ANALYSIS OF PARAMETERS OF TREES CORRESPONDING TO HUFFMAN CODES AND SUMS OF UNIT FRACTIONS

CLEMENS HEUBERGER, DANIEL KRENN, AND STEPHAN WAGNER

Abstract. For fixed $t \geq 2$, we consider the class of representations of 1 as sum of unit fractions whose denominators are powers of t or equivalently the class of canonical compact t-ary Huffman codes or equivalently rooted t-ary plane "canonical" trees.

We study the probabilistic behaviour of the height (limit distribution is shown to be normal), the number of distinct summands (normal distribution), the path length (normal distribution), the width (main term of the expectation and concentration property) and the number of leaves at maximum distance from the root (discrete distribution).

Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Austria
E-mail address: clemens.heuberger@aau.at

Institut für Optimierung und Diskrete Mathematik (Math B), TU Graz, Austria
E-mail address: krenn@math.tugraz.at

Department of Mathematical Sciences, Stellenbosch University, South Africa
E-mail address: swagner@sun.ac.za

C. Heuberger is supported by the Austrian Science Fund (FWF): S9606, that is part of the Austrian National Research Network "Analytic Combinatorics and Probabilistic Number Theory".

C. Heuberger and D. Krenn are supported by the Austrian Science Fund (FWF): W1230, Doctoral Program "Discrete Mathematics".

S. Wagner is supported by the National Research Foundation of South Africa, grant number 70560.

1. Introduction

Let $t \geq 2$ be an integer. We consider the following combinatorial classes which turn out to be equivalent. See Figure 1 for examples.

(1) Partitions of 1 into powers of t (representation of 1 as sum of unit fractions whose denominators are powers of t):

$$\mathcal{C}_{Partition} = \Big\{(x_1, \dots, x_r) \in \mathbb{Z}^r \;\Big|\; r \geq 0,\, 0 \leq x_1 \leq x_2 \leq \dots \leq x_r,\, \sum_{i=1}^{r} \frac{1}{t^{x_i}} = 1\Big\}.$$

The external size $|(x_1, \dots, x_r)|$ of such a representation $(x_1, \dots, x_r)$ is defined to be the number r of summands.

(2) Canonical compact t-ary Huffman codes:

$$\mathcal{C}_{Code} = \{C \subseteq \{1, \dots, t\}^* \mid C \text{ is prefix-free, compact and canonical}\}.$$

Here,

- $\{1, \dots, t\}^*$ denotes the set of finite words over the alphabet $\{1, \dots, t\}$,
- a code C is said to be prefix-free if no word in C is a proper prefix of any other word in C,
- a code C is said to be compact if the following property holds: if w is a proper prefix of a word in C, then for every letter $a \in \{1, \dots, t\}$, wa is a prefix of a word in C,
- a code C is said to be canonical if the lexicographic ordering of its words corresponds to a non-decreasing ordering of the word lengths. This condition corresponds to taking equivalence classes with respect to permutations of the alphabet (at each position in the words).

The external size $|C|$ of a code C is defined to be the cardinality of C.

If $C \in \mathcal{C}_{Code}$ with $C = \{w_1, \dots, w_r\}$ and the property that $\mathsf{length}(w_i) \leq \mathsf{length}(w_{i+1})$ for all i, then $(\mathsf{length}(w_1), \dots, \mathsf{length}(w_r)) \in \mathcal{C}_{Partition}$. This is a bijection between $\mathcal{C}_{Code}$ and $\mathcal{C}_{Partition}$ preserving the external size.

(3) Canonical rooted t-ary trees:

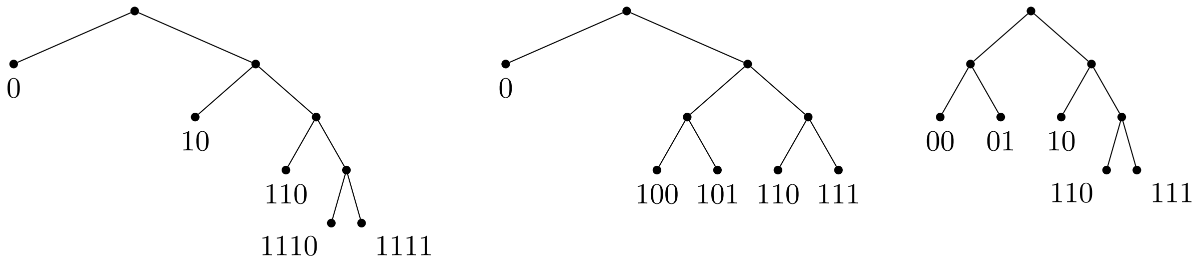
$$\mathcal{C}_{Tree} = \{T \text{ rooted } t\text{-ary plane tree} \mid T \text{ is canonical}\}.$$

Here,

- t-ary means that each vertex has no or t children,
- plane tree means that an ordering "from left to right" of the children of each vertex is specified,
- canonical means that the following holds for all k: if the vertices of depth (i.e., distance to the root) k are denoted by $v_1, \dots, v_K$ from left to right, then $\deg(v_i) \leq \deg(v_{i+1})$ holds for all i.

The external size $|T|$ of a tree is given by the number of its leaves, i.e., the number of vertices of degree 1.

If $C \in \mathcal{C}_{Code}$, then a tree $T \in \mathcal{C}_{Tree}$ can be constructed such that the vertices of T are given by the prefixes of the words in C, the root is the vertex corresponding to the empty word, and the children of a proper prefix w of a code word are given from left to right by wa for $a = 1, \dots, t$. This is a bijection between $\mathcal{C}_{Code}$ to $\mathcal{C}_{Tree}$ preserving the external size.



$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} \qquad 1 = \frac{1}{2} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \qquad 1 = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8}$$

FIGURE 1. All elements of external size 5 (and internal size 4) in $\mathcal{C}_{Tree}$, $\mathcal{C}_{Code}$ and $\mathcal{C}_{Partition}$ for $t = 2$.

Further formulations, details and remarks can be found in [9]. We will simply speak of an element in the class $\mathcal{C}$ when the particular interpretation as an element of $\mathcal{C}_{Partition}$, $\mathcal{C}_{Code}$ or $\mathcal{C}_{Tree}$ is not relevant. Our proofs will use the tree model, therefore, $\mathcal{C}_{Tree}$ is abbreviated as $\mathcal{T}$.

The external size of an element in $\mathcal{C}$ is always congruent to 1 modulo $t-1$. This can easily be seen in the tree model, where the number of leaves r and the number of internal vertices n are connected by the identity

$$r = 1 + n(t-1).$$

Therefore, we will from now on consider the *internal size:* for a tree $T \in \mathcal{C}_{Tree}$ the internal size of T is the number $n(T)$ of internal vertices, for a code $C \in \mathcal{C}_{Code}$ the internal size is the number of proper prefixes of words of C, and for a partition $(x_1, \ldots, x_r) \in \mathcal{C}_{Partition}$ the internal size is defined to be $(r-1)/(t-1)$. We will omit the word "internal" and will always use the variable n to denote the size.

The asymptotics of the number of elements in $\mathcal{C}$ of size n has been studied by various authors, cf. again [9]. In that paper, building upon a generating function approach by Flajolet and Prodinger [11], the following result has been obtained:

Theorem I ([9])**.** *For $t \geq 2$, the number of elements of size n in $\mathcal{C}$ can be estimated as*

$$R\rho^{n+1} + \Theta(\rho_2^n),$$

where $\rho > \rho_2$ and R are positive real constants depending on t with asymptotic expansions (as $t \to \infty$),

$$\rho = 2 - \frac{1}{2^{t+1}} + O\left(\frac{t}{2^{2t}}\right), \qquad \rho_2 = 1 + \frac{\log 2}{t} + O\left(\frac{1}{t^2}\right), \qquad R = \frac{1}{8} + \frac{t-2}{2^{t+5}} + O\left(\frac{t^2}{2^{2t}}\right).$$

In fact, all O-constants can be made explicit and more terms of the asymptotic expansions in t of ρ, ρ_2 and R can be given.

The purpose of this contribution is to study the probabilistic behaviour of various parameters of a random element in $\mathcal{C}$ of size n (all elements considered to be equally likely):

(1) The *height* $h(T)$ of a tree $T \in \mathcal{C}_{Tree}$ is defined to be the maximum distance of a leaf from the root. In the interpretation as a code, this is the maximum length of a code word. In a representation of 1 as a sum of unit fractions, this corresponds to the largest denominator used (more precisely, to the largest exponent of the denominator).

The height is discussed in Section 3. It is asymptotically normally distributed with mean $\sim \mu_h n$ and variance $\sim \sigma_h^2 n$, where

$$\mu_h = \frac{1}{2} + \frac{t-2}{2^{t+3}} + O\Big(\frac{t^2}{2^{2t}}\Big) \quad \text{and} \quad \sigma_h^2 = \frac{1}{4} + \frac{-t^2+5t-2}{2^{t+4}} + O\Big(\frac{t^3}{2^{2t}}\Big),$$

cf. Theorem III.

(2) The *number of distinct summands* of a representation $(x_1, \dots, x_r)$ of 1 as sum of unit fractions is denoted by $d(x_1, \dots, x_r)$. In the tree model, this corresponds to the cardinality $d(T)$ of the set of depths of leaves in a tree $T \in \mathcal{C}_{\mathit{Tree}}$. In the code model, this is the number of distinct lengths of code words.

The number $d(T)$ is studied in Section 4. It is asymptotically normally distributed with mean $\sim \mu_d n$ and variance $\sim \sigma_d^2 n$, where

$$\mu_d = \frac{1}{2} + \frac{t-4}{2^{t+3}} + O\Big(\frac{t^2}{2^{2t}}\Big) \quad \text{and} \quad \sigma_d^2 = \frac{1}{4} + \frac{-t^2+9t-14}{2^{t+4}} + O\Big(\frac{t^2}{2^{2t}}\Big),$$

cf. Theorem IV.

(3) The *maximum number of equal summands* of a representation $(x_1, \dots, x_r)$ of 1 as sum of unit fractions is denoted by $w(x_1, \dots, x_r)$. In the code model, this is the maximum number of code words of equal length; in the tree model, this is the "leaf-width" $w(T)$, the maximum number of leaves on the same level.

The number $w(T)$ is studied in Section 7. We prove that $\mathbb{E}(w(T)) = \mu_w \log n + O(\log\log n)$ with $\mu_w = 1/(t\log 2) + O(1/t^2)$ and a concentration property, cf. Theorem VII.

(4) The *(total) path length* $\ell(T)$ of a tree $T \in \mathcal{C}_{\mathit{Tree}}$ is defined to be the sum of the depths of all vertices of the tree. In our context, it is perhaps most natural to consider the *external path length* $\ell_{\mathit{external}}(T)$, though, which is the sum of depths over all leaves of the tree, as this parameter corresponds to the sum of lengths of code words in a code $C \in \mathcal{C}_{\mathit{Code}}$. Likewise, the *internal path length* $\ell_{\mathit{internal}}(T)$ is the sum of depths over all non-leaves. Clearly, we have $\ell_{\mathit{external}}(T) + \ell_{\mathit{internal}}(T) = \ell(T)$, and the relations

$$\ell_{\mathit{external}}(T) = \frac{t-1}{t}\ell(T) + n(T) \quad \text{and} \quad \ell_{\mathit{internal}}(T) = \frac{1}{t}\ell(T) - n(T)$$

for t-ary trees are easily proven. Therefore, all distributional results for any one of those parameters immediately cover all three. The total path length turns out to be asymptotically normally distributed as well (see Theorem VI), with mean $\sim \mu_{\mathit{tpl}} n^2$ and variance $\sim \sigma_{\mathit{tpl}}^2 n^3$. The coefficients have asymptotic expansions

$$\mu_{\mathit{tpl}} = \frac{t}{2}\cdot\mu_h = \frac{t}{4} + \frac{t(t-2)}{2^{t+4}} + O\Big(\frac{t^3}{2^{2t}}\Big) \quad \text{and} \quad \sigma_{\mathit{tpl}} = \frac{t^2}{12} + \frac{-t^4+5t^3+2t^2}{3\cdot 2^{t+4}} + O\Big(\frac{t^5}{2^{2t}}\Big).$$

The path length is studied in Section 6; its analysis is based on a generating function approach for the moments, combined with probabilistic arguments to obtain the central limit theorem.

(5) The *number of leaves on the last level* (i.e., maximum distance from the root) of a tree $T \in \mathcal{C}_{\mathit{Tree}}$ is denoted by $m(T)$. This corresponds to the number of code words of maximum length and to the number of smallest summands in a representation of 1 as a sum of unit fractions.

This parameter may appear to be the least interesting of the parameters we study. However, it is a natural technical parameter when constructing generating functions for the other parameters. From these generating functions, the probabilistic behaviour of $m(T)$ can be read off without too much effort, so we do include these results in Section 5.

The limit distribution of $m(T)$ is a discrete distribution with mean $2t + o(1)$ and variance $2t^2 + o(1)$, cf. Theorem V.

A noteworthy feature of the results listed above is the fact that the distributions we observe are quite different from those that one obtains for other probabilistic random tree models, specifically Galton–Watson trees (which include, amongst others, random t-ary trees), but also recursive trees and general families of increasing trees, see [5] for a general reference. Specifically,

- the asymptotic order of the height of a random Galton–Watson tree of order n is only $\sqrt{n}$, and it is known that the limiting distribution (which is sometimes called a Theta distribution) coincides with the distribution of the maximum of a Brownian excursion [10]. The height of random recursive trees (or other families of increasing trees) is even only of order $\log n$, and heavily concentrated around its mean, see [4].
- The path length of random Galton–Watson trees is of order $n^{3/2}$, and it follows an Airy distribution (like the area under a Brownian excursion) in the limit [18]. For recursive trees, the path length is of order $n \log n$ with a rather unusual limiting distribution [14].
- While the height of our canonical trees is greater than that of Galton–Watson trees, precisely the opposite holds for the width (as one would expect): it is of order $\sqrt{n}$ for Galton-Watson trees [6, 19], with the same limiting distribution as the height, as opposed to only $\log n$ in our setting. For recursive trees, the width is even of order $n/\sqrt{\log n}$, see [7].

Indeed, the structure of our canonical t-ary trees is comparable to that of *compositions*: counting the number of internal vertices on each level from the root, we obtain a restricted composition (see the series of papers by Bender and Canfield [1, 2, 3] on recent results concerning compositions with various local restrictions), in which each summand is at most t times the previous one. In the limit $t \to \infty$, one obtains compositions of n starting with a 1 in this way.

Last in this introduction a remark on the notations of the error terms: In all our major results those error terms have an explicit O-constant. The error functions $\varepsilon_j(\dots)$ that appear there are real functions which fulfil $|\varepsilon_j(\dots)| \le 1$ for all values of the indicated parameters. Those constants were calculated with the computer algebra system Sage [16].

2. The Generating Function

The height $h(T)$, the cardinality $d(T)$ of the set of different depths of leaves and the number $m(T)$ of leaves on the last level of a tree $T \in \mathcal{T}$ of size $n = n(T)$ can be analysed by studying a multivariate generating function $H(q, u, v, w)$, where q labels the size $n(T)$, u labels the number $m(T)$ of leaves on the last level, v labels the cardinality $d(T)$ of the set of depths of leaves and w labels the height $h(T)$.

Theorem II. *The generating function*

$$H(q, u, v, w) := \sum_{T \in \mathcal{T}} q^{n(T)} u^{m(T)} v^{d(T)} w^{h(T)}$$

can be expressed as

$$H(q,u,v,w) = a(q,u,v,w) + b(q,u,v,w)\frac{a(q,1,v,w)}{1-b(q,1,v,w)} \tag{2.1}$$

with

$$a(q,u,v,w) = \sum_{j=0}^{\infty} vq^{[j]}u^{t^j}w^j \prod_{i=1}^{j} \frac{1-v-q^{[i]}u^{t^i}}{1-q^{[i]}u^{t^i}},$$

$$b(q,u,v,w) = \sum_{j=1}^{\infty} \frac{vq^{[j]}u^{t^j}w^j}{1-q^{[j]}u^{t^j}} \prod_{i=1}^{j-1} \frac{1-v-q^{[i]}u^{t^i}}{1-q^{[i]}u^{t^i}}, \tag{2.2}$$

where $[j] := 1 + t + \cdots + t^{j-1}$.

Proof. The proof of Theorem II follows ideas of Flajolet and Prodinger [11], (see also [9]), which we only sketch briefly. Details can be found in Appendix A. One first considers

$$H_h(q,u,v) := [w^h]H(q,u,v,w) = \sum_{\substack{T\in\mathcal{T}\\ h(T)=h}} q^{n(T)}u^{m(T)}v^{d(T)}$$

for some $h \geq 0$. A tree T' of height $h+1$ arises from a tree T of height h by replacing j of its $m(T)$ leaves on the last level by internal vertices with t succeeding leaves respectively, where $1 \leq j \leq m(T)$. If $j < m(T)$, then $d(T') = d(T) + 1$; otherwise, we have $d(T') = d(T)$. For the generating function H_h, this translates to the recursion

$$\begin{aligned} H_{h+1}(q,u,v) &= \sum_{\substack{T\in\mathcal{T}\\ h(T)=h}} \Bigg(\sum_{j=1}^{m(T)-1} q^{n(T)+j}u^{jt}v^{d(T)+1} + q^{n(T)+m(T)}u^{m(T)t}v^{d(T)} \Bigg) \\ &= r(q,u,v)H_h(q,1,v) + s(q,u,v)H_h(q,qu^t,v) \end{aligned} \tag{2.3}$$

with

$$r(q,u,v) = \frac{qu^tv}{1-qu^t}, \qquad s(q,u,v) = \frac{1-v-qu^t}{1-qu^t},$$

and initial value $H_0(q,u,v) = uv$. This further means that

$$H(q,u,v,w) = uv + wr(q,u,v)H(q,1,v,w) + ws(q,u,v)H(q,qu^t,v,w),$$

and this functional equation can be solved by iteration. One obtains

$$H(q,u,v,w) = a(q,u,v,w) + b(q,u,v,w)H(q,1,v,w),$$

and (2.1) results by plugging in $u = 1$ and solving for $H(q,1,v,w)$. □

Next we recall results on the singularities of $H(q,1,1,1)$, see Proposition 10 of [9].

Lemma 2.1 ([9])**.** *The generating function $H(q,1,1,1)$ has exactly one singularity $q = q_0$ with $|q| < 1 - \frac{0.72}{t}$. This singularity q_0 is a simple real pole. For $t \geq 4$, we have*

$$q_0 = \frac{1}{2} + \frac{1}{2^{t+3}} + \frac{t+4}{2^{2t+5}} + \frac{3t^2+23t+38}{2^{3t+8}} + \frac{7t^3}{100\cdot 2^{4t}}\varepsilon_1(t).$$

For $t \in \{2,3\}$, the values are given in Table 1. Furthermore, let

$$Q = \frac{1}{2} + \frac{\log 2}{2t} + \frac{0.06}{t^2}$$

t	q_0	Q	t	q_0	Q
2	$0.5573678720\dots$	0.71317958	4	$0.5090030531\dots$	0.59306918
3	$0.5206401166\dots$	0.63074477	5	$0.5042116835\dots$	0.57200784

TABLE 1. Constants q_0 and Q for $2 \le t \le 5$.

for $t \ge 6$ and Q be given by Table 1 for $2 \le t \le 5$. Then q_0 is the only singularity of $H(q, 1, 1, 1)$ with $|q| \le q_0/Q$.

Using this result, we will be able to apply singularity analysis to all our generating functions in the coming sections.

3. The Height

We start our analysis with the height $h(T)$ of our canonical trees $T \in \mathcal{T}$. We show that the height is asymptotically (for large sizes $n = n(T)$) normally distributed and calculate its mean and variance. We will do this by means of the generating function $H(q, u, v, w)$ defined in Section 2.

So let us have a look at the bivariate generating function

$$H(q, 1, 1, w) = \sum_{T \in \mathcal{T}} q^{n(T)} w^{h(T)} = \frac{a(q, 1, 1, w)}{1 - b(q, 1, 1, w)}$$

for the height. We consider its denominator

$$D(q, w) := 1 - b(q, 1, 1, w) = \sum_{0 \le j} (-1)^j w^j \prod_{i=1}^{j} \frac{q^{[i]}}{1 - q^{[i]}}.$$

From Lemma 2.1 we know that $D(q, 1)$ has a simple zero q_0. Expanding $D(q, w)$ around $(q_0, 1)$ and using Theorem IX.9 (meromorphic singularity perturbation) from the book of Flajolet and Sedgewick [12] yields the desired results for the height without much effort. They are stated precisely in the following theorem.

Theorem III. *The height is asymptotically normally distributed. Its mean is $\mu_h n + O(1)$ and its variance is $\sigma_h^2 n + O(1)$ with*

$$\mu_h = \frac{1}{2} + \frac{t-2}{2^{t+3}} + \frac{2t^2 + 3t - 8}{2^{2t+5}} + \frac{9t^3 + 45t^2 + 2t - 88}{2^{3t+8}} + \frac{0.044t^4}{2^{4t}} \varepsilon_2(t)$$

and

$$\sigma_h^2 = \frac{1}{4} + \frac{-t^2 + 5t - 2}{2^{t+4}} + \frac{-4t^3 + 4t^2 + 27t - 7}{2^{2t+6}} + \frac{0.058t^4}{2^{3t}} \varepsilon_3(t)$$

for $t \ge 3$. In the case $t = 2$ we have $\mu_h = 0.5662757699\dots$ and $\sigma_h^2 = 0.2665499010\dots$.

We calculated the values of the constants μ_h and σ_h^2 numerically for $2 \le t \le 30$. Those values can be found in Table 3 in Appendix B, where a complete proof of Theorem III is given as well.

4. The Number of Distinct Depths of Leaves

In this section we study the number of distinct depths of leaves $d(T)$ of our canonical trees $T \in \mathcal{T}$, motivated by the interpretation as the number of distinct code lengths in Huffman codes. This parameter is also asymptotically normally distributed, and the approach is essentially the same as for the height, based on the generating function $H(q, u, v, w)$ from Section 2. To analyse the parameter $d(T)$, we look at the bivariate generating function

$$H(q, 1, v, 1) = \sum_{T \in \mathcal{T}} q^{n(T)} v^{d(T)} = \frac{a(q, 1, v, 1)}{1 - b(q, 1, v, 1)}$$

for the number of distinct depths of leaves. Again, we consider its denominator

$$D(q, v) := 1 - b(q, 1, v, 1) = 1 - \sum_{1 \le j} \frac{v}{1 - q^{[j]}} \prod_{i=1}^{j-1} \frac{1 - v - q^{[i]}}{1 - q^{[i]}}$$

and proceed as in the previous section. Lemma 2.1 tells us the existence of a simple zero q_0 of $D(q, 1)$. Again, we expand the denominator $D(q, v)$ around $(q_0, 1)$ and use Theorem IX.9 from the book of Flajolet and Sedgewick [12]. This results in the following theorem.

Theorem IV. *The number of distinct depths of leaves is asymptotically normally distributed. Its mean is $\mu_d n + O(1)$ and its variance is $\sigma_d^2 n + O(1)$ with*

$$\mu_d = \frac{1}{2} + \frac{t-4}{2^{t+3}} + \frac{2t^2 - t - 14}{2^{2t+5}} + \frac{9t^3 + 27t^2 - 76t - 144}{2^{3t+8}} + \frac{0.046t^4}{2^{4t}} \varepsilon_4(t)$$

and

$$\sigma_d^2 = \frac{1}{4} + \frac{-t^2 + 9t - 14}{2^{t+4}} + \frac{-4t^3 + 20t^2 + 3t - 54}{2^{2t+6}} + \frac{0.056t^4}{2^{3t}} \varepsilon_5(t)$$

for $t \ge 2$.

Again, as in the previous section, we calculated the values of the constants μ_d and σ_d^2 numerically for $2 \le t \le 30$, and they are given in Table 4 in Appendix C, where the proof of Theorem IV is detailed as well.

5. The Number of Leaves on the Last Level

For analysing the parameter $m(T)$ counting the number of leaves of maximum depth (labelled by the variable u in the generating function $H(q, u, v, w)$), we note that for fixed $|u| \le 1$, the dominant simple pole q_0 of $H(q, 1, 1, 1)$ is also the dominant singularity of $H(q, u, 1, 1)$ and is still a simple pole. Therefore, $m(T)$ tends to a discrete limiting distribution, we refer again to the book of Flajolet and Sedgewick [12, Section IX.2]. Note that the number $m(T)$ is always divisible by t by construction.

Theorem V. *Let q_0 and Q be as described in Lemma 2.1. Set $p_m = [u^{mt}]b(q_0, u, 1, 1)$ for $m \ge 1$. Then, for a random tree $T \in \mathcal{T}$ of size n, we have*

$$\mathbb{P}(m(T) = mt) = p_m + O(Q^n)$$

for $m \ge 1$.

Furthermore, we have

$$\mathbb{E}(m(T)) = 2t - \frac{t^2 - t}{2^{t+1}} - \frac{t^3 + 6t^2 - 5t}{2^{2t+3}} - \frac{3t^4 + 32t^3 + 61t^2 - 56t}{2^{3t+8}} + O\left(\frac{t^5}{2^{4t}} + Q^n\right)$$

and

$$\mathbb{V}(m(T)) = 2t^2 - \frac{t^4-3t^2}{2^{t+1}} - \frac{t^5+13t^4-3t^3-17t^2}{2^{2t+3}} + O\left(\frac{t^6}{2^{3t}} + Q^n\right).$$

The proof can be found in Appendix D. This theorem (slightly generalised) is a very useful tool in proving the central limit theorem for the path length in the following section.

6. The Path Length

This section is devoted to the analysis of the path length, as defined in the introduction. While the external path length is most natural in the setting of Huffman codes, it is more convenient to work with the total and the internal path length. As it was pointed out in the introduction, the three are essentially equivalent as they are (deterministically) related by simple linear equations.

We first use a generating functions approach to determine the asymptotic behaviour of the mean and variance. Let us define a generating function L_r for the r-th moment of the total path length as follows:

$$L_r(q,u,w) := \sum_{T\in\mathcal{T}} \ell(T)^r q^{n(T)} u^{m(T)} w^{h(T)}.$$

Note that $L_0(q,u,w) = H(q,u,1,w)$ in the notation of the previous sections. We are specifically interested in L_1 and L_2. In analogy to the approach we used to determine a formula for $H(q,u,v,w)$, we obtain a functional equation for $L_r(q,u,w)$ by first introducing

$$L_{r,h}(q,u) = [w^h]L_r(q,u,w) = \sum_{\substack{T\in\mathcal{T}\\ h(T)=h}} \ell(T)^r q^{n(T)} u^{m(T)}.$$

Replacing j leaves of depth h by internal vertices, thus creating tj new leaves of depth $h+1$, increases the total path length by $tj(h+1)$. Thus we get

$$\begin{aligned} L_{1,h+1}(q,u) &= \sum_{\substack{T\in\mathcal{T}\\ h(T)=h+1}} (h+1)m(T)q^{n(T)}u^{m(T)} + \sum_{\substack{T\in\mathcal{T}\\ h(T)=h}} \sum_{j=1}^{m(T)} \ell(T) q^{n(T)+j} u^{jt} \\ &= (h+1)u\frac{\partial}{\partial u} L_{0,h+1}(q,u) + \frac{qu^t}{1-qu^t}\left(L_{1,h}(q,1) - L_{1,h}(q,qu^t)\right). \end{aligned}$$

Define, for the sake of convenience, the linear operators $\Phi_u = u\frac{\partial}{\partial u}$, $\Phi_w = w\frac{\partial}{\partial w}$ and $\Phi_q = q\frac{\partial}{\partial q}$ acting on our generating functions. Then we obtain

$$L_1(q,u,w) = \Phi_u\Phi_w L_0(q,u,w) + \frac{qu^t w}{1-qu^t}\left(L_1(q,1,w) - L_1(q,qu^t,w)\right).$$

Likewise, one gets a functional equation for $L_2(q,u,w)$:

$$L_2(q,u,w) = 2\Phi_u\Phi_w L_1(q,u,w) - \Phi_u^2\Phi_w^2 L_0(q,u,w) + \frac{qu^t w}{1-qu^t}\left(L_2(q,1,w) - L_2(q,qu^t,w)\right).$$

Both functional equations can be solved by means of iteration in the same way as the functional equation for the generating function $H(q,u,v,w)$ that we used in previous sections, see Appendix E for details. In order to determine the asymptotic behaviour of mean and variance, one only needs to find the expansion around the dominating singularity q_0 and apply singularity analysis. The main term of the mean is easy to guess: assuming that the vertices are essentially uniformly distributed along the entire height, it is natural to conjecture that $\ell(T)$ is typically around $tn(T)h(T)/2$ and thus of quadratic order. This is indeed true, and the variance turns out to be of cubic order (terms of degree 4 cancel, as one would expect). The details are rather lengthy and given in the appendix.

In order to prove convergence to the Gaussian distribution, a different, more probabilistic approach is needed. Standard theorems from analytic combinatorics no longer apply since the path length grows faster than, for example, the height, so that mean and variance no longer have linear order.

We number the internal vertices of a random canonical t-ary tree of size n from 1 to n in a natural top-to-bottom, left-to-right way, starting at the root. Let $X_{k,n}$ denote the depth of the k-th internal vertex in a random tree $T \in \mathcal{T}$ of order n. Moreover, set $Y_{k,n} = X_{k+1,n} - X_{k,n} \in \{0,1\}$. In words, $Y_{k,n}$ is 1 if the $(k+1)$-th internal vertex has greater distance from the root than the k-th, and 0 otherwise. It is clear that the height can be expressed as

$$h(T) = 1 + \max_k X_{k,n} = 1 + X_{n,n} = 1 + \sum_{k=1}^{n-1} Y_{k,n},$$

which would indeed be an alternative approach to the central limit theorem for the height. More importantly, though, the internal path length can also be expressed in terms of the random variables $Y_{k,n}$:

$$\ell_{internal}(T) = \sum_{k=1}^{n} X_{k,n} = \sum_{k=1}^{n} \sum_{j=1}^{k-1} Y_{j,n} = \sum_{j=1}^{n-1} (n-j) Y_{j,n}.$$

Now

$$n^{-1} \ell_{internal}(T) = \sum_{j=1}^{n-1} \frac{n-j}{n} Y_{j,n}$$

can be seen as a sum of $n-1$ bounded random variables $Z_{j,n} = \frac{n-j}{n} Y_{j,n}$. An advantage of this decomposition over other possible decompositions (e.g., by counting the number of vertices at different depths) is that the number of variables is not random. The $Z_{j,n}$ are neither identically distributed (which is not a major issue) nor independent. Fortunately, however, they are almost independent in that they satisfy a so-called "strong mixing condition". Let $\mathcal{F}_{s_1}$ be the σ-algebra induced by the random variables $Z_{1,n}, Z_{2,n}, \ldots, Z_{s_1,n}$, and let $\mathcal{G}_{s_2}$ be the σ-algebra induced by the random variables $Z_{s_2,n}, Z_{s_2+1,n}, \ldots, Z_{n-1,n}$. There exist constants κ and λ such that

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \kappa e^{-\lambda(s_2 - s_1)} \tag{6.1}$$

for all $1 \leq s_1 < s_2 \leq n$ and all events $A \in \mathcal{F}_{s_1}$ and $B \in \mathcal{G}_{s_2}$. The main idea is simple: events $A \in \mathcal{F}_{s_1}$ describe the shape of the random tree T up to the s_1-th internal vertex, while events $B \in \mathcal{G}_{s_2}$ describe the shape of the random tree T from the s_2-th internal vertex on. The probabilities of such events can be calculated by means of the generating function approach explained in Section 2, and the exponential error terms that one obtains through this approach (as in Theorem V) yield the estimate (6.1) above. A more detailed explanation can be found in the appendixonce again.

t	μ_w	t	μ_w	t	μ_w
2	1.7107...	9	0.1804...	16	0.0961...
3	0.7660...	10	0.1603...	17	0.0901...
4	0.4936...	11	0.1442...	18	0.0848...
5	0.3650...	12	0.1311...	19	0.0801...
6	0.2902...	13	0.1202...	20	0.0759...
7	0.2411...	14	0.1109...	21	0.0721...
8	0.2063...	15	0.1030...	22	0.0686...

TABLE 2. Values of μ_w for $2 \le t \le 22$.

Once the stated mixing condition has been proven, one can apply general central limit theorems for sums of random variables with strong mixing conditions, here specifically a result of Sunklodas [17, Theorem 1]. Putting everything together, we get

Theorem VI. *The total path length (as well as the internal and external path lengths) is asymptotically normal distributed. Its mean is asymptotically $\mu_{tpl}n^2 + O(n)$ and its variance is asymptotically $\sigma^2_{tpl}n^3 + O(n^2)$ with*

$$\mu_{tpl} = \frac{t}{2}\mu_h = \frac{t}{4} + \frac{t^2-2t}{2^{t+4}} + \frac{2t^3+3t^2-8t}{2^{2t+6}} + \frac{9t^4+45t^3+2t^2-88t}{2^{3t+9}} + O\Big(\frac{t^4}{2^{4t}}\Big)$$

and

$$\sigma^2_{tpl} = \frac{t^2}{12} + \frac{-t^4+5t^3+2t^2}{3\cdot 2^{t+4}} + \frac{-4t^5+4t^4+t^3+14t^2}{3\cdot 2^{2t+6}} + O\Big(\frac{t^6}{2^{3t}}\Big)$$

for $t \ge 2$.

7. The Width

In this final section, we consider the width $w(T)$, the maximum number of leaves on the same level, for which we have the following theorem:

Theorem VII. *For a random $T \in \mathcal{T}$ of size n, we have*

$$\mathbb{E}(w(T)) = \mu_w \log n + O(\log\log n),$$

where μ_w is given by

$$\mu_w = \frac{1}{-(t-1)\log q_0} = \frac{1}{t\log(2)} + \frac{1}{t^2\log(2)} + \frac{1}{t^3\log(2)} + \frac{1}{t^4\log(2)} + \frac{1}{t^5\log(2)} + \frac{2}{t^6}\varepsilon_6(t)$$

for $t \ge 23$. For $2 \le t \le 22$, the values of μ_w are given in Table 2.

Furthermore, we have the concentration property

$$\mathbb{P}(|w(T) - \mu_w\log n| \ge 3\mu_w \log\log n) = O\left(\frac{1}{\log n}\right). \tag{7.1}$$

Once again, we only sketch the idea of the proof here, details can be found in Appendix F.

We consider the trees with width bounded by K. The corresponding generating function $W_K(q) = \sum_{\substack{T\in\mathcal{T}\\ w(T)\le K}} q^{n(T)}$ can be constructed by a suitable transfer matrix, and we quantify the obvious convergence of $W_K(q)$ to $H(q,1,1,1)$. The dominant singularity q_K of $W_K(q)$

is estimated by truncating the infinite positive eigenvector of an infinite transfer matrix corresponding to $H(q,1,1,1)$ and applying methods from Perron-Frobenius theory. Then the probability $\mathbb{P}(w(T) \le K)$ can be extracted from $W_K(q)$ using singularity analysis. Our key estimate states that the singularity q_K converges exponentially to q_0, from which the main term of the expectation as well as the concentration property are obtained quite easily. A more precise result on the distribution of the width would depend on a better understanding of the behaviour of q_K as $K \to \infty$, which seems to be quite complicated.

References

1. Edward A. Bender and E. Rodney Canfield, *Locally restricted compositions. I. Restricted adjacent differences*, Electron. J. Combin. **12** (2005), Research Paper 57, 27 pp.
2. ———, *Locally restricted compositions. II. General restrictions and infinite matrices*, Electron. J. Combin. **16** (2009), no. 1, Research Paper 108, 35.
3. ———, *Locally restricted compositions III. Adjacent-part periodic inequalities*, Electron. J. Combin. **17** (2010), no. 1, Research Paper 145, 9.
4. Michael Drmota, *The height of increasing trees*, Ann. Comb. **12** (2009), no. 4, 373–402.
5. ———, *Random trees*, SpringerWienNewYork, Vienna, 2009.
6. Michael Drmota and Bernhard Gittenberger, *On the profile of random trees*, Random Structures Algorithms **10** (1997), no. 4, 421–451.
7. Michael Drmota and Hsien-Kuei Hwang, *Profiles of random trees: correlation and width of random recursive trees and binary search trees*, Adv. in Appl. Probab. **37** (2005), no. 2, 321–341.
8. R. E. Eaves, *A sufficient condition for the convergence of an infinite determinant.*, SIAM J. Appl. Math. **18** (1970), 652–657.
9. Christian Elsholtz, Clemens Heuberger, and Helmut Prodinger, *The number of Huffman codes, compact trees, and sums of unit fractions*, arXiv:1108.5964v1 [math.CO].
10. Philippe Flajolet, Zhicheng Gao, Andrew Odlyzko, and Bruce Richmond, *The distribution of heights of binary trees and other simple trees*, Combin. Probab. Comput. **2** (1993), no. 2, 145–156.
11. Philippe Flajolet and Helmut Prodinger, *Level number sequences for trees*, Discrete Math. **65** (1987), no. 2, 149–156.
12. Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.
13. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics. A foundation for computer science*, second ed., Addison-Wesley, 1994.
14. Hosam M. Mahmoud, *Limiting distributions for path lengths in recursive trees*, Probab. Engrg. Inform. Sci. **5** (1991), no. 1, 53–59.
15. Bruce Richmond and Arnold Knopfmacher, *Compositions with distinct parts*, Aequationes Math. **49** (1995), no. 1–2, 86–97.
16. William A. Stein et al., *Sage Mathematics Software (Version 5.2)*, The Sage Development Team, 2012, `http://www.sagemath.org`.
17. Ĭ. Sunklodas, *The rate of convergence in the central limit theorem for strongly mixing random variables*, Litovsk. Mat. Sb. **24** (1984), no. 2, 174–185.
18. Lajos Takács, *On the total heights of random rooted trees*, J. Appl. Probab. **29** (1992), no. 3, 543–556.
19. ———, *Limit distributions for queues and random rooted trees*, J. Appl. Math. Stochastic Anal. **6** (1993), no. 3, 189–216.

Appendix A. Supplement to Section 2, "The Generating Function"

Proof of Theorem II. As it was already mentioned in Section 2, we first consider

$$H_h(q,u,v) := [w^h]H(q,u,v,w) = \sum_{\substack{T\in\mathcal{T}\\ h(T)=h}} q^{n(T)}u^{m(T)}v^{d(T)}$$

for some $h \geq 0$.

A tree T' of height $h+1$ arises from a tree T of height h by replacing j of its $m(T)$ leaves on the last level by internal vertices with t succeeding leaves respectively, where $1 \leq j \leq m(T)$. If $j = m(T)$, then all old leaves become internal vertices, so that $d(T') = d(T)$; otherwise, at least one of them becomes a new leaf, meaning that we have a new level that contains one or more leaves, hence $d(T') = d(T) + 1$. For the generating function, this translates to the following functional equation:

$$\begin{aligned} H_{h+1}(q,u,v) &= \sum_{\substack{T\in\mathcal{T}\\ h(T)=h}} \left(\sum_{j=1}^{m(T)-1} q^{n(T)+j}u^{jt}v^{d(T)+1} + q^{n(T)+m(T)}u^{m(T)t}v^{d(T)} \right) \\ &= \sum_{\substack{T\in\mathcal{T}\\ h(T)=h}} q^{n(T)}v^{d(T)} \left(qu^t v \frac{1-(qu^t)^{m(T)}}{1-qu^t} + (1-v)(qu^t)^{m(T)} \right) \\ &= r(q,u,v)H_h(q,1,v) + s(q,u,v)H_h(q,qu^t,v), \end{aligned} \tag{A.1}$$

where we set

$$r(q,u,v) = \frac{qu^t v}{1-qu^t}, \qquad s(q,u,v) = \frac{1-v-qu^t}{1-qu^t}.$$

Note that the initial value is given by $H_0(q,u,v) = uv$. Now set

$$\mathcal{D}_0 := \{(q,u,v,w) \in \mathbb{C}^4 \mid |q| < 1/5, |u| \leq 1, |v-1| < 1/5, |w| \leq 1\}.$$

We note that if $(q,u,v,w) \in \mathcal{D}_0$, we have

$$|r(q,u,v)| \leq \frac{3}{10}, \qquad |s(q,u,v)| \leq \frac{1}{2}.$$

This and (A.1) imply that $|H_h(q,u,v)| \leq (4/5)^h$ holds for $h \geq 0$ and $(q,u,v,w) \in \mathcal{D}_0$. This implies that $H(q,u,v,w) = \sum_{h\geq 0} H_h(q,u,v)w^h$ converges uniformly for $(q,u,v,w) \in \mathcal{D}_0$.

Multiplying (A.1) by w^{h+1} and summing over all $h \geq 0$ yields the functional equation

$$H(q,u,v,w) = uv + wr(q,u,v)H(q,1,v,w) + ws(q,u,v)H(q,qu^t,v,w). \tag{A.2}$$

We iterate this functional equation and obtain

$$\begin{aligned} H(q,u,v,w) = a_k(q,u,v,w) &+ b_k(q,u,v,w)H(q,1,v,w) \\ &+ c_k(q,u,v,w)H(q,q^{[k+1]}u^{t^{k+1}},v,w) \end{aligned} \tag{A.3}$$

for $k \geq 0$ with

$$a_k(q,u,v,w) = v\sum_{j=0}^{k} q^{[j]}u^{t^j}w^j \prod_{i=0}^{j-1} s(q,q^{[i]}u^{t^i},v),$$

$$b_k(q,u,v,w) = \sum_{j=0}^{k} r(q, q^{[j]}u^{t^j}, v) w^{j+1} \prod_{i=0}^{j-1} s(q, q^{[i]}u^{t^i}, v),$$

$$c_k(q,u,v,w) = w^{k+1} \prod_{i=0}^{k} s(q, q^{[i]}u^{t^i}, v).$$

Let now

$$\mathcal{D} = \{(q,u,v,w) \in \mathbb{C}^4 \mid |q| < |u|^{1-t}, |w| \cdot |1-v| < 1\}.$$

For $(q,u,v,w) \in \mathcal{D}$, we have $\lim_{k\to\infty} q^{[k]}u^{t^k} = 0$ and $\lim_{k\to\infty} ws(q, q^{[k]}u^{t^k}, v) = |w| \cdot |v-1| < 1$ and the limits

$$a(q,u,v,w) := \lim_{k\to\infty} a_k(q,u,v,w) = v \sum_{j=0}^{\infty} q^{[j]}u^{t^j} w^j \prod_{i=0}^{j-1} s(q, q^{[i]}u^{t^i}, v),$$

$$b(q,u,v,w) := \lim_{k\to\infty} b_k(q,u,v,w) = \sum_{j=0}^{\infty} r(q, q^{[j]}u^{t^j}, v) w^{j+1} \prod_{i=0}^{j-1} s(q, q^{[i]}u^{t^i}, v)$$

exist. As $\lim_{k\to\infty} c_k(q,u,v,w) = 0$ for these $(q,u,v,w) \in \mathcal{D}$, the limit of (A.3) for $k \to \infty$ is

$$H(q,u,v,w) = a(q,u,v,w) + b(q,u,v,w)H(q,1,v,w). \tag{A.4}$$

Setting $u = 1$ in (A.4) yields (2.1). □

We also state a simplified expression and a functional equation for $b(q,u,v,w)$ in the case $v = 1$, $w = 1$:

Lemma A.1. *We have*

$$b(q,u,1,1) = \sum_{j=1}^{\infty} (-1)^{j-1} \prod_{i=1}^{j} \frac{q^{[i]}u^{t^i}}{1 - q^{[i]}u^{t^i}} = \frac{qu^t}{1-qu^t}(1 - b(q, qu^t, 1, 1)).$$

In particular, the coefficient $[u^j]b(q,u,1,1)$ *vanishes if* j *is not a multiple of* t.

Proof of Lemma A.1. This is an immediate consequence of (2.2). □

Appendix B. Supplement to Section 3, "The Height"

This section is, as the title reveals, a supplement to our discussion of the height. It contains numerically calculated values for the constants of Theorem III and the proof of this theorem. We start with the latter. Note that a brief sketch of the proof was already given in Section 3.

Proof of Theorem III. Throughout this proof the notations of Section 3 are used. Further, we make use of Theorem IX.9 of Flajolet and Sedgewick [12] and apply that theorem to the function $H(q,1,1,w)$.

Recall the notation $D(q,w)$ as the denominator of $H(q,1,1,w)$ and let q_0 be the zero of $D(q,1)$ according to Lemma 2.1. Set

$$c_{ij} = \left.\frac{\partial^{i+j}}{\partial q^i \partial w^j} D(q,w)\right|_{q = q_0,\ w = 1}.$$

Then the expectation of $h(T)$ is asymptotically normally distributed and we can obtain the mean $\mu_h n + O(1)$ with

$$\mu_h = \frac{c_{01}}{c_{10} q_0},$$

and the variance to $\sigma_h^2 n + O(1)$ with

$$\sigma_h^2 = \frac{c_{01}^2 c_{20} q_0 + c_{01} c_{10}^2 q_0 - 2\, c_{01} c_{10} c_{11} q_0 + c_{02} c_{10}^2 q_0 + c_{01}^2 c_{10}}{c_{10}^3 q_0^2}.$$

To calculate the coefficients c_{ij} we need derivatives of $D(q,w)$. In order to avoid working with infinite sums, we use the approximations

$$D_K(q,w) := \sum_{0 \le k < K} (-1)^k w^k \prod_{j=1}^{k} \frac{q^{[j]}}{1 - q^{[j]}}.$$

Lemma B.1 shows that the error made by using those approximations is small. For the calculations themselves, Sage [16] was used. □

Lemma B.1. *Let* $i \in \{0,1,2\}$ *and* $j \in \mathbb{N}_0$*, and let* $q \in \mathbb{C}$ *with* $1/2 \le |q| \le 1/r_3$*, where* $r_3 = 1 + \frac{\log 2}{t} - \frac{\log 2 - \log^2 2}{2t^2}$*. Then*

$$\frac{\partial^{i+j}}{\partial q^i \partial w^j} \bigl(D(q,w) - D_4(q,w)\bigr)\Big|_{w=1} = O\left(\frac{1}{2^{t^2}}\right).$$

Proof. The result was shown for $i \in \{0,1\}$ and $j = 0$ in [9]. Here we follow the proof of Lemma 9 of that article. We first note that it is sufficient to show the result for $j = 0$, since that derivative results in a polynomial in k, which is asymptotically smaller than the factor t^k which appears.

Now set

$$f_j(q) := \frac{q^{[j]}}{1 - q^{[j]}}.$$

We obtain

$$\frac{\partial}{\partial q}\left(\prod_{j=1}^{k} f_j(q)\right) = \frac{1}{q} \prod_{j=1}^{k} f_j(q) \left(\sum_{j=1}^{k} \frac{[j]}{1 - q^{[j]}}\right)$$

for its first derivative and

$$\frac{\partial^2}{\partial q^2}\left(\prod_{j=1}^{k} f_j(q)\right) = \frac{1}{q^2} \prod_{j=1}^{k} f_j(q) \left(\left(\sum_{j=1}^{k} \frac{[j]}{1 - q^{[j]}}\right)^2 - \sum_{j=1}^{k} \frac{[j]}{1 - q^{[j]}} + \sum_{j=1}^{k} \frac{[j]^2 q^{[j]}}{\bigl(1 - q^{[j]}\bigr)^2}\right)$$

for its second. As in [9], we can find the bounds

$$\left|\prod_{j=1}^{k} f_j(1/z)\right| \le \frac{t}{2^{-1 + t(k-1)/2 + (k-3)t^2}}$$

and

$$\left|\sum_{j=1}^{k} \frac{[j]}{1 - (1/z)^{[j]}}\right| \le 4kt^k.$$

Therefore, we also deduce that

$$\left| \sum_{j=1}^{k} \frac{[j]^2 q^{[j]}}{\left(1-q^{[j]}\right)^2} \right| \leq \left(\sum_{j=1}^{k} \frac{[j]}{1-(1/|z|)^{[j]}} \right)^2 \leq (4kt^k)^2.$$

This yields the bound

$$\left| \frac{\partial^{i+j}}{\partial q^i \partial w^j} \big(D(q,w) - D_4(q,w) \big) \Big|_{w=1} \right| \leq |z|^2 \sum_{k=4}^{\infty} \frac{t}{2^{(k-3)t^2+(k-1)t/2-1}} \left(2(4kt^k)^2 + 4kt^k \right)$$
$$\leq \sum_{k=4}^{\infty} \frac{k^2 t^{2k+1}}{2^{(k-3)t^2+(k-1)t/2-9}} \leq \sum_{k=4}^{\infty} \frac{c}{2^{(k-3)t^2}}$$

for some positive constant c. Since the last sum in the previous inequality is $O(2^{-t^2})$, the result follows. □

The end of this section contains the following: For $t \leq 30$ we calculated the constants of Theorem III numerically. The computer algebra software Sage [16] was used for this purpose. The results can be found in Table 3.

t	μ_h	σ_h^2
2	0.5662757699172865	0.2665499010273937
3	0.5330981433252730	0.2636253024859229
4	0.5216132420088969	0.2465916388296734
5	0.5137644953351326	0.2404182925457133
6	0.5084950082063058	0.2396633993739495
7	0.5051047365215813	0.2411570855092153
8	0.5030001253275541	0.2432575483836213
9	0.5017308605343554	0.2452173961787763
10	0.5009832278618641	0.2467757623911674
11	0.500551313637743	0.2479077234990245
12	0.5003057656286383	0.2486821135530906
13	0.5001680187030247	0.2491894707701658
14	0.5000916023570357	0.2495111461587043
15	0.5000496052425100	0.2497099052572736
16	0.5000267068978588	0.2498301915991255
17	0.5000143062444377	0.2499017551259219
18	0.5000076297101404	0.2499437283128117
19	0.5000040532034994	0.2499680504612380
20	0.5000021457914275	0.2499819989727347
21	0.5000011324949086	0.2499899266916567
22	0.500000596048271	0.2499943971277963
23	0.5000003129248821	0.2499969005482699
24	0.5000001639129082	0.2499982938141369
25	0.500000085681714	0.2499990649513116
26	0.5000000447034934	0.2499994896349970
27	0.5000000232830670	0.2499997224658077
28	0.5000000121071942	0.2499998495913860
29	0.500000006286428	0.2499999187421003
30	0.5000000032596291	0.2499999562278376

TABLE 3. Numerical values of the constants in mean and variance of the height for small values of t, cf. Theorem III. It would be possible to calculate the values with even higher accuracy.

APPENDIX C. SUPPLEMENT TO SECTION 4, "THE NUMBER OF DISTINCT DEPTHS OF LEAVES"

Similar to the previous supplementary section, this section contains explicitly calculated values for the constants of Theorem IV and a detailed proof of this theorem, following the proof sketch that was given in Section 4. We start with the latter. The ideas used are very similar to the ones in the analysis of the height.

Proof of Theorem IV. Throughout this proof the notations of Section 4 are used. Again, as with the heights, we make use of Theorem IX.9 of Flajolet and Sedgewick [12] and apply that theorem to the function $H(q,1,v,1)$.

Again, we use the notation $D(q,v)$ for the denominator of $H(q,1,v,1)$ and let q_0 be the zero of $D(q,1)$ according to Lemma 2.1. We expand $D(q,v)$ around $(q_0,1)$ and can then calculate the main term of mean and variance from the coefficients of that series. The required formulas can be found in the proof of Theorem III in Appendix B.

Again, to calculate the coefficients we need derivatives of $D(q,v)$ and we use the approximations

$$D_K(q,v) := 1 - \sum_{1\le k<K} \frac{v}{1-q^{[k]}} \prod_{j=1}^{k-1} \frac{1-v-q^{[j]}}{1-q^{[j]}}.$$

Lemma C.1 shows that the error in this approximation is small. Again, for the calculations themselves, Sage [16] was used. □

Lemma C.1. *Let $i,j\in\{0,1,2\}$, and let $q\in\mathbb{C}$ with $1/2\le|q|\le 1/r_3$, where $r_3 = 1+\frac{\log 2}{t} - \frac{\log 2-\log^2 2}{2t^2}$. Then*

$$\frac{\partial^{i+j}}{\partial q^i \partial v^j}\bigl(D(q,v)-D_4(q,v)\bigr)\bigg|_{v=1} = O\left(\frac{1}{2^{t^2}}\right).$$

Proof. The proof is similar to the proof of Lemma B.1. □

The end of this section contains numerically calculated values for the constants of Theorem III. We used the computer algebra software Sage [16], and the results can be found in Table 4.

Appendix D. Supplement to Section 5, "The Number of Leaves on the Last Level"

Proof of Theorem V. Let $q_1 = 1-\frac{0.72}{t}$. Then singularity analysis shows that the probability generating function $p_n(u)$ of $m(T)$ is given by

$$p_n(u) = b(q_0,u,1,1) + O(Q^n),$$

uniformly for $|u|\le 1$.

The limiting distribution follows from [12, Theorem IX.2]. Expectation and variance follow upon differentiating $b(q_0,u,1,1)$ with respect to u and inserting the asymptotic expression for q_0. □

Appendix E. Supplement to Section 6, "The Path Length"

Here we provide some more details of our analysis of the total (internal, external) path length, starting with the generating functions. Recall that we defined the generating function $L_r(q,u,w)$ for the r-th moment of the total path length:

$$L_r(q,u,w) = \sum_{T\in\mathcal{T}} \ell(T)^r q^{n(T)} u^{m(T)} w^{h(T)}.$$

t	μ_d	σ_d^2
2	0.4042366935349558	0.2491723144610512
3	0.4868358747318154	0.2900504810033636
4	0.5024585834463688	0.2741245386044700
5	0.5050331954313614	0.2607084552774208
6	0.5043408269340329	0.2530808413030350
7	0.5030838633817897	0.2495578056054625
8	0.5020050053196333	0.2483362931739360
9	0.5012375070905983	0.2482103208441572
10	0.5007377066674932	0.2485046286268309
11	0.5004288693844008	0.2488904008073738
12	0.5002446296853791	0.2492332759318571
13	0.5001374740872935	0.2494951950687874
14	0.5000763363460676	0.2496791536316180
15	0.5000419739265400	0.2498015045792620
16	0.5000228916911940	0.2498797960254888
17	0.500012398761189	0.2499284618053178
18	0.500006676000353	0.2499580344990146
19	0.5000035763570187	0.2499756801559131
20	0.5000019073704041	0.2499860521721408
21	0.5000010132849795	0.2499920724820041
22	0.5000005364434586	0.2499955296224207
23	0.500000283122517	0.2499974965964656
24	0.5000001490117357	0.2499986067389993
25	0.500000078231130	0.2499992288642147
26	0.5000000409782024	0.2499995753167091
27	0.5000000214204217	0.2499997671693008
28	0.5000000111758715	0.2499998728744530
29	0.5000000058207663	0.2499999308492945
30	0.5000000030267984	0.2499999625142652

TABLE 4. Values of the constants in mean and variance of the number of distinct depths of leaves for small values of t, cf. Theorem IV. It would be possible to calculate the values with even higher accuracy.

In particular, $L_0(q, u, w)$ is the ordinary generating function for all trees, where u marks the number of leaves on the highest level and w the height. From the recursive characterisation of canonical trees, we got the identity

$$L_0(q, u, w) = u + \frac{qu^t}{1 - qu^t} \left(L_0(q, 1, w) - L_0(q, qu^t, w) \right),$$

from which we obtained, by means of iteration, an explicit formula for L_0, namely

$$L_0(q, u, w) = a_0(q, u, w) + b(q, u, w) L_0(q, 1, w)$$

and in particular,

$$L_0(q, 1, w) = \frac{a_0(q, 1, w)}{1 - b(q, 1, w)},$$

where

$$a_0(q, u, w) = \sum_{j=0}^{\infty} (-1)^j w^j q^{[j]} u^{t^j} \prod_{i=1}^{j} \frac{q^{[i]} u^{t^i}}{1 - q^{[i]} u^{t^i}}$$

and

$$b(q, u, w) = \sum_{j=1}^{\infty} (-1)^{j-1} w^j \prod_{i=1}^{j} \frac{q^{[i]} u^{t^i}}{1 - q^{[i]} u^{t^i}}.$$

Likewise, the functional equations one obtains for L_1 and L_2 can be solved by means of iteration: one has

$$L_1(q,u,w) = \Phi_u\Phi_w L_0(q,u,w) + \frac{qu^t}{1-qu^t}\left(L_1(q,1,w) - L_1(q,qu^t,w)\right),$$

and thus

$$L_1(q,u,w) = a_1(q,u,w) + b(q,u,w)L_1(q,1,w),$$

and in particular

$$L_1(q,1,w) = \frac{a_1(q,1,w)}{1-b(q,1,w)}$$

with

$$a_1(q,u,w) = \sum_{j=0}^{\infty}(-1)^j w^j (\Phi_u\Phi_w L_0)(q, q^{[j]}u^{t^j}, w)\prod_{i=1}^{j}\frac{q^{[i]}u^{t^i}}{1-q^{[i]}u^{t^i}}.$$

Finally,

$$L_2(q,u,w) = 2\Phi_u\Phi_w L_1(q,u,w) - \Phi_u^2\Phi_w^2 L_0(q,u,w) + \frac{qu^t}{1-qu^t}\left(L_2(q,1,w) - L_2(q,qu^t,w)\right),$$

and thus

$$L_2(q,u,w) = a_2(q,u,w) + b(q,u,w)L_2(q,1,w),$$

and in particular

$$L_2(q,1,w) = \frac{a_2(q,1,w)}{1-b(q,1,w)}$$

with

$$a_2(q,u,w) = \sum_{j=0}^{\infty}(-1)^j w^j \Big(2(\Phi_u\Phi_w L_1)(q, q^{[j]}u^{t^j}, w) - (\Phi_u^2\Phi_w^2 L_0)(q, q^{[j]}u^{t^j}, w)\Big)\prod_{i=1}^{j}\frac{q^{[i]}u^{t^i}}{1-q^{[i]}u^{t^i}}.$$

Substituting back, we get an explicit expression for $L_1(q,1,w)$:

$$\begin{aligned} L_1(q,1,w) = {} & \frac{a_0(q,1,w)(\Phi_w b)(q,1,w)}{(1-b(q,1,w))^3}\sum_{j=0}^{\infty}(-1)^j w^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[j]},w) \\ & + \frac{a_0(q,1,w)}{(1-b(q,1,w))^2}\sum_{j=0}^{\infty}(-1)^j w^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u\Phi_w b)(q,q^{[j]},w) \\ & + \frac{(\Phi_w a_0)(q,1,w)}{(1-b(q,1,w))^2}\sum_{j=0}^{\infty}(-1)^j w^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[j]},w) \\ & + \frac{1}{1-b(q,1,w)}\sum_{j=0}^{\infty}(-1)^j w^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u\Phi_w a_0)(q,q^{[j]},w). \end{aligned}$$

The dominant term in this sum is the first one, with a triple pole at the dominant singularity q_0. The second and third term, however, are also relevant in the calculation of the variance, where one further term in the asymptotic expansion is needed in view of the inevitable cancellation in the main term. Singularity analysis immediately yields the asymptotic behaviour of the mean: since the pole is of cubic order, the order of the mean is quadratic, i.e., it is asymptotically equal to $\mu_{tpl}n^2$, where the constant μ_{tpl} is given by

$$\mu_{tpl} = \frac{(\Phi_w b)(q_0,1,1)}{2(\Phi_q b)(q_0,1,1)^2}\sum_{j=0}^{\infty}(-1)^j(\Phi_u b)(q_0,q_0^{[j]},1)\prod_{i=1}^{j}\frac{q_0^{[i]}}{1-q_0^{[i]}}.$$

Plugging in the definition of b as a sum, it is possible to simplify this further: one has

$$(\Phi_u b)(q,u,1) = \sum_{k=1}^{\infty}(-1)^{k-1}\prod_{h=1}^{k}\frac{q^{[h]}u^{t^h}}{1-q^{[h]}u^{t^h}}\sum_{h=1}^{k}\frac{t^h}{1-q^{[h]}u^{t^h}}$$

by logarithmic differentiation and thus

$$\begin{aligned}(\Phi_u b)(q,q^{[j]},1) &= \sum_{k=1}^{\infty}(-1)^{k-1}\prod_{h=1}^{k}\frac{q^{[h]+t^h[j]}}{1-q^{[h]+t^h[j]}}\sum_{h=1}^{k}\frac{t^h}{1-q^{[h]+t^h[j]}}\\ &= \sum_{k=1}^{\infty}(-1)^{k-1}\prod_{i=j+1}^{j+k}\frac{q^{[i]}}{1-q^{[i]}}\sum_{h=1}^{k}\frac{t^h}{1-q^{[h+j]}}\end{aligned}$$

since $[h]+t^h[j]=[h+j]$ by definition. Plugging in, we find

$$\mu_{tpl} = \frac{(\Phi_w b)(q_0,1,1)}{2(\Phi_q b)(q_0,1,1)^2}\sum_{j=0}^{\infty}\sum_{k=1}^{\infty}(-1)^{j+k-1}\prod_{i=1}^{j+k}\frac{q_0^{[i]}}{1-q_0^{[i]}}\sum_{h=1}^{k}\frac{t^h}{1-q_0^{[h+j]}}.$$

Substituting $\ell = j+k$ and interchanging the order of summation, we arrive at

$$\begin{aligned}\mu_{tpl} &= \frac{(\Phi_w b)(q_0,1,1)}{2(\Phi_q b)(q_0,1,1)^2}\sum_{\ell=1}^{\infty}(-1)^{\ell-1}\prod_{i=1}^{\ell}\frac{q_0^{[i]}}{1-q_0^{[i]}}\sum_{k=1}^{\ell}\sum_{h=1}^{k}\frac{t^h}{1-q_0^{[h+\ell-k]}}\\ &= \frac{(\Phi_w b)(q_0,1,1)}{2(\Phi_q b)(q_0,1,1)^2}\sum_{\ell=1}^{\infty}(-1)^{\ell-1}\prod_{i=1}^{\ell}\frac{q_0^{[i]}}{1-q_0^{[i]}}\sum_{r=1}^{\ell}\sum_{h=1}^{r}\frac{t^h}{1-q_0^{[r]}}\\ &= \frac{(\Phi_w b)(q_0,1,1)}{2(\Phi_q b)(q_0,1,1)^2}\sum_{\ell=1}^{\infty}(-1)^{\ell-1}\prod_{i=1}^{\ell}\frac{q_0^{[i]}}{1-q_0^{[i]}}\sum_{r=1}^{\ell}\frac{t[r]}{1-q_0^{[r]}}.\end{aligned}$$

Noting now that

$$(\Phi_q b)(q,1,1) = \sum_{\ell=1}^{\infty}(-1)^{\ell-1}\prod_{i=1}^{\ell}\frac{q^{[i]}}{1-q^{[i]}}\sum_{r=1}^{\ell}\frac{[r]}{1-q^{[r]}},$$

which can be seen by another logarithmic differentiation, we can replace the sum in the expression for μ_{tpl} above by $t\cdot(\Phi_q b)(q_0,1,1)$, which finally yields

$$\mu_{tpl} = \frac{t}{2}\cdot\frac{(\Phi_w b)(q_0,1,1)}{(\Phi_q b)(q_0,1,1)},$$

and the fraction is precisely μ_h since the generating function of the mean height is

$$\frac{a_0(q,1,1)(\Phi_w b)(q,1,1)}{(1-b(q,1,1))^2}+\frac{(\Phi_w a_0)(q,1,1)}{1-b(q,1,1)},$$

of which the first term dominates (yet another application of singularity analysis). This means that we have proven the identity $\mu_{tpl} = t\mu_h/2$.

For the variance, one also needs the asymptotic behaviour of $L_2(q,1,1)$ at the dominant singularity. Only the terms of pole order 4 and 5 (i.e., highest and second-highest) are needed: they are

$$\begin{aligned}
L_2(q,1,1) &= \frac{6a_0(q,1,1)(\Phi_w b)(q,1,1)^2}{(1-b(q,1,1))^5}\left(\sum_{j=0}^{\infty}(-1)^j w^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[j]},1)\right)^2 \\
&+ \frac{4a_0(q,1,1)(\Phi_w b)(q,1,1)^2}{(1-b(q,1,1))^4}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}\left([j+1](\Phi_u^2 b)(q,q^{[j]},1)+\sum_{r=1}^{j}\frac{t[r]}{1-q^{[r]}}(\Phi_u b)(q,q^{[j]},1)\right) \\
&+ \frac{8a_0(q,1,1)(\Phi_w b)(q,1,1)}{(1-b(q,1,1))^4}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[j]},1)\sum_{k=0}^{\infty}(-1)^k \prod_{i=1}^{k}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u\Phi_w b)(q,q^{[k]},1) \\
&+ \frac{6(\Phi_w a_0)(q,1,1)(\Phi_w b)(q,1,1)}{(1-b(q,1,1))^4}\left(\sum_{j=0}^{\infty}(-1)^j w^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[j]},1)\right)^2 \\
&+ \frac{2a_0(q,1,1)(\Phi_w b)(q,1,1)^2}{(1-b(q,1,1))^4}\left(\sum_{j=0}^{\infty}(-1)^j w^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[j]},1)\right)^2 \\
&+ \frac{2a_0(q,1,1)(\Phi_w b)(q,1,1)}{(1-b(q,1,1))^4}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[j]},1)\sum_{k=1}^{\infty}(-1)^k k \prod_{i=1}^{k}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u b)(q,q^{[k]},1) \\
&- \frac{2a_0(q,1,1)(\Phi_w b)(q,1,1)^2}{(1-b(q,1,1))^4}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q^{[i]}}{1-q^{[i]}}(\Phi_u^2 b)(q,q^{[j]},1)
\end{aligned}$$

Applying singularity analysis to the highest- and second-highest order terms of both L_1 and L_2 yields the variance: the terms of order n^4 cancel (as one would expect), and one finds that the variance is asymptotically $\sigma_{tpl}^2 n^3$, where

$$\begin{aligned}
\sigma_{tpl}^2 &= \frac{F(q_0)^2(\Phi_q^2 b)(q_0,1,1)}{(\Phi_q b)(q_0,1,1)^5} - \frac{F(q_0)(\Phi_q F)(q_0)}{(\Phi_q b)(q_0,1,1)^4} \\
&- \frac{(\Phi_w b)(q_0,1,1)^2}{3(\Phi_q b)(q_0,1,1)^3}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q_0^{[i]}}{1-q_0^{[i]}}(\Phi_u^2 b)(q_0,q_0^{[j]},1) \\
&+ \frac{2(\Phi_w b)(q_0,1,1)^2}{3(\Phi_q b)(q_0,1,1)^3}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q_0^{[i]}}{1-q_0^{[i]}}\left([j+1](\Phi_u^2 b)(q_0,q_0^{[j]},1)+\sum_{r=1}^{j}\frac{t[r]}{1-q^{[r]}}(\Phi_u b)(q_0,q_0^{[j]},1)\right) \\
&+ \frac{(\Phi_w b)(q_0,1,1)}{3(\Phi_q b)(q_0,1,1)^3}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q_0^{[i]}}{1-q_0^{[i]}}(\Phi_u b)(q_0,q_0^{[j]},1)\sum_{k=0}^{\infty}(-1)^k \prod_{i=1}^{k}\frac{q_0^{[i]}}{1-q_0^{[i]}}(\Phi_u\Phi_w b)(q_0,q_0^{[k]},1) \\
&+ \frac{(\Phi_w b)(q_0,1,1)^2}{3(\Phi_q b)(q_0,1,1)^3}\left(\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q_0^{[i]}}{1-q_0^{[i]}}(\Phi_u b)(q_0,q_0^{[j]},1)\right)^2 \\
&+ \frac{(\Phi_w b)(q_0,1,1)}{3(\Phi_q b)(q_0,1,1)^3}\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j}\frac{q_0^{[i]}}{1-q_0^{[i]}}(\Phi_u b)(q_0,q_0^{[j]},1)\sum_{k=1}^{\infty}(-1)^k k \prod_{i=1}^{k}\frac{q_0^{[i]}}{1-q_0^{[i]}}(\Phi_u b)(q_0,q_0^{[k]},1)
\end{aligned}$$

and the function $F(q)$ is given by

$$F(q) = (\Phi_w b)(q,1,1) \sum_{j=0}^{\infty} (-1)^j \prod_{i=1}^{j} \frac{q^{[i]}}{1-q^{[i]}} (\Phi_u b)(q, q^{[j]}, 1).$$

We determined numerical values of these constants as in the previous sections, they are given in Table 5.

t	μ_{tpl}	σ^2_{tpl}
2	0.5746406730225036	0.636553899565319
3	0.7996893802701904	0.9538514746097371
4	1.043226570739454	1.424940599745666
5	1.284411238386164	2.078739994014109
6	1.525485024618925	2.926628748193911
7	1.767866577825535	3.972171302166417
8	2.012000501310217	5.210807673614956
9	2.257788872404600	6.634216448921346
10	2.504916139309320	8.23405080979501
11	2.753032225007583	10.00388584911538
12	3.001834593771830	11.93967669304990
13	3.251092121569661	14.03939441023803
14	3.500641216499250	16.30239232264572
15	3.750372039318825	18.72881526046276
16	4.000213655182871	21.31916858572890
17	4.250121603077721	24.07405283275217
18	4.500068667391264	26.99402565883372
19	4.750038505433244	30.07954611947160
20	5.000021457914275	33.33096498586472
21	5.250011891196540	36.74853674754146
22	5.500006556530974	40.33243901952973
23	5.750003598636143	44.08279205182939
24	6.000001966954898	47.99967527333957
25	6.250001071021418	52.08314008372820
26	6.500000581145415	56.33321917051825
27	6.750000314321405	60.74993301181408
28	7.000000169500719	65.33329426993452
29	7.250000091153200	70.08331068457584
30	7.500000048894436	74.99998693820047

TABLE 5. Values of the constants in mean and variance of the total path length t, cf. Theorem VI. It would be possible to calculate the values with even higher accuracy.

Finally, let us describe in some more detail how the central limit theorem is obtained. Recall that $X_{k,n}$ is the (random) depth of the k-th vertex, and that $Y_{k,n} = X_{k+1,n} - X_{k,n}$. The internal path length is given by

$$\ell_{internal}(T) = \sum_{j=1}^{n-1} (n-j) Y_{j,n},$$

and thus

$$n^{-1} \ell_{internal}(T) = \sum_{j=1}^{n-1} \frac{n-j}{j} Y_{j,n}.$$

Setting $Z_{j,n} = \frac{n-j}{n} Y_{j,n}$, we obtain a decomposition for the random variable $n^{-1}\ell_{internal}(T)$:

$$n^{-1} \ell_{internal}(T) = \sum_{j=1}^{n-1} Z_{j,n}.$$

The point behind this rescaling is that the $Z_{j,n}$ are bounded now, so that they have bounded third absolute moments (and generally bounded moments of any order), which is one of the conditions to make Theorem 1 of Sunklodas [17] applicable. Another condition is that the variance of the sum grows at least linearly, which is satisfied in view of our considerations above (the variance of $\ell(T)$ is of cubic order, so the variance of the rescaled random variable is still of linear order). In Sunklodas' paper, the variables are also assumed to have expectance 0, which we could of course achieve by subtracting the mean from each $Z_{j,n}$.

The main criterion is the strong mixing inequality that was already mentioned in Section 6. Let two events $A \in \mathcal{F}_{s_1}$ in the σ-algebra generated by $Z_{1,n}, \dots, Z_{s_1,n}$ and $B \in \mathcal{G}_{s_2}$ in the σ-algebra generated by $Z_{s_2,n}, Z_{s_2+1,n}, \dots, Z_{n-1,n}$ be given. The event A consists of a collection of possible shapes of the random tree T up to the s_1-th vertex v_{s_1}, and likewise B consists of a collection of possible shapes of the random tree T from the s_2-th vertex v_{s_2} onwards.

Let H_0 be the number of vertices with label $> s_1$ on the same level as v_{s_1}, and let H_1 be the number of vertices on the following level. For any possible shape that is allowed in the event A, there is only a limited number of possibilities for H_0 and H_1. Likewise, we define K_0 to be the number of vertices on the same level as v_{s_2}, but with lower label, and K_1 the number of vertices on the previous level. The part between the levels of v_{s_1} and v_{s_2} (excluding the levels on which these two vertices are located) can be regarded as a canonical *forest*, which is defined like a canonical tree, but with H_1 different roots and K_1 different leaves on the last level.

It is not complicated to modify our generating functions approach that we used to obtain Theorem V to the case of several roots. Let the generating function for this purpose be $H_h(q,u)$, where h is the number of roots, q marks the size and u the number of leaves on the last level. Then it follows that

$$H_h(q,u) = a_h(q,u) + \frac{a_h(q,1)b(q,u)}{1-b(q,1)},$$

where

$$a_h(q,u) = \sum_{j=0}^{\infty} (-1)^j q^{h[j]} u^{ht^j} \prod_{i=1}^{j} \frac{q^{[i]}u^{t^i}}{1-q^{[i]}u^{t^i}},$$

$$b(q,u) = \sum_{j=1}^{\infty} (-1)^{j-1} \prod_{i=1}^{j} \frac{q^{[i]}u^{t^i}}{1-q^{[i]}u^{t^i}}.$$

The number of canonical t-ary forests with h roots, k leaves on the last level and r internal vertices is $[q^r u^k]H_h(q,u)$. Singularity analysis yields a distributional result analogous to Theorem V, with an error term that is even uniform in h (note that $a_h(q,u)$ is bounded as a function of h in the relevant region!), but unfortunately not in k: one has

$$\frac{[q^r u^{mt}]H_h(q,u)}{[q^r]H_h(q,1)} = p_m(1 + O(Q_1^{-m} Q_2^r)),$$

where p_m is defined as in Theorem V and $0 < Q_1, Q_2 < 1$. However, p_m decreases exponentially in m as well, which we can use to our advantage: it is also true that

$$\frac{[q^r u^{mt}]H_h(q,u)}{[q^r]H_h(q,1)} = O(Q_3^m)$$

for some real number $0 < Q_3 < 1$.

Note that $[q^r u^k]H_h(q,u)$ gives the number of ways to fill a "gap" of r vertices, starting with h roots and ending with k leaves. This can be applied to the part of our tree T between the vertices v_{s_1} and v_{s_2}, the part between the root v_1 and v_{s_2} (where we just set $h = 1$) as well as the part from v_{s_1} to v_n (where we can sum over all k, which amounts to taking $[q^r]H_h(q,1)$).

The estimate above implies the following: the event that K_1, the number of vertices on the level before v_{s_2}, is greater than Mt, has probability $O(Q_3^{\delta(s_2-s_1)})$ if $M = \delta(s_2 - s_1)$ for some suitably chosen δ. Conditioned on the event that this is not the case, however, the difference of the probability of $A \cap B$ and the product of the probabilities of A and B is small:

$$\mathbb{P}(A \cap B | K_1 \le Mt) = \mathbb{P}(A|K_1 \le Mt)\mathbb{P}(B|K_1 \le Mt)\Big(1 + O(Q_1^{-\delta(s_2-s_1)}Q_2^{s_2-s_1})\Big).$$

Combining the two, we arrive at

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = O\Big(Q_1^{-\delta(s_2-s_1)}Q_2^{s_2-s_1} + Q_3^{\delta(s_2-s_1)}\Big),$$

and if δ is chosen sufficiently small, but fixed, then both terms decrease exponentially in $s_2 - s_1$, proving the strong mixing condition and thus the central limit theorem.

Appendix F. Supplement to Section 7, "The Width"

This appendix is devoted to the proof of Theorem VII.

Apart from the width $w(T)$, we also need the "inner width" $w^*(T)$ defined to be

$$w^*(T) := \max_{0 \le k < h(T)} L_T(k)$$

for a recursive construction, where $L_T(k)$ denotes the number of leaves at level k. By definition, the inner width $w^*(T)$ does not take the leaves on the last level into account.

For $K > 0$, we are interested in the generating function

$$W_K(q) := \sum_{\substack{T \in \mathcal{T} \\ w(T) \le K}} q^{n(T)}.$$

We represent $W_K(q)$ in terms of the generating functions

$$W_{K,r} := \sum_{\substack{T \in \mathcal{T} \\ w^*(T) \le K \\ m(T) = tr}} q^{n(T)}$$

for $r \ge 0$ such that

$$W_K(q) = 1 + \sum_{r=1}^{\lfloor K/t \rfloor} W_{K,r}.$$

Here, the summand 1 corresponds to the tree of order 1. For all other trees, the number $m(T)$ of leaves on the last level is clearly a multiple of t.

In order to compute $W_{K,r}$ recursively, we will do so for $1 \le r \le N(K)$ with $N(K) := \lceil K/(t-1) \rceil - 1$. Thus we consider the column vector

$$\mathbf{W}_K(q) := (W_{K,1}(q), \ldots, W_{K,N(K)})^T.$$

We consider the "transfer matrix"

$$M_K(q) := \left(q^r \left[\frac{r}{t} \le s \le \frac{r+K}{t}\right]\right)_{\substack{1\le r\le N(K)\\ 1\le s\le N(K)}},$$

where the Iversonian notation

$$[expr] = \begin{cases} 1 & \text{if } expr \text{ is true,}\\ 0 & \text{if } expr \text{ is false}\end{cases}$$

popularised by Graham, Knuth and Patashnik [13] has been used.

We now express $\mathbf{W}_K(q)$ in terms of $M_K(q)$:

Lemma F.1. *For $K \ge t$, we have*

$$\mathbf{W}_K(q) = (I - M_K(q))^{-1} \begin{pmatrix} q\\ 0\\ \vdots\\ 0\end{pmatrix}. \tag{F.1}$$

Proof. As in the proof of Theorem II, a tree T' of height $h+1 \ge 2$, inner width at most K and $m(T') = rt$ arises from a tree T of height h, inner width at most K and $m(T) = st$ by replacing r of the st leaves of T on the last level by inner vertices with t succeeding leaves each. We obviously have $r \le st$. In order to ensure that $w^*(T') \le K$, we have to ensure that $st - r \le K$. We rewrite these two inequalities as

$$\frac{r}{t} \le s \le \frac{r+K}{t}. \tag{F.2}$$

If we have $r \le N(K)$, we have $r < K/(t-1)$ and therefore $s < K/(t-1)$ by (F.2), i.e., $s \le N(K)$. This justifies our choice of $N(k)$. This construction yields s new inner vertices in T'.

There is only one tree T' of height < 2, inner width at most K and $m(T') = rt$, namely the star of order $t+1$ for $r = 1$ which has one internal vertex (the root).

Translating these considerations into the language of generating functions yields

$$W_{K,r}(q) = q[r=1] + \sum_{s=1}^{N(k)} q^r \left[\frac{r}{t} \le s \le \frac{r+K}{t}\right] W_{k,s}(q).$$

Rewriting this in vectorial form yields (F.1). □

In order to get asymptotic expressions for the coefficients of $\mathbf{W}_K$, we have to find the singularities of $(I - M_K(q))^{-1}$ as a meromorphic function in q. A value q is a singularity of $(I - M_K(q))^{-1}$ if and only if it is a zero of the determinant $\det(I - M_K(q))$, which holds if and only if 1 is an eigenvalue of $M_K(q)$. In the next lemma, we collect a few results connecting $M_K(q)$ with Perron–Frobenius theory.

Lemma F.2. *Let $K \ge t$ and $q > 0$. Then*

1. *the matrix $M_K(q)$ is a non-negative, irreducible, primitive matrix;*
2. *the function $q \mapsto \lambda_{\max}(M_K(q))$ mapping q to the spectral radius of $M_K(q)$ is a strictly increasing function from $(0,\infty)$ to $(0,\infty)$;*
3. *if $M_K(q)x \le x$ or $M_K(q)x \ge x$ holds componentwise for some positive vector x, then $\lambda_{\max}(M_K(q)) \le 1$ or $\lambda_{\max}(M_K(q)) \ge 1$, respectively.*

Proof. (1) The matrix $M_K(q)$ is non-negative by definition. We note that $\frac{r}{t} \le r-1$ holds for all $r \ge 2$ and $r+1 \le \frac{r+K}{t}$ holds for all $r < N(K)$. This implies that all subdiagonal, diagonal and superdiagonal elements of $M_K(q)$ are positive. Thus $M_K(q)$ is irreducible. As all diagonal elements are positive, it is also primitive.

(2) By Perron–Frobenius theory, the spectral radius is the largest eigenvalue. For $k \ge 1$, set $a_k(q) = (1,\dots,1)M_K(q)^k(1,\dots,1)^T$ and assume that $q_1 < q_2$. As $a_k(q)$ is $q^{kN(K)}$ times a polynomial in q with positive integer coefficients, we have $a_k(q_2) > (q_2/q_1)^{kN(K)}a_k(q_1)$. This implies that $\lim_{k\to\infty} a_k(q_2)/a_k(q_1) = +\infty$.

On the other hand, $a_k(q_j) \sim c_j\lambda_{\max}(M_K(q_j))^k$ for $j \in \{1,2\}$ and suitable positive constants c_1, c_2. As

$$+\infty = \lim_{k\to\infty} \frac{a_k(q_2)}{a_k(q_1)} = \lim_{k\to\infty} \frac{c_2}{c_1}\left(\frac{\lambda_{\max}(M_K(q_2)))}{\lambda_{\max}(M_K(q_1)))}\right)^k,$$

we conclude that $\lambda_{\max}(M_K(q_2)) > \lambda_{\max}(M_K(q_1))$.

(3) Assume that $M_K(q)x \le x$ for some positive x. Iterating this equation and multiplying with x^T from the left yields

$$x^TM_K(q)^kx \le x^Txx$$

for all $k \ge 1$. As $x^TM_K(q)^kx \sim c\lambda_{\max}(M_K(q))^k$ for some positive constant c and $k \to \infty$, we conclude that $\lambda_{\max}(M_K(q)) \le 1$.

The same argument can be used for the case $M_K(q)x \ge x$, too.

□

We consider the infinite matrix

$$M_\infty(q) := \left(q^r\left[\frac{r}{t} \le s\right]\right)_{\substack{1\le r\\ 1\le s}}$$

and the infinite determinant $\det(I - M_\infty(q))$ which is defined to be the limit of the principal minors $\det([r=s] - q^r\left[\frac{r}{t} \le s\right])_{\substack{1\le r\le N\\ 1\le s\le N}}$ when N tends to ∞, cf. Eaves [8]. For $|q| < 1$, this infinite determinant converges by Eaves' sufficient condition.

We now show that the infinite determinant is indeed the denominator of the generating function $H(q,1,1,1)$.

Lemma F.3. *We have*

$$\det(I - M_\infty(q)) = 1 - b(q,1,1,1)$$

where $b(q,u,1,1)$ is given in Lemma A.1.

Proof. When expanding the infinite determinant, we take the 1 on the diagonal in almost all rows and some other entry in rows $a_1 < a_2 < \dots < a_k$ for some k. These other entries have to come from $-M_\infty(q)$. Extracting the sign for these rows, we get

$$\begin{aligned}\det(I - M_\infty(q)) &= \sum_{k\ge0}(-1)^k \sum_{1\le a_1<a_2<\dots<a_k} \det(q^{a_i}[a_i \le ta_j])_{1\le i,j\le k}\\ &= \sum_{k\ge0}(-1)^k \sum_{1\le a_1<a_2<\dots<a_k} q^{a_1+\dots+a_k}\det([a_i \le ta_j])_{1\le i,j\le k}.\end{aligned}$$

We trivially have $a_i \le ta_j$ for $j \ge i$, so all entries above and on the diagonal of $([a_i \le ta_j])_{1\le i,j\le k}$ are 1. If $a_2 \le ta_1$, the first and the second row of $([a_i \le ta_j])_{1\le i,j\le k}$ are identical, so the determinant vanishes. Therefore, we only have to consider summands with $a_2 > ta_1$.

In this case, we clearly have $a_i > ta_1$ for all $i \geq 2$, i.e., the first column of $([a_i \leq ta_j])_{1\leq i,j\leq k}$ is $(1, 0, \ldots, 0)^T$. Repeating this argument, we see that only summands with $a_{j+1} > ta_j$ for $1 \leq j < k$ contribute to the determinant. In this case, the matrix $([a_i \leq ta_j])_{1\leq i,j\leq k}$ equals $([j \geq i])_{1\leq i,j\leq k}$ and has determinant 1.

We therefore obtained the representation

$$\det(I - M_\infty(q)) = \sum_{k\geq 0}(-1)^k \sum_{\substack{a_1,\ldots,a_k\\ \forall j: a_{j+1}>ta_j}} q^{a_1+\cdots+a_k}.$$

With the change of variables $a_1 =: b_k$ and $a_{j+1} - ta_j =: b_{k-j}$ for $1 \leq j < k$, we obtain

$$\begin{aligned}\det(I - M_\infty(q)) &= \sum_{k\geq 0}(-1)^k \sum_{b_1,\ldots,b_k\geq 1} q^{b_1[1]+\cdots+b_k[k]}\\ &= \sum_{k\geq 0}(-1)^k \prod_{j=1}^{k}\Bigl(\sum_{b_j\geq 1}(q^{[j]})^{b_j}\Bigr) = 1 - b(q, 1, 1, 1).\end{aligned}$$

□

If K tends to infinity, we do expect $W_K(q)$ to tend to $H(q, 1, 1, 1)$, as the restriction on the width becomes meaningless. We will need a slightly stronger result: we also need convergence of the numerator and the denominator of $W_K(q)$ given by (F.1) and Cramer's rule to the numerator $a(q, 1, 1, 1)$ and the denominator $1 - b(q, 1, 1, 1)$ of $H(q, 1, 1, 1)$, respectively. We prove this in two steps: first, we prove that the numerator and the denominator of $W_K(q)$ given by (F.1) and Cramer's rule tend to the corresponding infinite determinants.

Lemma F.4. *For $|q| < 1$, we have*

$$\det(I - M_K(q)) = \det(I - M_\infty(q)) + O(q^{K/(2t)}).$$

The same conclusion holds when the s-th column of both $I - M_K(q)$ and $I - M_\infty(q)$ are replaced by the vector $(q, 0, \ldots)^T$ with $K - 1$ and infinitely many zeroes, respectively. The estimate still holds for derivatives with respect to q.

Proof. The infinite determinant $\det(I - M_\infty(q))$ consists of summands

$$\pm\prod_{s\in S} q^{\pi(s)} = \pm q^{\sum_{s\in S}\pi(s)}$$

where $\pi : \mathbb{N} \to \mathbb{N}$ is a bijection such that there are only finitely many non-fixed points s of π and S is a finite subset of $\mathbb{N}$ containing all non-fixed points of π. Note that the complement of S corresponds to those columns where 1 has been chosen on the diagonal in the expansion of the determinant. Not all (π, S) will actually occur due to the Iversonian expression in the definition of $M_\infty(q)$.

For every $k \in \mathbb{N}$, there is a bijection from the set

$$\Bigl\{(\pi, S) \mid \pi : \mathbb{N} \to \mathbb{N} \text{ bijective, } S \subseteq \mathbb{N} \text{ finite such that } \{s \in \mathbb{N} \mid \pi(s) \neq s\} \subseteq S \text{ and } \sum_{s\in S}\pi(s) = k\Bigr\}$$

to the set

$$\Big\{(x_1,\ldots,x_j)\in\mathbb{N}^j \mid j\in\mathbb{N},\ \sum_{i=1}^{j}x_i=k \text{ with pairwise distinct } x_i\Big\}$$

of compositions of k with distinct parts: the set S can be recovered as the set of summands in the composition, the permutation π can be recovered from the order of the summands.

As there are at most $\exp(2\sqrt{k}\log k)$ compositions of k with distinct parts by a result of Richmond and Knopfmacher [15], there are at most that many summands $\pm q^k$ in the infinite determinant $\det(I-M_\infty(q))$.

The difference between $\det(I-M_\infty(q))$ and $\det(I-M_K(q))$ consists of those summands which do not choose the 1 on the diagonal in some row $>N(K)$ or choose some column s in some row r with $s>(r+K)/t$. In the latter case, the 1 on the diagonal cannot be chosen in row s, so that the exponent of q in this summand is at least $r+s>K/t$. So all summands in the difference are of the form $\pm q^k$ for some $k\ge K/t$. By triangle inequality and the above estimates, we obtain

$$|\det(I-M_\infty(q))-\det(I-M_K(q))|\le\sum_{k\ge K/t}\exp(2\sqrt{k}\log k)q^k=O(q^{K/(2t)}).$$

The argument does not change if the s-th column of both matrices is replaced by the column vector $(q,0,\ldots,0)^T$.

Differentiating the determinant can be done term by term. The error term does not change as the bound $O(q^{K/(2t)})$ is weak enough. □

The second step in the proof of the convergence of numerator and denominator of $W_K(q)$ consists of the following simple lemma.

Lemma F.5. *The denominator* $\det(I-M_K(q))$ *of* $W_K(q)$ *converges to* $1-b(q,1,1,1)$ *with error* $O(q^{K/(2t)})$. *The numerator* $\det(I-M_K(q))W_K(q)$ *of* $W_K(q)$ *converges to* $a(q,1,1,1)$ *with the same error. The same is true for derivatives with respect to* q.

Proof. The first statement is simply the combination of Lemmata F.4 and F.3.

As a formal power series, $W_K(q)$ converges to $H(q,1,1,1)$ as $[q^n]W_K(q)=[q^n]H(q,1,1,1)$ holds for $n\le(K-1)/(t-1)$, as a tree with n internal states has at most $1+n(t-1)$ leaves and therefore width at most $1+n(t-1)$.

As $1-b(q,1,1,1)$ has no root with $|q|<1/2$ by Lemma 2.1, $W_K(q)$ converges to $H(q,1,1,1)$ for $|q|<1/2$. As the denominator is already known to converge to the denominator $1-b(q,1,1,1)$ of $H(q,1,1,1)$, we conclude that the numerators (which are already known to converge to some infinite determinant) actually have to converge to $a(q,1,1,1)$.

Taking derivatives with respect to q does not change the argument by Lemma F.4. □

In order to obtain information on the roots of $\det(I-M_K(q))$ and therefore the singularities of $\mathbf{W}_K(q)$, we approximate the Perron–Frobenius eigenvector of $M_K(q)$ by the eigenvector of the infinite matrix $M_\infty(q)$. In the following lemma it turns out that we actually met this infinite eigenvector earlier.

Lemma F.6. *For* $r\ge1$, *we have*

$$q^r\Big(1-\sum_{j=1}^{\lceil r/t\rceil-1}[u^{jt}]b(q,u,1,1)\Big)=[u^{rt}]b(q,u,1,1). \tag{F.3}$$

In particular, $(p_r)_{r\ge 1}$ *as defined in Theorem V is a right eigenvector of* $M_\infty(q_0)$ *to the eigenvalue* 1, *i.e.,*

$$M_\infty(q_0)\cdot(p_r)_{r\ge1} = (p_r)_{r\ge1}. \tag{F.4}$$

Proof. Multiplying the left hand side of (F.3) with u^{rt} and summing over $r\ge1$ yields

$$\begin{aligned}\frac{qu^t}{1-qu^t} - \sum_{\substack{r\ge1\\ j\ge1\\ jt<r}}(qu^t)^r[u^{jt}]b(q,u,1,1) &= \frac{qu^t}{1-qu^t} - \sum_{j=1}^{\infty}[u^{jt}]b(q,u,1,1)\sum_{r=jt+1}^{\infty}(qu^t)^r\\ &= \frac{qu^t}{1-qu^t} - \frac{qu^t}{1-qu^t}\sum_{j=1}^{\infty}(qu^t)^{jt}[u^{jt}]b(q,u,1,1)\\ &= \frac{qu^t}{1-qu^t}(1-b(q,qu^t,1,1)) = b(q,u,1,1),\end{aligned}$$

which concludes the proof of (F.3).

Setting $q=q_0$ in (F.3) and noting that $1=b(q_0,1,1,1)=\sum_{r\ge1}p_r$ yields (F.4). □

We now use the fact that $(p_r)_{r\ge1}$ is an eigenvector of $M_\infty(q)$ to derive bounds for its entries.

Proposition F.7. *All* p_r, $r\ge1$, *are positive and we have*

$$\frac{1}{r}q_*^r \ll_t p_r \ll_t r^2q_*^r.$$

with

$$q_* = q_0^{1+\frac{1}{t-1}}.$$

Proof. By Theorem V, the p_r are limits of probabilities and therefore non-negative.

By the eigenvalue equation (F.4), we have

$$p_r \ge q_0^r p_{\lceil r/t\rceil}$$

for all $r\ge1$. Iterating this, we get

$$\begin{aligned}p_r &\ge q_0^{\sum_{j=0}^{\lfloor \log_t r\rfloor-1}\lceil r/t^j\rceil} p_{\lceil r/t^{\lfloor \log_t r\rfloor}\rceil} \gg_t q_0^{\sum_{j=0}^{\lfloor \log_t r\rfloor-1}(1+r/t^j)}\\ &\ge q_0^{\log_t r+\sum_{j=0}^{\infty}r/t^j} = r^{\log_t q_0}q_0^{r(1+1/(t-1))}.\end{aligned}$$

As $q_0\ge 1/t$ by Lemma 2.1, we have $\log_t q_0\ge -1$ and the lower bound follows.

To prove the upper bound, we proceed in two steps. In a first step, we note that the eigenvalue equation (F.4) together with the fact that $\sum_{r\ge1}p_r=1$ yields the weaker upper bound

$$p_r = q_0^r\sum_{s\ge\lceil r/t\rceil}p_s \le q_0^r\sum_{s\ge1}p_s = q_0^r.$$

In a second step, we use induction on r and assume that $p_s\le cs^2q_*^s$ for $s<r$ for some constant c depending on t. Then the eigenvalue equation (F.4) yields

$$\begin{aligned}p_r &\le q_0^r\sum_{s\ge\lceil r/t\rceil}p_s \le cq_0^r\sum_{\lceil r/t\rceil\le s<r}s^2q_*^s + q_0^r\sum_{r\le s}q_0^s \le cq_0^r\sum_{\lceil r/t\rceil\le s}s^2q_*^s + \frac{1}{1-q_0}q_0^{2r}\\ &= cq_0^r\left(\frac{\lceil r/t\rceil^2}{1-q_*} + \frac{2q_*\lceil r/t\rceil}{(1-q_*)^2} + \frac{q_*(1+q_*)}{(1-q_*)^3}\right)q_*^{\lceil r/t\rceil} + \frac{1}{1-q_0}q_0^{2r}\end{aligned}$$

$$\leq cq_0^r \left(\frac{(r+t)^2}{t^2(1-q_*)} + \frac{2q_*(r+t)}{t(1-q_*)^2} + \frac{q_*(1+q_*)}{(1-q_*)^3} \right) q_*^{r/t} + \frac{1}{1-q_0} q_0^{2r}.$$

As $t^2(1-q_*) > 1$ for $t \geq 2$ (cf. Lemma 2.1), we obtain

$$p_r \leq cr^2 q_0^r q_*^{r/t} = cr^2 q_0^{r\left(1+\frac{1}{t}\left(1+\frac{1}{t-1}\right)\right)} = cr^2 q_*^r$$

for sufficiently large r.

□

Lemma F.8. *The generating function $W_K(q)$ has a unique singularity q_K with $|q_K| \leq 0.6$ for $K \geq c_1$ for a suitable positive constant c_1 depending on t. It is a simple pole and a zero of* $\det(I - M_K(q))$*. Furthermore*

$$q_0 + c_2 \frac{1}{K} q_0^{K/(t-1)} \leq q_K \leq q_0 + c_3 K^2 q_0^{K/(t-1)}$$

for suitable positive constants c_2, c_3 depending on t.

Proof. In the following, c_4, c_5, … denote suitable constants depending on t.

As $H(q,1,1,1)$ has a unique pole q with $|q| \leq 0.6$ by Lemma 2.1 and numerator and denominator of $W_K(q)$ tend to the numerator and denominator of $H(q,1,1,1)$ respectively by Lemma F.5, $W_K(q)$ also has a a unique pole with $|q| \leq 0.6$ for sufficiently large K.

We set $x_K = (p_1, \ldots, p_{N(K)})^T$. If we find a q such that $M_K(q)x_K \geq x_K$, then Lemma F.2 implies that $\lambda_{\max}(M_K(q)) \geq 1$ and $q_K < q$.

We therefore consider the r-th row of $M_K(q)x_K$ for some $1 \leq r \leq N(K)$. We have

$$\begin{aligned}(M_K(q)x_K)_r &= q^r \sum_{\frac{r}{t} \leq s \leq \frac{r+K}{t}} p_s \geq q^r \sum_{\frac{r}{t} \leq s < \frac{r+K}{t}} p_s = q^r \left(\frac{p_r}{q_0^r} - \frac{p_{r+K}}{q_0^{r+K}} \right) \\ &= p_r \left(\frac{q}{q_0} \right)^r \left(1 - \frac{p_{r+K}}{p_r q_0^K} \right)\end{aligned}$$

by the eigenvalue equation (F.4). By Proposition F.7, we have

$$\frac{p_{r+K}}{p_r q_0^K} \leq c_4 r(r+K)^2 \frac{q_*^{r+K}}{q_*^r q_0^K} = c_4 r(r+K)^2 q_0^{K/(t-1)} \leq c_5 K^3 q_0^{K/(t-1)}.$$

Therefore, we have

$$\sqrt[r]{1 - \frac{p_{r+K}}{p_r q_0^K}} = \frac{1}{\left(1 - \frac{p_{r+K}}{p_r q_0^K}\right)^{-1/r}} \geq \frac{1}{1 + \frac{2p_{r+K}}{r p_r q_0^K}} \geq \frac{1}{1 + c_6 K^2 q_0^{K/(t-1)}}.$$

This means that for $q = q_0 + c_7 K^2 q_0^{K/(t-1)}$, we have $M_K(q)x_K \geq x_K$, as requested.

The proof of the lower bound runs along the same lines.

□

Proof of Theorem VII. By singularity analysis, we have

$$\mathbb{P}(w(T) \leq K) = (1 + O(0.6^{K/2t})) \left(\frac{q_K}{q_0} \right)^{-n-1} (1 + O(0.99^n))$$

for $K \geq c_8$.

We now estimate

$$\mathbb{E}(w(T)) = \sum_{K\geq 0}(1-\mathbb{P}(w(T)\leq K)). \tag{F.5}$$

We use the abbreviation $S := 1/q_0^{t-1} > 1$.

First, we consider the summands of (F.5) with $S^K \leq n/\log^2 n$. By Lemma F.8, we have

$$\left(\frac{q_K}{q_0}\right)^n \geq \left(1+c_9\frac{1}{S^K\log_S n}\right)^n \geq \left(1+c_{10}\frac{\log n}{n}\right)^n \geq c_{10}\log n.$$

We conclude that these summands of (F.5) contribute $\log_S n + O(\log\log n)$. In particular, the above estimates imply that

$$\mathbb{P}(w(T)-\log_S n \leq -2\log_S\log n) = O(1/\log n). \tag{F.6}$$

Now, we consider the summands of (F.5) with $n/\log^2 n < S^K \leq n\log^3 n$. These are $O(\log\log n)$ summands with each trivially contributing at most 1, so the total contribution is $O(\log\log n)$.

Next, we consider the summands of (F.5) with $n\log^3 n < S^K \leq n^{4t\log S}$. We now have

$$\frac{q_k}{q_0} \leq 1+c_{11}\frac{\log^2 n}{S^K} \leq 1+c_{11}\frac{1}{n\log n}$$

and therefore

$$\mathbb{P}(w(T)\leq K) \geq (1+O(n^{-|\log_S 0.6|/(2t)}))\exp\left(-n\log\left(\frac{q_k}{q_0}\right)\right) \geq 1-c_{12}\frac{1}{\log n}.$$

The total contribution of these summands is therefore $O(1)$. In particular, the above estimates imply that

$$\mathbb{P}(w(T)-\log_S n \geq 3\log_S\log n) = O(1/\log n). \tag{F.7}$$

Next, we consider the summands of (F.5) with $n^{4t\log S} < S^K \leq S^{tn}$. This time, we have

$$\frac{q_k}{q_0} \leq 1+c_{13}\frac{n^2}{n^4}$$

and therefore

$$\mathbb{P}(w(T)\leq K) = (1+O(n^{-2|\log 0.6|}))\exp\left(-n\log\left(\frac{q_k}{q_0}\right)\right) \geq 1-c_{14}\frac{1}{n}.$$

The total contribution of these summands is therefore $O(1)$.

Finally, we note that all summands with $K > tn$ vanish: any tree with n internal nodes has at most width tn.

Collecting all terms, we obtain

$$\mathbb{E}(w(T)) = \log_S n + O(\log\log n) = \frac{\log n}{-(t-1)\log q_0} + O(\log\log n).$$

Combining (F.6) and (F.7) immediately yields the concentration property (7.1). ◻