Electronic Journal of Probability

Electron. J. Probab. **24** (2019), no. 2, 1–44.
ISSN: 1083-6489 https://doi.org/10.1214/19-EJP263

Heavy subtrees of Galton-Watson trees with an application to Apollonian networks

Luc Devroye* Cecilia Holmgren† Henning Sulzbach‡

Abstract

We study heavy subtrees of conditional Galton-Watson trees. In a standard Galton-Watson tree conditional on its size being n, we order all children by their subtree sizes, from large (heavy) to small. A node is marked if it is among the k heaviest nodes among its siblings. Unmarked nodes and their subtrees are removed, leaving only a tree of marked nodes, which we call the k-heavy tree. We study various properties of these trees, including their size and the maximal distance from any original node to the k-heavy tree. In particular, under some moment condition, the 2-heavy tree is with high probability larger than cn for some constant $c > 0$, and the maximal distance from the k-heavy tree is $O(n^{1/(k+1)})$ in probability. As a consequence, for uniformly random Apollonian networks of size n, the expected size of the longest simple path is $\Omega(n)$. We also show that the length of the heavy path (that is, $k = 1$) converges (after rescaling) to the corresponding object in Aldous' Brownian continuum random tree.

Keywords: branching processes; fringe trees; spine decomposition; binary trees; continuum random tree; Brownian excursion; exponential functionals; Apollonian networks.
AMS MSC 2010: Primary 60J80, Secondary 60J85; 05C80.
Submitted to EJP on November 8, 2017, final version accepted on January 3, 2019.
Supersedes arXiv:1701.02527.

1 Introduction

We study Galton-Watson trees of size n. More precisely, we have a generic random variable ξ defined by

$$\mathbf{P}(\xi = i) = p_i, \quad 0 \leq i < \infty,$$

*School of Computer Science, McGill University, 3480 University Street, H3A 0E9 Montreal, QC, Canada.
E-mail: lucdevroye@gmail.com
†Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden.
E-mail: cecilia.holmgren@math.uu.se
‡School of Computer Science, McGill University, 3480 University Street, H3A 0E9 Montreal, QC, Canada.
Present address: School of Mathematics, University of Birmingham, Birmingham B15 2TT, Great Britain.
E-mail: henning.sulzbach@gmail.com

where $(p_i)_{i\geq 0}$ is a fixed probability distribution. Throughout the paper, we assume that

$$\mathbf{E}\left[\xi\right] = 1, \quad \text{and } 0 < \sigma^2 := \mathbf{E}\left[\xi^2\right] - 1 < \infty. \tag{1.1}$$

The random variable ξ is used to define a critical Galton-Watson process (see, e.g. [11]). In a standard construction, we label the nodes of the Galton-Watson tree in *preorder*, that is, by sorting them as they appear first in the depth first traversal. See Figure 1 for an example. If $\xi_1, \xi_2, \ldots$ are independent copies of ξ, then we assign ξ_i children to node i. In particular, we define the size of a tree $\mathcal{T}$ as

$$|\mathcal{T}| = \min\{t \geq 1 : S_t = -1\}, \quad \text{where } S_n := \sum_{i=1}^{n} (\xi_i - 1), \quad n \geq 0. \tag{1.2}$$

This is a Galton-Watson tree. Given $|\mathcal{T}| = n$, $\mathcal{T}$ is a conditional Galton-Watson tree. The associated random walk $(S_i)_{0\leq i\leq n}$ with $S_n = -1$ and $S_i \geq 0$ for all $0 \leq i \leq n$ is called *Łukasiewicz path*. (We extend this walk to a continuous function S_t by linear interpolation. See Figure 1.)

The family of conditional Galton-Watson trees has gained importance in the literature because it encompasses the simply-generated trees introduced by Meir and Moon [47], which are basically ordered rooted trees (of a given size) that are uniformly chosen from a class of trees. For example, when $p_0 = p_2 = 1/4, p_1 = 1/2$, the conditional Galton-Watson tree corresponds to a binary tree of size n chosen uniformly at random. When $(p_i)_{i\geq 0}$ is Poisson(1), then we obtain a random labeled rooted tree, also called a Cayley tree.

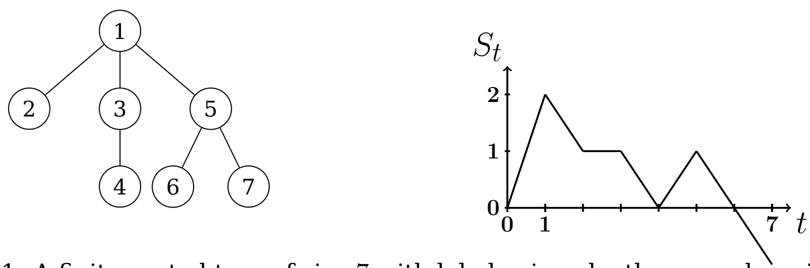

Figure 1: A finite rooted tree of size 7 with labels given by the preorder with associated Łukasiewicz path.

1.1 The asymptotic behaviour of Galton-Watson trees

In order to put the results of this paper into perspective, we shortly discuss the two main approaches towards limit theorems on conditional Galton-Watson trees with respect to their global and local behavior.

First, thanks to Aldous' groundbreaking work [5, 6, 7], it is well-known that conditional Galton-Watson trees converge (in a suitable sense and as random metric spaces endowed with the graph distance) after rescaling of edge-lengths by $\sqrt{n}$ in distribution to the Brownian continuum random tree. In this context, see also the work of Le Gall [43] and Marckert and Mokkadem [46]. We review these results in more detail in Section 6.

Second, as n grows large, the Galton-Watson tree in the vicinity of the root is described by the so-called *size-biased Galton-Watson tree* in the sense of Aldous-Steele (or Benjamini-Schramm [13]) convergence [9]. This infinite (but locally finite) random tree was introduced by Kesten [40] and is related to the so-called *spine decomposition* of the Galton-Watson tree. Compare Lyons, Pemantle and Peres [44], Lyons and Peres

[45, Chapter 12], Aldous and Pitman [8, Section 2.5] and Janson [36, Section 7]. More details and precise statements in this context are presented in Section 3.

The present paper looks at a less natural decomposition of the conditional Galton-Watson tree, but one that has far-reaching applications in computer science and the study of random networks, more precisely, random Apollonian networks.

1.2 Heavy subtrees and main results

One can reorder all sets of siblings by subtree size, from large to small, where ties are broken by considering the preorder index. For a node v in the (conditional or not) Galton-Watson tree distinct from its root, we denote by ρ_v the rank in its ordering (for example, $\rho_v = 1$ means that v has the largest subtree among its siblings). No rank is defined for the root. Let $A_v = (v_1, \ldots, v_d = v)$ be the sequence of ancestors of v if v is at distance $d \geq 1$ from the root. (The root does not appear in this sequence, and the node $v_i, 1 \leq i \leq d$, has distance i from the root.) We define the maximal rank

$$\rho_v^* = \max(\rho_{v_1}, \ldots, \rho_{v_d}).$$

For a fixed integer k, we define the *k-heavy Galton-Watson tree* as the tree formed by the root and all nodes v in the conditional Galton-Watson tree with $\rho_v^* \leq k$. In particular, as nodes in the k-heavy tree have rank at most k, they have out-degree k or less. For $k = 1$, we obtain a path, which we call the *heavy path*—just follow the path from the root down, always going to the largest subtree, taking the oldest branch in case of a tie.

Our main interest is the study of the case $k = 2$, the 2-heavy Galton Watson tree. We show that it captures a huge chunk of the Galton-Watson tree by proving the following result:

Theorem 1.1. *Consider a Galton-Watson tree whose offspring distribution satisfies* (1.1) *with* $\mathbf{E}\left[\xi^5\right] < \infty$ *conditional on having size* n *where* $\mathbf{P}(S_n = -1) > 0$.

(i) *There exists a constant* $\kappa > 0$ *such that*

$$\lim_{n\to\infty} \mathbf{P}\left(\text{Size of the 2-heavy tree} \geq \kappa n\right) = 1. \quad (1.3)$$

(ii) *There exists a constant* $\nu > 0$ *such that*

$$\mathbf{E}\left[\text{Size of the 2-heavy tree}\right] \geq \nu n.$$

Since the number of nodes of degree i in a conditional Galton-Watson tree is in probability asymptotic to np_i, it is easy to see that the size of the 2-heavy tree cannot be more than

$$n\left(1 - \sum_{i\geq 3}(i-2)p_i + o(1)\right), \quad (1.4)$$

so that there is no hope of replacing κn by $n - o(n)$ in (1.3). In fact, we believe that the size of the 2-heavy tree satisfies a law of large numbers when rescaled by n^{-1} as $n \to \infty$ with a limiting constant depending on the distribution of ξ. The condition $\mathbf{E}\left[\xi^5\right] < \infty$ is of technical nature, and we believe that the statement holds under the finite variance assumption (1.1). A related interesting statistic is the maximal size of *any* binary subtree of the conditional Galton-Watson tree. The lower bounds in Theorem 1.1 and the upper bound (1.4) also apply to this quantity, which we think deserves further studies.

We also study the maximal distance to the k-heavy trees. For a non-empty (connected or not) subgraph A of the conditional Galton-Watson tree, we call the maximal distance to A

$$d_{\max}(A) := \max_{v\notin A} \min_{w\in A} \operatorname{dist}(v, w),$$

where $\mathrm{dist}(\cdot,\cdot)$ refers to path distance between vertices in the conditional Galton-Watson tree. (This definition makes sense in any finite ordered tree.) The maximal distance to the k-heavy tree measures to some extent how pervasive the k-heavy tree is. In the next theorem, we let $\mathcal{H}_k$ denote the k-heavy subtree. Further, we write A_k for the set of all subtrees of the conditional Galton-Watson tree in which every node has at most k children. Observe that A_k is in general much larger than the collection of all subtrees of $\mathcal{H}_k$.

Theorem 1.2. *Consider a Galton-Watson tree whose offspring distribution satisfies* (1.1) *conditional on having size n where $\mathbf{P}(S_n = -1) > 0$. Let $k \geq 2$.*

(i) *If $\mathbf{E}\left[\xi^{k+3}\right] < \infty$, then, for any $\varepsilon > 0$, there exists a constant $C_* > 0$ such that*

$$\mathbf{P}\left(d_{\max}(\mathcal{H}_k) \leq C_* n^{1/(k+1)}\right) \geq 1-\varepsilon. \tag{1.5}$$

(ii) *If $\mathbf{E}\left[\xi^{k+2}\right] < \infty$ and $\sum_{\ell \geq k+1} p_\ell > 0$, then, for any $\varepsilon > 0$, there exists $c_* > 0$ such that*

$$\mathbf{P}\left(\inf_{T \in A_k} d_{\max}(T) \geq c_* n^{1/(k+1)}\right) \geq 1-\varepsilon. \tag{1.6}$$

The theorem shows that, under appropriate moment conditions on ξ, the k-heavy subtree exhausts the entire tree asymptotically optimally since every k-ary subtree leaves out nodes of distance order $n^{1/(k+1)}$ away. (Here, the choice of a k-ary subtree can even depend on the realization of the conditional Galton-Watson tree.) In particular, under the fifth moment condition from Theorem 1.1, the maximal distance from the 2-heavy tree is $\Theta(n^{1/3})$, a result that cannot possibly be deduced from standard continuum random tree results for conditional Galton-Watson trees [5, 6, 7, 43]. In Proposition 5.2 in Section 5 we give some results on necessary moment conditions on ξ to guarantee tightness of the sequence $n^{-1/(k+1)} \inf_{T \in A_k} d_{\max}(T)$, $n \geq 1$.

We finally study the length L_n of the heavy path.

Theorem 1.3. *Consider a Galton-Watson tree whose offspring distribution satisfies* (1.1) *conditional on having size n where $\mathbf{P}(S_n = -1) > 0$.*

(i) *There exists a non-negative random variable L_∞ such that, as $n \to \infty$, in distribution and with convergence of all moments,*

$$\frac{L_n}{\sqrt{n}} \to L_\infty.$$

(ii) *For $k \geq 0$, let $P_n(k)$ be the size of the subtree rooted at the node on level k on the heavy path. There exists a random decreasing process $P_\infty(t), t \in [0,1]$, with càdlàg paths such that, in distribution, in the Skorokhod topology on the set of càdlàg functions,*

$$\left(\frac{P_n(\lfloor t\sqrt{n}\rfloor)}{n}\right)_{0 \leq t \leq 1} \to P_\infty.$$

(iii) *For $0 \leq \ell \leq n$, let $Q_n(\ell) = \inf\{k \geq 0 : P_n(k) \leq \ell\}$. There exists a random continuous decreasing function $Q_\infty(t), t \in [0,1]$ such that, in distribution, on the space of continuous functions on $[0,1]$,*

$$\left(\frac{Q_n(\lfloor tn\rfloor)}{\sqrt{n}}\right)_{0 \leq t \leq 1} \to Q_\infty.$$

In Section 6, we discuss more detailed properties of L_∞ including the existence of a density (see (6.13)), a characterization of its distribution by a stochastic fixed-point equation (see (6.15)) and its (negative) moments (see (6.11) and (6.12)). Theorem 6.12 contains a more precise statement of Theorem 1.3 identifying the limiting random variables as functionals of a Brownian excursion. In particular, and as opposed to the k-heavy trees for $k \geq 2$, the heavy path can be studied using the global picture sketched in Section 1.1 above, and the distributions of the scaling limits L_∞, P_∞ and Q_∞ depend only on σ. The proof of Theorem 1.3 further reveals that the convergences in (i), (ii) and (iii) are joint and that the limiting objects are natural statistics in the continuum random tree. In this context, we draw connections to self-similar fragmentation processes studied by Bertoin [15, 16] and exploit results from his work.

Further, we study the tail behaviour of L_∞ near 0 and ∞ in more detail. In particular, we note that, at 0, it grows more slowly than any polynomial but much faster than the theta law which is the scaling limit of the tree height H_n, see (5.3) in Section 5. (H_n is equal to the maximal distance of a node from the root in τ_n.) Thus, the obvious inequality $L_n \leq H_n$ is loose. This is formulated in Proposition 6.2.

1.3 Apollonian networks

In 1930, Birkhoff [20] introduced a model of a planar graph that became known as an Apollonian network, a name coined by Andrade et al. [10] in 2005. Suggested as toy models of social and physical networks with remarkable properties, they are recursively defined by starting with three vertices that form a triangle in the plane. Given a collection of triangles in a triangulation, choose one (either at random, or following an algorithm), place a new vertex in its center, and connect it with the three vertices of the triangle. So, in each step, we create three new edges, one new point, and three new triangles (which replace an old one). After n steps, we have $3+n$ vertices, and $3+3n$ edges in the graph. This is an Apollonian network. One can also define a corresponding evolutionary tree: start with the original triangle as the root of a tree. In a typical step, select a leaf node of the tree (which corresponds to a triangle) and attach to it three children. This tree has a one-to-one relationship with the Apollonian network. It has $1+2n$ leaves (after n steps) and $1+3n$ vertices. (In particular, the n non-leaves in the tree correspond to the nodes in the Apollonian network lying strictly inside the initial triangle.) See Figure 2 for an illustration.

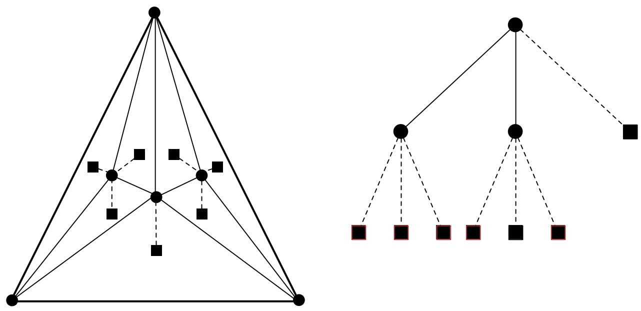

Figure 2: Apollonian network of size 3 with evolutionary tree. Leaves are drawn as boxes. Note that the outer three vertices in the network have no counterparts in the tree.

Random Apollonian networks. The most frequently studied random Apollonian network (see Zhou, Yan and Wang [53]) is one in which each triangle (in the network)—or, equivalently, each leaf in the tree—is chosen uniformly at random for splitting, leading to a so-called split tree [28]. Asymptotically, its height after n steps is bounded almost surely by $c \log n$ for a suitable constant $c > 0$ [22]. Typical distances, the diameter and node degrees in the network have recently been studied in a number or papers using probabilistic, combinatorial and analytic methods [3, 31, 42, 34, 26].

For this paper, the work on the longest simple path in the Apollonian network is most relevant. The asymptotic behavior of its length $\mathcal{L}_n^{\mathrm{rec}}$ is still not well understood today. A series of papers in recent years including [34] and [31] have culminated in the work of Collevecchio, Mehrabian and Wormald [25] who showed that $\mathcal{L}_n^{\mathrm{rec}}$ is with high probability at most of order $n^{1-\varepsilon}$ where ε can be chosen $4 \cdot 10^{-8}$.

Our main motivation to study k-heavy trees was to understand the length of the longest simple path $\mathcal{L}_n^{\mathrm{unif}}$ in the probabilistic model where we generate a random ordered tree of size $1 + 3n$ in which each non-leaf node has three children, such that all trees are equally likely. This corresponds to a conditional Galton-Watson tree (of size $1 + 3n$) with $p_0 = 2/3$ and $p_3 = 1/3$. We call the random network with this underlying evolutionary tree the *uniform* Apollonian network. With methods from analytic combinatorics, Darrasse and Soria [27] studied the degree distribution in this network. With similar techniques, Bodini, Darrasse and Soria [21] investigated typical distances. Relying on more probabilistic arguments, Albenque and Marckert [3] proved that a uniform Apollonian network possesses the *same* scaling limit as its evolutionary tree, namely the Brownian continuum random tree. In particular, typical distances and the diameter of the graph grow proportionally to the square root of the number of nodes. As a result, uniform Apollonian networks reveal a strikingly different behaviour than random recursive ones.

The length of the longest simple path $\mathcal{L}_n^{\mathrm{unif}}$ in a uniform Apollonian network is bounded from below by the size of any binary subtree embedded in the evolutionary Galton-Watson tree divided by two. (In fact, this is a deterministic bound valid in any Apollonian network, and the argument is essentially given in [31, Section 4], albeit in a different language. For the reader's convenience, we reproduce the proof in Appendix A.) In particular, $\mathcal{L}_n^{\mathrm{unif}}$ is larger than half the size of the 2-heavy tree. Therefore, by Theorem 1.1, and contrary to the situation in recursive Apollonian networks, the length of the longest simple path is not sublinear: there exists $c > 0$ such that

$$\lim_{n\to\infty} \mathbf{P}\left(\mathcal{L}_n^{\mathrm{unif}} \geq cn\right) = 1. \tag{1.7}$$

Similarly to (1.4), any c satisfying the last display is bounded away from 1. This follows from Lemma 3.1 in [31] stating that any simple path in an Apollonian network visits at most eight grandchildren of any vertex in the evolutionary tree and the fact that there is a positive proportion of nodes in a conditional Galton-Watson tree with $p_0 = 2/3$ and $p_3 = 1/3$ with nine grandchildren.

1.4 Notation

Throughout the paper, we use

- $h = \gcd\{i : p_i > 0, i > 0\}$,
- $\alpha = h/(\sigma\sqrt{2\pi})$,
- $I = \{n \geq 1 : \mathbf{P}(S_n = -1) > 0\}$,
- $I_n = \{1 \leq k \leq n : \mathbf{P}(S_{n-k} = 0) > 0\}$, for $n \in I$.

Here, for a set of integers J, $\gcd(J)$ denotes the greatest common divisor of all elements in J. From Bézout's lemma, it follows that $I = (\mathbb{N}h + 1)\backslash A$ for some finite set $A \subseteq \mathbb{N}$. We write

- $\mathcal{T}$ for a generic realization of the unconditional Galton-Watson tree,
- $\mathcal{T}_1, \mathcal{T}_2, \ldots$ for a sequence of independent copies of $\mathcal{T}$,

- $\tau_n, n \in I$, for $\mathcal{T}$ conditional on having size n.

$\mathcal{T}$ and τ_n are considered as ordered rooted trees. For $v \in \tau_n$, we let

- $\xi(v)$ be the number of children of v,
- $N(v)$ be the size of the subtree rooted at v,
- $H(v)$ be the height of the subtree rooted at v,
- $N_k(v)$ be the size of the k-th largest subtree rooted at a child of v, abbreviating $N_k(v) = 0$ if $k > \xi(v)$,
- $N_{k+}(v) = N_k(v) + N_{k+1}(v) + \dots$

We write $\varnothing$ for the root and abbreviate $\xi_\varnothing = \xi(\varnothing), N = N(\varnothing), H = H(\varnothing), N_k = N_k(\varnothing)$ and $N_{k+} = N_{k+}(\varnothing)$. (To increase readability, we suppress the dependence on n of these quantities in the notation.) For $n \in I$, if the context requires the indication of the size of the tree, we also write $\xi_\varnothing(n), \mathcal{N}(n), \mathcal{H}(n), \mathcal{N}_k(n)$ and $\mathcal{N}_{k+}(n)$ for the corresponding quantities in τ_n when referring to the root $\varnothing$. Finally, we let

- $Y_i = |\{v \in \tau_n : N(v) = i\}|, \quad i \geq 1.$

In all sections of this work with the exception of Section 6.2, all constants except $c, c_1, c_2, \dots, C, C_1, C_2, \dots$ carry fixed values. The values of constants used multiple times may vary between two results or proofs but not within. Here, constants $C, C_1, C_2, \dots > 0$ are meant to carry large values, whereas $c, c_1, c_2, \dots > 0$ are typically small.

1.5 Outline

The paper is organized as follows: first, in Section 2, we recall standard material on the size of the Galton-Watson tree $\mathcal{T}$ as well as recent results about the number of fringe trees in τ_n due to Janson [37]. We then state some related preliminary bounds in Lemma 2.2 and Corollary 2.3 for later purposes. In Section 3 we study the distribution of the subtree sizes of the children of the root in the conditional Galton-Watson tree. Most notably, we provide bounds on the corresponding distribution functions in Theorem 3.4. Apart from applying these bounds in subsequent sections, we think they are of independent interest. In Section 4 we study the 2-heavy tree and prove Theorem 1.1. Here, the proof of (1.3), the main part of the work, relies on a second moment argument. Section 5 is devoted to the proof of Theorem 1.2. While the upper bound in part (i) follows rather straightforwardly from our tools derived in earlier sections, the lower bound in (ii) relies on deeper results on the concentration of the number of fringe trees in [37]. Finally, in Section 6 we study the heavy path. The techniques used in this section differ substantially from the remaining content of the paper. In particular, Section 6.2 can be read independently of the remainder of this work.

Acknowledgements

The research of Luc Devroye was in part supported by an NSERC Discovery grant. The research of Cecilia Holmgren was supported by a grant of the Swedish Research Council, which allowed her visit at McGill University in November 2015 during which most of the work was carried out, and grants of the Ragnar Söderberg Foundation and the Knut and Alice Wallenberg Foundation. The research of Henning Sulzbach was supported by a Feodor Lynen Research Fellowship of the Alexander von Humboldt Foundation. The authors would like to thank Louigi Addario-Berry and Christina Goldschmidt for valuable discussions in particular regarding the arguments involved in the study of the heavy path. Further, the authors also thank Alexander Iksanov for helpful comments and for drawing our attention to Aurzada [12] leading to an improvement of the statement in Lemma 6.3 in an earlier version of this work. The authors also thank two anonymous referees for valuable comments.

2 Preliminary results and fringe trees

Let us start by recovering some classical results which have proved fruitful in the analysis of conditional Galton-Watson trees. Throughout this section we use the notation $\alpha, h, I, I_n, \mathcal{T}, \tau_n$ and Y_k as introduced in Section 1.4. Recall the following well-known identity going back to Dwass [30] (compare also Janson [36, Theorem 15.5] and the discussion therein),

$$\mathbf{P}(|\mathcal{T}| = n) = \frac{\mathbf{P}(S_n = -1)}{n}. \tag{2.1}$$

More generally, for independent copies $\mathcal{T}_1, \mathcal{T}_2, \ldots$ of $\mathcal{T}$,

$$\mathbf{P}(|\mathcal{T}_1| + \ldots + |\mathcal{T}_k| = n) = \frac{k}{n}\mathbf{P}(S_n = -k). \tag{2.2}$$

In this context, we cite a classical result for sums of independent integer random variables applied to the sequence S_n. By Petrov [49, Theorem VII.1] or Kolchin [41, Theorem 1.4.2], as $n \to \infty$,

$$\sup_{x \in \mathbb{Z}h - n} \left| \mathbf{P}(S_n = x) - \frac{\alpha}{\sqrt{n}} \exp\left(-\frac{x^2}{2\sigma^2 n}\right) \right| = o(n^{-1/2}). \tag{2.3}$$

In particular, for $x = o(\sqrt{n})$ with $x \in \mathbb{Z}h - n$, as $n \to \infty$,

$$\mathbf{P}(S_n = x) \sim \frac{\alpha}{\sqrt{n}}. \tag{2.4}$$

Similarly, as $n \to \infty, n \in \mathbb{N}h + 1$,

$$\mathbf{P}(S_n = -1) \sim \frac{\alpha}{\sqrt{n}}. \tag{2.5}$$

By summation, using (2.1) and (2.5), as $t \to \infty$,

$$\mathbf{P}(|\mathcal{T}| \geq t) \sim \frac{2\alpha}{h\sqrt{t}}. \tag{2.6}$$

For $n \in I, 1 \leq k \leq n$, the study of Y_k is closely related to the analysis of a random fringe subtree τ_n^* of the conditional Galton-Watson tree τ_n, that is, a subtree of τ_n rooted at a uniformly chosen node. For example, we have $\mathbf{E}[Y_k] = n \cdot \mathbf{P}(|\tau_n^*| = k)$. The study of random fringe subtrees was initiated by Aldous [4], who showed that, under assumption (1.1),

$$\tau_n^* \xrightarrow{d} \mathcal{T}, \quad \text{that is, } \mathbf{P}(\tau_n^* = \mathbf{t}) \to \mathbf{P}(\mathcal{T} = \mathbf{t}) \tag{2.7}$$

for all finite ordered rooted trees $\mathbf{t}$. In particular, $\mathbf{P}(|\tau_n^*| = k) \to \mathbf{P}(|\mathcal{T}| = k)$ as $n \to \infty, n \in \mathbb{N}h + 1$ for $k \in I$ fixed. For generalizations of Aldous' results, see Bennies and Kersting [14] and Janson [36]. More recently, Janson [37, Theorem 1.5] obtained finer results on subtree counts, in particular estimates and asymptotic expansions for the variance and a central limit theorem. We summarize special cases of his results in the following proposition. The exact expressions for mean and variance are contained in [37, Lemma 5.1] and [37, Lemma 6.1]. The uniform estimate on the variance (2.9) follows from [37, Theorem 6.7].

Proposition 2.1 (Janson [37])**.** *Let $n \in I$ and $1 \leq k \leq n$. We have*

$$\mathbf{E}[Y_k] = \frac{n\mathbf{P}(S_k = -1)\mathbf{P}(S_{n-k} = 0)}{k\mathbf{P}(S_n = -1)} = n\mathbf{P}(|\mathcal{T}| = k)\frac{\mathbf{P}(S_{n-k} = 0)}{\mathbf{P}(S_n = -1)}, \tag{2.8}$$

and, for $1 \leq k \leq (n-1)/2$*,*

$$\mathbf{E}\left[Y_k(Y_k-1)\right] = \frac{n(n-2k+1)\mathbf{P}\left(S_k=-1\right)^2\mathbf{P}\left(S_{n-2k}=1\right)}{k^2\mathbf{P}\left(S_n=-1\right)},$$

while $\mathbf{E}\left[Y_k(Y_k-1)\right]=0$ *for* $k > (n-1)/2$*. For fixed* $k \in I$*, as* $n \to \infty, n \in \mathbb{N}h+1$*,*

$$\frac{\mathrm{Var}(Y_k)}{n} \to \theta^2, \quad \theta^2 = \mathbf{P}\left(|\mathcal{T}|=k\right)\left[1+\mathbf{P}\left(|\mathcal{T}|=k\right)(1-2k-\sigma^{-2})\right] > 0,$$

and,

$$\frac{Y_k - n\mathbf{P}\left(|\mathcal{T}|=k\right)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,\theta^2),$$

where $\mathcal{N}(0,\theta^2)$ *denotes a normal random variable with variance* θ^2 *and mean* 0*.*

Finally, uniformly in $1 \leq k \leq n$*, as* $n \to \infty, n \in \mathbb{N}h+1$*,*

$$\mathrm{Var}(Y_k) = O(n). \tag{2.9}$$

It follows from (2.4), (2.5) and (2.8) that, as $n \to \infty$, for $k = k(n) = o(n)$ and $k \to \infty$ with $n \in \mathbb{N}h+1$, $k \in I_n \cap I$, we have

$$\mathbf{E}\left[Y_k\right] \sim \frac{\alpha n}{k^{3/2}}. \tag{2.10}$$

Many arguments in this manuscript rely on bounds on the mean such as those given below.

Lemma 2.2. *There exists a constant* $n_0 \geq 1$ *such that, for all* $n \geq n_0$*,* $n \in I$*,* $k \in I_n$*,*

$$\mathbf{E}\left[Y_k\right] \leq \begin{cases} 4\sqrt{2}\alpha(n-k)^{-1/2} & \text{for } n/2 \leq k \leq n-n_0, \\ 2\sqrt{2}\alpha n k^{-3/2} & \text{for } n_0 \leq k \leq n/2, \\ \lfloor n/k \rfloor & \text{for } 1 \leq k \leq n. \end{cases}$$

Similarly, there exist constants $n_1 \geq 1$ *and* $\varsigma > 0$*, such that, for all* $n \geq n_1$*,* $n \in I$*,* $k \in I_n \cap I$*,*

$$\mathbf{E}\left[Y_k\right] \geq \begin{cases} \alpha n k^{-3/2}/2 & \text{for } n_1 \leq k \leq n/2, \\ \varsigma n & \text{for } 1 \leq k \leq n_1. \end{cases}$$

Proof. By an application of (2.4) and (2.5) to (2.8), there exists $n_0 \geq 1$, such that, for all $n_0 \leq k \leq n-n_0$,

$$\frac{\mathbf{E}\left[Y_k\right]}{n} \leq 2\alpha k^{-3/2}\sqrt{\frac{n}{n-k}} \leq \begin{cases} 2\sqrt{2}\alpha k^{-3/2} & \text{for } n_0 \leq k \leq n/2, \\ \frac{4\sqrt{2}\alpha}{n\sqrt{n-k}} & \text{for } n/2 \leq k \leq n-n_0. \end{cases}$$

This shows the first two upper bounds. The third follows immediately from the deterministic bound $Y_k \leq \lfloor n/k \rfloor$. The first lower bound follows analogously. The second lower bound follows from (2.1) and (2.10), since, for $k \in I$, we have $\mathbf{E}\left[Y_k\right]/n \to \mathbf{P}\left(|\mathcal{T}|=k\right) = \mathbf{P}\left(S_k=-1\right)/k$. ☐

Corollary 2.3. *There exists a constant* $C > 0$ *such that, for all* $M \geq n_0$ *and* $n \geq M, n \in I$*, with* n_0 *as in Lemma 2.2, we have*

$$\sum_{k=M}^{n} \mathbf{E}\left[Y_k\right]\log k \leq Cn\frac{\log M}{\sqrt{M}} + (2+4\sqrt{2}\alpha)n^{3/4}\log n.$$

Proof. By applications of the upper bound in the previous theorem, we have

$$\sum_{k=M}^{\lfloor n/2\rfloor} \mathbf{E}\left[Y_k\right]\log k \le 2\sqrt{2}\alpha n \sum_{k=M}^{\lfloor n/2\rfloor} k^{-3/2}\log k \le Cn\frac{\log M}{\sqrt{M}},$$

$$\sum_{k=\lceil n/2\rceil}^{\lfloor n-\sqrt{n}\rfloor} \mathbf{E}\left[Y_k\right]\log k \le \frac{\log n}{\sqrt{n-n+\sqrt{n}}} \sum_{k=\lceil n/2\rceil}^{\lfloor n-\sqrt{n}\rfloor} 4\sqrt{2}\alpha \le 4\sqrt{2}\alpha n^{3/4}\log n,$$

$$\sum_{k=\lceil n-\sqrt{n}\rceil}^{n} \mathbf{E}\left[Y_k\right]\log k \le n\log n \sum_{k=\lceil n-\sqrt{n}\rceil}^{n} \frac{1}{k} \le 2\sqrt{n}\log n.$$

The claim follows by summing the three terms. □

3 Subtrees of the root: local convergence

We want to understand the properties of the subtree sizes of a node in a Galton-Watson tree conditional on having size n when these trees are ordered from large to small. This section has key inequalities that will be needed throughout the paper.

Let us give more details on the size-biased Galton-Watson tree mentioned in Section 1.1. Its construction goes as follows: Let $\zeta_1, \zeta_2, \dots$ be an infinite sequence of independent random variables drawn from the size-biased distribution $(ip_i)_{i\ge 0}$. Associate ζ_i with the i-th node on a one-sided infinite path (the spine). To every node i on the path assign $(\zeta_i - 1)$ children off the path, and make each child the root of an independent (unconditional) Galton-Watson tree. The ordered infinite size-biased Galton-Watson is obtained by choosing a uniform ordering on the children of every node on the infinite spine. A formulation of the local convergence result discussed in Section 1.1 is given in the following well-known proposition which is equivalent to Lemma 1 in Devroye [29] and closely related to Lemma 1.14 in Kesten [40]. (The convergence of $\xi_\varnothing$ had already been obtained by Kennedy [39].) Here, by $S_\downarrow$, we denote the set of non-negative integer valued sequences $x_1, x_2, \dots$ with $x_1 \ge x_2 \ge \dots$ and only finitely many non-zero elements. For $k \ge 1$ and $1 \le i \le k$, and real-valued random variables $X_1, \dots, X_k$, denote by $X_{(i:k)}$ the $(k-i+1)$-st order statistic. That is, $X_{(k:k)} \le X_{(k-1:k)} \le \dots \le X_{(1:k)}$. (For random trees $\mathcal{T}_1, \dots, \mathcal{T}_k$, we simplify the notation and write $|\mathcal{T}_{(i:k)}|$ for the size of i-th largest tree.)

Proposition 3.1. *Let ζ have the size-biased distribution $(ip_i)_{i\ge 0}$. For $k \ge 1$, let N_k be the size of the k-th largest subtree of a child of the root in τ_n. Then, as $n \to \infty$, $n \in \mathbb{N}h + 1$, in distribution on $S_\downarrow$,*

$$(N_2, N_3, \dots) \to (|\mathcal{T}_{(1:\zeta-1)}|, \dots, |\mathcal{T}_{(\zeta-1:\zeta-1)}|, 0, 0, \dots),$$

where $\mathcal{T}_1$, $\mathcal{T}_2, \dots, \zeta$ are independent and $\mathcal{T}_1, \mathcal{T}_2, \dots$ are copies of the unconditional Galton-Watson tree $\mathcal{T}$. In distribution, $\xi_\varnothing \to \zeta$, where $\xi_\varnothing$ is the number of children of the root of τ_n. For $k \in \mathbb{N}$ with $\mathbf{E}\left[\xi^{k+1}\right] < \infty$, we have $\mathbf{E}\left[\xi_\varnothing^k\right] \to \mathbf{E}\left[\zeta^k\right] < \infty$.

Remark 3.2. As $n - 1 = N_1 + N_2 + \dots$, it follows that, in the notation of the proposition, $n - N_1 \to 1 + \sum_{k=1}^{\zeta-1} |\mathcal{T}_k|$ in distribution as $n \to \infty$, $n \in \mathbb{N}h + 1$.

We are interested in tail bounds on $N_k, k \ge 2$. The order is suggested by the behaviour of the limiting random variable. Note that, in point (ii) of the following proposition, we write $f(t) = \Omega(g(t))$ as $t \to \infty$ for functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to (0, \infty)$ meaning that there exists a constant $c > 0$ such that, for all t sufficiently large, $|f(t)| \ge cg(t)$.

Proposition 3.3. *Let ζ have the size-biased distribution $(ip_i)_{i\ge 0}$. Let $k \ge 1$ and assume that $\mathcal{T}_1$, $\mathcal{T}_2, \dots, \zeta$ are independent where $\mathcal{T}_1, \mathcal{T}_2, \dots$ are copies of the unconditional Galton-Watson tree $\mathcal{T}$.*

(i) *If* $\mathbf{E}\left[\xi^{k+1}\right] < \infty$*, then, as* $t \to \infty$*,*

$$\mathbf{P}\left(|\mathcal{T}_{(k:\zeta-1)}| \geq t\right) = O(t^{-k/2}).$$

(ii) *If* $\sum_{\ell \geq k+1} p_\ell > 0$*, then, as* $t \to \infty$*,*

$$\mathbf{P}\left(|\mathcal{T}_{(k:\zeta-1)}| \geq t\right) = \Omega(t^{-k/2}).$$

(iii) *Finally, if* $k \geq 2$ *and* $\mathbf{E}\left[\xi^{k+1}\right] = \infty$*, then*

$$\lim_{t\to\infty} t^{k/2}\mathbf{P}\left(|\mathcal{T}_{(k:\zeta-1)}| \geq t\right) = \infty.$$

Proof. We have

$$\mathbf{P}\left(|\mathcal{T}_{(k:\zeta-1)}| \geq t\right) \leq \sum_{\ell \geq k} p_{\ell+1}(\ell+1)\binom{\ell}{k}\mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_k| \geq t\right).$$

By (2.6), the right-hand side is asymptotically equivalent to

$$\left(\frac{2\alpha}{h\sqrt{t}}\right)^k \sum_{\ell \geq k} p_{\ell+1}(\ell+1)\binom{\ell}{k}.$$

Since $\mathbf{E}\left[\xi^{k+1}\right] < \infty$, the term is of order $t^{-k/2}$.

For (ii), choose $\ell \geq k$ with $p_{\ell+1} > 0$. Then,

$$\mathbf{P}\left(|\mathcal{T}_{(k:\zeta-1)}| \geq t\right) \geq p_{\ell+1}(\ell+1)\mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_k| \geq t\right) \sim \left(\frac{2\alpha}{h\sqrt{t}}\right)^k p_{\ell+1}(\ell+1).$$

Again, the right hand side is of order $t^{-k/2}$.

For (iii), since $\mathbf{E}\left[\xi^{k+1}\right] = \infty$, for any $C > 0$, find K sufficiently large such that $\sum_{\ell=k}^{K} p_\ell(\ell+1)\binom{\ell}{k} \geq C$. Then

$$\begin{aligned}\mathbf{P}\left(|\mathcal{T}_{(k:\zeta-1)}| \geq t\right) &\geq \mathbf{P}\left(|\mathcal{T}_{(k:\zeta-1)}| \geq t, |\mathcal{T}_{(k+1:\zeta-1)}| < t\right)\\ &= \sum_{\ell \geq k} p_{\ell+1}(\ell+1)\binom{\ell}{k}\mathbf{P}\left(|\mathcal{T}| \geq t\right)^k \mathbf{P}\left(|\mathcal{T}| < t\right)^{\ell-k}\\ &\geq C\mathbf{P}\left(|\mathcal{T}| \geq t\right)^k \mathbf{P}\left(|\mathcal{T}| < t\right)^K.\end{aligned}$$

As $t \to \infty$, using (2.6), the right hand side is equivalent to $C(2\alpha h^{-1})^k t^{-k/2}$. As C was chosen arbitrarily, the final assertion of the proposition follows. □

The next theorem give corresponding results for the conditional Galton-Watson tree. In this context, recall that we write $\xi_\varnothing$ for the number of children of the root in τ_n.

Theorem 3.4. *Let* $k \geq 2$*.*

(i) *If* $\mathbf{E}\left[\xi^{k+1}\right] < \infty$*, then there exists a constant* $\beta_k > 0$*, such that, for all* $t \geq 1, n \in I$*,*

$$\mathbf{P}\left(N_k \geq t\right) \leq \beta_k t^{(1-k)/2}. \tag{3.1}$$

Writing $N_{k+} = N_k + N_{k+1} + \ldots$*, if* $\mathbf{E}\left[\xi^{(3k+1)/2}\right] < \infty$*, a corresponding bound holds for* $\mathbf{P}\left(N_{k+} \geq t\right)$ *with* β_k *replaced by some larger constant* β_{k+}*. Similarly, bounds of the same form are valid for* $\mathbf{E}\left[\xi_\varnothing \mathbf{1}_{\{N_k \geq t\}}\right]$ *if* $\mathbf{E}\left[\xi^{k+2}\right] < \infty$*, and for* $\mathbf{E}\left[\xi_\varnothing \mathbf{1}_{\{N_{k+} \geq t\}}\right]$ *if* $\mathbf{E}\left[\xi^{(3k+3)/2}\right] < \infty$*.*

(ii) *If* $\sum_{\ell \geq k} p_\ell > 0$*, then, for any* $0 < \varepsilon < 1$*, there exist constants* $\beta^* > 0$ *and* $n_2 \geq 1$ *both depending only on* k *and* ε*, such that, for all* $n \geq n_2, n \in I$*, and* $1 \leq t \leq (1-\varepsilon)n/k$*,*

$$\mathbf{P}(N_k \geq t) \geq \beta^* t^{(1-k)/2}.$$

(iii) *Finally, if* $k \geq 3$ *and* $\mathbf{E}\left[\xi^k\right] = \infty$*, then, for any sequence* ω_n *tending to infinity and* $\varepsilon < 1/(k+1)$*, we have*

$$\lim_{n\to\infty} \inf_{\omega_n \leq t \leq \varepsilon n} \mathbf{P}(N_k \geq t) \cdot t^{\frac{k-1}{2}} = \infty. \tag{3.2}$$

Remark 3.5. The proof of Theorem 3.4 (i) shows the following stronger result: for $k \geq 2$, there exists a constant $C > 0$ such that, for all $n \in I, \ell \geq k$ and $t \geq 1$,

$$\mathbf{P}(N_k \geq t, \xi_\varnothing = \ell) \leq C p_\ell \ell^{k+1} t^{(1-k)/2}. \tag{3.3}$$

Lemma 5.3 is the only result in this work that requires this stronger bound.

Remark 3.6. Since $N_k \xrightarrow{d} \mathcal{T}_{(k-1:\zeta-1)}$ and the moment condition on this random variable in order to have tails decaying as in Proposition 3.3 (i) is tight for $k \geq 3$, it is reasonable to conjecture that a tail bound such as (3.1) holds if and only if $\mathbf{E}\left[\xi^k\right] < \infty$. (3.2) shows that the latter is indeed necessary.

From Theorem 3.4 we deduce the following corollary using the well-known formula $\mathbf{E}[X] = \int_0^\infty \mathbf{P}(X > t)\, dt$ for a non-negative random variable X.

Corollary 3.7. *As* $n \to \infty$*,* $n \in \mathbb{N}h + 1$*,*

(i) *if* $\mathbf{E}\left[\xi^3\right] < \infty$*, then* $\mathbf{E}[N_2] = \Theta(\sqrt{n})$ *and* $\mathbf{E}\left[\sqrt{N_2}\right] = \Theta(\log n)$*,*
(ii) *if* $\mathbf{E}\left[\xi^{7/2}\right] < \infty$*, then* $\mathbf{E}[N_{2+}] = \Theta(\sqrt{n})$*,*
(iii) *if* $\mathbf{E}\left[\xi^4\right] < \infty$*, then* $\mathbf{E}[N_3] = O(\log n)$*,*
(iv) *if* $\mathbf{E}\left[\xi^5\right] < \infty$*, then* $\mathbf{E}[N_{3+}] = O(\log n), \mathbf{E}\left[(N_{3+})^2\right] = O(\sqrt{n})$ *and* $\mathbf{E}[N_4] = O(1)$*,*
(v) *if* $\mathbf{E}\left[\xi^{13/2}\right] < \infty$*, then* $\mathbf{E}[N_{4+}] = O(1)$*.*

If $\sum_{\ell \geq 3} p_\ell > 0$*, then big-O in (iii) can be replaced by* Θ*.*

The remainder of this section is devoted to the proofs of Theorem 3.4. In this context the following two observations are useful.

From (2.2) and (2.3), it follows that there exists $\omega_1 > 0$ such that

$$\sup_{k>0, k\in\mathbb{N}h-n} \frac{1}{k}\mathbf{P}(|\mathcal{T}_1| + \ldots + |\mathcal{T}_k| = n) \leq \omega_1 n^{-3/2}. \tag{3.4}$$

Similarly, there exist $n_5 \in \mathbb{N}$ and $\omega_2 > 0$ such that, for all $n \geq n_5$ and $k \leq \sqrt{n}$ with $n - k \in \mathbb{N}h$,

$$\frac{1}{k}\mathbf{P}(|\mathcal{T}_1| + \ldots + |\mathcal{T}_k| = n) \geq \omega_2 n^{-3/2}. \tag{3.5}$$

Lemma 3.8. *Let* $\mathcal{T}_1, \mathcal{T}_2, \ldots$ *be independent realizations of the Galton-Watson tree* $\mathcal{T}$*. For all* $\ell, t, n \geq 1$ *and* $1 \leq k < \ell$*,*

$$\mathbf{P}(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_\ell| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_\ell| = n) \leq \frac{\omega_1^\ell 16^{\ell-1}}{n^{3/2} t^{(\ell-1)/2}},$$

and

$$\mathbf{P}(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_k| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_\ell| = n) \leq \frac{\omega_1^{k+1} 16^k (\ell - k)}{n^{3/2} t^{(k-1)/2}} \frac{1}{\sqrt{\min(kt, \ell - k)}}.$$

Lemma 3.9. *Let $\mathcal{T}_1, \mathcal{T}_2, \ldots$ be independent realizations of the Galton-Watson tree $\mathcal{T}$. Let $2 \leq \ell < m$ and $0 < \varepsilon < 1/2$.*

(i) *There exist $n_6, n_7 \geq 1$ and $c_1 > 0$ depending only on ℓ and ε, such that, for $n \geq n_6, n - \ell \in \mathbb{N}h$ and $n_7 \leq t \leq (1-\varepsilon)n/\ell$, we have*

$$\mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_\ell| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_\ell| = n\right) \geq \frac{c_1}{n^{3/2} t^{(\ell-1)/2}}.$$

(ii) *There exist $n_8, n_9 \geq 1$ and $c_2 > 0$ depending only on ℓ, m and ε, such that, for $n \geq n_8, n - m \in \mathbb{N}h$ and $n_9 \leq t \leq (1-\varepsilon)n/\ell$, we have*

$$\mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_\ell| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_m| = n\right) \geq \frac{c_2}{n^{3/2} t^{(\ell-1)/2}}.$$

The two lemmas rely on the following simple results.

Lemma 3.10. *Let $n \geq 2, a, b \geq 1$ and $\varepsilon \in (0,1)$.*

(i) *We have*

$$\sum_{k=a}^{n-b} (k(n-k))^{-3/2} \leq \frac{16}{n^{3/2}\sqrt{\min(a,b)}}.$$

(ii) *If $a + b \leq (1-\varepsilon)n$, then*

$$\sum_{k=a}^{n-b} (k(n-k))^{-3/2} \geq \frac{\varepsilon}{n^{3/2}\sqrt{\min(a,b)}}.$$

Proof. (i) The statement is trivial if $a + b > n$, so we shall assume $a + b \leq n$. By symmetry, the sum is bounded from above by

$$2 \sum_{k=\min(a,b)}^{\lfloor \frac{n}{2} \rfloor} (k(n-k))^{-3/2} \leq 2\left(\frac{2}{n}\right)^{3/2} \sum_{k=\min(a,b)}^{\infty} k^{-3/2}.$$

If $\min(a,b) = 1$, then the sum on the right hand side is equal to $\zeta(3/2) < \sqrt{8}$ which shows the claim. Otherwise, using the monotonicity of $x \mapsto x^{-3/2}$, we have the bound $\sum_{k=\min(a,b)}^{\infty} k^{-3/2} \leq \int_{\min(a,b)-1}^{\infty} x^{-3/2} dx$ which is easily seen to be bounded by $\sqrt{8/\min(a,b)}$. This concludes the proof.

(ii) Since $n - b \geq a + \varepsilon n$, the sum is bounded from below by

$$\begin{aligned}\sum_{k=a}^{\lfloor a+\varepsilon n \rfloor} (k(n-k))^{-3/2} &\geq n^{-3/2} \sum_{k=a}^{\lfloor a+\varepsilon n \rfloor} k^{-3/2} \geq n^{-3/2} \int_a^{a+\varepsilon n} x^{-3/2} dx \\ &= 2n^{-3/2}\left(\frac{1}{\sqrt{a}} - \frac{1}{\sqrt{a+\varepsilon n}}\right) \geq n^{-3/2}\frac{\varepsilon}{\sqrt{a}},\end{aligned}$$

where we used $a \leq (1-\varepsilon)n$ and $\sqrt{1-x} \geq 1 - x/2$ for $x \in [0,1]$. The claim follows from a symmetry argument. ☐

Proof of Lemma 3.8. Let $S_\ell := \{(x_1, \ldots, x_\ell) : x_1, \ldots, x_\ell \geq t, x_1 + \ldots + x_\ell \leq n - t\}$. Using (3.4), we have

$$\begin{aligned}&\mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_\ell| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_\ell| = n\right) \\ &= \sum_{(k_1,\ldots,k_{\ell-1}) \in S_{\ell-1}} \mathbf{P}\left(|\mathcal{T}_1| = k_1, \ldots, |\mathcal{T}_{\ell-1}| = k_{\ell-1}, |\mathcal{T}_\ell| = n - k_1 - \ldots - k_{\ell-1}\right) \\ &\leq \omega_1^\ell \sum_{(k_1,\ldots,k_{\ell-1}) \in S_{\ell-1}} (k_1 \cdots k_{\ell-1}(n - k_1 - \ldots - k_{\ell-1}))^{-3/2}.\end{aligned}$$

Applying Lemma 3.10 (i) multiple times,

$$\begin{aligned}
&\sum_{(k_1,\dots,k_{\ell-1})\in S_{\ell-1}} (k_1\cdots k_{\ell-1}(n-k_1-\ldots-k_{\ell-1}))^{-3/2}\\
&= \sum_{(k_1,\dots,k_{\ell-2})\in S_{\ell-2}} (k_1\cdots k_{\ell-2})^{-3/2} \sum_{t\le k_{\ell-1}\le n-t-k_1-\ldots-k_{\ell-2}} (k_{\ell-1}(n-k_1-\ldots-k_{\ell-1}))^{-3/2}\\
&\le \frac{16}{\sqrt{t}} \sum_{(k_1,\dots,k_{\ell-2})\in S_{\ell-2}} (k_1\cdots k_{\ell-2}(n-k_1-\ldots-k_{\ell-2}))^{-3/2}\\
&\le \frac{16^{\ell-1}}{n^{3/2}t^{(\ell-1)/2}}.
\end{aligned}$$

This shows the first inequality. Next, using the first inequality, (3.5), and Lemma 3.10 (i),

$$\begin{aligned}
&\mathbf{P}\left(|\mathcal{T}_1|\ge t,\dots,|\mathcal{T}_k|\ge t, |\mathcal{T}_1|+\ldots+|\mathcal{T}_\ell|=n\right)\\
&= \sum_{j=\lceil kt\rceil}^{n-(\ell-k)} \mathbf{P}\left(|\mathcal{T}_1|\ge t,\dots,|\mathcal{T}_k|\ge t, |\mathcal{T}_1|+\ldots+|\mathcal{T}_k|=j\right)\mathbf{P}\left(|\mathcal{T}_{k+1}|+\ldots+|\mathcal{T}_\ell|=n-j\right)\\
&\le \frac{\omega_1^{k+1}16^{k-1}(\ell-k)}{t^{(k-1)/2}} \sum_{j=\lceil kt\rceil}^{n-(\ell-k)} (j(n-j))^{-3/2}\\
&\le \frac{\omega_1^{k+1}16^{k}(\ell-k)}{t^{(k-1)/2}n^{3/2}} \frac{1}{\sqrt{\min(kt,\ell-k)}}.
\end{aligned}$$

This concludes the proof. □

Proof of Lemma 3.9. We assume $h=1$ to keep the focus of attention on the main arguments. The modifications in the general case $h>1$ are standard. Let $\ell\ge 2$ and $\varepsilon'\in(0,1/2)$. Let $\varepsilon\in(0,1)$ such that $(1-\varepsilon)^{\ell-1}=1-\varepsilon'$ and $t\ge n_5$. Further, let $n\ge t\ell$. By (3.5), we have

$$\begin{aligned}
&\mathbf{P}\left(|\mathcal{T}_1|\ge t,\dots,|\mathcal{T}_\ell|\ge t, |\mathcal{T}_1|+\ldots+|\mathcal{T}_\ell|=n\right)\\
&= \sum_{(k_1,\dots,k_{\ell-1})\in S_{\ell-1}} \mathbf{P}\left(|\mathcal{T}_1|=k_1,\dots,|\mathcal{T}_{\ell-1}|=k_{\ell-1}, |\mathcal{T}_\ell|=n-k_1-\ldots-k_{\ell-1}\right)\\
&\ge \omega_2^{\ell} \sum_{(k_1,\dots,k_{\ell-1})\in S_{\ell-1}} (k_1\cdots k_{\ell-1}(n-k_1-\ldots-k_{\ell-1}))^{-3/2}. \qquad (3.6)
\end{aligned}$$

For $x>0$, let $g_1(x)=x$ and, inductively, $g_{j+1}(x)=(x+g_j(x))/(1-\varepsilon), j\ge 1$. Note that, for $x>0, j\ge 1$, we have $g_j(x)\le g_{j+1}(x)$ and $jx\le g_j(x)\le jx/(1-\varepsilon)^{j-1}$. For $s\in[1,\infty), m,i,j\ge 1$, it follows from Lemma 3.10 (ii), that

$$\begin{aligned}
&\sum_{\substack{k_1,\dots,k_j\ge s\\ \sum_p k_p\le m-g_i(s)}} (k_1\cdots k_j(m-k_1-\ldots-k_j))^{-3/2}\\
&\qquad\ge \frac{\varepsilon}{\sqrt{s}} \sum_{\substack{k_1,\dots,k_{j-1}\ge s\\ \sum_p k_p\le m-g_{i+1}(s)}} (k_1\cdots k_{j-1}(m-k_1-\ldots-k_{j-1}))^{-3/2}.
\end{aligned}$$

Let $n\ge \ell n_5(1-\varepsilon)^{1-\ell}$ and $t\le(1-\varepsilon)^{\ell-1}n/\ell$ such that $g_\ell(t)\le n$. Then, starting from (3.6) and applying the last inequality recursively shows that

$$\mathbf{P}\left(|\mathcal{T}_1|\ge t,\dots,|\mathcal{T}_\ell|\ge t, |\mathcal{T}_1|+\ldots+|\mathcal{T}_\ell|=n\right)\ge n^{-3/2}\omega_2^{\ell}\left(\frac{\varepsilon}{t}\right)^{\frac{\ell-1}{2}}.$$

By Bernoulli's inequality, $\varepsilon \geq \varepsilon'/(\ell-1)$ which concludes the proof of the first statement.

To show the second inequality, note that

$$\begin{aligned}&\mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_\ell| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_m| = n\right) \\ &\geq \sum_{j=\lceil \ell t(1-\varepsilon/2)^{-1}\rceil}^{n} \mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_\ell| \geq t, \sum_{\ell=1}^{\ell} |\mathcal{T}_i| = j\right) \mathbf{P}\left(|\mathcal{T}_{\ell+1}| + \ldots + |\mathcal{T}_m| = n-j\right).\end{aligned}$$

Now, let $C := \max(n_7(\ell, \varepsilon/2), n_6(\ell, \varepsilon/2)(1-\varepsilon/2)/\ell)$ and $n \geq C\ell(1-\varepsilon/2)^{-1} + 1$. Let $C \leq t \leq (1-\varepsilon)n/\ell$. Then, by part (i) of the lemma, the first factor in the sum is bounded from below by $c_1(\ell, \varepsilon/2) t^{\frac{1-\ell}{2}} j^{-3/2}$. By (3.5), we can further bound the right hand side of the last display by

$$\frac{c_1(\ell, \varepsilon/2)}{t^{\frac{\ell-1}{2}}} \sum_{j=\lceil \ell t(1-\varepsilon/2)^{-1}\rceil}^{n-n_5-(m-\ell)^2} (j(n-j))^{-3/2}.$$

Upon increasing C if necessary, the sum is bounded from below by a term of the order $n^{-3/2}$. This concludes the proof. □

Proof of Theorem 3.4. (i) We may assume $n \in I$ and $t \geq 1$. First,

$$\mathbf{P}(N_k \geq t) \leq \sum_{\ell \geq k} p_\ell \binom{\ell}{k} \frac{\mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_k| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_\ell| = n-1\right)}{\mathbf{P}(|\mathcal{T}| = n)}.$$

By Lemma 3.8,

$$\mathbf{P}(N_k \geq t) \leq \frac{(16\omega_1)^{k+1}}{\mathbf{P}(|\mathcal{T}| = n)\, t^{(k-1)/2}(n-1)^{3/2}} \left[1 + \sum_{\ell \geq k+1} \frac{p_\ell \ell^k (\ell-k)}{(\min(kt, \ell-k))^{1/2}}\right].$$

Since $\mathbf{E}\left[\xi^{k+1}\right] < \infty$, the second factor in this display is bounded. Inequality (3.1) now follows by approximating $\mathbf{P}(|\mathcal{T}| = n)$ with the help of (2.1) and (2.5).

To move from N_k to N_{k+}, note that, for non-negative numbers $u_1, \ldots, u_n, t$, in order to have $u_1 + \ldots + u_n \geq t$, we need to have $\max(u_1, \ldots, u_n) \geq t/n$. Thus, $\mathbf{P}(N_{k+} \geq t) \leq \mathbf{P}\left(N_k \geq t(\xi_\varnothing - k + 1)^{-1}\right)$. As above,

$$\begin{aligned}\mathbf{P}(N_{k+} \geq t) \leq \mathbf{P}(\xi_\varnothing \geq \lfloor t+k \rfloor) + (\mathbf{P}(|\mathcal{T}| = n))^{-1} \cdot \\ \sum_{\ell=k}^{\lfloor t+k-1 \rfloor} p_\ell \binom{\ell}{k} \mathbf{P}\left(|\mathcal{T}_1| \geq \frac{t}{\ell-k+1}, \ldots, |\mathcal{T}_k| \geq \frac{t}{\ell-k+1}, \sum_{j=1}^{\ell} |\mathcal{T}_i| = n-1\right).\end{aligned}$$

The second summand is bounded from above by

$$\frac{(16\omega_1)^{k+1}}{\mathbf{P}(|\mathcal{T}| = n)\, t^{(k-1)/2}(n-1)^{3/2}} \left[1 + \sum_{\ell=k+1}^{\lfloor t+k-1 \rfloor} \frac{p_\ell \ell^k (\ell-k+1)^{(k+1)/2}}{\min(kt/(\ell-k+1), \ell-k))^{1/2}}\right].$$

Since $\mathbf{E}\left[\xi^{(3k+1)/2}\right] < \infty$, using the same ideas as above, the last term is at most of order $t^{(1-k)/2}$. Further, by Markov's inequality, using Proposition 3.1,

$$\mathbf{P}(\xi_\varnothing \geq \lfloor t+k \rfloor) \leq \frac{\mathbf{E}\left[\xi_\varnothing^{(k-1)/2}\right]}{\lfloor t+k \rfloor^{(k-1)/2}} = O\left(\mathbf{E}\left[\xi^{(k+1)/2}\right] \lfloor t+k \rfloor^{(1-k)/2}\right).$$

The claim follows.

(ii) Let $\ell \geq 2$ and $\lambda = \min\{i \geq \ell : p_i > 0\}$. Then,

$$\mathbf{P}\left(N_\ell \geq t\right) \geq p_\lambda (\mathbf{P}\left(|\mathcal{T}| = n\right))^{-1} \mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_\ell| \geq t, |\mathcal{T}_1| + \ldots + |\mathcal{T}_\lambda| = n - 1\right).$$

The assertion follows from Lemma 3.9 together with (2.1) and (2.5).

(iii) Let $K \geq k+2$ be an integer. We suppose that $h = 1$ for the sake of presentation. Using the first statement in Lemma 3.9, there exists $c > 0$ depending only on the offspring distribution and k, ε but not on n or K, such that, for all sufficiently large $n \in I$ and $\omega_n \leq t \leq \varepsilon n$, we have

$$\begin{aligned}
&\mathbf{P}\left(N_k \geq t\right) \mathbf{P}\left(|\mathcal{T}| = n\right) \\
&\geq \sum_{\ell=k+1}^{K} p_\ell \binom{\ell}{k} \mathbf{P}\left(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_k| \geq t, |\mathcal{T}_{k+1}| \leq \frac{t}{\ell-k}, \ldots, |\mathcal{T}_\ell| \leq \frac{t}{\ell-k}, \sum_{j=1}^{\ell} |\mathcal{T}_j| = n-1\right) \\
&= \sum_{\ell=k+1}^{K} p_\ell \binom{\ell}{k} \sum_{\substack{0 \leq c_{k+1}, \ldots, c_\ell \\ \leq t/(\ell-k)}} \mathbf{P}\Bigg(|\mathcal{T}_1| \geq t, \ldots, |\mathcal{T}_k| \geq t, \\
&\qquad\qquad \sum_{j=1}^{k} |\mathcal{T}_j| = n - 1 - \sum_{j=k+1}^{\ell} c_j\Bigg) \prod_{m=k+1}^{\ell} \mathbf{P}\left(|\mathcal{T}| = c_m\right) \\
&\geq c n^{-3/2} t^{(1-k)/2} \sum_{\ell=k+1}^{K} p_\ell \binom{\ell}{k} \sum_{0 \leq c_{k+1}, \ldots, c_\ell \leq t/(\ell-k)} \prod_{m=k+1}^{\ell} \mathbf{P}\left(|\mathcal{T}| = c_m\right) \\
&= c n^{-3/2} t^{(1-k)/2} \sum_{\ell=k+1}^{K} p_\ell \binom{\ell}{k} \left(\mathbf{P}\left(|\mathcal{T}| \leq \frac{t}{\ell-k}\right)\right)^{\ell-k}.
\end{aligned}$$

Using the asymptotic expansion of $\mathbf{P}\left(|\mathcal{T}| = n\right)$ stemming from (2.1) and (2.5), it follows that

$$\liminf_{n\to\infty} \inf_{\omega_n \leq t \leq \varepsilon n} \mathbf{P}\left(N_k \geq t\right) \cdot t^{(k-1)/2} \geq c\alpha^{-1} \sum_{\ell=k+1}^{K} p_\ell \binom{\ell}{k}.$$

The assertion (3.2) follows since the right hand side becomes arbitrarily large as $K \to \infty$. ☐

4 The 2-heavy tree

Let $\mathbb{T}$ be a *fixed* finite ordered rooted tree whose root shall be labeled $\varnothing$. As in Section 1.2, to each node $v \neq \varnothing$, we assign the rank ρ_v where $\rho_v = i$ if its subtree is the i-th largest among all the subtrees rooted at its siblings. Ties are broken by the original order in the tree. If v has distance $k \geq 1$ from $\varnothing$, let $v_0 := \varnothing, v_1, \ldots, v_{k-1}, v_k = v$ be the nodes on the path connecting the root to v where v_i has depth i. The path from $\varnothing$ to v has nodes of indices $\rho_{v_1}, \ldots, \rho_{v_k} = \rho_v$. It is called the index sequence of v and denoted by $R(v)$. We define $R(\varnothing) = \emptyset$ as the empty word. For a set A of words of finite lengths over the alphabet $\mathbb{N}$, we set $\mathcal{V}(A) = \{v \in \mathbb{T} : R(v) \in A\}$. Further, for $B \subseteq \mathbb{N}$, we write B^* for the set of all finite length (even 0) words with symbols drawn from B. For example, $\mathcal{V}(\{1\}^*)$ is the set of nodes in $\mathbb{T}$ that have all their ancestors and itself of index 1 plus the root. Of course, these nodes form the heavy path. Furthermore, we recover the k-heavy tree $\mathcal{V}(\{1, \ldots, k\}^*)$ of $\mathbb{T}$ by removing from $\mathbb{T}$ all nodes of index strictly larger than k and their subtrees. For $k = 2$, we obtain the 2-heavy tree. The 2-heavy Galton-Watson tree is denoted by $\mathcal{B}_n$ and its size by B_n. It is tempting to think that B_n is increasing in probability or, at least, in mean. The following example shows that this is not the case. Let $p_0, p_2, p_5 > 0$ with $p_0 + p_2 + p_5 = 1$. Then, on the one hand, almost surely, τ_5 is binary and $B_5 = 5$. On the other hand, almost surely, τ_6 consists of the root with five children. Thus $B_6 = 3$. Note that this issue cannot be avoided by imposing an aperiodicity condition such as $p_i > 0$ for all i.

Proof of Theorem 1.1 (ii). We show by induction that there exist $\nu_1, \nu_2 > 0$ such that

$$\mathbf{E}[B_n] \geq \nu_1 n + \nu_2 \sqrt{n} - 1/2 \tag{4.1}$$

for all $n \in I$. First, since $B_i = i$ for $i \in \{0,1,2,3\} \cap I$, we need to have

$$\nu_1 + \nu_2 \leq 3/2, \quad 2\nu_1 + \sqrt{2}\nu_2 \leq 5/2, \quad 3\nu_1 + \sqrt{3}\nu_2 \leq 7/2. \tag{4.2}$$

Assume (4.1) holds for all $j \in I \cap \{0,1,\ldots,n-1\}$ with $n \geq 4$ and $n \in I$. Then, with $b(n) = \mathbf{E}[B_n]$,

$$\begin{aligned} b(n) &= 1 + \mathbf{E}[b(N_1)] + \mathbf{E}[b(N_2)] \\ &\geq \nu_1 \mathbf{E}[N_1 + N_2] + \nu_2 \mathbf{E}\left[\sqrt{N_1} + \sqrt{N_2}\right] \\ &= \nu_1(n-1) - \nu_1 \mathbf{E}[N_{3+}] + \nu_2 \mathbf{E}\left[\sqrt{n-1-N_{2+}} + \sqrt{N_2}\right] \\ &\geq \nu_1(n-1) - \nu_1 \mathbf{E}[N_{3+}] + \nu_2\sqrt{n} - \nu_2 - \nu_2 \mathbf{E}[N_{2+}]/\sqrt{n-1} + \nu_2 \mathbf{E}\left[\sqrt{N_2}\right]. \end{aligned}$$

Here, in the last step, we have used that $1 - x \leq \sqrt{1-x}$ for all $x \in [0,1]$. By Corollary 3.7, there exist strictly positive constants C_1, c_2, C_3, such that

$$\mathbf{E}[N_{3+}] \leq C_1 \log n, \quad \mathbf{E}\left[\sqrt{N_2}\right] \geq c_2 \log n, \quad \mathbf{E}[N_{2+}] \leq C_3\sqrt{n-1}.$$

Thus,

$$b(n) \geq -\nu_1 - \nu_2 - \nu_2 C_3 + (\nu_2 c_2 - \nu_1 C_1)\log n + \nu_1 n + \nu_2 \sqrt{n}.$$

From here, the claim $b(n) \geq \nu_1 n + \nu_2\sqrt{n} - 1/2$ follows if both

$$1/2 - \nu_1 - \nu_2(1 + C_3) \geq 0, \quad \text{and,} \quad \nu_2 c_2 - \nu_1 C_1 \geq 0.$$

The last expression and all inequalities in (4.2) can simultaneously be satisfied by choosing $\nu_2 = \nu_1 C_1/c_2$ and

$$\nu_1 \leq \min\left\{\frac{3}{2(1 + C_1/c_2)}, \frac{5}{2(2 + \sqrt{2}C_1/c_2)}, \frac{7}{2(3 + \sqrt{3}C_1/c_2)}, \frac{1}{2(1 + C_1(1 + C_3)/c_2)}\right\}. \quad \square$$

To prove the first part of Theorem 1.1, we return to the setting of a fixed ordered rooted tree $\mathbb{T}$. For a node $v \in \mathbb{T}$ define by $n(v)$ the size of the subtree rooted at v. (We use $n(v)$ rather than $N(v)$ to emphasize that we work in a fixed tree $\mathbb{T}$.) Let $\mathbb{B} = \mathcal{V}(\{1,2\}^*)$ denote the 2-heavy tree in $\mathbb{T}$. For $M \geq 2$, let $\mathbb{T}_1$ be the binary subtree of $\mathbb{B}$ containing all nodes with subtree sizes (with respect to the original tree $\mathbb{T}$) at least M. That is, the vertex set of $\mathbb{T}_1$ is $\{v \in \mathbb{B} : n(v) \geq M\}$. Then, let $\mathcal{V}_2$ be the set of nodes in $\mathbb{B}$ with graph distance 1 from $\mathbb{T}_1$. By construction, $n(v) \leq M - 1$ for $v \in \mathcal{V}_2$. Furthermore, let $\mathcal{V}_4$ be the subset of nodes $v \in \mathbb{T}$ which are in a subtree rooted at a node in $\mathcal{V}_2$. (In particular, $\mathcal{V}_2 \subseteq \mathcal{V}_4$.) Next, let $\mathcal{V}_3$ be the set of nodes in $\mathbb{T}$ which are neither in $\mathbb{T}_1$ nor in $\mathcal{V}_4$. In particular, $|\mathbb{T}_1| + |\mathcal{V}_3| + |\mathcal{V}_4| = |\mathbb{T}|$. See Figure 3 for an illustration.

Note that, by construction, $|\mathcal{V}_4| \leq (M-1)|\mathcal{V}_2|$ and $|\mathbb{B}| \geq |\mathbb{T}_1| + |\mathcal{V}_2|$. Thus,

$$(M-1)|\mathbb{B}| \geq |\mathbb{T}_1| + (M-1)|\mathcal{V}_2| = |\mathbb{T}| - |\mathcal{V}_3| - |\mathcal{V}_4| + (M-1)|\mathcal{V}_2| \geq |\mathbb{T}| - |\mathcal{V}_3|.$$

We arrive at the useful inequality,

$$|\mathbb{B}| \geq \frac{|\mathbb{T}| - |\mathcal{V}_3|}{M-1}, \tag{4.3}$$

noting that the set $\mathcal{V}_3$ and its size depend on the choice of M.

Proof of Theorem 1.1 (i). Let $\mathbb{T}_1(\tau_n)$ ($\mathcal{V}_3(\tau_n)$, respectively) be $\mathbb{T}_1$ ($\mathcal{V}_3$, respectively) in the conditional Galton-Watson tree τ_n and recall that these quantities depend on M. We show that, for some (in fact, all sufficiently) large M, there exists $0 < \gamma = \gamma(M) < 1$ such that $\mathbf{P}\left(|\mathcal{V}_3(\tau_n)| \geq \gamma n\right) \to 0$ as $n \to \infty$. In particular, by (4.3),

$$\mathbf{P}\left(B_n \leq n(1-\gamma)/(M-1)\right) \to 0$$

proving the theorem.

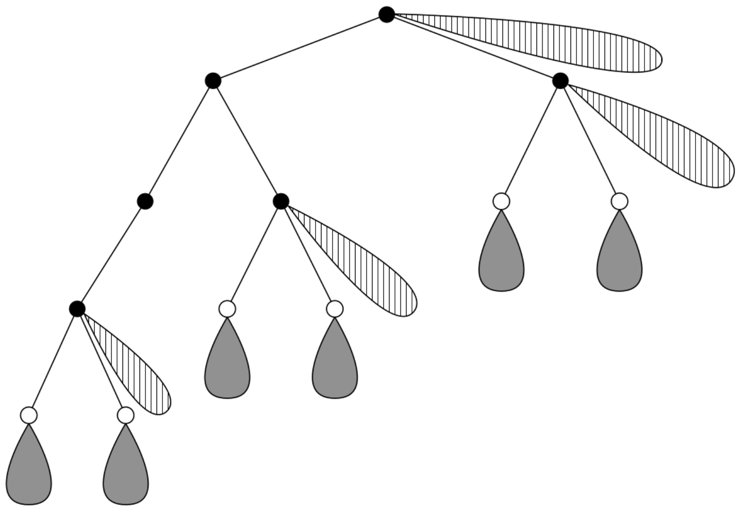

Figure 3: Instance of the construction underlying the proof of Theorem 1.1. Black-filled nodes form $\mathbb{T}_1$, non-filled nodes constitute $\mathcal{V}_2$, dashed subtrees indicate $\mathcal{V}_3$, and $\mathcal{V}_4$ is represented by the solid subtrees merged with $\mathcal{V}_2$.

If $p_0 + p_1 + p_2 = 1$, then $\mathcal{B}_n = \tau_n$, and the statement of the theorem is trivial. Thus, we shall assume $p_3 + p_4 + \ldots > 0$. Recall that, for a node $v \in \tau_n$, we write $N(v)$ for its subtree size, $N_k(v)$ for the size of the k-th largest subtree rooted at a child of v and $N_{k+}(v) = N_k(v) + N_{k+1}(v) + \ldots$ v is dropped in the notation when considering the root. For $1 \leq k \leq n$, $n \in I$, set $W_k = \sum_{v \in \tau_n, N(v)=k} N_{3+}(v)$. It is crucial to note that, given the random set $\{v \in \tau_n : N(v) = k\}$, the terms in this sum are independent and distributed like N_{3+} in the tree τ_k. By construction,

$$|\mathcal{V}_3(\tau_n)| = \sum_{v \in \mathbb{T}_1(\tau_n)} N_{3+}(v) \leq \sum_{k=M}^{n} W_k.$$

Let $\omega_n = \lfloor n^{1/5} \rfloor$. (Any integer sequence tending to infinity which is $o(n^{1/4})$ would work.) For $n \geq 0$, denote by $\mathcal{N}_{3+}(n)$ a generic random variable with the distribution of N_{3+} in τ_n. (We set $\mathcal{N}_{3+}(n) := 0$ for $n \notin I$.) By Markov's inequality and Corollary 3.7 (iii), there exists $C_1 > 0$ such that, for all $n \in I$ and $\gamma > 0$,

$$\mathbf{P}\left(\sum_{k=\omega_n}^{n} W_k \geq \gamma n/2\right) \leq \frac{2}{n\gamma}\sum_{k=\omega_n}^{n} \mathbf{E}\left[W_k\right] = \frac{2}{n\gamma}\sum_{k=\omega_n}^{n} \mathbf{E}\left[Y_k\right]\mathbf{E}\left[\mathcal{N}_{3+}(k)\right] \\ \leq \frac{2C_1}{n\gamma}\sum_{k=\omega_n}^{n} \mathbf{E}\left[Y_k\right]\log k.$$

By Corollary 2.3, there exists a constant $C_2 > 0$ such that, for all n sufficiently large, we have

$$\mathbf{P}\left(\sum_{k=\omega_n}^{n} W_k \geq \gamma n/2\right) \leq \frac{4C_2}{\gamma}\frac{\log \omega_n}{\sqrt{\omega_n}}$$

for all $\gamma > 0$. For all $\gamma > 0$, the right hand side of this display tends to zero as $n \to \infty$. Let $Z_1, Z_2, \dots$ be independent copies of $\mathcal{N}_{3+}(k)$ and $W_k^* = \sum_{j=1}^{2\mathbf{E}[Y_k]} Z_j$. Then,

$$\begin{aligned}\mathbf{P}\left(\sum_{k=M}^{\omega_n} W_k \geq \gamma n/2\right) &\leq \mathbf{P}\left(\sum_{k=M}^{\omega_n} W_k \geq \gamma n/2, W_k \leq 2\mathbf{E}\left[W_k^*\right] \text{for all } M \leq k \leq \omega_n\right) \\ &\quad + \mathbf{P}\left(W_k > 2\mathbf{E}\left[W_k^*\right] \text{for some } M \leq k \leq \omega_n\right) \\ &\leq \mathbf{1}_{[\gamma n/8,\infty)}\left(\sum_{k=M}^{\omega_n} \mathbf{E}\left[Y_k\right]\mathbf{E}\left[\mathcal{N}_{3+}(k)\right]\right) \\ &\quad + \omega_n \sup_{1\leq k\leq \omega_n} \mathbf{P}\left(W_k \geq 2\mathbf{E}\left[W_k^*\right]\right). \end{aligned} \tag{4.4}$$

Further, for any $1 \leq k \leq n$, $k \in I_n$,

$$\begin{aligned}\mathbf{P}\left(W_k \geq 2\mathbf{E}\left[W_k^*\right]\right) &\leq \mathbf{P}\left(W_k \geq 2\mathbf{E}\left[W_k^*\right], Y_k \leq 2\mathbf{E}\left[Y_k\right]\right) + \mathbf{P}\left(Y_k \geq 2\mathbf{E}\left[Y_k\right]\right) \\ &\leq \mathbf{P}\left(W_k^* \geq 2\mathbf{E}\left[W_k^*\right]\right) + \mathbf{P}\left(Y_k \geq 2\mathbf{E}\left[Y_k\right]\right).\end{aligned}$$

We use Chebyshev's inequality to bound both summands in the last expression. Applying (2.3) to (2.8) and using (2.9) shows the existence of a constant $C_3 > 0$ such that, for all $n \in I, 1 \leq k \leq \omega_n$, $k \in I_n$,

$$\mathbf{P}\left(Y_k \geq 2\mathbf{E}\left[Y_k\right]\right) \leq \frac{\mathrm{Var}(Y_k)}{\mathbf{E}\left[Y_k\right]^2} \leq C_3\frac{\omega_n^3}{n}.$$

By similar arguments also relying on Corollary 3.7 (iv), there exists $C_4 > 0$ such that, for all $n \in I, 1 \leq k \leq \omega_n$, $k \in I_n$,

$$\mathbf{P}\left(W_k^* \geq 2\mathbf{E}\left[W_k^*\right]\right) \leq \frac{\mathrm{Var}(W_k^*)}{\mathbf{E}\left[W_k^*\right]^2} \leq \frac{\mathbf{E}\left[\mathcal{N}_{3+}(k)^2\right]}{2\mathbf{E}\left[Y_k\right]\mathbf{E}\left[\mathcal{N}_{3+}(k)\right]^2} \leq C_4\frac{\omega_n^2}{n}.$$

Here, we have also used the fact that $\liminf_{n\to\infty, n\in\mathbb{N}h+1} \mathbf{E}\left[\mathcal{N}_{3+}(n)\right] > 0$. Hence, the second summand in (4.4) converges to zero as $n \to \infty$. By Corollaries 2.3 and 3.7, there exists a constant $C_5 > 0$ (depending on the offspring distribution but not on M or n) with $\sum_{k=M}^{\omega_n} \mathbf{E}\left[Y_k\right]\mathbf{E}\left[\mathcal{N}_{3+}(k)\right] \leq C_5 n \log M/\sqrt{M}$. Choosing M large enough such that $c_6 := 8C_5 \log M/\sqrt{M} < 1$, the first summand in (4.4) is identically 0, and we deduce

$$\mathbf{P}\left(|\mathcal{V}_3(\tau_n)| \geq \gamma n\right) \to 0$$

for any $\gamma \in (c_5, 1)$. □

5 Distances

The aim of this section is to prove Theorem 1.2. The following result is closely related to this theorem.

Theorem 5.1. *Consider a Galton-Watson tree whose offspring distribution satisfies* (1.1) *conditional on having size* n. *Let* $k \geq 3$.

(i) *If* $\mathbf{E}\left[\xi^{k+1}\right] < \infty$, *then, for any* $\varepsilon > 0$, *there exists* $C_1 > 0$ *such that, for all* $n \in I$,

$$\mathbf{P}\left(\max_{v\in\tau_n} N_k(v) \leq C_1 n^{2/k}\right) \geq 1 - \varepsilon. \tag{5.1}$$

One can replace $N_k(v)$ *by* $N_{k+}(v)$ *in this result upon possibly increasing* C_1 *if* $\mathbf{E}\left[\xi^{(3k+1)/2}\right] < \infty$.

(ii) *If* $\mathbf{E}\left[\xi^{k+1}\right] < \infty$ *and* $\sum_{\ell\geq k} p_\ell > 0$, *then, for any* $\varepsilon > 0$, *there exists* $c_1 > 0$ *such that, for all* $n \in I$,

$$\mathbf{P}\left(\max_{v\in\tau_n} N_k(v) \geq c_1 n^{2/k}\right) \geq 1 - \varepsilon. \tag{5.2}$$

Let us briefly discuss this result and Theorem 1.2. First of all, the lower bounds (1.6) and (5.2) are much harder to obtain than the upper bounds (1.5) and (5.1), where (1.6) follows very easily from (5.2) from known tail bounds on the height of τ_n (see (5.4)). Second, in light of Theorem 3.4 (ii), the moment conditions imposed in (1.6) and (5.2) are somewhat unexpected. Indeed, we believe that these results are valid under the finite variance assumption on the offspring distribution in (1.1). However, since our proof uses the second moment method and involves suitable bounds on variances which crucially rely on the estimates in Theorem 3.4 (i), we cannot remove these conditions. Third, similarly to statement of Theorem 3.4 (iii), we can make the following two statements about the necessity of moment conditions in order to have tightness of the sequence $n^{-2/k}\max_{v\in\tau_n} N_k(v), n \geq 1$.

(i) For $k \geq 3$, if $\mathbf{E}\left[\xi^k\right] = \infty$, then, for any $C > 0$, as $n \to \infty, n \in \mathbb{N}h + 1$, we have $\mathbf{E}\left[|\{v \in \tau_n : N_k(v) \geq Cn^{2/k}\}|\right] \to \infty$.

(ii) For $k \geq 5$, if $\mathbf{E}\left[\xi^k\right] = \infty$ and $\mathbf{E}\left[\xi^{k-1}\right] < \infty$, then $n^{-2/k}\max_{v\in\tau_n} N_k(v)$ tends to infinity in probability as $n \to \infty, n \in \mathbb{N}h + 1$. (That is, for any $C > 0$, $\mathbf{P}\left(\max_{v\in\tau_n} N_k(v) \geq Cn^{2/k}\right) \to 1$.)

These claims lead to the following proposition accompanying Theorem 1.2, where we recall that A_k stands for the set of subtrees of τ_n in which every node has at most k children.

Proposition 5.2. *Consider a Galton-Watson tree whose offspring distribution satisfies* (1.1) *conditional on having size* n.

(i) *Let* $k \geq 2$, $\mathbf{E}\left[\xi^{k+1}\right] = \infty$ *and* $C > 0$. *For* $T \in A_k$, *let* Y_T *be the number of nodes in* τ_n *which have graph distance at least* $Cn^{1/(k+1)}$ *from* T. *Then,* $\mathbf{E}\left[\inf_{T\in A_k} Y_T\right] \to \infty$.

(ii) *For* $k \geq 4$, *if* $\mathbf{E}\left[\xi^{k+1}\right] = \infty$ *and* $\mathbf{E}\left[\xi^k\right] < \infty$, *then* $n^{-1/(k+1)}\inf_{T\in A_k} d_{\max}(T)$ *tends to infinity in probability as* $n \to \infty, n \in \mathbb{N}h + 1$.

At this point, it is necessary to discuss results on the height H_n of τ_n, that is, the maximal distance of a node from the root, in more detail. In accordance with Aldous' theory on conditional Galton-Watson trees, the scaling limit of H_n is given by the maximum of a Brownian excursion. More precisely,

$$\frac{\sigma H_n}{\sqrt{2n}} \xrightarrow{d} H_\infty,$$

where H_∞ has the theta distribution. That is,

$$\mathbf{P}\left(H_\infty \le x\right) = \sum_{j=-\infty}^{\infty} (1-2j^2x^2)\exp(-j^2x^2), \quad x>0. \tag{5.3}$$

In this generality, this limit theorem goes back to Kolchin [41, Theorem 2.4.3]. In the case of Cayley trees, (5.3) had already been discovered by Rényi and Szerekes [50] and for full binary trees, that is $p_0 = p_2 = 1/2$, by Flajolet and Odlyzko [33].

Moreover, there are universal upper bounds that will be useful in this section: there exists $\delta \in (0, \sigma^2/2]$, such that

$$\sup_{n\ge 1} \mathbf{P}\left(\frac{H_n}{\sqrt{n}} \ge x\right) \le \exp(-\delta x^2), \quad x>0. \tag{5.4}$$

This is Theorem 1.2 in Addario-Berry, Devroye and Janson [2].

The rest of this section is devoted to the proofs of Theorems 1.2 and 5.1, Proposition 5.2 and the two statements (i), (ii) above.

5.1 Upper bounds

Recall the definition of the index sequences $R(v), v \in \tau_n$ from Section 4, that $H(v)$ denotes the height of the subtree rooted at v in τ_n and that we write B^* for the set of all finite length words with symbols from a set $B \subseteq \mathbb{N}$. As $v \mapsto R(v)$ maps from τ_n to $\mathbb{N}^*$, we shall define families of random variables $H^*(y), N^*(y), y \in \mathbb{N}^*$ by $H^*(R(v)) := H(v)$ and $N^*(R(v)) := N(v)$ for $v \in \tau_n$ and $H^*(y) = N^*(y) = 0$ if $y \notin \{R(v) : v \in \tau_n\}$. In particular, for $\ell \ge 1$, $H^*(\ell)$ describes the height of the subtree rooted at the child with rank ℓ of the root in τ_n.

Lemma 5.3. *Let $k \ge 2$ and $\mathbf{E}\left[\xi^{k+2}\right] < \infty$. Then, there exists a constant $C > 0$ such that, for all $n \in I$ and $t \ge 1$,*

$$\mathbf{P}\left(\sup_{\ell \ge k} H^*(\ell) \ge t\right) \le Ct^{1-k}.$$

Proof. Let $\{\mathcal{H}_i(n) : n \in I, i \ge 1\}$ be a family of independent random variables where each $\mathcal{H}_i(n)$ is distributed like the height of τ_n. Furthermore, assume that the family is independent of τ_n. Using (5.4), we have

$$\begin{aligned}
\mathbf{P}\left(\sup_{\ell \ge k} H^*(\ell) \ge t\right) &= \mathbf{E}\left[\mathbf{P}\left(\sup_{k \le \ell \le \xi_\varnothing} H^*(\ell) \ge t \middle| \xi_\varnothing, N_1, N_2, \dots\right)\right] \\
&\le \mathbf{E}\left[\xi_\varnothing \sup_{k \le \ell \le \xi_\varnothing} \mathbf{P}\left(\mathcal{H}_\ell(N_\ell) \ge t \middle| \xi_\varnothing, N_1, N_2, \dots\right)\right] \\
&\le \mathbf{E}\left[\xi_\varnothing \exp(-\delta t^2/N_k)\right] \\
&= \sum_{\ell=k}^{\infty} \int_0^\ell \mathbf{P}\left(N_k \ge -\delta t^2/\log(s/\ell), \xi_\varnothing = \ell\right) ds.
\end{aligned}$$

By inequality (3.3) in Remark 3.5, there exists $C_1 > 0$ such that the right-hand side of the last display is bounded from above by

$$\sum_{\ell=k}^{\infty} C_1 p_\ell \ell^{k+2} \int_0^1 \left(-\frac{\delta t^2}{\log s}\right)^{(1-k)/2} ds \leq C_1 \Gamma((k+1)/2) \mathbf{E}\left[\xi^{k+2}\right] \delta^{(1-k)/2} t^{1-k}.$$

Here, $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ denotes the Gamma function. This concludes the proof. □

Proposition 5.4. *Let $k \geq 2$ and $\mathbf{E}\left[\xi^{k+1}\right] < \infty$. Then, there exists a constant $C > 0$ such that, for all $n \in I$ and $t \geq 1$,*

$$\mathbf{P}\left(\max_{v \in \tau_n} N_k(v) \geq t\right) \leq C \frac{n}{t^{k/2}}.$$

The bound also holds for $N_{k+}(v)$ if $\mathbf{E}\left[\xi^{(3k+1)/2}\right] < \infty$ upon possibly increasing C.

The proposition immediately yields statement (5.1) in Theorem 5.1.

Proof. The left hand side is zero for $t \geq \lceil n/2 \rceil$. Thus, we assume $t \leq \lceil n/2 \rceil - 1$. Note that, for all nodes $v \in \tau_n$ with $N(v) \geq \lceil n/2 \rceil$, we must have $v \in \mathcal{V}(\{1\}^*)$. Hence, there are at most $|\mathcal{V}(\{1\}^*)|$ of them in the tree. Thus, by Theorem 3.4 (i), writing $w_1, \ldots, w_n$ for the nodes of τ_n listed in preorder,

$$\begin{aligned}
\mathbf{P}\left(\max_{v \in \tau_n} N_k(v) \geq t\right) &\leq \mathbf{E}\left[|\{v \in \tau_n : N_k(v) \geq t\}|\right] \\
&= \sum_{i=1}^{n} \mathbf{P}\left(N_k(w_i) \geq t\right) \\
&\leq \beta_k t^{(1-k)/2} \mathbf{E}\left[|\{v \in \tau_n : N(v) \geq t\}|\right] \\
&\leq \beta_k t^{(1-k)/2} \left(\sum_{\ell=t}^{\lceil n/2 \rceil - 1} \mathbf{E}\left[Y_\ell\right] + \mathbf{E}\left[|\mathcal{V}(\{1\}^*)|\right]\right) \\
&\leq C_1 \left(\frac{n}{t^{k/2}} + \frac{\sqrt{n}}{t^{(k-1)/2}}\right),
\end{aligned}$$

where C_1 can be chosen independently of t and n by Lemma 2.2 and the fact that $\mathbf{E}\left[|\mathcal{V}(\{1\}^*)|\right] = O(\sqrt{n})$. The same argument applies to $N_{k+}(v)$. □

In order to transfer the result to distances, we need a tighter bound when restricting to nodes on the heavy path. Recall that, for a finite of infinite deterministic set of words A over the alphabet $\mathbb{N}$, we write $\mathcal{V}(A)$ for the random set of nodes $v \in \tau_n$ with $R(v) \in A$.

Lemma 5.5. *Let $k \geq 2$ and $\mathbf{E}\left[\xi^{k+1}\right] < \infty$. Then, for any deterministic finite of infinite set $\mathcal{A} \subseteq \mathbb{N}^*$ and $t \geq 1$,*

$$\mathbf{P}\left(\max_{v \in \mathcal{V}(\mathcal{A})} N_k(v) \geq t\right) \leq \beta_k t^{(1-k)/2} \mathbf{E}\left[|\mathcal{V}(\mathcal{A})|\right],$$

with β_k as in (3.1). If $\mathbf{E}\left[\xi^{(3k+1)/2}\right] < \infty$, then the bound also holds with k replaced by $k+$ (and β_k by β_{k+}). Furthermore, if $\mathbf{E}\left[\xi^{k+2}\right] < \infty$, then there exists a constant $C > 0$ such that,

$$\mathbf{P}\left(\max_{v \in \mathcal{V}(\mathcal{A}), \ell \geq k} H^*(R(v)\ell) \geq t\right) \leq C t^{1-k} \mathbf{E}\left[|\mathcal{V}(\mathcal{A})|\right].$$

Here, and subsequently, $R(v)\ell \in \mathbb{N}^*$ denotes the concatenation of $R(v)$ and ℓ.

As the average height of τ_n is well-known to be of order $\sqrt{n}$, we deduce the following corollary:

Corollary 5.6. *Let $k \geq 2$ and $\mathbf{E}\left[\xi^{k+1}\right] < \infty$. Then, there exists a constant $C_1 > 0$ such that*

$$\mathbf{P}\left(\max_{v \in \mathcal{V}(\{1\}^*)} N_k(v) \geq t\right) \leq C_1 \frac{\sqrt{n}}{t^{(k-1)/2}}.$$

If $\mathbf{E}\left[\xi^{(3k+1)/2}\right] < \infty$, then the same results hold with $N_k(v)$ replaced by $N_{k+}(v)$ upon possibly increasing C_1. Finally, if $\mathbf{E}\left[\xi^{k+2}\right] < \infty$, then there exists $C_2 > 0$, such that

$$\mathbf{P}\left(\max_{v \in \mathcal{V}(\{1\}^*), \ell \geq k} H^*(R(v)\ell) \geq t\right) \leq C_2 \frac{\sqrt{n}}{t^{k-1}}.$$

Proof of Lemma 5.5. For $\ell \geq 0$, let $\mathcal{A}_{\ell,n}$ be the subset of $\mathcal{A}$ of vectors of length ℓ where each entry is bounded from above by n. We have

$$\mathbf{P}\left(\max_{v \in \mathcal{V}(\mathcal{A})} N_k(v) \geq t\right) \leq \sum_{\ell=0}^{n} \mathbf{P}\left(\max_{v \in \mathcal{V}(\mathcal{A}_{\ell,n})} N_k(v) \geq t\right). \tag{5.5}$$

We denote the elements of $\mathcal{A}_{\ell,n}$ by $y_1, \ldots, y_K$, $K = K(\ell) \leq n^\ell$. Let $\{\mathcal{N}_k^{(i)}(j) : i \geq 1, j \in I\}$ be a family of independent random variables where each $\mathcal{N}_k^{(i)}(j)$ is distributed like N_k in the tree τ_j. Then, using (3.1),

$$\begin{aligned}
&\mathbf{P}\left(\max_{v \in \mathcal{V}(\mathcal{A}_{\ell,n})} N_k(v) \geq t\right) \\
&= \sum_{0 \leq n_1, \ldots, n_K \leq n} \mathbf{P}\left(\max_{v \in \mathcal{V}(\mathcal{A}_{\ell,n})} N_k(v) \geq t \,\middle|\, \bigcap_{j=1}^{K} \{N^*(y_j) = n_j\}\right) \mathbf{P}\left(\bigcap_{j=1}^{K} \{N^*(y_j) = n_j\}\right) \\
&= \sum_{0 \leq n_1, \ldots, n_K \leq n} \mathbf{P}\left(\max_{1 \leq j \leq K} \mathcal{N}_k^{(j)}(n_j) \geq t\right) \mathbf{P}\left(\bigcap_{j=1}^{K} \{N^*(y_j) = n_j\}\right) \\
&\leq \sum_{0 \leq n_1, \ldots, n_K \leq n} |\{1 \leq j \leq K : n_j \geq t\}| \sup_{1 \leq i \leq K} \mathbf{P}\left(\mathcal{N}_k^{(1)}(n_i) \geq t\right) \mathbf{P}\left(\bigcap_{j=1}^{K} \{N^*(y_j) = n_j\}\right) \\
&\leq \beta_k t^{(1-k)/2} \mathbf{E}\left[|\{v \in \mathcal{V}(\mathcal{A}_{\ell,n}) : N(v) \geq t\}|\right].
\end{aligned}$$

Plugging the bound into (5.5) gives

$$\mathbf{P}\left(\max_{v \in \mathcal{V}(\mathcal{A})} N_k(v) \geq t\right) \leq \beta_k t^{(1-k)/2} \mathbf{E}\left[|\{v \in \mathcal{V}(\mathcal{A}) : N(v) \geq t\}|\right] \leq \beta_k t^{(1-k)/2} \mathbf{E}\left[|\mathcal{V}(\mathcal{A})|\right].$$

The same proof works for $N_{k+}(v)$. Similarly, one obtains the result for the heights upon replacing $N_k(v)$ by $\max_{\ell \geq k} H^*(R(v)\ell)$ and using Lemma 5.3. □

Proposition 5.7. *Let $k \geq 2$ and $\mathbf{E}\left[\xi^{k+2}\right] < \infty$. Then, there exists a constant $C > 0$ such that, for $t \geq 1$, $n \in I$,*

$$\mathbf{P}\left(\max_{v \in \tau_n, \ell \geq k} H^*(R(v)\ell) \geq t\right) \leq C\left(\frac{n}{t^k} + \frac{\sqrt{n}}{t^{k-1}}\right).$$

Part (i) of Theorem 1.2 follows immediately from the proposition.

Proof. We may assume $t \geq n_0$ with n_0 as in Lemma 2.2. For $k \geq 1, n \in I$, let $(\mathcal{H}(n), \mathcal{N}_k(n), \bar{\xi}(n))$ be distributed like $(H, N_k, \xi_\varnothing)$ in τ_n. Using (5.4), we have

$$\begin{aligned}
&\mathbf{P}\left(\max_{v\in\tau_n\setminus\mathcal{V}(\{1\}^*),\ell\geq k} H^*(R(v)\ell)\geq t\right)\\
&\leq \mathbf{E}\left[|\{v\in\tau_n\setminus\mathcal{V}(\{1\}^*) : H^*(R(v)\ell)\geq t \text{ for some } \ell\geq k\}|\right]\\
&= \sum_{i=1}^{n}\mathbf{P}\left(H^*(R(i)\ell)\geq t \text{ for some } \ell\geq k, i\notin\mathcal{V}(\{1\}^*)\right)\\
&\leq \sum_{i=1}^{n}\sum_{\ell=k}^{n}\sum_{m=t}^{\lceil n/2\rceil-1}\sum_{j=t}^{m}\mathbf{P}\left(\mathcal{H}(j)\geq t\right)\mathbf{P}\left(\mathcal{N}_\ell(m)=j, \ell\leq\bar{\xi}(m)\right)\mathbf{P}\left(N(i)=m\right)\\
&\leq \sum_{i=1}^{n}\sum_{\ell=k}^{n}\sum_{m=t}^{\lceil n/2\rceil-1}\mathbf{E}\left[\exp\left(-\frac{\delta t^2}{\mathcal{N}_\ell(m)}\right)\mathbf{1}_{t\leq\mathcal{N}_\ell(m),\ell\leq\bar{\xi}(m)}\right]\mathbf{P}\left(N(i)=m\right).
\end{aligned}$$

The expectation in the last display is bounded by

$$\begin{aligned}
&\int_0^{e^{-\delta t^2/m}}\mathbf{P}\left(\exp\left(-\frac{\delta t^2}{\mathcal{N}_\ell(m)}\right)\geq x, \ell\leq\bar{\xi}(m)\right)dx\\
&=\int_0^{e^{-\delta t^2/m}}\mathbf{P}\left(\mathcal{N}_k(m)\geq\frac{\delta t^2}{\log 1/x}, \ell\leq\bar{\xi}(m)\right)dx.
\end{aligned}$$

By Theorem 3.4 (i), there exists $C_1 > 0$ such that

$$\begin{aligned}
\sum_{\ell=k}^{n}\int_0^{e^{-\delta t^2/m}}\mathbf{P}\left(\mathcal{N}_k(m)\geq\frac{\delta t^2}{\log 1/x}, \ell\leq\bar{\xi}(m)\right)dx &\leq \int_0^{e^{-\delta t^2/m}}\mathbf{E}\left[\bar{\xi}(m)\mathbf{1}_{\mathcal{N}_k(m)\geq\frac{\delta t^2}{\log 1/x}}\right]dx\\
&\leq C_1\frac{t^{1-k}}{\delta^{(k-1)/2}}\int_0^{e^{-\frac{\delta t^2}{m}}}\left(\log\frac{1}{x}\right)^{\frac{k-1}{2}}dx\\
&\leq C_2 m^{(1-k)/2}e^{-\delta t^2/m}.
\end{aligned}$$

Here, $C_2 > 1$ denotes a constant which is independent of m, t and n. Summarizing and using Lemma 2.2, we obtain

$$\begin{aligned}
\mathbf{P}\left(\max_{v\in\tau_n\setminus\mathcal{V}(\{1\}^*),\ell\geq k} H^*(R(v)\ell)\geq t\right) &\leq C_2\sum_{m=t}^{\lceil n/2\rceil-1}\mathbf{E}\left[Y_m\right]\left(m^{(1-k)/2}e^{-\delta t^2/m}+e^{-\delta t}\right)\\
&\leq 2\sqrt{2}C_2\alpha n\sum_{m=t}^{\lceil n/2\rceil-1}m^{-1-k/2}e^{-\delta t^2/m}\leq C_3 n t^{-k},
\end{aligned}$$

for some $C_3 > 0$. Together with Corollary 5.6 for the maximum over nodes on the heavy path, this concludes the proof. □

5.2 Lower bounds

Our lower bounds rely on a variant of the second moment method which requires sufficiently tight upper bounds on variances (or second moments). To this end, we use Lemma 6.1 in Janson [37] and introduce the notation used in this work. Denote by $\mathfrak{T}$ the set of all ordered rooted trees. For a function $f : \mathfrak{T} \to \mathbb{R}$, let F be defined by

$$F(\mathbb{T}) := F(f, \mathbb{T}) := \sum_{v\in\mathbb{T}} f(\mathbb{T}_v), \quad \mathbb{T}\in\mathfrak{T}.$$

Here $\mathbb{T}_v$ denotes the subtree in $\mathbb{T}$ rooted at v. For $k \geq 1$, we abbreviate $f_k(\mathbb{T}) := f(\mathbb{T})\mathbf{1}_{|\mathbb{T}|=k}$. Note that $F(f_k, \tau_n) = Y_k$ for $f = \mathbf{1}$, where $\mathbf{1}$ denotes the function mapping every tree to 1. Then, for $1 \leq m \leq k \leq n/2$,

$$\mathrm{Cov}(F(f_k, \tau_n), F(f_m, \tau_n)) = I_1(f, k, m) + I_2(f, k, m) + I_3(f, k, m),$$

where

$$\begin{aligned} I_1(f,k,m) =& \frac{n\mathbf{P}\left(S_{n-k}=0\right)\mathbf{P}\left(S_k=-1\right)}{k\mathbf{P}\left(S_n=-1\right)}\mathbf{E}\left[f_k(\tau_k)F(f_m,\tau_k)\right], \\ I_2(f,k,m) =& \frac{n(n-k-m+1)}{mk}\mathbf{P}\left(S_k=-1\right)\mathbf{P}\left(S_m=-1\right)\mathbf{E}\left[f_k(\tau_k)\right]\mathbf{E}\left[f_m(\tau_m)\right]\cdot \\ & \left(\frac{\mathbf{P}\left(S_{n-k-m}=1\right)}{\mathbf{P}\left(S_n=-1\right)} - \frac{\mathbf{P}\left(S_{n-k}=0\right)}{\mathbf{P}\left(S_n=-1\right)}\frac{\mathbf{P}\left(S_{n-m}=0\right)}{\mathbf{P}\left(S_n=-1\right)}\right), \end{aligned}$$

and

$$\begin{aligned} I_3(f,k,m) =& -\frac{n(k+m-1)}{mk}\frac{\mathbf{P}\left(S_{n-k}=0\right)}{\mathbf{P}\left(S_n=-1\right)}\frac{\mathbf{P}\left(S_{n-m}=0\right)}{\mathbf{P}\left(S_n=-1\right)}\cdot \\ & \mathbf{P}\left(S_k=-1\right)\mathbf{P}\left(S_m=-1\right)\mathbf{E}\left[f_k(\tau_k)\right]\mathbf{E}\left[f_m(\tau_m)\right]. \end{aligned}$$

Note that, by the crucial Lemma 6.2 in [37], cancellation effects in $I_2(f,k,m)$ cause this term to be of the order n (for m, k fixed), rather than n^2. Below, we only need upper bounds on the variance which allows us to neglect $I_3(f,k,m)$. For $i = 1, 2$, we set $I_i(k,m) = I_i(\mathbf{1}, k, m)$.

For $1 \leq t \leq n$, $t \in \mathbb{N}$, we define

$$\mathcal{Y}_t = |\{v \in \tau_n : t \leq N(v) \leq 2t\}| = \sum_{\ell=t}^{2t} Y_\ell.$$

From Lemma 2.2, we know that there exists a constant $K_1 > 0$ depending only on the offspring distribution such that, for all $1 \leq t \leq n/4$, we have

$$K_1^{-1}\frac{n}{\sqrt{t}} \leq \mathbf{E}\left[\mathcal{Y}_t\right] \leq K_1\frac{n}{\sqrt{t}}. \tag{5.6}$$

Proposition 5.8. *There exists a constant $C > 0$, such that, for all $1 \leq t \leq (n-1)/4, t \in \mathbb{N}$ and $n \in I$, we have*

$$\mathrm{Var}(\mathcal{Y}_t) \leq Cn.$$

In particular, for any sequence $t = t(n) = o(n)$, we have, as $n \to \infty$, in probability,

$$\frac{\mathcal{Y}_t}{\mathbf{E}\left[\mathcal{Y}_t\right]} \to 1.$$

Proof. We use the notation introduced above with the function $f = \mathbf{1}$. Obviously,

$$\mathrm{Var}(\mathcal{Y}_t) = \sum_{k,m=t}^{2t}\mathrm{Cov}(Y_k, Y_m) \leq 2\sum_{m=t}^{2t}\sum_{k=m}^{2t}\mathrm{Cov}(Y_k, Y_m).$$

In the following, $C_i, i \geq 1$, denote constants independent of k, m, t and n, whose precise values are of no relevance. For $m \leq k$, by the local limit theorem (2.3), we have $I_1(k,m) \leq C_1 n m^{-3/2}(\max(1, k-m))^{-1/2}$. Thus,

$$\sum_{m=t}^{2t}\sum_{k=m}^{2t} I_1(k,m) \leq C_2 n \sum_{m=t}^{2t} m^{-3/2}\sqrt{2t-m} \leq C_3 n.$$

By Lemma 6.2 in [37], for $t \le m \le k \le 2t$,

$$I_2(k,m) \le C_4 n^2 ((km)^{-3/2} \left(\frac{1}{n} + \frac{k+m}{n^{3/2}} + \frac{km}{n^2} \right) \le C_5 t^{-2}(nt^{-1} + \sqrt{n} + t). \quad (5.7)$$

Hence, $\sum_{m=t}^{2t} \sum_{k=m}^{2t} I_2(k,m) \le C_6(nt^{-1} + \sqrt{n} + t) \le C_7 n$. This finishes the proof. □

For $\ell \ge 2$ let $n_\ell(\mathbb{T})$ denotes the size of the ℓ-th largest subtree of a child of the root in $\mathbb{T}$. For example, in our notation, we have $N_\ell = n_\ell(\tau_n)$. For $t > 0$, let $g_\ell(\mathbb{T}) = \mathbf{1}_{n_\ell(\mathbb{T}) \ge t}$. (We suppress t in the notation.) For $t > 0$, let $t' = \lfloor (\ell+2)t \rfloor$ and define

$$V_t = |\{v \in \tau_n : t' \le N(v) \le 2t', N_\ell(v) \ge t\}|.$$

Then,

$$\mathbf{E}[V_t] = \sum_{i=t'}^{2t'} \mathbf{P}(\mathcal{N}_\ell(i) \ge t)\, \mathbf{E}[Y_i], \quad (5.8)$$

where, as before, we write $\mathcal{N}_\ell(i)$ for a random variable distributed like N_ℓ in τ_i.

Proposition 5.9. *Let $\ell \ge 2$.*

(i) *If $\mathbf{E}\left[\xi^{\ell+1}\right] < \infty$, then, there exists a constant $C_1 > 0$, such that, for $n \in I$ sufficiently large and $t \le n/4$,*

$$\mathbf{E}[V_t] \le C_1 \frac{n}{t^{\ell/2}}.$$

(ii) *If $\sum_{m \ge \ell} p_m > 0$, then, there exist constants $C_2, K_2 > 0$, such that, for $n \in I$ sufficiently large, and $C_2 \le t \le n/(4\ell)$,*

$$\mathbf{E}[V_t] \ge K_2^{-1} \frac{n}{t^{\ell/2}}.$$

(iii) *If $\mathbf{E}\left[\xi^{\ell+1}\right] < \infty$, then there exists a constant $K_3 > 0$ such that, for all $n \in I, 1 \le t < (n-1)/(4(\ell+2))$, we have*

$$\mathrm{Var}(V_t) \le (1 + K_3 t^{(3-\ell)/2})\mathbf{E}[V_t] + K_3 \left(nt^{-\ell} + \sqrt{n} t^{1-\ell} + t^{2-\ell}\right).$$

Proof. The bounds on the mean in (i) and (ii) immediately follow from (5.8) and the bounds in (5.6) using the tail bounds in Theorem 3.4 (i). In (iii), we may assume $\sum_{m \ge \ell} p_m > 0$, since, otherwise, $V_t = 0$ almost surely. We then have

$$\mathrm{Var}(V_t) \le \sum_{m=t'}^{2t'} (I_1(g_\ell, m, m) + I_2(g_\ell, m, m)) + 2 \sum_{m=t'}^{2t'} \sum_{k=m+1}^{2t'} (I_1(g_\ell, k, m) + I_2(g_\ell, k, m)),$$

where

$$I_2(g_\ell, k, m) = I_2(k,m)\mathbf{P}(\mathcal{N}_\ell(k) \ge t)\, \mathbf{P}(\mathcal{N}_\ell(m) \ge t).$$

and

$$I_1(g_\ell, k, m) = \frac{n\mathbf{P}(S_{n-k} = 0)}{k\mathbf{P}(S_n = -1)} \mathbf{P}(S_k = -1)\, \mathbf{E}\left[\mathbf{1}_{\tilde{N}_\ell \ge t} |\{v \in \tau_k : \tilde{N}(v) = m, \tilde{N}_\ell(v) \ge t\}|\right],$$

where the tilde on the right-hand side indicates that the quantities are considered in the tree τ_k. Combining the bounds in Theorem 3.4 (i) and (5.7), there exists $C_4 > 0$ such that

$$\sum_{m=t'}^{2t'} \sum_{k=m}^{2t'} I_2(g_\ell, k, m) \le C_4 (n + \sqrt{n} t + t^2) t^{-\ell}.$$

Next, again using Theorem 3.4 (i),

$$\begin{aligned}
&\sum_{m=t'}^{2t'} \sum_{k=m+1}^{2t'} I_1(g_\ell, k, m) \\
&= \mathbf{E}\left[|\{(v,w) \in \tau_n^2 : t' \le N(v), N(w) \le 2t', N_\ell(v) \ge t, N_\ell(w) \ge t, w \in (\tau_n)_v, w \ne v\}|\right] \\
&\le \beta_\ell t^{(1-\ell)/2} \mathbf{E}\left[|\{(v,w) \in \tau_n^2 : t' \le N(v) \le 2t', N_\ell(v) \ge t, w \in (\tau_n)_v, w \ne v\}|\right] \\
&\le 2\beta_\ell(\ell+1) t^{(3-\ell)/2} \mathbf{E}\left[|\{v \in \tau_n : t' \le N(v) \le 2t', N_\ell(v) \ge t\}|\right] \\
&= 2\beta_\ell(\ell+1) t^{(3-\ell)/2} \mathbf{E}\left[V_t\right].
\end{aligned}$$

Here, we have used $(\tau_n)_v$ for the subtree in τ_n rooted at v. Finally, $\sum_{m=t'}^{2t'} I_1(g_\ell, m, m) = \mathbf{E}[V_t]$. This concludes the proof. □

Proofs of Theorem 1.2 and Theorem 5.1. As already indicated, the upper bounds (5.1) and (1.5) follow immediately from Propositions 5.4 and 5.7. For the lower bound in (5.2), let $\ell \ge 3$, and note that, by Chebyshev's inequality, using the bounds in Proposition 5.9, for t and n sufficiently large with $t \le (n-1)/(4(\ell+2))$,

$$\mathbf{P}(V_t = 0) \le \mathbf{P}(|V_t - \mathbf{E}[V_t]| \ge \mathbf{E}[V_t]) \le \frac{\mathrm{Var}(V_t)}{\mathbf{E}[V_t]^2} \le \frac{1+K_3}{\mathbf{E}[V_t]} + K_2^2 K_3 \left(\frac{1}{n} + \frac{t}{n^{3/2}} + \frac{t^2}{n^2}\right).$$

Now, (5.2) follows from Proposition 5.4 (iii) upon choosing $t = cn^{2/\ell}$ with $c > 0$ sufficiently small. For the lower bound in (1.6) note that, for $\varepsilon > 0$, there exists $n_3 > 0$ such that, for all $n \ge n_3$, we have $\mathbf{P}(H \ge \varepsilon\sqrt{n}) \ge 1 - \varepsilon$. Hence, for $n_3 \le m \le n$,

$$\begin{aligned}
&\mathbf{P}\left(\max_{v \in \tau_n} \min_{1 \le i \le \ell} H^*(R(v)i) \ge \varepsilon\sqrt{m}\right) \\
&\ge \mathbf{P}\left(\max_{v \in \tau_n} \min_{1 \le i \le \ell} H^*(R(v)i) \ge \varepsilon\sqrt{m}, \max_{v \in \tau_n} N_\ell(v) \ge m\right) \\
&\ge \sum_{j=1}^{n} \sum_{m'=m}^{n} \mathbf{P}\left(\min_{1 \le i \le \ell} H^*(R(j)i) \ge \varepsilon\sqrt{m}, N_\ell(j) = m', N_\ell(j') < m \text{ for all } 1 \le j' < j\right) \\
&\ge (1-\varepsilon)^\ell \mathbf{P}\left(\max_{v \in \tau_n} N_\ell(v) \ge m\right).
\end{aligned}$$

Hence, the lower bound in (1.6) follows from the lower bound in (5.2) upon choosing $m = c_1 n^{2/\ell}$ in the last display with $c_1 > 0$ sufficiently small. □

Proof of Proposition 5.2 and the preceding claims (i), (ii). We start with claim (i) and let $t = t(n) = Cn^{2/k}$ for some $C > 0$. With this choice, it is clearly sufficient to show that $\mathbf{E}[V_t] \to \infty$ as $n \to \infty$. By (3.2), for any $C_1 > 0$ and all $n \in I$ sufficiently large, we have

$$\mathbf{P}(\mathcal{N}_k(i) \ge t) \ge C_1 t^{(1-k)/2}$$

for all $i = t', \dots, 2t'$. Using this bound, (5.8) and (5.6), for all n large enough, we obtain

$$\mathbf{E}[V_t] \ge C_1 K_1^{-1} n t^{-k/2} = C_1 K_1^{-1} C.$$

As K_1 and C are fixed and C_1 was chosen arbitrarily, we deduce the assertion $\mathbf{E}[V_t] \to \infty$. To show claim (ii) again set $t = t(n) = Cn^{2/k}$. By following the steps in the proof of Proposition 5.9, there exists a constant $C_1 > 0$ independent of n and C such that

$$\begin{aligned}
\mathrm{Var}(V_t) \le C_1 &\sup_{t' \le m \le 2t'} \mathbf{P}(\mathcal{N}_k(m) \ge t)^2 \left(nt^{-3} + \sqrt{n}t^{-2} + t^{-1}\right) \\
&+ \left(1 + C_1 t \sup_{t' \le m \le 2t'} \mathbf{P}(\mathcal{N}_k(m) \ge t)\right) \mathbf{E}[V_t].
\end{aligned}$$

Since $\mathbf{E}\left[\xi^{k-1}\right] < \infty$, using Theorem 3.4 (i), we can bound $\mathbf{P}\left(N_k \geq y\right) \leq \mathbf{P}\left(N_{k-2} \geq y\right) \leq \beta_{k-2} y^{(3-k)/2}$ for all $y > 0$. As $\mathbf{E}\left[V_t\right] \to \infty$ by claim (i) and $k \geq 5$, it is straightforward to verify that $\mathrm{Var}(V_t) = o(\mathbf{E}\left[V_t\right]^2)$. This concludes the proof by the second moment argument in the proof of Theorem 5.1.

The arguments to deduce Proposition 5.2 are very similar to those necessary to obtain Theorem 1.2 from Theorem 5.1 and therefore only sketched. First of all, to show part (ii) note that the subtrees rooted at the children of rank $1, \ldots, k$ of a node v with $N_k(v) \geq C n^{2/k}$ all have heights at least $C\varepsilon n^{1/k}$ for some small ε with high probability depending only on ε and not on C. Since for any large C such nodes v exist with high probability by claim (ii), for any $T \in A_k$, we find nodes at least $C\varepsilon n^{1/k}$ away from T. In fact, this argument also yields at least $C\varepsilon n^{1/k}/2 - 1$ many nodes which have graph distance $C\varepsilon n^{1/k}/2 - 1$ from T. This simple observation explains why claim (i) implies Proposition 5.2 (i): in fact, $\liminf_{n\to\infty} \mathbf{E}\left[|\{v \in \tau_n : N_k(v) \geq C n^{2/k}\}|\right] > 0$ for all $C > 0$ would be sufficient in this context. □

6 The heavy path

In this section, we study the heavy path $\mathcal{V}(\{1\}^*)$ in the conditional Galton-Watson tree τ_n. We set $L_n = |\mathcal{V}(\{1\}^*)| - 1$ as in the introduction. Recall from Section 1.1, that the scaling limit of conditional Galton-Watson trees is Aldous' continuum random tree. More precisely, define the *depth-first process* (or *contour function*) $(D_i)_{0\leq i\leq 2n-2}$ by $D_i = d(f(i))$, where $f(i), 0 \leq i \leq 2n-2$, denotes the node visited in the i-th step of the depth first traversal, and $d(v)$ measures the distance of a node v from the root. We extend the process to a continuous function on $[0, 2n-2]$ by linear interpolation. Endowing the space of continuous functions with the supremum norm, we have,

$$\left(\frac{D_{t(2n-2)}}{\sqrt{n}}\right)_{0\leq t\leq 1} \xrightarrow{d} \frac{2}{\sigma} \cdot \mathbf{e}, \tag{6.1}$$

where $\mathbf{e}$ is a standard Brownian excursion. This is Aldous's Theorem 2 [6]. As already indicated in the introduction, the heavy path can be defined in the continuum random tree making use of its definition based on a Brownian excursion. Therefore, using (6.1), convergence of $L_n/\sqrt{n}$ boils down to an application of the continuous mapping theorem. The technical steps in this context leading to Theorem 1.3 are intricate and of entirely different flavour than the arguments in the rest of the paper. Therefore, we defer the proof of this theorem to Section 6.2.

It turns out that L_∞ can be represented as an exponential functional of a subordinator $\xi(t), t \geq 0$, that is, $L_\infty = 2\sigma^{-1}\int_0^\infty e^{-\frac{1}{2}\xi(t)}dt$. Such quantities have applications in various fields such as self-similar Markov processes and mathematical finance. We refer to Bertoin and Yor [18] for a survey. In particular, as worked out in detail in Section 6.2, the existence of a density for L_∞ as well as the formula for the moments follow from general results on exponential functionals due to Carmona, Petit and Yor [24].

Remark 6.1. As stated in Theorem 1.3, we also prove functional limit theorems (after rescaling) for the quantities

$$P_n(k) = \{N(v) : v \text{ has distance } k \text{ from the root and } R(v) = 1\ldots1\}, \quad k \geq 1$$

and $Q_n(\ell) = \inf\{k \geq 0 : P_n(k) \leq \ell\}, 1 \leq \ell \leq n$. See (6.6) and (6.7) in Theorem 6.12. The limiting random variables can be expressed in terms of the subordinator ξ involving a random time-change.

It is natural to compare L_n to the height H_n. In particular, since $L_n \leq H_n$, the bound (5.4) on the tail of H_n also applies to the right tail of L_n. For the limiting behaviour, we have

$$\lim_{x\to\infty} \frac{-\log \mathbf{P}(H_\infty > x)}{x^2} = 1, \quad \lim_{x\to 0} -x^2 \log \mathbf{P}(H_\infty \le x) = \pi^2.$$

Our next result shows that the decay of the distribution function of T_∞ is considerably slower at 0. Still, all its derivatives vanish at 0.

Proposition 6.2. *We have*

$$\lim_{x\to 0} \frac{-\log \mathbf{P}(L_\infty \le x)}{\log^2 x} = \frac{2}{\log 2}, \quad \lim_{x\to\infty} \frac{-\log \mathbf{P}(L_\infty > x)}{x^2} = \frac{1}{2}.$$

The proof of the first part of the proposition relies on sandwiching the random variable $L_n/\sqrt{n}$ between two quantities admitting series representations of the form $\sum_{i=0}^\infty \rho^i Z_i$ for some $0<\rho<1$ and a sequence of independent and identically distributed random variables $Z_1, Z_2, \dots$ It is presented in Section 6.1. The second claim shown in Section 6.2 uses a result due to Rivero [51] on the right tail decay of exponential functionals of subordinators (see (6.14)). (There are also general results on the left tail decay of exponential functionals. Compare, e.g. Pardo, Rivero and van Schaik [48] and the references given therein. We did however not find any result in the literature covering our case.)

6.1 Proof of Proposition 6.2 for $x \to 0$

The proof uses the following lemma.

Lemma 6.3 ([12], Theorem 4)**.** *Let $Z_1, Z_2, \dots$ be a sequence of non-negative, independent and identically distributed random variables with $\mathbf{E}[\log \max(Z_1, 1)] < \infty$. Further, assume that $0 < \liminf_{t\to 0} t^{-\alpha}\mathbf{P}(Z_1 \le t) \le \limsup_{t\to 0} t^{-\alpha}\mathbf{P}(Z_1 \le t) < \infty$ for some $\alpha > 0$. Then, for all $0<\rho<1$,*

$$\lim_{t\to 0} \frac{-\log \mathbf{P}\left(\sum_{i=0}^\infty \rho^i Z_i \le t\right)}{\log^2 t} = \frac{\alpha}{2\log 1/\rho}.$$

Note that, for exponentially distributed $Z_1, Z_2, \dots$, the lemma coincides with [17, Proposition 3].

Proof of Proposition 6.2 (Lower bound). Assume $h=1$ for the sake of presentation. Fix $0<\delta<1/2$ such that $(1-\delta)^i n \notin \mathbb{N}$ for all $i, n \ge 1$ and set $c := 1/(1-\delta)$. For $i \ge 0$, let $e_i \in \{1\}^*$ be the vector of $1'$s of length i and $\sigma_i := \sigma_i(n) := \inf\{j \ge 0 : N^*(e_j) \le (1-\delta)^i n\}$. Then $L_n = \sigma_{\lceil \log_c n\rceil} - 1$. Let $0 < \varepsilon < 1/2 - \delta$ and β^* be as in Theorem 3.4 (ii) with $k=2$ and the chosen ε. The crucial observation is that there exist $C_1, C_2 > 0$ such that, for all $n \ge C_1$ and $j \le \lfloor \log_c n - C_2 \rfloor$, we have, stochastically,

$$\sigma_j \le \sum_{i=1}^j G_i, \tag{6.2}$$

where $G_1, G_2, \dots$ is a sequence of independent geometrically distributed random variables on $\{1, 2, \dots\}$ and G_i has success parameter $\beta^*/\sqrt{\delta(1-\delta)^{i-1}n}$. Taking (6.2) for granted, we obtain, in a stochastic sense,

$$L_n = \sigma_{\lfloor \log_c n - C_2\rfloor} + (L_n - \sigma_{\lfloor \log_c n - C_2\rfloor}) \le (1-\delta)^{-C_2-1} + \sum_{i=1}^{\lfloor \log_c n - C_2\rfloor} G_i.$$

A simple direct computation using nothing but $1+x \le e^x, x\in\mathbb{R}$, shows that a geometrically distributed random variable with success probability $0<p<1$ is stochastically smaller than $1 + E/p$ where E has the standard exponential distribution. It follows that, in probability,

$$L_n \leq \sum_{i=1}^{\lceil \log_c n \rceil} (1 + \sqrt{\delta n}(\beta^*)^{-1}(1-\delta)^{i/2} E_i) + (1-\delta)^{-C_2-1},$$

where $E_1, E_2, \ldots$ is a sequence of independent random variables each of which having the standard exponential distribution. Hence, in probability,

$$\frac{\beta^*(L_n - (1-\delta)^{-C_2-1} - \lceil \log_c n \rceil)}{\sqrt{\delta n}} \leq \sum_{i=0}^{\infty} (1-\delta)^{i/2} E_i.$$

It follows that $L_\infty \leq \frac{\sqrt{\delta}}{\beta^*} \sum_{i=0}^{\infty} (1-\delta)^{i/2} E_i$ stochastically. From here, the lower bound on the limit inferior follows from Lemma 6.3 since we can choose δ arbitrarily close to $1/2$.

It remains to prove the bound (6.2). Let $t \in \mathbb{N}$, $j \geq 0$ and $n \in I$. Then,

$$\begin{aligned}
&\mathbf{P}(\sigma_{j+1} \geq t) - \mathbf{P}(\sigma_j \geq t) \\
&= \sum_{k=0}^{t-1} \sum_{\ell=\lceil (1-\delta)^{j+1} n \rceil}^{\lfloor (1-\delta)^j n \rfloor} \mathbf{P}(\sigma_{j+1} \geq t | N^*(e_k) = \ell, \sigma_j = k)\, \mathbf{P}(N^*(e_k) = \ell, \sigma_j = k) \\
&= \sum_{k=0}^{t-1} \sum_{\ell=\lceil (1-\delta)^{j+1} n \rceil}^{\lfloor (1-\delta)^j n \rfloor} \mathbf{P}\left(\tilde{N}(\ell, t-k-1) > (1-\delta)^{j+1} n\right) \mathbf{P}(N^*(e_k) = \ell, \sigma_j = k),
\end{aligned}$$

where $(\tilde{N}(\ell, i))_{i \geq 0}$ is distributed like $(N^*(e_i))_{i \geq 0}$ but in the tree τ_ℓ. For any $(1-\delta)^{j+1} n < m \leq \ell \leq \lfloor (1-\delta)^j n \rfloor$ and $i \geq 1$, we have

$$\mathbf{P}\left(\tilde{N}(\ell, i) \leq (1-\delta)^{j+1} n | \tilde{N}(\ell, i-1) = m\right) = \mathbf{P}\left(\tilde{N}_{2+}(m) \geq m - (1-\delta)^{j+1} n\right).$$

Now, we specify C_1, C_2 as follows: first, let $n_2' \geq n_2$ with n_2 as in Theorem 3.4 (ii) (with $k = 2$ and the chosen ε) such that $p \in I$ for all $p \geq n_2'$. Then, let C_2 be large enough such that $(1-\delta)^{2-C_2} \geq n_2'$ and set $C_1 = c^{C_2+2}$. It follows that, for all $n \geq C_1, j \leq \lfloor \log_c n - C_2 \rfloor$ and m satisfying $(1-\delta)^{j+1} n < m \leq \lfloor (1-\delta)^j n \rfloor$, we have $m \geq n_2'$ and $m - (1-\delta)^{j+1} n \leq (1-\varepsilon) m/2$. Thus, by Theorem 3.4 (ii), the right hand side of the last display is bounded from below by $\beta^*(m - (1-\delta)^{j+1} n)^{-1/2} \geq \beta^*(\delta(1-\delta)^j n)^{-1/2}$. Since $(\tilde{N}(\ell, i))_{i \geq 1}$ is a Markov chain, we have

$$\begin{aligned}
&\mathbf{P}(\sigma_{j+1} \geq t) - \mathbf{P}(\sigma_j \geq t) \\
&\leq \sum_{k=0}^{t-1} \sum_{\ell=\lceil (1-\delta)^{j+1} n \rceil}^{\lfloor (1-\delta)^j n \rfloor} \left(1 - \frac{\beta^*}{\sqrt{\delta(1-\delta)^j n}}\right)^{t-k-1} \mathbf{P}(N^*(e_k) = \ell, \sigma_j = k) \\
&\leq \sum_{k=0}^{t-1} \left(1 - \frac{\beta^*}{\sqrt{\delta(1-\delta)^j n}}\right)^{t-k-1} \mathbf{P}(\sigma_j = k).
\end{aligned}$$

Hence, $\mathbf{P}(\sigma_{j+1} \geq t) \leq \mathbf{P}(\sigma_j + G_{j+1} \geq t)$ where σ_j and G_{j+1} are independent. Iterating the argument concludes the proof. □

Proof of Proposition 6.2 (Upper bound). First of all, as it will become clear in the formulation of Theorem 6.12 in the next section, the scaling limit $\sigma L_\infty / 2$ does not depend on the offspring distribution. Hence, we may assume that $p_0 = p_2 = 1/2$. In particular,

$\sigma = 1$. Next, let $\{U_{i,j} : i, j \geq 1\}$ be a family of independent random variables with the uniform distribution on $[0,1]$. Let $2 < a' < a$ be non-algebraic. For $i \geq 1$, define

$$Q_i = |\{j \geq 0 : N^*(e_j) \in (na^{-i}, na^{-i+1}]\},$$
$$R_i = \min\left\{ t \in \mathbb{N} : \sum_{j=1}^{t} \beta_2^2 U_{i,j}^{-2} \geq n(m_i - a^{-i}) \right\}, \quad m_i = \frac{a^{-i+1}}{a'}.$$

Fix $k \in \mathbb{N}$ (large). We will show that for all n sufficiently large, stochastically,

$$\sum_{i=1}^{k} Q_i \geq \sum_{i=1}^{k} R_i. \tag{6.3}$$

For now, let us use this bound to conclude the proof of the proposition. Note that the random variable $U_{1,1}^{-2}$ is in the domain of attraction of a non-negative stable distribution with index $1/2$. More precisely, for some $c > 0$,

$$n^{-2} \sum_{j=1}^{n} U_{1,j}^{-2} \xrightarrow{d} \mathcal{S}, \quad \log \mathbf{E}\left[e^{i\lambda \mathcal{S}}\right] = -c|\lambda|^{1/2}(1 - i \text{ sign}(t)).$$

The limit law is the Levy distribution with density $\sqrt{c/(2\pi)}x^{-3/2}e^{-c/(2x)}$ on $[0, \infty)$. A straightforward computation shows that $\mathcal{S}^{-1/2}$ is distributed like $c^{-1/2}|\mathcal{N}|$, where $\mathcal{N}$ has the standard normal distribution. In particular, for any $x > 0$, as $n \to \infty$,

$$\mathbf{P}\left(R_i/\sqrt{n} \geq x\right) \to \mathbf{P}\left((m_i - a^{-i})^{1/2}(c\beta_2^2)^{-1/2}|\mathcal{N}| \geq x\right).$$

It follows that, for $x > 0$,

$$\begin{aligned}
\mathbf{P}(L_\infty \leq x) = \lim_{n\to\infty} \mathbf{P}\left(L_n \leq x\sqrt{n}\right) &\leq \limsup_{n\to\infty} \mathbf{P}\left(\sum_{i=1}^{k} Q_i \leq x\sqrt{n}\right) \\
&\leq \lim_{n\to\infty} \mathbf{P}\left(\sum_{i=1}^{k} R_i \leq x\sqrt{n}\right) \\
&= \mathbf{P}\left(\sum_{i=1}^{k} (c\beta_2^2)^{-1/2}(m_i - a^{-i})^{1/2}|\mathcal{N}_i| \leq x\right) \\
&= \mathbf{P}\left((c\beta_2^2)^{-1/2}(a/a' - 1)^{1/2} \sum_{i=1}^{k} a^{-i/2}|\mathcal{N}_i| \leq x\right),
\end{aligned}$$

where $\mathcal{N}_1, \mathcal{N}_2, \ldots$ are independent standard normal random variables. Since the left hand side does not depend on k, we may substitute $k = \infty$ on the right hand side. Lemma 6.3 concludes the proof since we can choose $a > 2$ arbitrarily.

It remains to prove (6.3). To this end, for $i \geq 1$, define $P_i = \max\{N(j) : N(j) \in [nm_i, na^{-i+1}]\}$. Subsequently, assume that $n \geq 4a^k a'/(a' - 2)$. Then, since for all non-leaves $v \in \tau_n$, we have $N^*(R(v)1) \geq (N(v) - 1)/2$, a simple computation shows that the quantities $P_1, \ldots, P_k$ are well-defined. Let $t > 0$. Then,

$$\begin{aligned}
\mathbf{P}\left(\sum_{i=1}^{k} Q_i \geq t\right) &= \sum_{x=\lceil nm_k \rceil}^{\lfloor na^{-k+1} \rfloor} \mathbf{P}\left(\sum_{i=1}^{k} Q_i \geq t, P_k = x\right) \\
&= \sum_{x=\lceil nm_k \rceil}^{\lfloor na^{-k+1} \rfloor} \sum_{\ell \geq 0} \mathbf{P}\left(Q_k \geq t - \ell, \sum_{i=1}^{k-1} Q_i = \ell \,\middle|\, P_k = x\right) \mathbf{P}(P_k = x).
\end{aligned}$$

Observe that, conditionally on $P_k = x$, the random variables $(Q_1, \dots, Q_{k-1}), Q_k$ are independent. Hence,

$$\begin{aligned}
&\mathbf{P}\left(\sum_{i=1}^{k} L_i \geq t\right) \\
&= \sum_{\ell \geq 0} \sum_{x=\lceil nm_k \rceil}^{\lfloor na^{-k+1} \rfloor} \mathbf{P}\left(Q_k \geq t-\ell \middle| P_k = x\right) \mathbf{P}\left(\sum_{i=1}^{k-1} Q_i = \ell \middle| P_k = x\right) \mathbf{P}\left(P_k = x\right).
\end{aligned}$$

The crucial observation is that, conditionally on $P_k = x$, the random variable Q_k is stochastically larger than R_k. To see this, note that, by Theorem 3.4 (i), we know that $N_2 \geq \beta_2^2 U_{1,1}^{-2}$ in probability. Hence, for any $\lceil nm_k \rceil \leq x \leq \lfloor na^{-k+1} \rfloor$ and $y \geq 1$, using the notation from the previous proof, we deduce

$$\begin{aligned}
\mathbf{P}\left(Q_k \geq y | P_k = x\right) &= \mathbf{P}\left(\tilde{N}(x,y) > na^{-k}\right) \\
&\geq \mathbf{P}\left(\sum_{j=1}^{y-1} \beta_2^2 U_{1,j}^{-2} < n(m_k - a^{-k})\right) = \mathbf{P}\left(R_k \geq y\right).
\end{aligned}$$

We conclude

$$\begin{aligned}
\mathbf{P}\left(\sum_{i=1}^{k} Q_i \geq t\right) &\geq \sum_{\ell \geq 0} \mathbf{P}\left(R_k \geq t-\ell\right) \sum_{x=\lceil nm_k \rceil}^{\lfloor na^{-k+1} \rfloor} \mathbf{P}\left(\sum_{i=1}^{k-1} Q_i = \ell \middle| P_k = x\right) \mathbf{P}\left(P_k = x\right) \\
&= \sum_{\ell \geq 0} \mathbf{P}\left(R_k \geq t-\ell\right) \mathbf{P}\left(\sum_{i=1}^{k-1} Q_i = \ell\right) \\
&= \mathbf{P}\left(\sum_{i=1}^{k-1} Q_i + R_k \geq t\right).
\end{aligned}$$

Iterating gives the desired claim and finishes the proof. □

6.2 Proof of Theorem 1.3 and further results

To keep this section self-contained, let us recall some definitions. For a discrete ordered rooted tree $\mathbb{T}$, the heavy path is defined as the unique path from the root to a leaf which always continues in the largest subtree. Here, ties are broken considering the preorder index. It is easy to read off the length of the heavy path from the depth-first search process encoding $\mathbb{T}$ since each excursion above a level corresponds to a subtree. Thus, starting with the interval $I_0 := [0, 2|\mathbb{T}| - 2]$ at time 0, given the interval I_i at time $i \geq 0$, I_{i+1} is chosen as the largest subinterval of I_i corresponding to an excursion above level $i + 1$. We now extend the concept to arbitrary continuous excursions. To this end, let

$$\mathcal{C}_{\text{ex}} := \{f : [0,1] \to \mathbb{R}_0^+ \text{ continuous} : f(0) = f(1) = 1\}.$$

We always consider $\mathcal{C}_{\text{ex}}$ endowed with the topology induced by the supremum norm $\|f\| = \sup_{t \in [0,1]} |f(t)|$.

Superlevel sets for excursions. Let $\mathcal{V}$ be the space of open subsets of $[0,1]$, where open refers to the subspace topology of $[0,1]$ in $\mathbb{R}$. For $O_1, O_2 \in \mathcal{V}$, we define $d(O_1, O_2) = d_{\mathrm{H}}(O_1^c, O_2^c)$, where d_{H} denotes the Hausdorff distance. For $O \in \mathcal{V}$ and a $\mathcal{V}$-valued sequence $O_n, n \geq 0$, we have $d(O_n, O) \to 0$ if any only if $\lambda(O_n \Delta O) \to 0$ where $A \Delta B := A \backslash B \cup B \backslash A$ and λ denotes the Lebesgue measure on $[0,1]$. $(\mathcal{V}, d)$ is a compact metric space (hence Polish). Every element of $\mathcal{V}$ uniquely decomposes in at most countably many disjoint open intervals.

For a function $f \in \mathcal{C}_{\text{ex}}$ and $t \geq 0$, the superlevel set $\mathcal{P}_f(t) = \{s \in [0,1] : f(s) > t\}$ is open. The $\mathcal{V}$-valued process $\mathcal{P}_f := \mathcal{P}_f(t), t \geq 0$ has the following properties:

(i) $\mathcal{P}_f(t) \subseteq \mathcal{P}_f(s)$ for $0 \leq s \leq t$,

(ii) $\mathcal{P}_f$ is right-continuous, that is, $\mathcal{P}_f(t) = \lim_{s \downarrow t} \mathcal{P}_f(s)$ for all $t \geq 0$,

(iii) $\mathcal{P}_f(t) = \emptyset$ for all t large enough, and

(iv) $x \in \partial \mathcal{P}_f(t) \Rightarrow x \notin \partial \mathcal{P}_f(s)$ for all $0 \leq t < s$.

Here, and subsequently, ∂O denotes the boundary of an open set $O \subseteq [0,1]$. Conversely, for every $\mathcal{V}$-valued process $\mathcal{P}_t, t \geq 0$ satisfying (i)–(iii), we can define

$$f_{\mathcal{P}}(t) = \sup\{s \geq 0 : t \in \mathcal{P}_s\},$$

and observe that $\mathcal{P}_t = \mathcal{P}_{f_{\mathcal{P}}}(t)$ for all $t \geq 0$. Note that $f_{\mathcal{P}}$ is lower semi-continuous. (A non-negative function on $[0,1]$ is lower semi-continuous if and only if $\mathcal{P}_f(t)$ is open for all $t \geq 0$.) Further, for all $f_{\mathcal{P}}(t) \leq s < f_{\mathcal{P}}(t-)$, we have $t \in \partial \mathcal{P}_f(s)$. In particular, it easily follows that $f_{\mathcal{P}} \in \mathcal{C}_{\text{ex}}$ if and only if $\mathcal{P}_t, t \geq 0$ satisfies (iv). Letting $\mathcal{W}$ be the set of $\mathcal{V}$-valued processes satisfying (i)–(iv), the map $f \mapsto \mathcal{P}_f$ is a bijection between $\mathcal{C}_{\text{ex}}$ and $\mathcal{W}$.

The heavy path construction. For $O \in \mathcal{V}$, let $\mathfrak{m}(O)$ denote the interval with largest length in O. In case several intervals qualify, we choose the smallest of them with respect to the order $\preceq$ defined for intervals I, I' by

$$I \preceq I' :\Leftrightarrow \inf I \leq \inf I'.$$

For a $\mathcal{V}$-valued process $\mathcal{P}$, we define a process $\mathcal{P}^*_t, t \geq 0$ with $\mathcal{P}^*_t \subseteq \mathcal{P}_t$ for all $t \geq 0$ as follows: set $\mathcal{P}^*_0 = \mathcal{P}_0$ and $T_0 = 0$. Then, inductively, for $n \geq 0$, given T_n and $\mathcal{P}^*_t$ for all $t \leq T_n$, let

$$\begin{aligned} T_{n+1} &= \inf\{t > T_n : \mathfrak{m}(\mathcal{P}^*_{T_n} \cap \mathcal{P}_t) \leq 2^{-(n+1)}\}, \\ \mathcal{P}^*_t &= \mathfrak{m}(\mathcal{P}^*_{T_n} \cap \mathcal{P}_t), \quad T_n < t < T_{n+1}, \\ \mathcal{P}^*_{T_{n+1}} &= \mathfrak{m}\left(\lim_{s \uparrow T_{n+1}} \mathcal{P}^*_s \cap \mathcal{P}_t\right). \end{aligned}$$

$T_\infty := \lim_{n \to \infty} T_n$ is finite and bounded by $\inf\{t \geq 0 : \mathcal{P}_t = \emptyset\}$. For $t \geq T_\infty$, we set $\mathcal{P}^*_t = \emptyset$. Then, $\mathcal{P}^* \in \mathcal{W}$ and $\mathcal{P}^*_t$ is an interval for all $t \geq 0$. We also define $t_* = \lim_{n \to \infty} \inf \mathcal{P}^*_{T_n}$ and $t^* = \lim_{n \to \infty} \sup \mathcal{P}^*_{T_n}$. We call $\mathcal{P}$ trivial if $\mathcal{P}_t = \emptyset$ for all $t \geq 0$. For a non-trivial process $\mathcal{P}_t, t \geq 0$, two scenarios are possible:

(i) $T_n < T_\infty$ for all $n \geq 1$. Then, $\mathcal{P}^*_t$ is continuous at T_∞ and $t_* = t^*$.

(ii) $T_n = T_\infty$ for some $n \geq 1$. Then, $\mathcal{P}^*_t$ is discontinuous at T_∞ and $t_* < t^*$.

For $f \in \mathcal{C}_{\text{ex}}$, write $\mathcal{P}^*_f$ for $\mathcal{P}^*$ and T^f_∞ for T_∞ when $\mathcal{P} = \mathcal{P}_f$. If f is the depth-first search process of a discrete ordered rooted tree rescaled on the unit interval then T^f_∞ is the length of the corresponding heavy path. For a discussion of the heavy path in a general real tree, see the end of this section.

Remark 6.4. The sequence $T_n, n \geq 0$ arising in the heavy path construction plays no role in the sequel. We could replace the sequence $2^{-(n+1)}, n \geq 0$ in its definition by any monotonically decreasing sequence $\alpha_n, n \geq 0$ with $\alpha_n \to 0$ and $\alpha_n \geq 2^{-(n+1)}$. This leaves $\mathcal{P}^*$ and T_∞ invariant. In fact, we could also let α_n depend on $\mathcal{P}$ by setting $\alpha_n = \frac{1}{2}\lambda(\mathcal{P}^*_{T_n})$.

Unfortunately, some technical issues arise in this construction. The map $O \to \lambda(\mathfrak{m}(O))$ is continuous, and so is $(O, O') \mapsto O \cap O'$. Similarly, the map $O \mapsto \inf O$ ($O \mapsto \sup O$, respectively) is measurable and continuous at $O \in \mathcal{V}$ if and only if $0 \in O$ ($1 \in O$, respectively). The map $O \to \mathfrak{m}(O)$ is measurable and continuous at $O \in \mathcal{V}$ if only if the largest interval in O is unique. For any fixed $t \geq 0$, the map $f \to \mathcal{P}_f(t)$ is not continuous on $\mathcal{C}_{\text{ex}}$. The set $\mathcal{W}$ is not closed when endowing the set of all $\mathcal{V}$-valued processes with the topology of uniform convergence on compact sets. The following important lemma contains a positive result in the converse direction. Here and subsequently, we recall the definition of the modulus of continuity of a continuous function f on $[0,1]$:

$$\omega_f(\varepsilon) = \sup_{|s-t| \leq \varepsilon} |f(t) - f(s)|, \quad \varepsilon > 0.$$

By the Arzela-Ascoli theorem, for a family of continuous functions (f_i) on $[0,1]$, we have $\sup_i \omega_{f_i}(\varepsilon) \to 0$ as $\varepsilon \to 0$ if (f_i) is relatively compact. (In other words, the family is uniformly equicontinuous.)

Lemma 6.5. *Let $f, f_n, n \geq 1$ be continuous excursions. Suppose that, uniformly on compact sets, we have $d((\mathcal{P}_{f_n}(t), \mathcal{P}_f(t)) \to 0$. Then, $\|f_n - f\| \to 0$.*

Proof. For ease of notation, abbreviate $\mathcal{P}_n := \mathcal{P}_{f_n}, n \geq 1$ and $\mathcal{P} := \mathcal{P}_f$. Fix $\varepsilon > 0$ and n large enough such that $d((\mathcal{P}_n)_t, \mathcal{P}_t) \leq \varepsilon$ for all $0 \leq t \leq \|f\|$. Fix $t \in [0,1]$ and let $x_n = f_n(t)$. Suppose that $t \in \partial\mathcal{P}_n(x_n)$. Then, there exists $t'_n \in f^{-1}(\{x_n\})$ with $|t'_n - t| \leq \varepsilon$ and $t'_n \in \partial\mathcal{P}(x_n)$. This implies $|f(t) - f_n(t)| \leq \omega_f(\varepsilon)$. If $t \notin \partial\mathcal{P}_n(x_n)$, then $f_n = x_n$ on some closed interval I_n containing t which may choose maximal. If $\sup I_n < t + 2\varepsilon$, then, since $\sup I_n \in \partial\mathcal{P}_n(x_n)$, we have $|f_n(t) - f(t)| \leq |f_n(\sup I_n) - f(\sup I_n)| + |f(\sup I_n) - f(t)| \leq 2\omega_f(2\varepsilon)$ from the first part of the proof. The same bound follows if $\inf I > t - 2\varepsilon$. Now, assume $[t-2\varepsilon, t+2\varepsilon] \subseteq I$. Then, we must have $f \geq x_n$ on $[t-\varepsilon, t+\varepsilon]$. If $f(t) \neq f_n(t)$, since $f > x_n$ is not possible on the entire interval $[t-2\varepsilon, t+2\varepsilon]$, there exists $t'_n \in [t-2\varepsilon, t+2\varepsilon]$ with $t'_n \in \partial\mathcal{P}(x_n)$. As above, this implies $|f(t) - f_n(t)| \leq \omega_f(2\varepsilon)$. Since f is continuous, we have $\omega_f(\varepsilon) \to 0$ as $\varepsilon \to 0$ finishing the proof. ☐

The Skorokhod space. Let (S, d) be a Polish space. By $\mathcal{D}_S$ we denote the set of càdlàg functions with values in S. A function $f : [0, \infty) \to S$ is called càdlàg if, for all $t \geq 0$, it is right-continuous at t and, for all $t > 0$, the left limit $f(t-) := \lim_{s \uparrow t} f(s)$ exists. $\mathcal{D}_S$ is endowed with the Skorokhod topology: a sequence $f_n, n \geq 1$ converges to a function f if and only if there exists a sequence of strictly increasing continuous functions $\lambda_n : [0, \infty) \to [0, \infty)$ such that $\lambda_n \to$ id uniformly on $[0, \infty)$ and $f_n \circ \lambda_n \to f$ uniformly on compact sets. For details on $\mathcal{D}_s$, we refer to Billingsley's book [19]. Again, one can easily check that $f \mapsto \mathcal{P}_f$ is not continuous on $\mathcal{C}_{\text{ex}}$. Further, $\mathcal{W} \subseteq \mathcal{D}_{\mathcal{V}}$ is not closed. The following lemma is crucial.

Lemma 6.6. *The set $\mathcal{W} \subseteq \mathcal{D}_{\mathcal{V}}$ endowed with its relative topology is Polish. $\mathcal{W}$ is measurable with respect to the Borel-σ-algebra on $\mathcal{D}_{\mathcal{V}}$. Further, the map $f \mapsto \mathcal{P}_f$ from $\mathcal{C}_{\text{ex}}$ to $\mathcal{D}_{\mathcal{V}}$ is measurable.*

Proof. Let us first show that $\mathcal{P} \mapsto f_{\mathcal{P}}$ is continuous regarded as map $\mathcal{W} \to \mathcal{C}_{\text{ex}}$. To this end, let $\mathcal{P}, \mathcal{P}_n, n \geq 1$ be elements in $\mathcal{W}$ with $\mathcal{P}_n \to \mathcal{P}$ in the Skorokhod topology. Choose a sequence $\lambda_n, n \geq 1$ of strictly increasing continuous bijections on $[0, \infty)$ with $\lambda_n \to$ id uniformly on $[0, \infty)$ and $\mathcal{P}_n \circ \lambda_n \to \mathcal{P}$ uniformly on compact sets. By Lemma 6.5, $\|f_{\mathcal{P}_n \circ \lambda_n} - f_{\mathcal{P}}\| \to 0$. Hence, it remains to show that $\|f_{\mathcal{P}_n \circ \lambda_n} - f_{\mathcal{P}_n}\| \to 0$. But for any $\mathcal{P}' \in \mathcal{W}$ and any strictly increasing bijection λ, we have $f_{\mathcal{P}' \circ \lambda} = \lambda^{-1} \circ f_{\mathcal{P}'}$. Thus, $\|f_{\mathcal{P}_n \circ \lambda_n} - f_{\mathcal{P}_n}\| \leq \sup_{t > 0} |\lambda(t) - t| \to 0$. This shows the claimed continuity.

In view of Lemma 6.5, for $\mathcal{P}, \mathcal{P}' \in \mathcal{W}$, define

$$d^*(\mathcal{P}, \mathcal{P}') = \|f_{\mathcal{P}} - f_{\mathcal{P}'}\| + d_{\text{sk}}(\mathcal{P}, \mathcal{P}'),$$

where d_{sk} denotes any complete metric generating the Skorokhod topology on $\mathcal{D}_{\mathcal{V}}$. (See [19] for an explicit construction.) From the continuity of $\mathcal{P} \mapsto f_{\mathcal{P}}0$, it follows that d^* generates the relative topology on $\mathcal{W}$. Since $(\mathcal{D}_{\mathcal{V}}, d_{\text{sk}})$ is separable, the same follows for $(\mathcal{W}, d^*)$. If $\mathcal{P}_n, n \geq 1$ is Cauchy with respect to d^*, then it is Cauchy with respect to d_{sk}. Hence, there exists a d_{sk}-limit $\mathcal{P}' \in \mathcal{D}_{\mathcal{V}}$. Further, by definition and completeness of the supremum norm, there exists $g \in \mathcal{C}_{\text{ex}}$ with $\|f_{\mathcal{P}_n} - g\| \to 0$. Clearly, this implies $g = f_{\mathcal{P}'}$ and $\mathcal{P}' \in \mathcal{W}$. Hence, $\mathcal{W}$ is complete with respect to d^*. By construction, the embedding $\mathcal{W} \to \mathcal{D}_{\mathcal{V}}$ is continuous. Both measurability of $\mathcal{W}$ and measurability of $f \mapsto \mathcal{P}_f$ now follow from the Lusin-Suslin theorem [38, Theorem 15.1]. □

Finally, one also has to verify measurability of the quantities arising in the construction of the heavy path.

Lemma 6.7. *The maps $\mathcal{P} \mapsto T_\infty$ and $\mathcal{P} \to \mathcal{P}^*$ are measurable.*

Proof. We keep track of more quantities in the construction. Set $\mathcal{P}_0^{(0)} = \mathcal{P}_0$. Inductively, for $n \geq 0$,

$$\begin{aligned}
\mathcal{P}_t^{(n)} &:= \mathfrak{m}(\mathcal{P}_{T_n}^{(n)}), \quad t < T_n, \\
\mathcal{P}_t^{(n)} &:= \mathfrak{m}(\mathcal{P}_{T_n}^{(n)} \cap \mathcal{P}_t), \quad t > T_n, \\
T_{n+1} &= \inf\{t > 0 : \lambda(\mathcal{P}_t^{(n)}) \leq 2^{-(n+1)}\}, \\
\mathcal{P}_{T_{n+1}}^{(n+1)} &:= \mathfrak{m}\left(\mathcal{P}_{T_{n+1}-}^{(n)} \cap \mathcal{P}_t\right).
\end{aligned}$$

(The third line is merely an observation.) Then, $\mathcal{P}_t^* = \sum_{i=0}^{\infty} \mathbf{1}_{[T_i, T_{i+1})}(t) \mathcal{P}_t^{(i)}$. In order to show that $\mathcal{P} \to \mathcal{P}^*$ is measurable, we need to verify that, for all $i \geq 0$, $\mathcal{P} \to \mathcal{P}^{(i)}$ is measurable and that T_i is a stopping-time with respect to the family of σ-algebras $\mathcal{F}_t = \sigma(\{\pi_s : 0 \leq s \leq t\})$. (This means that $\{T_i \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$, where $\pi_t : \mathcal{D}_{\mathcal{V}} \to \mathcal{V}, \pi_t(\mathcal{P}) = \mathcal{P}_t$.) This can be done by induction on i. Clearly, $\mathcal{P}_t^{(0)}$ is measurable. T_1 is a hitting-time of a closed set, therefore a stopping time by standard arguments. Further, it is well-known that $\mathcal{P} \to \mathcal{P}_T$ is measurable for any stopping-time T. Finally, the map $\mathcal{P} \mapsto \mathcal{P}_- := (\mathcal{P}_{t-}), t \geq 0$ is measurable. Hence, $\mathcal{P}_t^{(1)}$ is measurable. Now, proceed inductively. Measurability of $\mathcal{P} \mapsto T_\infty$ follows since T_∞ is the limit of measurable functions. □

Continuity properties. In this section, we discuss regularity of the map $f \mapsto \mathcal{P}_f^*$. It should be clear that if $\mathcal{P}_f^*$ is continuous at t and $\lambda(\mathcal{P}_f^*(t)) < 1/2$, then, for any sequence $f_n \to f$, we have $d(\mathcal{P}_{f_n}^*(t), \mathcal{P}_f^*(t)) \to 0$. This follows with not much work from, e.g. Proposition 5.1 in [35]. In general, such a statement is not true if $\lambda(\mathcal{P}_f^*(t)) > 1/2$. This leads us to consider suitable subsets of excursions and to study time-transformations on $\mathbb{R}$ to align points of discontinuity of $\mathcal{P}_{f_n}^*$ and $\mathcal{P}_f^*$ (Lemma 6.8).

For $f \in \mathcal{C}_{\text{ex}}$, define

$$M_f(x) = \{(s,t) : 0 \leq s < t \leq 1, f(s) = f(t) = x, f > x \text{ on } (s,t)\}, \quad x \geq 0.$$

Now, let

$$\mathcal{C}_{\text{ex}}^{(1)} = \{f \in \mathcal{C}_{\text{ex}} : \text{For all } 0 \leq x \leq \|f\| \text{ there exists at most one pair } (s,t) \in M_f(x) \\ \text{maximizing } t - s\}$$

and

$$\mathcal{C}_{\text{ex}}^{(2)} = \{f \in \mathcal{C}_{\text{ex}} : \text{For all } t \geq 0 \text{ the set } \{x : f(x) = t\} \text{ contains at most one} \\ \text{local minimum of } f\}.$$

In a Brownian excursion, all local minima are strict and pairwise distinct. Hence, for all $x \geq 0$, the set $M_f(x)$ contains at most two elements and $\mathbf{e} \in \mathcal{C}_{\text{ex}}^{(2)}$. It is well-known that every local minima t does not decompose the interval $(\sup\{s < t : \mathbf{e}(s) > \mathbf{e}(t)\}, \inf\{s > t : \mathbf{e}(s) > \mathbf{e}(t)\})$ equidistantly. Hence, $\mathbf{e} \in \mathcal{C}_{\text{ex}}^{(1)}$. For $f \in \mathcal{C}_{\text{ex}}$, define

$$\mathbf{m}_f(t) := \lambda(\mathcal{P}_f^*(t)), \quad t \geq 0, \quad \zeta_f(t) := \inf\{s > 0 : \mathbf{m}_f(s) \leq t\}, \quad t \in [0,1].$$

The map $t \mapsto \zeta_f(t)$ is continuous. Every point of discontinuity of $\mathcal{P}_f^*$ (or, equivalently, of $\mathbf{m}_f$) corresponds to an interval on which ζ_f is constant. For $f \in \mathcal{C}_{\text{ex}}$ let $M_f = \{\mathbf{m}_f(t) : t \geq 0\}$. Further, for $r \geq 0$ and $f \in \mathcal{C}_{\text{ex}}$, set

$$f_r^*(t) := (f(t) - f(\inf \mathcal{P}_f^*(r))\mathbf{1}_{\mathcal{P}_f^*(r)}(t).$$

Clearly, if $f \in \mathcal{C}_{\text{ex}}^{(1)}$ then $f_r^* \in \mathcal{C}_{\text{ex}}^{(1)}$, analogously for $\mathcal{C}_{\text{ex}}^{(2)}$. We now set $\mathcal{C}_{\text{ex}}^* = \mathcal{C}_{\text{ex}}^{(1)} \cap \mathcal{C}_{\text{ex}}^{(2)}$.

In the following lemma, recall that, for a càdlàg function f with values in a Polish space and $t > 0$, we have set $f(t-) := \lim_{s \uparrow t} f(s)$.

Lemma 6.8. *Let $f_n, n \geq 1$ be a sequence of continuous excursions and $f \in \mathcal{C}_{\text{ex}}^*$. Suppose that $\|f_n - f\| \to 0$. Let $r \in M_f$ with $\mathbf{m}_f(\zeta_f(r)-) \geq \frac{1}{2}\mathbf{m}_f(0)$. Then, there exists a sequence $r_n \to r$ with $\zeta_{f_n}(r_n) \to \zeta_f(r)$ such that*

$$d(\mathcal{P}_{f_n}^*(\zeta_{f_n}(r_n)), \mathcal{P}_f^*(\zeta_f(r)) \to 0, \tag{6.4}$$

$$d(\mathcal{P}_{f_n}^*(\zeta_{f_n}(r_n)-), \mathcal{P}_f^*(\zeta_f(r)-)) \to 0, \tag{6.5}$$

and $\|(f_n)_{\zeta_{f_n}(r_n)}^* - f_{\zeta_f(r)}^*\| \to 0$.

Proof. It is easy to see that, for any $r, s \in [0,1]$ and $f, g \in \mathcal{C}_{\text{ex}}$, we have

$$\begin{aligned} &|\|f_r^* - g_s^*\| - \|f-g\|| \leq \\ &\quad \omega_f(|\inf \mathcal{P}_f(\zeta_f(r)) - \inf \mathcal{P}_g(\zeta_g(s)|) + \omega_f(|\sup \mathcal{P}_f(\zeta_f(r)) - \sup \mathcal{P}_g(\zeta_g(s))|) \\ &\quad + \omega_g(|\inf \mathcal{P}_f(\zeta_f(r)) - \inf \mathcal{P}_g(\zeta_g(s))|) + \omega_g(|\sup \mathcal{P}_f(\zeta_f(r)) - \sup \mathcal{P}_g(\zeta_g(s))|). \end{aligned}$$

Hence, the final claim of the lemma is a direct implication of the remaining statements. If $\mathbf{m}_f$ is continuous at $\zeta_f(r)$, then we can simply choose $r_n = r$. In this case, if $r > \mathbf{m}_f(1/2)$, the assertions (6.4), (6.5) even hold for general $f \in \mathcal{C}_{\text{ex}}$. The interesting case is when $\mathbf{m}_f$ is discontinuous at $\zeta_f(r)$ which we assume from now on. Let $\alpha = \inf \mathcal{P}_f(\zeta_f(r)-)$ and $\beta = \sup \mathcal{P}_f(\zeta_f(r)-)$. Since $f \in \mathcal{C}_{\text{ex}}^{**}$ there exists a unique strict minimum x of f on (α, β) such that, either, i) $\mathcal{P}_f(\zeta_f(r)) = (\alpha, x)$, or, ii) $\mathcal{P}_f(\zeta_f(r)) = (x, \beta)$. We have $x \neq (\alpha+\beta)/2$ since $x \in \mathcal{C}_{\text{ex}}^*$. Let $\alpha' = (\alpha+x)/2, \beta' = (\beta+x)/2$ and $s_n = \inf\{f_n(s) : \alpha' < s < \beta'\}$. In case of i), let $x_n = \inf\{\alpha' < y < \beta' : f(y) = s_n\}$, while, for ii), we set $x_n = \sup\{\alpha' < y < \beta' : f(y) = s_n\}$. Now, let $r_n = \mathbf{m}_{f_n}(s_n)$. Then, for all n sufficiently large, there exist $\alpha_n < x_n < \beta_n$ such that $\mathcal{P}_{f_n}(s_n-) = (\alpha_n, \beta_n)$ and, for i), $\mathcal{P}_{f_n}(s_n) = (\alpha_n, x_n)$ while, for ii), $\mathcal{P}_{f_n}(s_n) = (x_n, \beta_n)$. We also have $\alpha_n \to \alpha, \beta_n \to \beta$ and $x_n \to x$. All statements follow readily. $\square$

Proposition 6.9. *The map $f \mapsto \mathcal{P}_f^*$ is continuous at every $f \in \mathcal{C}_{\text{ex}}^*$.*

Proof. Let $\varepsilon > 0$ be small. Let $f^{(0)} = f$ and, recursively, $f^{(\ell+1)} = (f^{(\ell)})_{\zeta_{f^{(\ell)}}(1-\varepsilon)}^*$. Define $s^{(\ell)} = \zeta_{f^{(\ell)}}(1-\varepsilon)$ and $\beta^{(\ell)} = \sum_{k=0}^{\ell} s^{(k)}$. Assume that $\|f_n - f\| \to 0$ for a sequence of continuous excursions $f_n, n \geq 1$. Denote by $r_n^{(0)}$ the sequence from Lemma 6.8 with $r = \mathbf{m}_f(1-\varepsilon)$. Set $s_n^{(0)} := \zeta_{f_n}(r_n^{(0)})$ and $f_n^{(0)} := f_n$. Then, for $\ell \geq 0$, inductively,

$f_n^{(\ell+1)} := (f_n^{(\ell)})^*_{s_n^{(\ell)}}$, where $s_n^{(\ell)} = \zeta_{f_n^{(\ell)}}(r_n^{(\ell)})$ and $r_n^{(\ell)}$ is the sequence from Lemma 6.8 based on the functions $f^{(\ell)}, f_n^{(\ell)}, n \geq 1$ and $r = \mathbf{m}_{f^{(\ell)}}(1-\varepsilon)$. Let $\beta_n^{(\ell)} = \sum_{k=0}^{\ell} s_n^{(k)}$.

By Lemma 6.8, we have $s_n^{(\ell)} \to s^{(\ell)}$, hence $\beta_n^{(\ell)} \to \beta^{(\ell)}$ for all $\ell \geq 0$. Now, fix K large and assume that n is sufficiently large such that $s_n^{(\ell)} > 0$ for all $0 \leq \ell \leq K$. Define $\lambda_n(\beta^{(\ell)}) = \beta_n^{(\ell)}$ for all $1 \leq \ell \leq K$, and linear on interval $[\beta^{(\ell)}, \beta^{(\ell+1)}]$, $0 \leq \ell \leq K-1$. Extend λ_n to a continuous bijection on $[0,\infty)$ by a straight line of slope one for $t \geq \beta^{(K)}$. Clearly, $\lambda_n \to \mathrm{id}$ uniformly on $[0,1]$. Fix $\varepsilon' > 0$. By Lemma 6.8, for all n sufficiently large, for all $1 \leq \ell \leq K$ we have $|d(\mathcal{P}^*_{f_n}(\beta_n^{(\ell)}), \mathcal{P}^*_f(\beta^{(\ell)}))| \leq \varepsilon'$ and $|d(\mathcal{P}^*_{f_n}(\beta_n^{(\ell)}-), \mathcal{P}^*_f(\beta^{(\ell)}-))| \leq \varepsilon'$. Hence, for those n,

$$\sup_{0 \leq t \leq \beta^{(K)}} |d(\mathcal{P}^*_{f_n} \circ \lambda_n(t), \mathcal{P}^*_f(t))| \leq \varepsilon + \varepsilon'.$$

By construction, for those large n,

$$\sup_{t > \beta^{(K)}} |d(\mathcal{P}^*_{f_n} \circ \lambda_n(t), \mathcal{P}^*_f(t))| \leq \omega_f((1-\varepsilon)^K) + \omega_{f_n}((1-\varepsilon)^K + 2\varepsilon').$$

Since we choose both $\varepsilon, \varepsilon'$ arbitrarily small and K arbitrarily large, this finishes the proof. □

Proposition 6.10. *The map $f \mapsto \zeta_f$ is continuous at every $f \in \mathcal{C}^*_{\mathrm{ex}}$. (Here, ζ_f is considered as an element in the space of continuous functions on $[0,1]$ endowed with the supremum norm.) In particular, $f \mapsto T^f_\infty$ is continuous at $f \in \mathcal{C}^*_{\mathrm{ex}}$.*

Proof. First of all, it is easy to see that, for all $\varepsilon > 0$ and $f \in \mathcal{C}_{\mathrm{ex}}$, we have $\omega_{\zeta_f}(\varepsilon) \leq \omega_f(\varepsilon)$. Now, suppose that $\|f_n - f\| \to 0$ with $\in \mathcal{C}^*_{\mathrm{ex}}$. For ease of notation, let us abbreviate $\zeta_n := \zeta_{f_n}, n \geq 1$ and $\zeta = \zeta_f$. Clearly, $\zeta_n(1) = 0$ for all n sufficiently large. Hence, by the Arzela-Ascoli theorem, (ζ_n) is relatively compact. It suffices to prove that, for any $t \in (0,1)$, we have $\zeta_n(t) \to \zeta(t)$. Assume that $\mathbf{m}_f$ is continuous at $\zeta_f(t)$. Then $\mathbf{m}_{f_n}(\zeta_f(t)) \to \mathbf{m}_f(\zeta_f(t)) = t$. Clearly, for every $\varepsilon > 0$ we can find $t - \varepsilon < s < t$ such that $\mathbf{m}_f$ is continuous at $\zeta_f(s)$. In particular, $\mathbf{m}_{f_n}(\zeta_f(s)) \to \mathbf{m}_f(\zeta_f(s)) = s$. This implies $\liminf \zeta_n(t) \geq \zeta(s)$. By continuity, it follows $\liminf \zeta_n(t) \geq \zeta(t)$. Similarly, one shows $\limsup \zeta_n(t) \leq \zeta(t)$.

Now assume that $\mathbf{m}_f$ is discontinuous at $\zeta_f(t)$. Then, there exist $t' \leq t \leq t''$ with $t' < t''$ such that ζ_f is constant on $[t', t'']$. For any $\varepsilon > 0$ there exists $t' - \varepsilon < s < t'$ such that $\mathbf{m}_f(s)$ is continuous at s. By the first part, this implies $\zeta_n(s) \to \zeta(s)$. By monotonicity, $\limsup \zeta_n(t') \leq \zeta(s)$. By continuity, this implies $\limsup \zeta_n(t') \leq \zeta(t')$. Similarly, $\liminf \zeta_n(t'') \geq \zeta(t'')$. Since $\zeta(t') = \zeta(t'')$ this implies $\zeta_n(x) \to \zeta(x)$ for all $x \in [t', t'']$. □

Remark 6.11. It is important to note that neither of the two propositions holds for general $f \in \mathcal{C}^{(1)}_{\mathrm{ex}}$ or $f \in \mathcal{C}^{(2)}_{\mathrm{ex}}$; both conditions are important.

We can now apply the continuous mapping theorem. The following result contains the first statement in Theorem 1.3. Note that the quantity L_∞ in Theorem 1.3 equals $2T^{\mathbf{e}}_\infty/\sigma$ here, while $P_\infty = \mathbf{m}_{\frac{2}{\sigma}\mathbf{e}}$ and $Q_\infty = 2\zeta_{\mathbf{e}}/\sigma$.

Theorem 6.12. *Consider a Galton-Watson tree whose offspring distribution satisfies* (1.1) *conditional on having size n with $n \in I$.*

(i) *Let L_n be the length of the corresponding heavy path. Then, in distribution,*

$$\frac{L_n}{\sqrt{n}} \to \frac{2}{\sigma} \cdot T^{\mathbf{e}}_\infty .$$

(ii) *For $k \geq 0$, let $P_n(k)$ be the size of the subtree rooted at the node on level k on the heavy path. In distribution, in the Skorokhod topology on $\mathcal{D}_{[0,\infty)}$,*

$$\frac{P_n(\lfloor \cdot \sqrt{n} \rfloor)}{n} \to \mathbf{m}_{\frac{2}{\sigma}\mathbf{e}} \, . \tag{6.6}$$

(iii) *For $0 \leq \ell \leq n$, let $Q_n(\ell) = \inf\{k \geq 0 : P_n(k) \leq \ell\}$. Then, in distribution, on the space of continuous functions on $[0,1]$,*

$$\frac{Q_n(\lfloor \cdot \, n \rfloor)}{\sqrt{n}} \to \frac{2}{\sigma} \cdot \zeta_{\mathbf{e}} \, . \tag{6.7}$$

The heavy path in the Brownian Continuum tree. Interval decompositions governed by a Brownian excursion can be studied with the help of self-similar fragmentations introduced by Bertoin [16]. We recall a version of Definition 2 in this work: a $\mathcal{V}$-valued process $F(t), t \geq 0$ with càdlàg paths is called *self-similar with index* $\alpha \in \mathbb{R}$, if

(1) $F(0) = [0,1]$, $F(t) \subseteq F(s)$ for all $t \geq s \geq 0$;
(2) $F(t)$ is continuous in probability at every $t \geq 0$;

further, given $F(t) = \cup I_j$ for $t \geq 0$ and disjoint open intervals $I_1, I_2, \ldots$,

(3) the processes $(F(t+s) \cap I_j)_{s \geq 0}, j \geq 1$ are stochastically independent;
(4) for all $j \geq 1$, $F(t+s) \cap I_j, s \geq 0$ is distributed like $F(|I_j|^\alpha s), s \geq 0$ rescaled to fit on I_j.

Bertoin [16] observes that $\mathcal{P}_{\mathbf{e}}(t), t \geq 0$ is a self-similar fragmentation process with $\alpha = -1/2$. Hence, the process $\mathcal{P}^*_{\mathbf{e}}(t), t \geq 0$ is also a self-similar process with $\alpha = -1/2$. For $t \geq 0$, let

$$\varrho_1(t) = \begin{cases} \inf\left\{u \geq 0 : \int_0^u \sqrt{\mathbf{m}_{\mathbf{e}}(r)} dr > t\right\}, & \text{if } t < \int_0^\infty \sqrt{\mathbf{m}_{\mathbf{e}}(r)} dr \\ \infty & \text{otherwise.} \end{cases}$$

It follows from [16, Theorem 2] that the $\mathcal{V}$-valued càdlàg process $H(\cdot) := \mathcal{P}_{\mathbf{e}}(\varrho_1(\cdot))$ is a *homogeneous* interval fragmentation, that is, a self-similar fragmentation process with index $\alpha = 0$. (Here, and subsequently, we abbreviate $\mathcal{P}_{\mathbf{e}}(\infty) = H(\infty) = \emptyset$.) Homogeneous fragmentation processes were studied in detail in another work of Bertoin [15]. In particular, by exploiting the connection between interval fragmentations and exchangeable partitions of the natural numbers [16, Lemmas 5 and 6], the arguments in the proof of Theorem 3 in [15] relying on a Poisson point process construction reveal that $\xi(\cdot) := -\log \lambda(H(\cdot))$ is a subordinator, that is, an increasing non-negative càdlàg process with stationary and independent increments. By [15, Theorem 2], (the distribution) of a homogeneous fragmentation process is characterized by a unique exchangeable partition measure which is determined by an *erosion coefficient* $c \geq 0$ and a *Lévy measure* ν on $S^* := \{x \in \mathbb{R}^{\mathbb{N}} : x_1 \geq x_2 \geq \ldots \geq 0, \sum_{i \geq 1} x_i \leq 1\} \setminus \{(1, 0, \ldots)\}$ with the property that $\int_{S^*} (1 - x_1) d\nu(x) < \infty$. We refer to [15] for a detailed discussion of this characterization and only use the following two results: first, by the arguments in [16, Section 4], for $\mathcal{P}^*_{\mathbf{e}}$, we have $c = 0$, and ν is concentrated on $\{(x, 0, \ldots) : 1/2 \leq x \leq 1\}$ where the projection on the first component denoted by ν_1 satisfies

$$\nu_1(dx) = 2(2\pi x^3(1-x)^3)^{-1/2} \mathbf{1}_{[1/2,1)}(x) dx.$$

Second, by the arguments in the proof of Theorem 3 in [15] the Laplace transform $\mathbf{E}\left[\exp(-q\xi(t))\right], t, q \geq 0$ is given by $\exp(-t\Phi(q))$ with

$$\Phi(q) = \int_{1/2}^1 (1 - x^q) d\nu_1(x). \tag{6.8}$$

In other words, the Lévy measure Π of ξ is given by

$$\Pi(dx) = 2(2\pi e^{-3x}(1-e^{-x})^3)^{-1/2}\mathbf{1}_{[0,\log 2)}(x)dx.$$

One can verify that

$$\begin{aligned}\Phi(q) &= \frac{4}{\sqrt{\pi}} \cdot {}_2F_1\left(-\frac{1}{2}, \frac{3}{2}-q; \frac{1}{2}; \frac{1}{2}\right) \\ &= \frac{4}{\sqrt{\pi}}\left(1-\left(\frac{1}{2}\right)^{3/2}\int_0^{1/2} t^{-3/2}((1-t)^{q-\frac{3}{2}}-1)dt\right), \end{aligned} \tag{6.9}$$

where ${}_2F_1$ denotes the standard hypergeometric function. In particular,

$$\Phi\left(\frac{1}{2}\right) = 2\sqrt{\frac{2}{\pi}}\left(\sqrt{2}-\log(1+\sqrt{2})\right).$$

The definition and properties of Φ extend to $q<0$. In particular, Φ is infinitely often differentiable on $\mathbb{R}$ and

$$\Phi'(0) = -\frac{4}{\sqrt{\pi}}\int_0^{1/2}\frac{\log(1-t)}{(t(1-t))^{3/2}}dt = 5.1525\ldots \tag{6.10}$$

Summarizing, we obtain the following result, which is closely related to [16, Corollary 2].

Proposition 6.13. *Let* $\xi(t), t\geq 0$ *be a subordinator with* $\mathbf{E}\left[\exp(-q\xi(t))\right] = \exp(-t\Phi(q))$ *as in* (6.8). *For* $t\geq 0$*, let*

$$\varrho_2(t) = \begin{cases}\inf\left\{u\geq 0 : \int_0^u e^{-\frac{1}{2}\xi(r)}dr > t\right\}, & \text{if } t<\int_0^\infty e^{-\frac{1}{2}\xi(r)}dr \\ \infty & \text{otherwise.}\end{cases}$$

Then, $\exp(-\xi(\varrho_2(t))), t\geq 0$ *and* $\mathbf{m}_{\mathbf{e}}(t), t\geq 0$ *are identically distributed.*

Clearly, we have, in distribution,

$$T_\infty^{\mathbf{e}} = \int_0^\infty \exp\left(-\frac{1}{2}\xi(t)\right)dt.$$

For an overview of results on exponential functionals of Lévy processes we refer to Bertoin and Yor's survey [18]. In particular, by results going back to Carmona, Petit and Yor [24] (see also [18, Theorem 2]), for $k\geq 1$,

$$\mathbf{E}\left[(T_\infty^{\mathbf{e}})^k\right] = \frac{k!}{\Phi(\frac{1}{2})\cdots\Phi(\frac{k}{2})}. \tag{6.11}$$

Since the jumps of ξ are bounded from above by $\log 2$, by [17, Proposition 2] (or [18, Theorem 3]) we also have $\mathbf{E}\left[\exp(q\xi(t))\right] = \exp(-t\Phi(-q))$ for $q\geq 0$, and

$$\mathbf{E}\left[(T_\infty^{\mathbf{e}})^{-k}\right] = (-1)^{k-1}\frac{\Phi'(0)}{2}\frac{\Phi\left(-\frac{1}{2}\right)\cdots\Phi\left(-\frac{(k-1)}{2}\right)}{(k-1)!}. \tag{6.12}$$

Next, by [24, Proposition 2.1], $T_\infty^{\mathbf{e}}$ admits a density $f(x)$ on $(0,\infty)$ which is infinitely often differentiable and

$$f(x) = \int_x^\infty \Pi\left(\left(\frac{\log(u/x)}{2},\infty\right)\right)f(u)du, \quad x>0. \tag{6.13}$$

Finally, from (6.9), using the substitution $t = v/(q - 3/2)$, it is straightforward to show that $\Phi(q) \sim \sqrt{8q}$ as $q \to \infty$. Hence, the decay of the right tail of the corresponding distribution is given by [51, Proposition 2]:

$$\lim_{x \to \infty} \frac{-\log \mathbf{P}\left(T_\infty^{\mathbf{e}} > x\right)}{x^2} = \frac{1}{2}. \tag{6.14}$$

A family of perpetuities. Let $0 < r < 1$. The dynamics of $\mathbf{m_e}(t), t \geq 0$ imply that

$$T_\infty^{\mathbf{e}} \stackrel{d}{=} \zeta_{\mathbf{e}}(r) + \sqrt{\mathbf{m_e}(r)} T_\infty^{\mathbf{e}^*}, \tag{6.15}$$

where $\mathbf{e}^*$ is an independent copy of $\mathbf{e}$. In particular, $T_\infty^{\mathbf{e}^*}, (\zeta_{\mathbf{e}}(r), \mathbf{m_e}(r))$ are independent while $\zeta_{\mathbf{e}}(r), \mathbf{m_e}(r)$ are defined using the *same* Brownian excursion $\mathbf{e}$. Hence, $T_\infty^{\mathbf{e}}$ is characterized by a family of perpetuities, one for each value of r. For more background on stochastic fixed-point equations of perpetuity type and a proof for the fact that (6.15) indeed determines the distribution of $T_\infty^{\mathbf{e}}$, we refer to Vervaat [52]. For all $0 < r < 1$, stochastically,

$$\sum_{k=0}^{\infty} \left(\frac{r}{2}\right)^{k/2} \zeta_{\mathbf{e}}^{(k)}(r) \leq T_\infty^{\mathbf{e}} \leq \sum_{k=0}^{\infty} r^{k/2} \zeta_{\mathbf{e}}^{(k)}(r), \tag{6.16}$$

where $\zeta_{\mathbf{e}}^{(0)}(r), \zeta_{\mathbf{e}}^{(1)}(r), \ldots$ are independent copies of $\zeta_{\mathbf{e}}(r)$. Similarly, in the proof of Proposition 6.2, we have shown that there exists a constant $C > 0$ and, for all $a > 2$ a constant $c > 0$ such that, stochastically,

$$c \sum_{k=1}^{\infty} a^{-k/2} |\mathcal{N}_k| \leq T_\infty^{\mathbf{e}} \leq C \sum_{k=0}^{\infty} 2^{-k/2} E_k, \tag{6.17}$$

where $\mathcal{N}_1, \mathcal{N}_2$ are independent standard normal random variables and $E_1, E_2, \ldots,$ are independent random variables with the standard exponential distribution. In fact, our proofs also revealed that, with the same constants c, C, a, in probability,

$$c a^{-1/2} |\mathcal{N}_1| \leq \zeta_{\mathbf{e}}(1/2) \leq C 2^{-1/2} E_1. \tag{6.18}$$

Note that the lower bound in (6.17) does not follow from (6.16) and (6.18) due to the factor $1/2$ in (6.16). Hence, the tail bound deduced from the discrete-time approach is stronger than the bound we could show relying only on the perpetuity (6.15).

Heavy trees in real trees. In the final paragraph, we give an outlook of the theory of heavy trees and the heavy path in the framework of real trees. We remain brief, as a full discussion of the topic would go far beyond the scope of this work. For background on real trees, we refer to Evans' book [32] and Le Gall's survey [43].

A metric space $(\mathcal{T}, d)$ is called a *real tree* if it satisfies the following two points:

- for every pair of points $a, b \in \mathcal{T}$ there exists a unique isometry $\varphi_{a,b} : [0, d(a,b)] \to \mathcal{T}$ for which $\varphi_{a,b}(0) = a$ and $\varphi_{a,b}(d(a,b)) = b$,
- if $q : [0,1] \to \mathcal{T}$ is a continuous and injective map with $q(0) = a, q(1) = b$, then $q([a,b]) = \varphi_{a,b}([0, d(a,b)])$.

In words, $(\mathcal{T}, d)$ is geodesic and loop-free and therefore generalizes the concept of a discrete tree to a continuous level. We use the shorthand notation $[a,b] := \varphi_{a,b}([0, d(a,b)])$ for the path between a and b in $\mathcal{T}$. Augmenting $(\mathcal{T}, d)$ by a probability measure μ on the Borel-σ-field on $\mathcal{T}$ and a unique vertex ρ (the root), the quadruple $(\mathcal{T}, d, \mu, \rho)$ becomes a rooted measured real tree. In the remainder we are only interested in cases when the spaces are compact. An important construction of such spaces is via continuous excursions $f \in \mathcal{C}_{\text{ex}}$ using the pseudometric

$$d_f(a,b) := f(a) + f(b) - 2\inf\{f(a) : a \wedge b \le s \le a \vee b\}\,.$$

(Again, this definition is reminiscent of the definition of the graph distance between two vertices in a discrete tree.) With $\mathcal{T}_f = [0,1]/\sim$ where $x \sim y$ if and only if $d_f(x,y) = 0$, $\mu = f_*(\text{Leb})$ (the pushforward measure) and $\rho = f(0)$, the tuple $(\mathcal{T}_f, d_f, \mu_f, \rho_f)$ is well-known to be a compact rooted measured real tree [32, Chapter 3]. For a given compact rooted measured real tree $(\mathcal{T}, d, \mu, \rho)$ and $x \in \mathcal{T}$, we call the number of connected components of $\mathcal{T} \setminus \{x\}$ the degree of x. (This number is at most countably infinite). We call a point $x \in \mathcal{T}$ a leaf if its degree is one, and write $\mathscr{L}$ for the set of leaves. Of particular interested in the theory of random trees are those satisfying $\text{supp}(\mu) = \mathcal{T}$, and we only consider these cases from now on.

Let $\mathscr{B}$ be the set of branching points of $\mathcal{T}$, that is, points with degree at least three. Set $\mathscr{B}^* = \mathscr{B} \cup \{\rho\}$. For each $b \in \mathscr{B}^*$ we may order the connected components $C_1^b, C_2^b, \dots$ of $\mathcal{T} \setminus \{b\}$ which do not contain the root ρ according to their μ-masses. Note that these masses are non-zero as μ has full support. (Since real trees are not ordered, a discussion of ties is technical and omitted here.) Let $\mathscr{B}^*_x = \mathscr{B}^* \cap [\rho, x]$. Clearly, there exists a unique leaf $\bar{x}$ such that, for all $b \in \mathscr{B}^*_{\bar{x}}$, we have $\bar{x} \in C_1^b$. The path $\ell(\mathcal{T}) = [\rho, \bar{x}]$ is the heavy path in $\mathcal{T}$ and $d(\rho, \bar{x})$ its length. Similarly, the k-heavy tree can be defined as $\bigcup_{x \in \mathscr{L}_k} [\rho, x]$ where $\mathscr{L}_k$ is the set of leafs x such that, for all $b \in \mathscr{B}^*_x$, we have $x \in \bigcup_{i=1}^k C_i^b$. (Starting with a Brownian excursion $\mathbf{e}$, we have $T_\infty^{\mathbf{e}} = d_{\mathbf{e}}(\rho_{\mathbf{e}}, \bar{x})$. As all branching points in the Brownian continuum tree have degree three, the 2-heavy tree is equal to the entire tree). We can generalize the definition of the functions in this section to the level of real trees. For example, for $0 < t \le d(\rho, \bar{x})$, we let $p_y \in [\rho, \bar{x}]$ be the unique element for which $d(\rho, p_y) = y$. Then, we set

$$\mathbf{m}_{\mathcal{T}}(t) = \mu(C_1^y).$$

(If $y \notin \mathscr{B}^*$, then C_1^y is to be understood as the unique component of $\mathcal{T} \setminus \{y\}$ which does not contain ρ.) The definition of the corresponding inverse $\zeta_{\mathcal{T}}$ remains unchanged: for $t \in [0,1]$, we set $\zeta_{\mathcal{T}}(t) = \inf\{s > 0 : \mathbf{m}_{\mathcal{T}}(s) \le t\}$. A discussion of continuity of these functions is more involved. First of all, it is necessary to change perspective and consider isometry classes of real trees (or metric spaces) with respect to the so-called *Gromov-Hausdorff-Prokhorov* distance. (For details and definitions, see [1] and [23].) Next, it is important to observe that the function $\mathbf{m}_{\mathcal{T}}$ is invariant under isometries and can therefore be defined for isometry classes. As for continuous functions, the maps $\mathbf{m}_{\mathcal{T}}, \zeta_{\mathcal{T}}$ are not continuous on the entire space of (equivalence classes) of real trees. Indeed, continuity of these functions can only be expected at (equivalence classes of) real trees $(\mathcal{T}, d, \mu, \rho)$ satisfying $C_1^b > C_2^b + C_3^b + \cdots$ for all $b \in \mathscr{B}^*$ (or, at least, for all $b \in [\rho, \bar{x}]$). This could be subject of future work.

A Appendix

For the sake of completeness, we state and prove the lemma connecting the length of the longest simple path in the Apollonian network and the size of the largest binary subtree in the underlying evolutionary tree. Essentially, this is a reproduction of the proof of Theorem 1.2 (a) in [31].

Lemma A.1. *Let G be an arbitrary Apollonian network with $3+n, n \ge 0$ vertices (outer vertices included) and $1+2n$ faces. Denote by $\mathcal{L}$ the number of vertices on the longest simple path in G. Let T be the corresponding evolutionary tree with n non-leaves and $1+2n$ leaves. Then, for any binary subtree B of T, we have $\mathcal{L} \ge (|B|+5)/2$.*

Proof. We follow the arguments in [31]. Label the three outer vertices in the network by $1, 2$ and 3 in an arbitrary way. Then, label the vertex inserted in the center of the initial triangle in the first step by $*$. It decomposes this triangle into three triangles $\Delta_1, \Delta_2, \Delta_3$ where Δ_i has vertices $\{1, 2, 3, *\} \setminus \{i\}$. Now, let $\mathcal{L}'$ be the largest number such that there exist paths in G containing $\mathcal{L}'$ *edges* from i to j avoiding k for all pairwise distinct $i, j, k \in \{1, 2, 3\}$. By construction $\mathcal{L} \geq \mathcal{L}' + 2$, and it suffices to prove that $\mathcal{L}' \geq (|B| + 1)/2$ for any binary subtree B of T. It is clearly equivalent to prove this inequality for a largest binary subtree B^*. As in [31], we proceed by induction over the number of vertices $3 + n, n \geq 0$ in the network. For $n = 0, 1, 2$ it is trivial to check that $\mathcal{L}' = (|B^*| + 1)/2$. We shall assume that the assertion is correct for all $m \leq n - 1$ and consider an Apollonian network G with $3 + n, n \geq 3$ vertices. For $i = 1, 2, 3$, let $\mathcal{L}'_i$ be defined as $\mathcal{L}$ in the sub-network based on the vertices $\{1, 2, 3, *\} \setminus \{i\}$ together with all vertices of G lying strictly inside the triangle Δ_i. By definition, for pairwise distinct $i, j, k \in \{1, 2, 3\}$, we find

(i) a simple path from i to $*$ inside Δ_j avoiding k containing at least $\mathcal{L}'_j$ many edges,
(ii) a simple path from i to $*$ inside Δ_k avoiding j containing at least $\mathcal{L}'_k$ many edges,
(iii) a simple path from $*$ to j inside Δ_k avoiding i containing at least $\mathcal{L}'_k$ many edges, and
(iv) a simple path from $*$ to j inside Δ_i avoiding k containing at least $\mathcal{L}'_i$ many edges.

Hence, by concatenating the paths in (i) and (iii), (i) and (iv), and (ii) and (iv), and since i, j, k were chosen arbitrarily, we obtain

$$\mathcal{L}' \geq \max\{\mathcal{L}'_1 + \mathcal{L}'_2, \mathcal{L}'_1 + \mathcal{L}'_3, \mathcal{L}'_2 + \mathcal{L}'_3\}.$$

Let B_i^* denote a largest binary subtree of the subtree rooted at the child of the root of the evolutionary tree corresponding to triangle Δ_i. By induction hypothesis, we have $\mathcal{L}'_i \geq (|B_i^*| + 1)/2$. Hence,

$$\mathcal{L}' \geq \frac{1}{2} \max\{|B_1^*| + |B_2^*|, |B_1^*| + |B_3^*|, |B_2^*| + |B_3^*|\} + 1 = (|B^*| + 1)/2. \qquad \square$$

References

[1] R. Abraham, J.-F. Delmas, and P. Hoscheit, *A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces*, Electron. J. Probab. **18** (2013), no. 14, 21 pp. MR-3035742

[2] L. Addario-Berry, L. Devroye, and S. Janson, *Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees*, Ann. Probab. **41** (2013), no. 2, 1072–1087. MR-3077536

[3] M. Albenque and J.-F. Marckert, *Some families of increasing planar maps*, Electron. J. Probab. **13** (2008), no. 56, 1624–1671. MR-2438817

[4] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, Ann. Appl. Probab. **1** (1991), no. 2, 228–266. MR-1102319

[5] D. Aldous, *The continuum random tree. I*, Ann. Probab. **19** (1991), no. 1, 1–28. MR-1085326

[6] D. Aldous, *The continuum random tree. II. An overview*, Stochastic Analysis (Durham, 1990), London Math. Soc. Lecture Note Ser., vol. 167, Cambridge Univ. Press, Cambridge, 1991, pp. 23–70. MR-1166406

[7] D. Aldous, *The continuum random tree. III*, Ann. Probab. **21** (1993), no. 1, 248–289. MR-1207226

[8] D. Aldous and J. Pitman, *Tree-valued Markov chains derived from Galton-Watson processes*, Ann. Inst. Henri Poincaré Probab. Stat. **34** (1998), no. 5, 637–686. MR-1641670

[9] D. Aldous and J. M. Steele, *The Objective Method: Probabilistic Combinatorial Optimization and Local Weak Convergence*, Probability on Discrete Structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 1–72. MR-2023650

[10] J. S. Andrade, H. J. Herrmann, R. F. S. Andrade, and L. R. da Silva, *Apollonian Networks: Simultaneously Scale-Free, Small World, Euclidean, Space Filling, and with Matching Graphs*, Phys. Rev. Lett. **94** (2005), 018702.

[11] K. B. Athreya and P. E. Ney, *Branching processes*, Springer-Verlag, New York-Heidelberg, 1972, Die Grundlehren der mathematischen Wissenschaften, Band 196. MR-0373040

[12] F. Aurzada, *A short note on small deviations of sequences of i.i.d. random variables with exponentially decreasing weights*, Statist. Probab. Lett. **78** (2008), no. 15, 2300–2307. MR-2462665

[13] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13 pp. MR-1873300

[14] J. Bennies and G. Kersting, *A random walk approach to Galton-Watson trees*, J. Theoret. Probab. **13** (2000), no. 3, 777–803. MR-1785529

[15] J. Bertoin, *Homogeneous fragmentation processes*, Probab. Theory Related Fields **121** (2001), no. 3, 301–318. MR-1867425

[16] J. Bertoin, *Self-similar fragmentations*, Ann. Inst. Henri Poincaré Probab. Stat. **38** (2002), no. 3, 319–340. MR-1899456

[17] J. Bertoin and M. Yor, *On the entire moments of self-similar Markov processes and exponential functionals of Lévy processes*, Ann. Fac. Sci. Toulouse Math. (6) **11** (2002), no. 1, 33–45. MR-1986381

[18] J. Bertoin and M. Yor, *Exponential functionals of Lévy processes*, Probab. Surv. **2** (2005), 191–212. MR-2178044

[19] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Inc., New York-London-Sydney, 1968. MR-0233396

[20] G. D. Birkhoff, *On the number of ways of colouring a map*, Proceedings of the Edinburgh Mathematical Society **2** (1930), 83–91.

[21] O. Bodini, A. Darrasse, and M. Soria, *Distances in random Apollonian network structures*, 20th Annual International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2008), Discrete Math. Theor. Comput. Sci. Proc., AJ, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008, pp. 307–318. MR-2721463

[22] N. Broutin and L. Devroye, *Large deviations for the weighted height of an extended class of trees*, Algorithmica **46** (2006), no. 3-4, 271–297. MR-2291957

[23] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001. MR-1835418

[24] P. Carmona, F. Petit, and M. Yor, *On the distribution and asymptotic results for exponential functionals of Lévy processes*, Exponential functionals and principal values related to Brownian motion, Bibl. Rev. Mat. Iberoamericana, Rev. Mat. Iberoamericana, Madrid, 1997, pp. 73–130. MR-1648657

[25] A. Collevecchio, A. Mehrabian, and N. Wormald, *Longest paths in random Apollonian networks and largest r-ary subtrees of random d-ary recursive trees*, J. Appl. Probab. **53** (2016), no. 3, 846–856. MR-3570098

[26] C. Cooper, A. Frieze, and R. Uehara, *The height of random k-trees and related branching processes*, Random Structures Algorithms **4** (2014), 675–702. MR-3275702

[27] A. Darrasse and M. Soria, *Degree distribution of random Apollonian network structures and Boltzmann sampling*, 2007 Conference on Analysis of Algorithms, AofA 07, Discrete Math. Theor. Comput. Sci. Proc., AH, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2007, pp. 313–324. MR-2509531

[28] L. Devroye, *Universal limit laws for depths in random trees*, SIAM J. Comput. **28** (1999), no. 2, 409–432. MR-1634354

[29] L. Devroye, *A note on the probability of cutting a Galton-Watson tree*, Electron. J. Probab. **16** (2011), no. 72, 2001–2019. MR-2851053

[30] M. Dwass, *The total progeny in a branching process and a related random walk.*, J. Appl. Probab. **6** (1969), 682–686. MR-0253433

[31] E. Ebrahimzadeh, L. Farczadi, P. Gao, A. Mehrabian, C. M. Sato, N. Wormald, and J. Zung, *On longest paths and diameter in random Apollonian networks*, Random Structures Algorithms **45** (2014), 703–725. MR-3351950

[32] S. N. Evans, *Probability and real trees*, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008, Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. MR-2351587

[33] P. Flajolet and A. Odlyzko, *The average height of binary trees and other simple trees*, J. Comput. System Sci. **25** (1982), no. 2, 171–213. MR-0680517

[34] A. Frieze and C. E. Tsourakakis, *Some properties of random Apollonian networks*, Internet Math. **10** (2014), no. 1-2, 162–187. MR-3274544

[35] C. Goldschmidt and B. Haas, *Behavior near the extinction time in self-similar fragmentations. I. The stable case*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 2, 338–368. MR-2667702

[36] S. Janson, *Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation*, Probab. Surv. **9** (2012), 103–252. MR-2908619

[37] S. Janson, *Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees*, Random Structures Algorithms **48** (2016), no. 1, 57–101. MR-3432572

[38] A. S. Kechris, *Classical Descriptive Set Theory*, Graduate Texts in Mathematics, vol. 156, Springer-Verlag, New York, 1995. MR-1321597

[39] D. P. Kennedy, *The Galton-Watson process conditioned on the total progeny*, J. Appl. Probab. **12** (1975), no. 4, 800–806. MR-0386042

[40] H. Kesten, *Subdiffusive behavior of random walk on a random cluster*, Ann. Inst. Henri Poincaré Probab. Stat. **22** (1986), no. 4, 425–487. MR-0871905

[41] V. F. Kolchin, *Random mappings*, Translation Series in Mathematics and Engineering, Optimization Software, Inc., Publications Division, New York, 1986, Translated from the Russian, With a foreword by S. R. S. Varadhan. MR-0865130

[42] I. Kolossváry, J. Komjáthy, and L. Vágó, *Degrees and distances in random and evolving Apollonian networks*, Adv. in Appl. Probab. **48** (2016), no. 3, 865–902. MR-3568896

[43] J.-F. Le Gall, *Random trees and applications*, Probab. Surv. **2** (2005), 245–311. MR-2203728

[44] R. Lyons, R. Pemantle, and Y. Peres, *Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes*, Ann. Probab. **23** (1995), no. 3, 1125–1138. MR-1349164

[45] R. Lyons and Y. Peres, *Probability on trees and networks*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 42, Cambridge University Press, New York, 2016. MR-3616205

[46] J.-F. Marckert and A. Mokkadem, *The depth first processes of Galton-Watson trees converge to the same Brownian excursion*, Ann. Probab. **31** (2003), no. 3, 1655–1678. MR-1989446

[47] A. Meir and J. W. Moon, *On the altitude of nodes in random trees*, Canad. J. Math. **30** (1978), no. 5, 997–1015. MR-0506256

[48] J. C. Pardo, V. Rivero, and K. van Schaik, *On the density of exponential functionals of Lévy processes*, Bernoulli **19** (2013), no. 5A, 1938–1964. MR-3129040

[49] V. V. Petrov, *Sums of independent random variables*, Springer-Verlag, New York-Heidelberg, 1975, Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. MR-0388499

[50] A. Rényi and G. Szekeres, *On the height of trees.*, J. Aust. Math. Soc. **7** (1967), 497–507. MR-0219440

[51] V. Rivero, *A law of iterated logarithm for increasing self-similar Markov processes*, Stoch. Stoch. Rep. **75** (2003), no. 6, 443–472. MR-2029617

[52] W. Vervaat, *On a stochastic difference equation and a representation of non-negative infinitely divisible random variables*, Adv. in Appl. Probab. **11** (1979), 750–783. MR-0544194

[53] T. Zhou, G. Yan, and B.-H. Wang, *Maximal planar networks with large clustering coefficient and power-law degree distribution*, Phys. Rev. E **71** (2005), 046141.