$$

which implies

$$\limsup_{n\to\infty} K_{t_n}(B) = 1 - \liminf_{n\to\infty} K_{t_n}(B^c) \leq 1 - K_t(B^c) = K_t(B)\,.$$

These together give

$$K_{t_n}(B) \to K_t(B)\,.$$

□

Remark 4.6. *As one of the opponents of this dissertation, Dr. Tamás F. Móri pointed out, Lemma 4.14 is in fact a consequence of the commonly known theorem which states that the pointwise convergence of the density functions implies the convergence in total variation of the according distributions, which in turn implies weak convergence. This theorem is usually cited as that of Scheffé (see [43]), although it can already be found in the work of Frigyes Riesz, from twenty years before (see [27]).*

Proof of Lemma 4.13. Let $\varphi : \mathbb{R}^+ \to \mathbb{R}$ be bounded and continuous and let η_n be a sequence of measures on $\mathbb{R}^+$ converging weakly to η. By the definition of P,

$$\begin{aligned}\int_{\mathbb{R}^+} \varphi\,\mathrm{d}(\eta_n P) &= \int_{\mathbb{R}^+\times\mathbb{R}^+} k(t, s)\varphi(s)\,\mathrm{d}(\eta_n(t) \times \mathrm{Leb}(s)) = \\ &= \int_{\mathbb{R}^+} \left[\int_{\mathbb{R}^+} k(t, s)\varphi(s)\,\mathrm{d}s\right] \mathrm{d}\eta_n(t).\end{aligned}$$

The function

$$\bar{\varphi}(t) := \int_{\mathbb{R}^+} k(t, s)\varphi(s)\,\mathrm{d}s$$

is obviously bounded, and also continuous: this is exactly the statement of Lemma 4.14. But then the weak convergence of η_n to η implies that

$$\int_{\mathbb{R}^+} \bar{\varphi}(t)\,\mathrm{d}\eta_n(t) \to \int_{\mathbb{R}^+} \bar{\varphi}(t)\,\mathrm{d}\eta(t),$$

so we have

$$\int_{\mathbb{R}^+} \varphi\,\mathrm{d}(\eta_n P) \to \int_{\mathbb{R}^+} \bar{\varphi}(t)\,\mathrm{d}\eta(t) = \int_{\mathbb{R}^+} \varphi\,\mathrm{d}(\eta P)$$

for every bounded continuous φ, which is exactly what we want to prove. □

4.8 Proof of Theorem 4.3: computation of the entropy

This section consists of the proof of Theorem 4.3, the explicit calculation for the entropy.

Proof of Theorem 4.3. We know that $\frac{1}{n}H_n = -\frac{1}{n}\sum_{|x|=n}\Delta_x \log \Delta_x$ converges almost surely to some constant h, and this constant is equal to the limit of the expected values. For this section we use the shorthand notation already introduced in (52),

$$T_x = e^{-\lambda^* \tau_x}. \tag{58}$$

To compute h, first observe that

$$\mathbf{E}\sum_{|x|=n}\Delta_x\Theta\log(\Delta_x\Theta) = \mathbf{E}\left(\sum_{|x|=n}\Theta\Delta_x\log\Delta_x\right) + \mathbf{E}\left((\Theta\log\Theta)\sum_{|x|=n}\Delta_x\right) = \mathbf{E}\left(\Theta\sum_{|x|=n}\Delta_x\log\Delta_x\right) + \mathbf{E}\left(\Theta\log\Theta\right),$$

where we have used that $\sum_{|x|=n}\Delta_x = 1$ by definition.

Next we observe that on the other hand, the same expression can be written as

$$\mathbf{E}\sum_{|x|=n}\Delta_x\Theta\log(\Delta_x\Theta) = \mathbf{E}\sum_{|x|=n}T_x\Theta_x\log\left(T_x\Theta_x\right) =$$
$$\mathbf{E}\left(\sum_{|x|=n}\Theta_x T_x\log\left(T_x\right)\right) + \mathbf{E}\left(\sum_{|x|=n}T_x\Theta_x\log\Theta_x\right) =$$
$$\sum_{|x|=n}\left(\mathbf{E}\Theta_x\right)\mathbf{E}\left(T_x\log T_x\right) + \sum_{|x|=n}\mathbf{E}\left(T_x\right)\mathbf{E}\left(\Theta_x\log\Theta_x\right) =$$
$$\left(\mathbf{E}\Theta\right)\mathbf{E}\sum_{|x|=n}\left(T_x\log T_x\right) + \mathbf{E}\left(\Theta\log\Theta\right)\mathbf{E}\left(\sum_{|x|=n}T_x\right),$$

where we have used that for any $x \in \mathcal{N}$, Θ_x and τ_x are independent. Recall that $\mathbf{E}\left(\sum_{|x|=n}T_x\right) = 1$.

Since (32) implies that $\mathbf{E}\left(\Theta\log\Theta\right) < \infty$, comparing the two formulae gives the conclusion

$$\mathbf{E}\left(\Theta\sum_{|x|=n}\Delta_x\log\Delta_x\right) = \left(\mathbf{E}\Theta\right)\mathbf{E}\left(\sum_{|x|=n}T_x\log T_x\right). \tag{59}$$

We compute the right-hand side with an induction,

$$A_n := \mathbf{E}\left(\sum_{|x|=n} T_x \log T_x\right) = \mathbf{E}\left(\sum_{|y|=n-1}\sum_{i=1}^{K} T_{yi}\log T_{yi}\right) =$$

$$\left(\mathbf{E}\sum_{i=1}^{K} e^{-\lambda^*(\tau_{yi}-\tau_y)}\right)\mathbf{E}\left(\sum_{|y|=n-1} T_y \log T_y\right) +$$

$$\left(\mathbf{E}\sum_{|y|=n-1} T_y\right)\mathbf{E}\left(\sum_{i=1}^{K} e^{-\lambda^*(\tau_{yi}-\tau_y)}\log e^{-\lambda^*(\tau_{yi}-\tau_y)}\right) =$$

$$A_{n-1} + \mathbf{E}\left(\sum_{i=1}^{K} T_i \log T_i\right),$$

so

$$A_n = n\mathbf{E}\left(\sum_{i=1}^{K} T_i \log T_i\right).$$

Now write this back to (59) to get

$$\mathbf{E}\left(\Theta\frac{1}{n}H_n\right) = (\mathbf{E}\Theta)\,\mathbf{E}\left(-\sum_{i=1}^{K} T_i \log T_i\right).$$

Since $\lim \frac{1}{n}H_n = h$ almost surely and $\mathbf{E}\Theta < \infty$, we can apply the dominated convergence theorem if we check that $\frac{1}{n}H_n$ is bounded. This follows from the standard upper bound for entropy of measures on the finite set $\{x \in \partial\mathcal{N} : |x| = n\}$, which has K^n elements, coming from Jensen's inequality:

$$H_n = -\sum_{|x|=n} \mu_n(\{x\})\log\mu_n(\{x\}) = \int\limits_{\{x\in\partial\mathcal{N}:|x|=n\}} \log\frac{1}{\mu_n(\{x\})}\,\mathrm{d}\mu_n(x) \overset{\text{Jensen}}{\leq}$$

$$\leq \log \int\limits_{\{x\in\partial\mathcal{N}:|x|=n\}} \frac{1}{\mu_n(\{x\})}\,\mathrm{d}\mu_n(x) = \log\sum_{|x|=n}\mu_n(\{x\})\frac{1}{\mu_n(\{x\})} = \log K^n = n\log K,$$

so $\frac{1}{n}H_n \leq \log K$. Now dominated convergence gives

$$h = \mathbf{E}\left(-\sum_{i=1}^{K} T_i \log T_i\right).$$

Recalling (58), the proof of the theorem is complete. □

Remark 4.7. *This value can be explicitly calculated, as soon as the weight function is given, since the τ_i variables are the sum of independent, exponentially distributed random variables with parameters $(w(j))_{j=0}^{i-1}$. Alternatively, with the function $\widehat{\rho}$ defined in (3) in Section 2.2.2,*

$$h = \lambda^* \frac{\mathrm{d}\widehat{\rho}(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=\lambda^*}.$$

References

[1] David Aldous. *Probability Approximations via the Poisson Clumping Heuristic*. Springer, 1989.

[2] David Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1(2):228–266, 1991.

[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[4] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electronic J. Probab.*, 6(paper 23):1–13, 2001.

[5] Julien Berestycki. Multifractal spectra of fragmentation processes. *Journal of Statistical Physics*, 113(3):411–430, 2003.

[6] Jean Bertoin. *Random fragmentation and coagulation processes*. Cambridge Univ Pr, 2006.

[7] J. D. Biggins. Martingale convergence in the branching random walk. *Journal of Applied Probability*, 14(1):25–37, 1977.

[8] 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.

[9] 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.

[10] Fan Chung, Shirin Handjani, and Doug Jungreis. Generalizations of Polya's urn problem. *Ann. Comb.*, 7(2):141–153, 2003.

[11] Steffen Dereich and Peter Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Probab.*, 14(43):1222–1267, 2009.

[12] 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.

[13] Joseph Leo Doob. *Stochastic Processes*. Wiley, 1953.

[14] Michael Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.

[15] T. Duquesne. Packing and Hausdorff measures of stable trees. *Lévy Matters I*, pages 93–136, 2010.

[16] Thomas Duquesne and Jean-Franois Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probability Theory and Related Fields*, 131:553–603, 2005.

[17] Richard Durrett. *Random Graph Dynamics*. Cambridge University Press, Cambridge, 2007. Cambridge Series in Statistical and Probabilistic Mathematics.

[18] Kenneth Falconer. *Techniques in Fractal Geometry*. Wiley, 1997.

[19] B. Haas and G. Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electronic Journal of Probability*, 9(paper 4):57, 2004.

[20] B. Haas and G. Miermont. Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees. 2010. Arxiv preprint, `http://arxiv.org/abs/1003.3632`.

[21] B. Haas, G. Miermont, J. Pitman, and M. Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *The Annals of Probability*, 36(5):1790–1837, 2008.

[22] 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.

[23] Peter Jagers and Olle Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, 16(2):221–259, 1984.

[24] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Statist.*, 37:1211–1223, 1966.

[25] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63(6):066123, May 2001.

[26] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85(21):4629–4632, Nov 2000.

[27] Norbert Kusolitsch. Why the theorem of Scheffé should be rather called a theorem of Riesz. *Period. Math. Hungar.*, 61(1-2):225–229, 2010.

[28] Jean-François Le Gall. Processus de branchement, arbres et superprocessus. In *Development of mathematics 1950–2000*, pages 763–793. Birkhäuser, Basel, 2000.

[29] László Lovász. *Very large graphs*. 2008. Arxiv preprint, `http://arxiv.org/abs/0902.0132`.

[30] R. Lyons. A simple path to Biggins' martingale convergence for branching random walk. In K.B. Athreya and P. Jagers, editors, *Classical and modern branching processes*, The IMA volumes in mathematics and its applications. Springer, 1997.

[31] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of l log l criteria for mean behavior of branching processes. *The Annals of Probability*, 23(3):1125–1138, 1995.

[32] T. F. Móri. On random trees. *Studia Sci. Math. Hungar.*, 39(1-2):143–155, 2002.

[33] T. F. Móri. The maximum degree of the Barabási–Albert random tree. *Comb. Probab. Computing*, 14:339–348, 2005.

[34] T. F. Móri. A surprising property of the Barabási-Albert random tree. *Studia Sci. Math. Hungar.*, 43:265–273, 2006.

[35] Olle Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrsch. Verw. Gebiete*, 57(3):365–395, 1981.

[36] Roberto Oliveira and Joel Spencer. Connectivity transitions in networks with superlinear preferential attachment. *Internet Math.*, 2(2):121–163, 2005.

[37] Peter Olofsson. The $x \log x$ condition for general branching processes. *J. Appl. Probab.*, 35(3):537–544, 1998.

[38] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Struct. Alg.*, 5(2):337–347, 1994.

[39] Anna Rudas. Random tree growth with general weight function. 2004. Arxiv preprint, `http://arxiv.org/abs/math/0410532`.

[40] Anna Rudas and Bálint Tóth. Random tree growth with branching processes - a survey. In B Bollobás, R Kozma, and D Miklós, editors, *Handbook of Large-Scale Random Networks*, volume 18 of *Bolyai Society Mathematical Studies*, chapter 4. Springer, 2009.

[41] Anna Rudas, Bálint Tóth, and Benedek Valkó. Random trees and general branching processes. *Random Struct. Algorithms*, 31(2):186–202, 2007.

[42] Anna Rudas and Imre Péter Tóth. Entropy and Hausdorff dimension in random growing trees. *Stochastics and Dynamics*, Accepted for publication: 2011. 12. 15. DOI Number: 10.1142/S0219493712500104.

[43] Henry Scheffé. A useful convergence theorem for probability distributions. *Ann. Math. Statist.*, 18(3):434–438, 1947.

[44] Robert T. Smythe and Hosam M. Mahmoud. A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.*, (51):1–29, 1994.

[45] 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.

[46] Remco van der Hofstad. *Random Graphs and Complex Networks*. 2011, http://www.win.tue.nl/~rhofstad/NotesRGCN2011.pdf. in preparation.

A Figures

The following figures visualize the difference in the random tree after 1000 discrete time steps, depending on the choice of the weight function w. From a repellent case ($w(k) = \frac{1}{k+1}$), through the constant, square root and linear cases, we gradually reach the extremely preferential case ($w(k) = (k+1)^{3/2}$), the last choice producing the "blow-up" phenomenon studied in [36].

The simulation was written in `Python` and the resulting data set is visualized through a certain spring-embedded algorithm in `Cytoscape`, both softwares are freeware (the layout used inside Cytoscape is called `yFiles:Organic`).

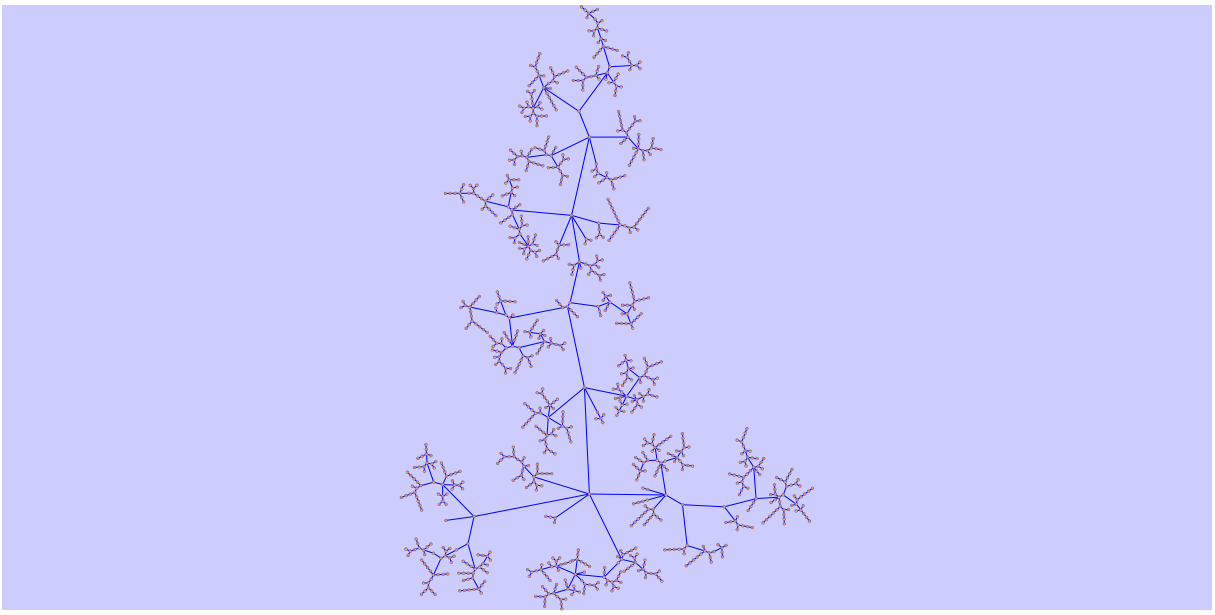

Figure 1: $w(k) = \frac{1}{k+1}$, 1000 discrete time steps

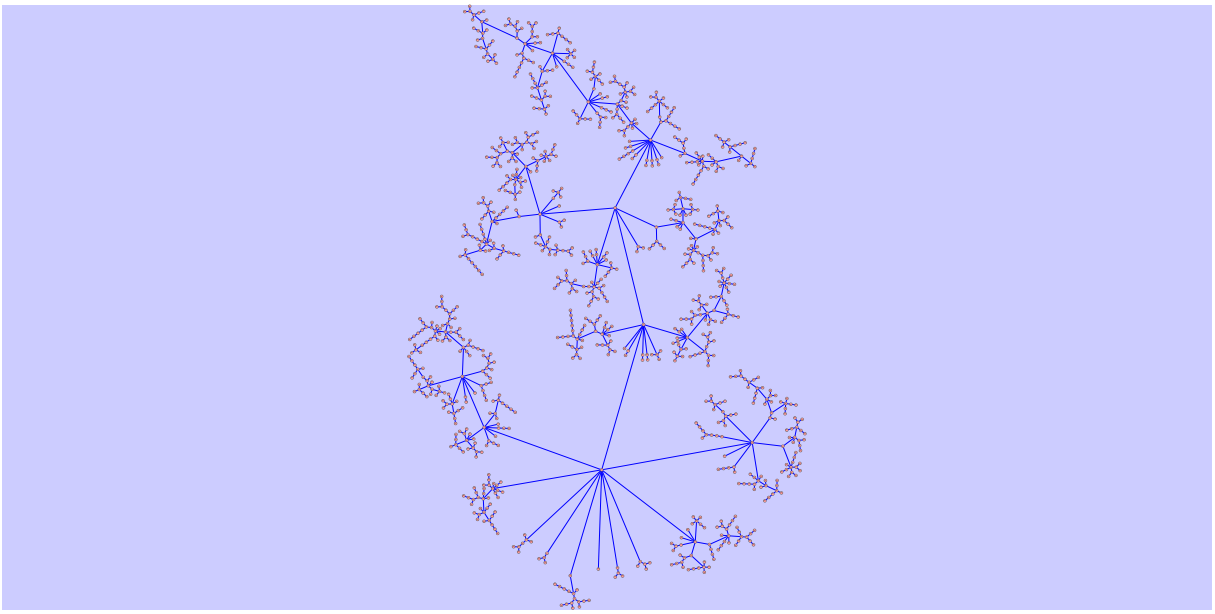

Figure 2: $w(k) = 1$, 1000 discrete time steps

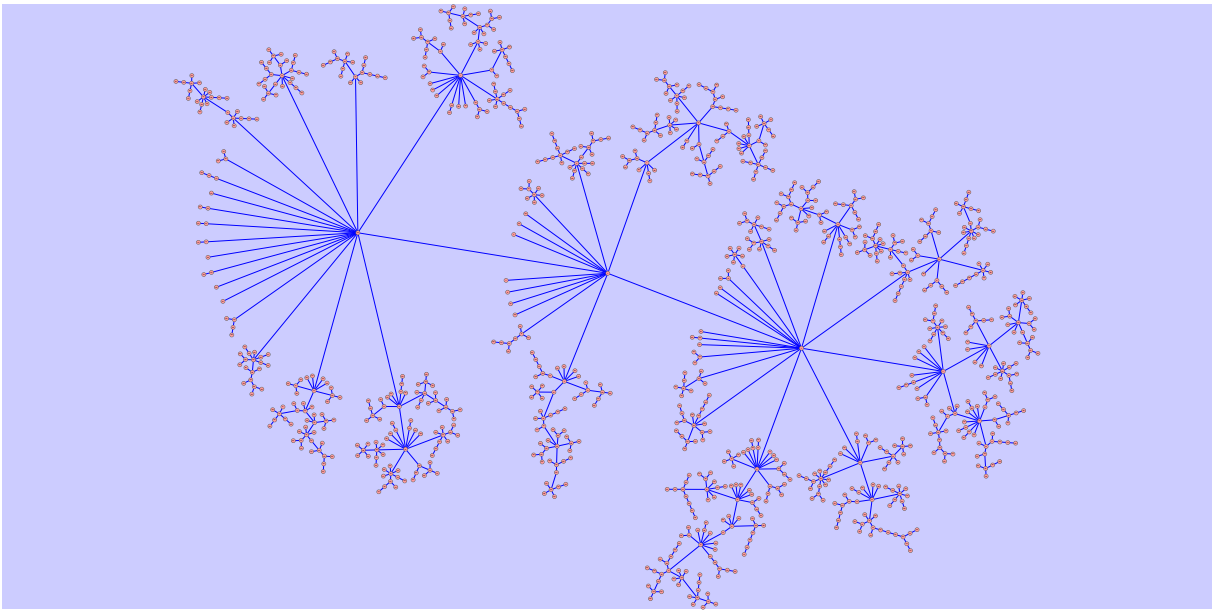

Figure 3: $w(k) = \sqrt{k+1}$, 1000 discrete time steps

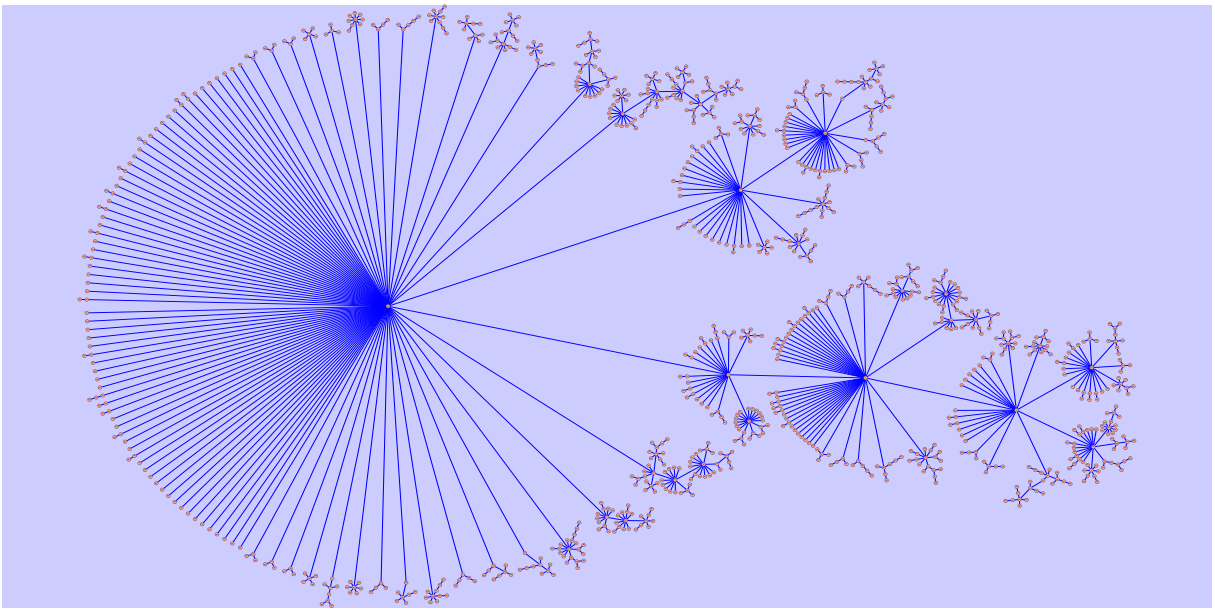

Figure 4: $w(k) = k + 1$, 1000 discrete time steps

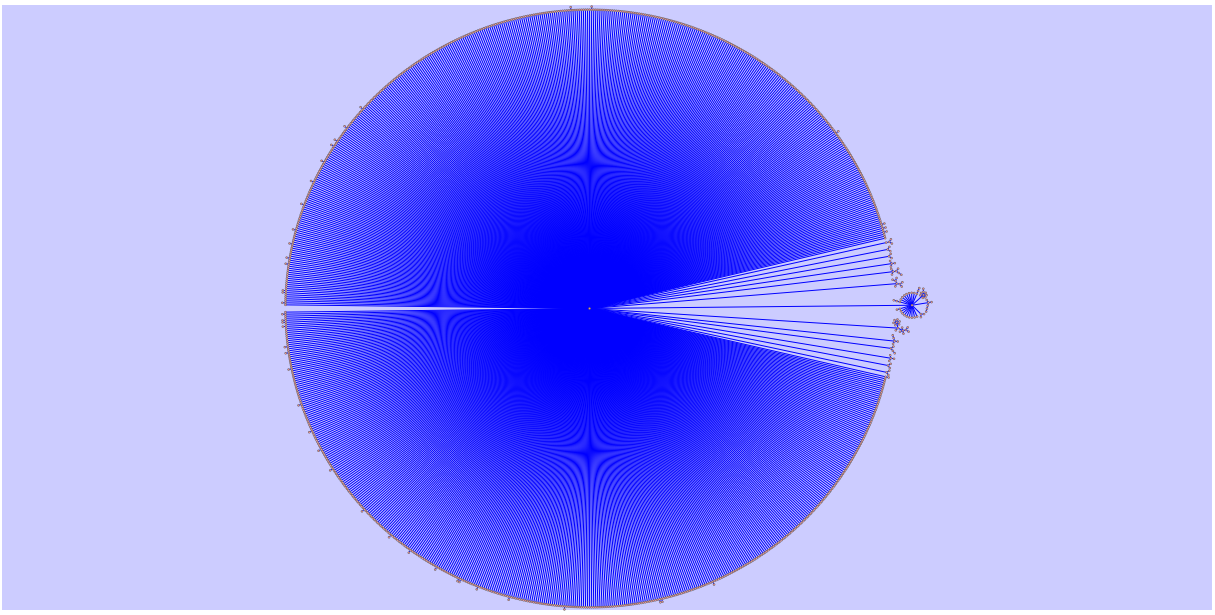

Figure 5: $w(k) = (k + 1)^{3/2}$, 1000 discrete time steps