The height of depth-weighted random recursive trees*

Kevin Leckey
Faculty of Statistics
Technische Universität Dortmund

Dieter Mitsche†
Institut Camille Jordan
Univ. Lyon, Univ. Jean Monnet

Nick Wormald‡
School of Mathematics
Monash University

Abstract

In this paper we introduce a model of depth-weighted random recursive trees (DRRT), created by recursively joining a new leaf to an existing vertex v. In this model, the probability of choosing v depends on its depth in the tree. In particular, we assume that there is a function f such that if v has depth k then its probability of being chosen is proportional to $f(k)$. We consider the expected value of the diameter of this model as determined by f, and for various increasing f we find expectations that range from polylogarithmic to linear.

1 Introduction

Recently there has been an explosion of models of randomly growing networks. One of the most interesting parameters of these networks, from the applications point of view, is the diameter. In preferential attachment models, new vertices are attached to randomly chosen vertices, but with some preference for attaching to vertices of high degree. In this paper, we study similar models, with the difference that the preference is based on the distance from a single vertex.

In order to understand networks where attachment preferences are based on distance, it is natural to begin with the simplest case of randomly growing trees. Moreover, Mehrabian [9] recently developed a general approach to giving bounds on the diameter of randomly generated networks, based on coupling with the random recursive tree,

*Part of this research was done while the second and third authors were in residence at Centre Recerca Matemàtica (CRM) during the program "Strategic Behavior and Phase Transitions in Random and Complex Combinatorial Structures" (April-June 2015).

†Research supported by IDEXLYON of Université de Lyon (Programme Investissements d'Avenir ANR16-IDEX-0005).

‡Research supported by the Australian Laureate Fellowships grant FL120100125.

whose diameter is well understood. Thus, understanding the tree case can lead to results for more general graphs. In [9] there is also a comprehensive survey of known results for the diameter of randomly generated networks. Open problem 2 in that paper asks for the diameter of a randomly growing tree when each new vertex is attached according to some specified probability distribution, and finishes with some discussion including cases where the probability depends on the distance from the root. Thus we are led to the following question: for some prescribed non-negative function f, build a tree randomly as follows. Start with the root vertex, and repeatedly add new leaves, at each step attaching the new leaf to a vertex v chosen randomly but non-uniformly, with weight proportional to $f(k)$ where k is the distance from v to the root. In this paper, we determine the expected height of such a tree for various classes of functions f.

In Section 2 we introduce the model of random trees precisely, and state our main results. We discuss an embedding of the random tree into a continuous-time process in Section 3. Section 4 considers functions f for which the expected height of the trees is polylogarithmic in the number of vertices, and Section 5 considers weight functions for which the expected height is almost linear. Open problems, some pertaining to gaps in the spectrum of our results, are in the final section.

2 Model and results

Model. Let $\mathbb{R}_+$ denote the set of all positive real numbers and let $f : \mathbb{N}_0 \to \mathbb{R}_+$ be an arbitrary function. A *depth-weighted random recursive tree (DRRT)* with n vertices and weight function f is a rooted tree $\mathcal{T}_n(f)$ generated according to the following recursive procedure:

- The tree $\mathcal{T}_1(f)$ consists of an isolated root vertex labelled 1;
- If $\mathcal{T}_j(f)$ is already constructed, assign a weight $w_j(v)$ to each vertex via

  $$w_j(v) = f(D_v),$$

  where D_v is the distance of v to the root.
- Choose a vertex U among all existing vertices at random according to their weights, i.e. choose v with probability $w_j(v)/\sum_x w_j(x)$.
- Define $\mathcal{T}_{j+1}(f)$ to be the tree obtained from $\mathcal{T}_j(f)$ by adding a new vertex labelled $j+1$ and a new edge between $j+1$ and U.

The vertex set of $\mathcal{T}_n(f)$ is then $\{1, \ldots, n\}$.

Definition 2.1. *Let $\mathcal{T}_n(f)$ be a DRRT with n vertices and some weight function f. Let D_v be the distance of v from the root. Then*

$$H_n(f) := \max_{v \in \{1,\ldots,n\}} D_v.$$

The parameter $H_n(f)$ is called the height *of $\mathcal{T}_n(f)$, and D_v is called the* depth *of v. We also use* level *i to denote the set of vertices at depth i.*

Notation. We use the standard Bachmann-Landau notation for the asymptotic behaviour of sequences: For sequences $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$, $a_n = \mathrm{O}(b_n)$ denotes the existence of a constant $C > 0$ such that there exists $n_0 \in \mathbb{N}$ so that for all $n \geq n_0$, $|a_n| \leq C|b_n|$. Moreover, we write $a_n = \Omega(b_n)$ if $b_n = \mathrm{O}(a_n)$, and $a_n = \Theta(b_n)$ if both $a_n = \mathrm{O}(b_n)$ and $a_n = \Omega(b_n)$. Furthermore, we use the abbreviation $[n] = \{1, \dots, n\}$. Finally, $x \wedge y$ denotes the minimum of two real numbers x, y.

Remark 2.2. *If $f \equiv c$ for some constant c, we use $\mathcal{T}_n(c)$ to denote the DRRT with constant weight function f. Note that $\mathcal{T}_n(1)$ (and also $\mathcal{T}_n(c)$ for arbitrary $c > 0$) coincides with the usual model for a random recursive tree. The height of such a tree has been studied quite thoroughly; see, e.g., [1, 3]. In particular, it is known [1, Corollary 1.3] that $\mathbb{E}[H_n(1)] = e \log n - 1.5 \log\log n + \mathrm{O}(1)$.*

A Hoppe tree is a slight variation of a random recursive tree. It was introduced and studied in [7]. Note that a Hoppe tree is a DRRT with a weight function given by $f(0) = \vartheta$ for some $\vartheta > 0$ and $f(k) = 1$ for $k \geq 1$. It is shown in [7, Theorem 2.2] that the height of a Hoppe tree is sharply concentrated around the mean of a random recursive tree. In particular the difference between the expected heights in both models remains bounded as the number of vertices tends to infinity.

The main objective of this article is to link the growth rate of f to the asymptotic behaviour of $H_n(f)$. More precisely, we derive bounds on $\mathbb{E}[H_n(f)]$ by different approaches depending on the type of weight function. Our main results are summarised in the following theorem.

Theorem 2.3. *Let f be a weight function.*

(a) If $\sup_k f(k) < \infty$ and $\inf_k f(k) > 0$, then $\mathbb{E}[H_n(f)] = \Theta(\log n)$.

(b) If $f(k) = (k+1)^\alpha$ for some $\alpha \geq 0$, then $\mathbb{E}[H_n(f)] = \Theta(\log n)$.

(c) If $f(k) = \exp(k^\beta)$ for some $\beta \in (0,1)$, then $\mathbb{E}[H_n(f)] = \mathrm{O}\left((\log n)^{1/(1-\beta)}\right)$.

(d) If $f(k) = c^k$ for some $c > 2$, then $\mathbb{E}[H_n(f)] = \Omega(n/\log n)$.

(e) If $f(k) = \exp(ak \log k)$ with $a > 1$, then $\mathbb{E}[H_n(f)] = \Theta(n)$.

It is apparent that the theorem only covers some particular functions $f(k)$ representative of various growth rates. Our main goal is to obtain a sense of where the growth of height has significant changes in its behaviour and to illustrate several different methods of argument. Note that the choice in (b) of $f(k) = (k+1)^\alpha$ instead of the more natural function $f(k) = k^\alpha$ is somewhat arbitrary, to avoid weight 0 for the root vertex in the first step of the algorithm. A close inspection of the proof shows that other choices such as for example $f(k) = k^\alpha + 1$ or $f(k) = \max\{k^\alpha, 1\}$ would — with very minor modifications of the proof — work as well.

The bounds provided by our methods are not sharp at all. In particular, bounded weight functions (as in (a)) will most probably lead to trees which do not differ much from random recursive trees in terms of height (cf. Remark 2.2 for the case of Hoppe trees). For instance, we believe that the restriction $c > 2$ and the extra $\log n$ denominator in (d) can be removed. In summary, we state the following conjecture.

Conjecture 2.4. *The results (a) and (d) in Theorem 2.3 can be replaced with:*

(a′) If $\sup_k f(k) < \infty$ *and* $\inf_k f(k) > 0$*, then* $\mathbb{E}[H_n(f)] = \mathbb{E}[H_n(1)] + \mathrm{O}(1)$*.*

(d′) If $f(k) = c^k$ *for some* $c > 1$*, then* $\mathbb{E}[H_n(f)] = \Theta(n)$*.*

We also conjecture the following coupling possibility to retrieve information on (increasing) weight functions that are not covered by Theorem 2.3.

Conjecture 2.5. *Let* f *and* g *be weight functions. Suppose* f *is increasing and satisfies*

$$\frac{f(k+1)}{f(k)} \geq \frac{g(k+1)}{g(k)}, \quad k \geq 0.$$

Then $\mathbb{P}(H_n(f) \geq x) \geq \mathbb{P}(H_n(g) \geq x)$ *for all* $n \geq 2$ *and* $x \geq 0$*. In particular,* $\mathbb{E}[H_n(f)] \geq \mathbb{E}[H_n(g)]$*.*

Note that it would not be sufficient to replace the condition on ratios by the simple condition $f(k) \geq g(k)$, since if g agreed with f everywhere except for $g(0) < f(0)$, then $\mathcal{T}_n(g)$ would tend to have slightly greater height than $\mathcal{T}_n(f)$.

Remark 2.6. *Note that if* f *is increasing then* $\mathbb{P}(H_{n+1}(f) = H_n(f) + 1) \geq \frac{1}{n}$*, since there is at least one vertex on the last non-empty level of the tree and the weight of this vertex is at least as large as the weight of the other vertices in the tree. This yields a lower bound* $\mathbb{E}[H_n(f)] \geq \log n + \mathrm{O}(1)$ *for increasing* f *without using the conjecture.*

As a special case of Conjecture 2.5, we have the following.

Proposition 2.7. *If* f *is an increasing function, then* $H_n(f)$ *stochastically dominates* $H_n(1)$*. In particular,* $\mathbb{E}[H_n(f)] \geq \mathbb{E}[H_n(1)]$*. Similarly, if* g *is a decreasing function, then* $H_n(1)$ *stochastically dominates* $H_n(g)$*, and* $\mathbb{E}[H_n(g)] \leq \mathbb{E}[H_n(1)]$*.*

Proof. We only show the first part, i.e. that $H_n(f)$ stochastically dominates $H_n(1)$, the second part being analogous. It is easy to couple the process building $\mathcal{T}_n(f)$ and that for $\mathcal{T}_n(1)$, such that, the vertex i added to $\mathcal{T}_n(f)$ at step i is at least as deep as i in $\mathcal{T}_n(1)$. Indeed, inductively for any d there are at least as many vertices of depth at least d in $\mathcal{T}_n(f)$ as in $\mathcal{T}_n(1)$, so their relative weight is at least as large, and the coupling is easy. Denoting by $N(n, \ell)$ the number of vertices of $\mathcal{T}_n(f)$ at level $\ell \geq 0$, and denoting by $N'(n, \ell)$ the number of vertices of $\mathcal{T}_n(1)$ at level $\ell \geq 0$, we have for every $\ell \geq 0$, that $\sum_{i \geq \ell} N(n, i) \geq \sum_{i \geq \ell} N'(n, i)$. In particular, this implies $\mathbb{E}[H_n(f)] \geq \mathbb{E}[H_n(1)]$. □

Note that this proposition improves the bound in Remark 2.6 since $\mathbb{E}[H_n(1)] \sim e \log n$. (See Remark 2.2.)

3 Continuous-time embedding

As a preparation for the proof of Theorem 2.3(a), we discuss a continuous-time embedding of a DRRT. This embedding is a straightforward generalisation of a well known approach for random recursive trees (cf., e.g., [1]). Note that it would be sufficient to only keep track of the profile (i.e. the number of vertices in each level of the tree) to derive the height of the tree. However, we embed the entire tree structure for easier comparison with previous work on random recursive trees.

The Ulam-Harris tree. The Ulam-Harris tree $\mathcal{T}^{UH}$ is a rooted tree with vertex set

$$V_{UH} = \{\emptyset\} \cup \bigcup_{n\geq 1} \mathbb{N}^n$$

and edges between $(v_1, \ldots, v_k)$ and $(v_1, \ldots, v_{k+1})$ for every $k \geq 0$ and $(v_1, \ldots, v_{k+1}) \in \mathbb{N}^{k+1}$ (here and subsequently, $(v_1, \ldots, v_k) := \emptyset$ for $k = 0$).

Similarly to the corresponding embedding of random recursive trees [1], we now assign a birth time to each vertex in V_{UH}. Afterwards, the random tree $\mathcal{T}_t^{\mathrm{CT}}$ will be the subtree of $\mathcal{T}^{UH}$ consisting of all vertices born until time $t \geq 0$.

Birth times. Fix a weight function f. Moreover, let

$$\{X^v = (E_1^v, E_1^v + E_2^v, E_1^v + E_2^v + E_3^v, \ldots) : v \in V_{UH}\}$$

be a family of independent copies of a sequence $X = (E_1, E_1 + E_2, \ldots)$ built from i.i.d. exponentially distributed random variables $E_1, E_2, \ldots$ with rate 1. Finally, let

$$X^v(f) = X^v/f(k) = \left(\frac{E_1^v}{f(k)}, \frac{E_1^v + E_2^v}{f(k)}, \ldots\right), \qquad k \geq 0,\, v \in \mathbb{N}^k, \tag{1}$$

and note that $(E_i^v/f(k))_{i\geq 1}$ is a sequence of i.i.d. exponentially distributed random variables with rate $f(k)$. The birth time $B_v(f)$ of a vertex $v \in V_{UH}$ is defined recursively as follows. The root of the tree is born at time zero, i.e. $B_\emptyset(f) = 0$. Additionally, the time difference $B_v(f) - B_{(v_1,\ldots,v_{k-1})}(f)$ between the birth of a vertex $(v_1, \ldots, v_{k-1})$ and its child $v = (v_1, \ldots, v_k)$ is determined by

$$B_v(f) = B_{(v_1,\ldots,v_{k-1})}(f) + X_{v_k}^{(v_1,\ldots,v_{k-1})}(f), \quad k \geq 1,\, v \in \mathbb{N}^k,$$

where $X_m^w(f) = \sum_{i=1}^m E_i^w/f(k)$.

The continuous-time embedding. For $t \in \mathbb{R}_+$ let $V_t^{CT}(f) = \{v \in V_{UH} : B_v(f) \leq t\}$. We define $\mathcal{T}_t^{\mathrm{CT}}(f)$ to be the subtree of $\mathcal{T}^{UH}$ induced by the vertex set $V_t^{CT}(f)$. Note that $\mathcal{T}_t^{\mathrm{CT}}(f)$ is indeed a tree (i.e. connected) by construction of the birth times.

The sequence $(\mathcal{T}_t^{\mathrm{CT}}(f))_{t\in\mathbb{R}_+}$ is called *continuous-time embedding* of $(\mathcal{T}_n(f))_{n\geq 1}$ (and is a special instance of the Crump-Mode-Jagers process, see [5] for details). This name is justified by the next lemma. In preparation, let

$$t_n(f) = \min\{t \geq 0 : |\{v \in V_{UH} : B_v(f) \leq t\}| = n\}. \tag{2}$$

Lemma 3.1. *The sequences $(\mathcal{T}_n(f))_{n\in\mathbb{N}}$ and $(\mathcal{T}_{t_n(f)}^{\mathrm{CT}}(f))_{n\in\mathbb{N}}$ of rooted trees are equal in distribution. In particular, if $H_t^{\mathrm{CT}}(f)$ denotes the height of $\mathcal{T}_t^{\mathrm{CT}}(f)$, $t \in \mathbb{R}_0^+$, then*

$$H_{t_n(f)}^{\mathrm{CT}}(f) \overset{d}{=} H_n(f).$$

Proof. By the memorylessness of the exponential distribution, after $n-1$ vertices have been born, the waiting time before the birth of the next child to a given vertex $v \in \mathcal{T}^{\mathrm{CT}}_{t_{n-1}(f)}(f)$ is an exponentially distributed variable with rate $f(D_v)$, where D_v denotes the unique integer k with $v \in \mathbb{N}^k$. These variables are independent for different vertices v. It is a well known fact that for independent, exponentially distributed random variables $Y_1, \ldots, Y_n$ with rates $\lambda_1, \ldots, \lambda_n$

$$\mathbb{P}\left(Y_i = \min_{j\in[n]} Y_j\right) = \frac{\lambda_i}{\lambda_1 + \ldots + \lambda_n}.$$

Hence the probability of the newborn vertex being a child of any given vertex v is proportional to $f(D_v)$. Thus, both tree sequences follow the same growth rule. □

For a random recursive tree (i.e. $f \equiv 1$) it is easy to check that $(t_i(1) - t_{i-1}(1))_{i\geq 2}$ is a sequence of independent, exponentially distributed random variables such that $\mathbb{E}[t_{i+1}(1) - t_i(1)] = 1/i$. In this special case Addario-Berry and Ford use the continuous-time embedding to derive the following result [1, Corollary 1.3]:

Theorem 3.2. *The height $H_n(1)$ of a random recursive tree on n vertices satisfies $\mathbb{E}[H_n(1)] = e\log n - \frac{3}{2}\log\log n + \mathrm{O}(1)$. Moreover, for all $c' < 1/(2e)$ there is a constant $C = C(c') > 0$ such that for all $n \geq 1$ and $k \geq 1$*

$$\mathbb{P}(|H_n(1) - \mathbb{E}[H_n(1)]| \geq k) \leq C\exp(-c'k).$$

These tail bounds will be useful for the proof of Theorem 2.3(a) in the next section. We also use the continuous-time embedding to obtain the following coupling of different weight functions:

Lemma 3.3. *Let f, g_1, g_2 be weight functions with $g_1 \leq f \leq g_2$. Then, for any $n, m_1, m_2 \in \mathbb{N}$ and $x > 0$,*

$$\mathbb{P}(H_n(f) \geq x) \leq \mathbb{P}(H_{m_2}(g_2) \geq x) + \mathbb{P}(t_n(g_1) \geq t_{m_2}(g_2)),$$
$$\mathbb{P}(H_n(f) \leq x) \leq \mathbb{P}(H_{m_1}(g_1) \leq x) + \mathbb{P}(t_n(g_2) \leq t_{m_1}(g_1)),$$

with birth-times $t_n(g_1)$ and $t_n(g_2)$ given in (2).

Proof. By construction of $\mathcal{T}^{\mathrm{CT}}_t$ and $g_1 \leq f \leq g_2$ we have that $\mathcal{T}^{\mathrm{CT}}_t(g_1)$ is a subgraph of $\mathcal{T}^{\mathrm{CT}}_t(f)$ and $\mathcal{T}^{\mathrm{CT}}_t(f)$ is a subgraph of $\mathcal{T}^{\mathrm{CT}}_t(g_2)$ at any time $t \in \mathbb{R}^+_0$. Hence, we also have $t_n(g_2) \leq t_n(f) \leq t_n(g_1)$ and $H^{\mathrm{CT}}_t(g_1) \leq H^{\mathrm{CT}}_t(f) \leq H^{\mathrm{CT}}_t(g_2)$. Therefore,

$$\mathbb{P}(H_n(f) \geq x) \leq \mathbb{P}(H^{\mathrm{CT}}_{t_n(f)}(g_2) \geq x) \leq \mathbb{P}(H_{m_2}(g_2) \geq x) + \mathbb{P}(t_n(f) \geq t_{m_2}(g_2)).$$

Thus, the first bound in the lemma follows from $t_n(f) \leq t_n(g_1)$. The second bound holds by similar arguments. □

4 Functions giving polylogarithmic height

This section contains (poly)logarithmic bounds on the height of DRRTs with subexponential weight functions, i.e. parts (a)-(c) in Theorem 2.3. We start with bounded weight functions. In this case we use concentration results and the continuous-time embedding to bound the difference between such DRRTs and classical random recursive trees.

4.1 Bounded weight functions

If f is a bounded function, a combination of Lemma 3.3 and Theorem 3.2 yields the following tail bounds, which are sufficient to obtain Theorem 2.3(a) (cf. Corollary 4.2).

Lemma 4.1. *Let f be a weight function such that*

$$c \leq f(k) \leq dc, \qquad k \geq 0,$$

for some constants $c > 0$ and $d > 1$. Furthermore, let $\mu(n)$ be the expected height of a random recursive tree, i.e.

$$\mu(n) = \mathbb{E}[H_n(1)] = e \log n - \frac{3}{2} \log \log n + \mathrm{O}(1).$$

Then, for all $c' < 1/(2e)$ and $M > 0$ there are constants $C, D > 0$ (with C being the same constant as in Theorem 3.2, and independent of M) such that for all $a \geq 1$

$$\mathbb{P}\left(H_n(f) \geq \mu\left(\left\lceil n^{d+M} \right\rceil\right) + a\right) \leq Ce^{-c'a} + Dn^{-M/(2(d-1))}, \tag{3}$$

$$\mathbb{P}\left(H_n(f) \leq \mu\left(\left\lfloor n^{1/(d+M)} \right\rfloor\right) - a\right) \leq Ce^{-c'a} + Dn^{-M/(2(d-1)(d+M))}. \tag{4}$$

Since $H_n(1) \leq n$, the following is immediate.

Corollary 4.2. *If $\sup_k f(k) < \infty$ and $\inf_k f(k) > 0$ then*

$$\mathbb{E}[H_n(f)] = \Theta(\log n).$$

As preparation for the proof of Lemma 4.1 we start with a simple concentration result to bound the birth times in the continuous-time embedding. These concentration results lead to a bound in Corollary 4.4 which will be useful for the proof of Lemma 4.1. The proofs of Lemma 4.3 and Corollary 4.4 are deferred to the end of the section.

Lemma 4.3. *Let $(Y_i)_{i\geq 1}$ be a sequence of independent, exponentially distributed random variables with $\mathbb{E}[Y_i] = 1/i$. For $1 \leq \ell \leq n$ let $S_{\ell,n} = \sum_{i=\ell}^{n} Y_i$. Then, for $x \geq 0$,*

$$\mathbb{P}(S_{\ell,n} \geq x) \leq \binom{n}{\ell} e^{-\ell x},$$

$$\mathbb{P}(S_{\ell,n} \leq x) \leq \binom{n}{\ell - 1} \exp\left(-(n - \ell + 1)\exp(-x)\right).$$

Corollary 4.4. *Let $(S_{\ell,n})_{\ell \leq n}$ be as in Lemma 4.3. Then, for all $\alpha > 1$ and $M > 0$ there is a constant $\beta = \beta(\alpha, M) > 0$ such that for all integers $j, k \in \mathbb{N}$ with $k \geq \lceil j^{\alpha+M} \rceil$*

$$\mathbb{P}((\alpha - 1)S_{1,j} \geq S_{j+1,k}) \leq \beta j^{-M/(2(\alpha-1))}.$$

Proof of Lemma 4.1. We start with (3): Let $m_2 = \lceil n^{d+M} \rceil$. Lemma 3.3 yields

$$\mathbb{P}\left(H_n(f) \geq \mu\left(m_2\right) + a\right) \leq \mathbb{P}\left(H_{m_2}(dc) \geq \mu\left(m_2\right) + a\right) + \mathbb{P}\left(t_n(c) \geq t_{m_2}(dc)\right). \tag{5}$$

Note that $H_{m_2}(dc)$ has the same distribution as the height of a random recursive tree (since constant factors in the weight function have no influence) and therefore, the first probability in (5) can be bounded by Theorem 3.2. For the second probability recall that, in a random recursive tree, $(t_i(1)-t_{i-1}(1))_{i\geq 2}$ is a sequence of independent, exponentially distributed random variables with $\mathbb{E}[t_{i+1}(1)-t_i(1)] = 1/i$. Also recall that $t_i(\lambda) = t_i(1)/\lambda$ for any $\lambda > 0$ by construction. After rewriting

$$\mathbb{P}(t_n(c) \geq t_{m_2}(dc)) = \mathbb{P}\left((d-1)t_n(1) \geq t_{m_2}(1) - t_n(1)\right), \tag{6}$$

it is easy to apply the concentration inequalities stated in Lemma 4.3 above (the explicit bound is given in Corollary 4.4). More precisely, note that $t_n(1)$ has the same distribution as $S_{1,n-1}$ in Corollary 4.4, since $t_{i+1}(1) - t_i(1) \sim \mathrm{Exp}(i)$ by construction, where $\mathrm{Exp}(i)$ denotes an exponentially distributed random variable with rate i. Hence, $t_{m_2}(1) - t_n(1)$ is distributed as S_{n,m_2-1} and Corollary 4.4 yields for every $M > 0$

$$\mathbb{P}\left((d-1)t_n(1) \geq t_{m_2}(1) - t_n(1)\right) \leq \beta_{d,M} n^{-M/(2(d-1))},$$

in which $\beta_{d,M}$ is a constant depending only on d and M. Thus (5) yields (3).

The other bound (4) holds by essentially the same arguments. The main difference is that the roles of n and $m_1 = \lfloor n^{1/(d+M)} \rfloor$ in (6) are exchanged and therefore, (6) is replaced by

$$\mathbb{P}(t_n(dc) \leq t_{m_1}(c)) = \mathbb{P}\left(t_n(1) - t_{m_1}(1) \leq (d-1)t_{m_1}(1)\right). \tag{7}$$

Thus Corollary 4.4 yields

$$\mathbb{P}(t_n(dc) \leq t_{m_1}(c)) \leq \beta_{d,M} m_1^{-M/(2(d-1))},$$

and the assertion follows by definition of m_1. □

Proof of Lemma 4.3. The proof is a straightforward generalisation of the bounds (1.1)-(1.2) in [1, Proof of Corollary 1.3], based on the following observation: If $(E_i)_{i\geq 1}$ denotes a sequence of i.i.d. exponentially distributed random variables with rate 1 then, by the memorylessness of the exponential distribution and the fact that $Y_n \overset{d}{=} \min\{E_1, \ldots, E_n\}$, $S_{\ell,n}$ has the same distribution as the ℓ-th largest element among $E_1, \ldots, E_n$. Thus,

$$\mathbb{P}(S_{\ell,n} \geq x) = \mathbb{P}\left(\bigcup_{I\subset[n], |I|=\ell} \left\{ \min_{j\in I} E_j \geq x \right\} \right) \leq \binom{n}{\ell} \exp(-\ell x)$$

by the union bound, which yields the first part of the claim. Similarly, $S_{\ell,n}$ also has the same distribution as the $(n-\ell+1)$-th smallest element among $E_1, \ldots, E_n$, and hence

$$\mathbb{P}(S_{\ell,n} \leq x) = \mathbb{P}\left(\bigcup_{I\subset[n], |I|=n-\ell+1} \left\{ \max_{j\in I} E_j \leq x \right\} \right) \leq \binom{n}{\ell-1} \left(1 - e^{-x}\right)^{n-\ell+1}$$

which, combined with $1 - y \leq \exp(-y)$, finishes the proof. □

Proof of Corollary 4.4. Since $\mathbb{P}((\alpha-1)S_{1,j} \geq S_{j+1,k})$ is decreasing in k, we may assume w.l.o.g. that $k = \lceil j^{\alpha+M} \rceil$. Note that for any $x > 0$

$$\mathbb{P}((\alpha-1)S_{1,j} \geq S_{j+1,k}) \leq \mathbb{P}(S_{1,j} \geq x/(\alpha-1)) + \mathbb{P}(S_{j+1,k} \leq x). \tag{8}$$

Now let $x = (\alpha - 1 + M/2)\log j$. Then Lemma 4.3 and the choice for k yield

$$\begin{aligned}\mathbb{P}(S_{j+1,k} \leq x) &\leq \binom{k}{j} \exp\left(-(k-j+1)j^{-(\alpha-1+M/2)}\right) \\ &\leq \left(\frac{k}{j}\right)^j \exp\left(-j^{1+M/2} + j\right) \\ &\leq \exp(-\beta_1 j^{1+M/2})\end{aligned}$$

for a suitable constant $\beta_1 = \beta_1(\alpha, M) > 0$. Lemma 4.3 also yields

$$\mathbb{P}(S_{1,j} \geq x/(\alpha-1)) \leq j\exp(-x/(\alpha-1)) = \exp\left(-\frac{M}{2(\alpha-1)}\log j\right).$$

Finally note that $\exp(-\beta_1 j^{1+M/2}) \leq \beta_2 \exp\left(-\frac{M}{2(\alpha-1)}\log j\right)$ for a suitable constant $\beta_2 = \beta_2(\alpha, M) > 0$. Therefore, (8) and the previous bounds imply the assertion. □

4.2 Polynomial weight functions

We continue with the proof of Theorem 2.3(b); i.e. we prove the following result.

Theorem 4.5. *Let $\alpha \geq 0$ and $f(k) = (k+1)^\alpha$, $k \geq 0$. Then,*

$$\mathbb{E}[H_n(f)] = \Theta(\log n).$$

Proof. The lower bound follows from Remark 2.6. For the upper bound define approximating functions f_h, $h \in \mathbb{N}$, as

$$f_h(k) = f(k \wedge h), \qquad k \geq 0.$$

Note that we can couple $(\mathcal{T}_n(f_h))_{n\geq 1}$ and $(\mathcal{T}_n(f))_{n\geq 1}$ in such a way that they coincide for all n with $H_n(f) \leq h$, since the weights of the vertices coincide up to height h. In particular, $H_n(f_h) \leq h$ implies $H_n(f_h) = H_n(f)$ in that coupling and therefore

$$\mathbb{P}(H_n(f_h) > h) = \mathbb{P}(H_n(f) > h) \quad \text{for every } h \in \mathbb{N}. \tag{9}$$

Now let $h = \lfloor C\log n \rfloor$ and $\ell = \lceil c \log n \rceil$ for some constants $0 < c < C$ chosen later in the proof. The main idea of the proof is to show $\mathbb{P}(H_n(f_h) > h) = \mathrm{O}(n^{-1})$ for suitable C by bounding the height of each subtree of $\mathcal{T}_n(f_h)$ on level ℓ. To this end, let $\{v_1, \ldots, v_K\} \subset [n]$ be the set of all vertices in $\mathcal{T}_n(f_h)$ that are at distance ℓ from the root and let $J_1, \ldots, J_K$ be the subtree sizes (total progeny) of the subtrees rooted at $v_1, \ldots, v_K$. Note that, conditioned on $(K, J_1, \ldots, J_K)$, the heights of the subtrees are distributed as the heights of K independent depth-weighted random recursive trees with sizes $J_1, \ldots, J_K$ and weight functions

$$g_h(k) = f_h(k+\ell), \quad k \geq 0.$$

However, since $H_n(g_h)$ and $H_n(\tilde{g}_h)$ have the same distribution for $\tilde{g}_h = g_h/f_h(\ell)$, we obtain

$$\mathbb{P}(H_n(f_h) > h) = \sum_{k,j_1,\dots,j_k} \mathbb{P}\left((K, J_1, \dots, J_K) = (k, j_1, \dots, j_k)\right) \mathbb{P}\left(\bigcup_{i=1}^{k} \{H_{j_i}(\tilde{g}_h) > h - \ell\}\right).$$

Note that

$$\mathbb{P}\left(H_{j_i}(\tilde{g}_h) > h - \ell\right) \le \mathbb{P}\left(H_n(\tilde{g}_h) > h - \ell\right).$$

Now choose $C = 2c$. Since $1 \le \tilde{g}_h \le (C/c)^\alpha$, we may apply Lemma 4.1 with $d = 2^\alpha$, $M = 4(2^\alpha - 1)$, $c' = 1/(3e)$, $a = 6e \log n$, and note that $\mu(\lceil n^{d+M} \rceil) \le e(d + M) \log n + O(1)$. Therefore, by choosing $c = e(d + M + 6)$, by Lemma 4.1,

$$\mathbb{P}\left(H_n(\tilde{g}_h) > h - \ell\right) \le \mathbb{P}\left(H_n(\tilde{g}_h) > \mu(\lceil n^{d+M} \rceil) + a + O(1)\right) = O(n^{-2}).$$

Hence, by the union bound and since the other probabilities in the above summation sum to 1, we get that $\mathbb{P}(H_n(f_h) > h) = \mathrm{O}(1/n)$. It follows that $\mathbb{P}(H_n(f) > h) = \mathrm{O}(1/n)$. Therefore, since clearly $H_n(f) \le n$, we have $\mathbb{E}[H_n(f)] = O(\log n)$ and the assertion follows. □

4.3 Subexponential weight functions

A simple adaptation of the previous proof strategy yields Theorem 2.3(c). Thus we continue with the proof of the following result.

Theorem 4.6. *Let $\beta \in (0, 1)$ and $f(k) = \exp(k^\beta)$, $k \ge 0$. Then,*

$$\mathbb{E}[H_n(f)] = \mathrm{O}\left((\log n)^{1/(1-\beta)}\right).$$

Proof. The proof is very similar to the one for polynomial weight functions. Consider the levels $\ell = \lceil c \log^\gamma n \rceil$ and $h = \ell + \lfloor C \log n \rfloor$ for $\gamma = 1/(1 - \beta)$ and some constants $C, c > 0$ chosen later in the proof. Once again, let $f_h(k) = f(k \wedge h)$ and consider the function

$$\tilde{g}_h(k) = f_h(k + \ell)/f_h(\ell), \quad k \ge 0.$$

Note that $1 \le \tilde{g}_h(k) \le f_h(h)/f_h(\ell)$ and that $f_h(h)/f_h(\ell) = \mathrm{O}(1)$, which can be seen as follows: Since by the generalised Bernoulli inequality, $(1 + x)^\beta \le 1 + \beta x$ for $x \ge 0$ and $\beta \in (0, 1)$, we have

$$(h/\ell)^\beta \le \left(1 + C/c\,(\log n)^{1-\gamma}\right)^\beta \le 1 + \frac{C\beta}{c}\,(\log n)^{1-\gamma}.$$

Hence

$$f_h(h)/f_h(\ell) = \exp\left(\ell^\beta \left((h/\ell)^\beta - 1\right)\right) \le \exp\left(\beta C c^{\beta-1} (\log n)^{\beta\gamma+1-\gamma} (1 + o(1))\right).$$

The choice for γ yields $\beta\gamma + 1 - \gamma = 0$. Thus $f_h(h)/f_h(\ell) \le \exp(\beta + o(1))$ when choosing $C = c^{1-\beta}$. Hence, the same arguments as in Theorem 4.5 yield that $\mathbb{P}(H_n(f) > h) = \mathrm{O}(n^{-1})$. Therefore, $\mathbb{E}[H_n(f)] \le h + \mathrm{O}(1)$ and the assertion follows. □

Remark 4.7. *Recall that Proposition 2.7 yields* $\mathbb{E}[H_n(f)] = \Omega(\log n)$ *for every increasing* f*. Thus, for* f *as in Theorem 4.6, we only obtain*

$$\mathbb{E}[H_n(f)] = \Omega\left(\log n\right), \qquad \mathbb{E}[H_n(f)] = \mathrm{O}\left((\log n)^{1/(1-\beta)}\right),$$

and it remains an open problem to determine the asymptotic order of $\mathbb{E}[H_n(f)]$*.*

5 Functions giving quasilinear height

This section contains the missing proofs for parts (d) and (e) of Theorem 2.3.

5.1 Exponential weight functions

Theorem 5.1. *Let* $c > 2$ *and let* $f(k) = c^k$*. Then*

$$\mathbb{E}[H_n(f)] = \Omega(n/\log n).$$

Proof. Let $(N(n,\ell))_{\ell=1,\ldots,n}$ denote the profile of $\mathcal{T}_n(f)$, that is $N(n,\ell)$ equals the number of vertices in $\mathcal{T}_n(f)$ with depth ℓ. We will prove the stronger statement that with high probability $N(n,\ell) \le C\log n$ for all $1 \le \ell \le n$ and a sufficiently large constant C. More formally, for a constant $C > 0$ chosen later in the proof, let

$$A_n = \bigcap_{\ell=1}^{n} \{N(n,\ell) \le C\log n\}.$$

Note that A_n implies $H_n(f) \ge n/(C\log n)$. Thus it is sufficient to show $\mathbb{P}(A_n) \to 1$ as $n \to \infty$, or equivalently $\mathbb{P}(A_n^c) \to 0$.

Let $s = \lceil C\log n\rceil$, and $r < s$ (where r will be chosen later), and define the event $I_n(i,j,\ell) = I(i,j,\ell)$ to be

$$(N(i,\ell) = r) \wedge (N(j,\ell) = s) \wedge (N(j,\ell-1) < s) \wedge (N(j,\ell+1) < s).$$

First note that if none of the $I(i,j,\ell)$ hold for $1 \le i < j \le n$ and $\ell = 1,\ldots,n$, then the event A_n holds: If A_n does not hold, then there is a first time j such that there exists ℓ with $N(j,\ell) = s$. Then $N(j,\ell-1) < s$ and $N(j,\ell+1) < s$, and choosing any i so that $N(i,\ell) = r$, we have that $I(i,j,\ell)$ holds. Thus, we can show $\mathbb{P}(A_n) \to 1$ by showing that, with probability tending to 1, none of the events $I(i,j,\ell)$ holds. By the union bound it is therefore sufficient to show that

$$\mathbb{P}(I(i,j,\ell)) = o(n^{-3}) \quad \text{uniformly in } 1 \le i < j \le n \text{ and } \ell = 1,\ldots,n.$$

To bound the probability of a fixed event $I(i,j,\ell)$, first note that a new vertex t can be added to a DRRT according to the following two-stage procedure. First decide, with the correct probability, whether or not the vertex selected for attachment of the new leaf has depth in the set $\{\ell-1,\ell\}$, and then decide which level to choose in the appropriate set ($\{\ell-1,\ell\}$, or $[n]\setminus\{\ell-1,\ell\}$, as the case may be). Of course, deciding to attach to a node whose depth is in $\{\ell-1,\ell\}$ is equivalent to deciding to insert the new node such that its depth, i.e. distance from the root, is in the set $\{\ell, \ell+1\}$. We next re-describe this two-stage procedure in terms of i.i.d. uniformly distributed random variables, in order to define a coupling later in the proof. Since only the evolution of the tree $\mathcal{T}_j(f)$ for $j \geq i$ will be relevant for our further analysis, suppose that the tree $\mathcal{T}_i(f)$ is already constructed. Let D_t denote the depth of vertex t, that is its distance to the root. The sequence $(\mathcal{T}_j(f))_{j>i}$ is constructed as follows. Let $\{U_j, W_j : j \geq 1\}$ be a family of i.i.d. random variables uniformly distributed on $[0,1]$. For $t > i$, suppose the tree $\mathcal{T}_{t-1}(f)$ is already constructed. To attach a new vertex t, first decide whether $D_t \in \{\ell, \ell+1\}$ or not by using W_{t-i}; that is, define a Bernoulli random variable $\mathcal{I}_t$ via

$$\mathcal{I}_t = 1 \quad \iff \quad W_{t-i} \leq \frac{c^{\ell-1}N(t-1,\ell-1) + c^{\ell}N(t-1,\ell)}{\sum_{x=0}^{\infty} c^x N(t-1,x)}.$$

Note that the fraction of weights in the above expression is just the proportion of the total weight of vertices contained in levels $\ell-1$ and ℓ. If $\mathcal{I}_t = 1$, then D_t will be chosen equal to either ℓ or $\ell+1$, and the next U_j not used so far is used to decide which one, in the following way. Let $K = K(t) = \sum_{j=i+1}^{t-1} \mathcal{I}_j + 1$ and define a Bernoulli random variable Z_K via

$$Z_K = 1 \quad \iff \quad U_K \leq \frac{N(t-1,\ell-1)}{N(t-1,\ell-1) + cN(t-1,\ell)}. \tag{10}$$

If $Z_K = 1$, insert the new vertex into level ℓ (by choosing a parent uniformly at random among vertices on level $\ell-1$). Otherwise, i.e. when $Z_K = 0$, insert the vertex into level $\ell+1$.

On the other hand, if $\mathcal{I}_t = 0$, deduce the level for vertex t by splitting the unit interval according to the weights of levels in $[n] \setminus \{\ell-1, \ell\}$, in a similar fashion using an additional set of independent uniform random variables.

Now let $\mathcal{J} = \{t > i : \mathcal{I}_t = 1\}$ and let $t_1 < t_2 < \cdots$ be the (random) elements of $\mathcal{J}$ in increasing order.

To bound the probability of $I(i,j,\ell)$, we let $\tau_s(\ell-1) = \min\{m : N(m,\ell-1) = s\}$, and first establish that

$$I(i,j,\ell) \text{ implies } \mathbf{1}_{\{N(i,\ell)=r\}} \cdot \sum_{k=1}^{2s-r} Z_k \mathbf{1}_{\{t_k < \tau_s(\ell-1)\}} \geq s - r. \tag{11}$$

To show this, we may assume that $N(i,\ell) = r$. We consider two cases: (a) $t_{2s-r} \geq \tau_s(\ell-1)$ and (b) $t_{2s-r} < \tau_s(\ell-1)$. In case (a), the above inequality states that the number of vertices added to the initial r vertices in level ℓ, up to the time $\tau_s(\ell-1)$, is at least $s-r$. This is implied by $I(i,j,\ell)$, in particular $(N(j,\ell) = s) \wedge (N(j,\ell-1) < s)$. In case (b), the summation now counts vertices added to the initial r in level ℓ, up to the time when $2s-r$ have been added to levels in $\{\ell, \ell+1\}$. If the inequality in (11) is violated, then level $\ell+1$ receives s vertices before level ℓ receives $s-r$ additional vertices, which violates $(N(j,\ell) = s) \wedge (N(j,\ell+1) < s)$ from event $I(i,j,\ell)$, for all j. Therefore we have (11).

To bound the upper tail of the sum in (11), let $p_k = s/(s + c(r+k))$ for $k \geq 0$. Note that for all integers $t \in [i+1, \tau_s(\ell-1)]$ and conditioned on $N(i,\ell) = r$, the right hand side of (10) can be bounded by

$$\frac{N(t-1,\ell-1)}{N(t-1,\ell-1)+cN(t-1,\ell)} \leq \frac{s}{s+cN(t-1,\ell)} = p_{L_t} \text{ with } L_t = \sum_{k:t_k<t} Z_k. \tag{12}$$

We now introduce a new sequence $(R_j)_{j\geq 1}$ of Bernoulli random variables based on the random variables $(U_j)_{j\geq 1}$ in (10) via

$$R_j = 1 \quad \Longleftrightarrow \quad U_j \leq p_{R^+_{j-1}}, \quad \text{where } R_0^+ := 0 \text{ and } R_m^+ := \sum_{q=1}^{m} R_q \text{ for } m \geq 1.$$

Using induction on m, we can show

$$\mathbf{1}_{\{N(i,\ell)=r\}} \cdot \sum_{j=1}^{m} Z_j \mathbf{1}_{\{t_j<\tau_s(\ell-1)\}} \leq \sum_{j=1}^{m} R_j = R_m^+ \quad \text{for all } m \in \mathbb{N} \tag{13}$$

as follows. First, if (13) is a strict inequality for m then the inequality trivially holds for $m+1$ since both sums have Bernoulli increments. On the other hand, if the sums in (13) are equal for m, then, conditioned on $N(i,\ell) = r$, it holds that $R_{m+1} \geq Z_{m+1}\mathbf{1}_{\{t_{m+1}<\tau_s(\ell-1)\}}$ by definition and by applying (12) to every $t \in [t_m, \tau_s(\ell-1)]$.

Now, using (11) and (13), it follows that

$$\mathbb{P}(I(i,j,\ell)) \leq \mathbb{P}(R^+_{2s-r} \geq s-r). \tag{14}$$

We first point out how to finish the proof if $c > 4$. Note that for all k, we have $p_k \leq p_0 = s/(s+cr)$, so R^+_{2s-r} is stochastically dominated by a sum of $2s-r$ i.i.d. copies of R_1. Choosing $r \sim s/2$ and $c > 4$, we have for some $\epsilon > 0$ that $p_0 < 1/3 - \epsilon$, and also $2s-r \sim 3s/2$. Since $s \sim C\log n$, a standard Chernoff bound, e.g. Hoeffding's inequality [2, Theorem 2.8], implies that for large enough C, we have $\mathbb{P}(R^+_{2s-r} \geq s-r) = o(n^{-3})$. Hence, using the union bound, with probability tending to 1, none of the events $I(i,j,\ell)$ holds, as required.

For $c > 2$ we require a little more work. Write $Y_k = r + R_k^+$ (with $R_0^+ = 0$ by convention) and observe that

$$\mathbb{E}[Y_{k+1} - Y_k \mid Y_k] = p_{R_k^+} = g(Y_k) := \frac{s}{s+cY_k}.$$

The aim is to transform $(Y_k)_{k\geq 0}$ into a supermartingale and apply an Azuma-type supermartingale bound to prove $\mathbb{P}(R^+_{2s-r} \geq s-r) = o(n^{-3})$. This can be done by the differential equation method as in [12, Section 5.2].

The differential equation essentially models Y_k by a continuous function $y(x)$ and suggests for us to choose a function h with $h'(y)g(y) = 1$ for all $y \geq 0$. Setting $h(y) = y + cy^2/(2s)$, we get $h'(y)y'(x) = h'(y)g(y) = 1$. Then, noting that $|Y_{k+1} - Y_k| \leq 1$ and $h''(y) = c/s$, a Taylor expansion yields

$$h(Y_{k+1}) - h(Y_k) = h'(Y_k)(Y_{k+1} - Y_k) + \mathcal{E}_k, \qquad |\mathcal{E}_k| \leq c/s.$$

Since $c/s \to 0$ as $n \to \infty$, $\mathbb{E}[Y_{k+1} - Y_k | Y_k] = g(Y_k)$ and $h'(y)g(y) = 1$, it follows that, for any fixed $\epsilon > 0$,

$$\mathbb{E}[h(Y_{k+1}) - h(Y_k) \mid Y_0, \dots, Y_k] = \mathbb{E}[h(Y_{k+1}) - h(Y_k) \mid Y_k] < 1 + \epsilon$$

for n sufficiently large. Therefore, the sequence $(V_k)_{k\geq 0} := \big(h(Y_k) - k(1+\epsilon)\big)_{k\geq 0}$ is a supermartingale with respect to $Y_0, Y_1, \dots$ if n is sufficiently large. Since $|Y_{k+1} - Y_k| \leq 1$, it follows that uniformly for $k < 2s - r$ we have $|V_{k+1} - V_k| = O(c(1+\epsilon))$. Hence by the standard Azuma-type supermartingale bound (e.g. [12, Lemma 4.2]) we have for $\alpha > 0$

$$\mathbb{P}(V_{2s-r} \geq V_0 + \alpha) \leq \exp(-\Theta(\alpha^2/s)),$$

where the bound in Θ depends on c and ϵ. Choosing $\alpha = \epsilon s$, and $C = C(\epsilon, c)$ in $s = \lceil C \log n \rceil$ sufficiently large, makes this probability $o(n^{-3})$. (The values of C and ϵ will be fixed below.) The likely event $V_{2s-r} \leq V_0 + \epsilon s$ implies that

$$h(Y_{2s-r}) - h(Y_0) \leq 2s - r + \epsilon(3s - r),$$

and we are at liberty to choose $r \sim \epsilon s$, which then gives $h(Y_{2s-r}) \leq 2s + O(\epsilon s)$. For any $c > 2$, we may choose ϵ sufficiently small that this implies $Y_{2s-r} < s$. Since $Y_{2s-r} = r + R^+_{2s-r}$ by definition, such a choice for ϵ therefore yields $\mathbb{P}(R^+_{2s-r} \geq s - r) = o(n^{-3})$. The bound in (14) yields $\mathbb{P}(I(i,j,\ell)) = o(n^{-3})$ and the assertion follows. □

5.2 Path-like trees

The result stated in Theorem 2.3(e) is an immediate consequence of a more general observation for rapidly growing weight functions. Note that if f is growing sufficiently fast then there is non-vanishing chance of $\mathcal{T}_n(f)$ being a path. More precisely, note that the probability for $\mathcal{T}_n(f)$ being a path of n vertices (i.e. $H_n = n - 1$) is given by

$$\prod_{k=1}^{n-2} \frac{f(k)}{\sum_{j=0}^{k} f(j)} = \exp\left(\sum_{k=1}^{n-2} \log\left(1 - \left(1 + \frac{f(k)}{\sum_{j=0}^{k-1} f(j)}\right)^{-1}\right)\right). \tag{15}$$

This leads to the following result.

Lemma 5.2. *Let f be a weight function such that*

$$\sum_{k=1}^{\infty} \left(1 + \frac{f(k)}{\sum_{j=0}^{k-1} f(j)}\right)^{-1} < \infty.$$

Then,

$$\liminf_{n\to\infty} \mathbb{P}(H_n = n-1) > 0.$$

In particular, $\mathbb{E}[H_n] = \Theta(n)$.

Proof. The probability $\mathbb{P}(H_n = n-1)$ is stated in (15). By assumption and the fact that $\log(1-x) \sim -x$ as $x \to 0$ (for explicit bounds use, e.g., $-x-x^2 \leq \log(1-x) \leq -x$ for $x \in [0,1/2]$) we obtain

$$\mathbb{P}(H_n = n-1) \geq \exp\left(\sum_{k=1}^{\infty} \log\left(1 - \left(1 + \frac{f(k)}{\sum_{j=0}^{k-1} f(j)}\right)^{-1}\right)\right) > 0,$$

which yields the first part of the assertion. The second part obviously follows from $\mathbb{E}[H_n] \geq (n-1)\mathbb{P}(H_n = n-1)$. $\square$

Corollary 5.3. *Let $a > 1$, $f(0) = 1$ and $f(k) = \exp(ak\log k)$ for $k \geq 1$. Then*

$$\liminf_{n\to\infty} \mathbb{P}(H_n = n-1) > 0$$

and, in particular, $\mathbb{E}[H_n] = \Theta(n)$.

Proof. We have

$$\sum_{j=0}^{k-1} f(j) \leq f(k)(k^{-a} + k^{-2a} + \cdots + k^{-ka}) < 2f(k)k^{-a}$$

and hence the condition in Lemma 5.2 holds, and the assertion follows. $\square$

Remark 5.4. *Note that $\mathbb{P}(H_n = n-1)$ is decreasing in n since $\{H_n = n-1\} \subset \{H_{n-1} = n-2\}$ for any n. Therefore, $(\mathbb{P}(H_n = n-1))_{n\geq 1}$ is a convergent sequence and* lim inf *in Lemma 5.2 can be replaced by a limit.*

Remark 5.5. *It should also be possible to show linear height for slower growing weight functions if we, instead on focusing on the probability of being a path, consider the probability of resulting in something 'path-like' (e.g. $H_n \sim n$ as $n \to \infty$ with positive probability by showing that $A_n = \{H_{n+1} = H_n\}$ occurs only a sublinear number of times).*

We believe that $f(k) = \exp(ak\log k)$ should lead to a 'path-like' tree for every $a > 0$ and thus $\mathbb{E}[H_n(f)] = \Theta(n)$ for these weight functions.

6 More open problems

The introduction of DRRTs raises a lot of questions that remain unanswered in this paper. Our main objective in this paper is to find relationships between increasing functions f and the rate of growth of the height of $\mathcal{T}_n(f)$. Our results concern polylogarithmic or quasilinear height. There is a gap in between that we have not touched upon: if $0 < \alpha < 1$, for what f is the expected roughly n^α? One could also ask what kind of 'penalty' (i.e. decreasing f) it takes to obtain a tree with $\mathbb{E}[H_n(f)] = o(\log n)$. The results in Theorem 2.3 have several other obvious gaps that would be interesting to fill. For a more difficult problem, one can also ask for the variance or limit law of the heights of these trees.

There are also a variety of other tree parameters to consider based on results for random recursive trees:

- What is the asymptotic behaviour of the *depth* D_n of vertex n, i.e. the distance of the n-th inserted vertex to the root (cf. [4] for random recursive trees)? A study of the depth could potentially be an easy way to obtain lower bounds on the height of tree.
- What is the total number of leaves in the tree $\mathcal{T}_n(f)$ (cf. [11] and [10] for random recursive trees)?

  It is not hard to check that, on average, half of the vertices in a random recursive tree are leaves. On the other hand, for rapidly increasing f (such as in Theorem 2.3(e)), one expects a 'path-like' tree and thus only very few leaves.
- One could also study the *total path length* $\sum_{j=1}^{n} D_j$ of the tree (cf. [8] for random recursive trees). A natural parameter to study would also be the total weight $\sum_{j=1}^{n} f(D_j)$ of all vertices in $\mathcal{T}_n(f)$.

Acknowledgment We wish to thank Abbas Mehrabian pointing us towards this topic, and the referees for suggesting improvements in the presentation and pointing out some glitches in the proofs.

References

[1] Addario-Berry, L. and Ford, K. Poisson-Dirichlet branching random walks. *Ann. Appl. Probab.* **23**, 283–307, 2013.

[2] Boucheron, S., Lugosi, G. and Massart, P. *Concentration inequalities: A nonasymptotic theory of independence.* Oxford University Press, 2013.

[3] Devroye, L. Branching processes in the analysis of the heights of trees. *Acta Informatica* **24** (3), 277–298, 1987

[4] Devroye, L. Applications of the theory of records in the study of random trees. *Acta Informatica* **26** (1-2), 123–130, 1988.

[5] Jagers, P. *Branching processes with biological applications.* John Wiley & Sons, London, 1975.

[6] Janson, S., Łuczak, T. and Rucinski, A. *Random graphs.* Wiley, New York, 2000.

[7] Leckey, K. and Neininger, R. Asymptotic analysis of Hoppe trees. *Journal of Applied Probability* **50** (1), 228–238, 2013.

[8] Mahmoud, H. Limiting Distributions for Path Lengths in Recursive Trees. *Probability in the Engineering and Informational Sciences* **5**(1), 53–59, 1991.

[9] Mehrabian, A. Justifying the small-world phenomenon via random recursive trees. *Random Structures & Algorithms* **50**, 201–224, 2017.

[10] Na, H. and Rapaport, A. Distribution of nodes of a tree by degree. *Mathematical Biosciences* **6**, 313–329, 1970.

[11] Najock, D. and Heyde, C.C. On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *Journal of Applied Probability* **19**(3), 675–680, 1982.

[12] Wormald, N. C. The differential equation method for random graph processes and greedy algorithms. In *Lectures on Approximation and Randomized Algorithms*, M. Karoński and H.J. Prömel (eds), 73–155. PWN, Warsaw, 1999.