SUB-GAUSSIAN TAIL BOUNDS FOR THE WIDTH AND HEIGHT OF CONDITIONED GALTON–WATSON TREES.

LOUIGI ADDARIO-BERRY, LUC DEVROYE, AND SVANTE JANSON

Abstract. We study the height and width of a Galton–Watson tree with offspring distribution ξ satisfying $\mathbb{E}\,\xi = 1$, $0 < \operatorname{Var}\xi < \infty$, conditioned on having exactly n nodes. Under this conditioning, we derive sub-Gaussian tail bounds for both the width (largest number of nodes in any level) and height (greatest level containing a node); the bounds are optimal up to constant factors in the exponent. Under the same conditioning, we also derive essentially optimal upper tail bounds for the number of nodes at level k, for $1 \le k \le n$.

1. Introduction

A Galton–Watson tree is the family tree of a Galton–Watson process, i.e., it is a random rooted tree, constructed recursively from the root, where each node has a random number of children and these random numbers are independent copies of some random variable ξ taking values in $\{0, 1, \dots\}$. We let $\mathcal{T}$ denote a (random) Galton–Watson tree. ($\mathcal{T}$ depends of course on ξ, or rather its distribution, but the offspring distribution ξ is fixed throughout the paper and is therefore not shown explicitly in the notation.) We view the children of each node as arriving in some random order, so that $\mathcal{T}$ is an ordered, or *plane* tree.

At times in the paper it will be useful to think of $\mathcal{T}$ as a subtree of the so-called *Ulam–Harris* tree $\mathcal{U}$: this is the tree with root $\varnothing$ whose non-root nodes correspond to finite sequences of integers $v_1 \dots v_k$, with $v_1 \dots v_k$ having parent $v_1 \dots v_{k-1}$ and children $\{v_1 \dots v_k i \ : \ i \in \{1, 2, \dots\}\}$. For a node v of $\mathcal{U}$ we think of vi as the i'th child of v. Any rooted plane tree T in which all nodes have at most countably many children can be viewed as a subtree of $\mathcal{U}$ by sending the root of T to the root $\varnothing$ of $\mathcal{U}$ and using the ordering of children in T to recursively define an embedding of T into $\mathcal{U}$ (see e.g. [23]).

We will study the *conditioned Galton–Watson tree* $\mathcal{T}_n$, which is the random tree $\mathcal{T}$ conditioned on having exactly n nodes. In symbols, $\mathcal{T}_n := (\mathcal{T} \mid |\mathcal{T}| = n)$, where, for any tree T, $|T|$ denotes its number of nodes. (We consider in the sequel only n such that $\mathbb{P}(|\mathcal{T}| = n) > 0$.) For examples of standard types of random trees that can be represented as conditioned

Date: 15 November, 2010.
2000 *Mathematics Subject Classification.* 60C05,60J80.

Galton–Watson trees for suitable ξ, see e.g. Devroye [7]. The conditioned Galton–Watson trees are essentially the same as the random *simply generated trees* [26], see e.g. [7] or [9].

As is well-known, the distribution of the tree $\mathcal{T}_n$ is not changed if ξ is replaced by another random variable ξ' whose distribution is replaced by *tilting* (or *conjugation*) [19]: $\mathbb{P}(\xi' = k) = ca^k\,\mathbb{P}(\xi = k)$, $k \ge 0$, for some $a > 0$ and normalizing constant c. (Necessarily, $c = \bigl(\mathbb{E}\,a^{\xi}\bigr)^{-1}$, and thus $\mathbb{E}\,a^{\xi} < \infty$.) We may, except in some exceptional cases, by a suitable tilting assume that $\mathbb{E}\,\xi = 1$, so that the branching process is critical. This turns out to be convenient, and we will in the sequel always make this assumption $\mathbb{E}\,\xi = 1$. We further assume that ξ has finite variance $\sigma^2 := \operatorname{Var}\xi < \infty$. We exclude the trivial case $\xi = 1$ a.s., i.e., we assume $\sigma^2 > 0$. (Equivalently, when $\mathbb{E}\,\xi = 1$, $\mathbb{P}(\xi = 0) > 0$.)

For a rooted tree T (deterministic or random), the *depth* $h(v)$ of a node v is its distance to the root; the root thus has depth 0. Let $Z_k(T)$ be the width at level k, i.e., the number of nodes at depth k, $k = 0, 1, \dots$. We define, as usual, the *width* of the tree by

$$W = W(T) := \max_{k\ge 0} Z_k(T), \tag{1.1}$$

and the *height* by

$$H = H(T) := \max\{h(v) : v \in T\} = \max\{k : Z_k(T) > 0\}. \tag{1.2}$$

It is well-known that the width and height of a conditioned Galton–Watson tree $\mathcal{T}_n$ both are of the order $\sqrt{n}$. More precisely, $n^{-1/2}W(\mathcal{T}_n)$ and $n^{-1/2}H(\mathcal{T}_n)$ both converge in distribution, as $n \to \infty$, see e.g. [1], [5], [10] and [9]; moreover, they converge jointly [5], [16],

$$\bigl(n^{-1/2}W(\mathcal{T}_n), n^{-1/2}H(\mathcal{T}_n)\bigr) \xrightarrow{\mathrm{d}} (\sigma W, \sigma^{-1}H) \tag{1.3}$$

for some limit variables W and H, that furthermore do not depend on the distribution of ξ. (W is the maximum of a Brownian excursion, and $H \overset{\mathrm{d}}{=} 2W$; see further [18].)

Two of the main results of the paper are to prove essentially optimal uniform sub-Gaussian upper tail bounds for both $W(\mathcal{T}_n)/\sqrt{n}$ and $H(\mathcal{T}_n)/\sqrt{n}$ for every offspring distribution ξ with finite variance. As an immediate consequence, the estimates $\mathbb{E}\,W(\mathcal{T}_n) = O(n^{1/2})$ and $\mathbb{E}\,H(\mathcal{T}_n) = O(n^{1/2})$ hold; even these much weaker statements are to our knowledge new at this level of generality. (For estimates assuming an exponential moment of ξ, see e.g. [13].)

We let $C_1, C_2, \dots, c_1, c_2, \dots$ denote positive constants that may depend on the distribution of ξ (and in particular on σ^2) but not on n or other parameters unless explicitly indicated. (We use C_i for "large" and c_i for "small" constants.) Proofs are given in Section 4.

Theorem 1.1. *Suppose that* $\mathbb{E}\,\xi = 1$ *and* $\operatorname{Var}\xi < \infty$*. Then*

$$\mathbb{P}\big(W(\mathcal{T}_n) \geq x\big) \leq C_1 e^{-c_1 x^2/n}$$

for all $x \geq 0$ *and* $n \geq 1$*.*

Theorem 1.2. *Suppose that* $\mathbb{E}\,\xi = 1$ *and* $0 < \operatorname{Var}\xi < \infty$*. Then*

$$\mathbb{P}(H(\mathcal{T}_n) \geq h) \leq C_2 e^{-c_2 h^2/n} \tag{1.4}$$

for all $h \geq 0$ *and* $n \geq 1$*.*

The condition $\operatorname{Var}\xi > 0$ excludes the case $\mathbb{P}(\xi = 1) = 1$, in which case $\mathcal{T}_n$ is a path of length n.

Corollary 1.3. *Suppose that* $\mathbb{E}\,\xi = 1$ *and* $0 < \operatorname{Var}\xi < \infty$*. Then* $\mathbb{E}\,W(\mathcal{T}_n) = O(n^{1/2})$ *and* $\mathbb{E}\,H(\mathcal{T}_n) = O(n^{1/2})$*. More generally, for every fixed* $r < \infty$*,* $\mathbb{E}(W(\mathcal{T}_n)^r) = O(n^{r/2})$ *and* $\mathbb{E}(H(\mathcal{T}_n)^r) = O(n^{r/2})$*.*

While our methods do not prove the convergence (1.3) of $W(\mathcal{T}_n)/\sqrt{n}$ and $H(\mathcal{T}_n)/\sqrt{n}$, we have thus as a corollary obtained tightness of them, and we believe that our argument might be the simplest proof of this tightness.

On the other hand, knowing the limit result (1.3), it follows from the fact that the bounds in Corollary 1.3 hold for every r that all moments (also joint) converge in (1.3). In particular, by the known formulas for the moments of W and $H \overset{\mathrm{d}}{=} 2W$ (see e.g. [3]), as $n \to \infty$,

$$\mathbb{E}\big(W(\mathcal{T}_n)^r\big)/n^{r/2} \to \sigma^r\,\mathbb{E}\,W^r = \sigma^r 2^{-r/2} r(r-1)\Gamma(r/2)\zeta(r), \tag{1.5}$$

$$\mathbb{E}\big(H(\mathcal{T}_n)^r\big)/n^{r/2} \to \sigma^{-r}\,\mathbb{E}\,H^r = \sigma^{-r} 2^{r/2} r(r-1)\Gamma(r/2)\zeta(r). \tag{1.6}$$

For joint moments, see [8] and [18]. These results are well-known if ξ is assumed to have an exponential moment, see e.g. [14] and [11], but to our knowledge they have not, even in the case $r = 1$, been proved before without extra conditions.

We emphasise that we obtain these bounds for higher moments of both $W(\mathcal{T}_n)$ and $H(\mathcal{T}_n)$, and even sub-Gaussian tail bounds for both variables, without assuming more than a finite second moment of ξ. This is somewhat surprising, at least for the width, since a ξ with a large tail will produce a very wide Galton–Watson tree $\mathcal{T}$ with comparatively large probability; the explanation is that if the tree has one generation that is very large, say of size m, then it will probably have many nodes (of order m^2) in later generations, so the conditioning on exactly n nodes makes this event very unlikely if $m \gg \sqrt{n}$. In other words, the bounds on the width hold, not because it is difficult for the Galton–Watson tree to get many branches, but because it is difficult to get rid of them in time.

Remark 1.4. We assume $\sigma^2 = \operatorname{Var}\xi < \infty$ throughout the paper. Since increasing σ makes the width larger and the height smaller (asymptotically at least), see e.g. (1.5)–(1.6), it is not reasonable to expect that the results for the width generalize to the case $\sigma^2 = \infty$. However, for the same reason it seems likely that the results for the height extend, but we have not investigated that and leave that as an open problem. In particular, we ask the following questions (assuming $\mathbb{E}\,\xi = 1$): Is $\mathbb{E}\,H(\mathcal{T}_n) = O(n^{1/2})$ also if $\sigma^2 = \infty$? Is $\mathbb{E}\,H(\mathcal{T}_n) = o(n^{1/2})$ if $\sigma^2 = \infty$?

Next we consider the width $Z_k(\mathcal{T}_n)$ at a given level k. Of course, $Z_k(\mathcal{T}_n) \le W(\mathcal{T}_n)$, so the results above for $W(\mathcal{T}_n)$ immediately imply the same bounds for $Z_k(\mathcal{T}_n)$, uniformly in k. In particular,

$$\mathbb{E}\,Z_k(\mathcal{T}_n) = O\big(n^{1/2}\big). \tag{1.7}$$

For $k \asymp n^{1/2}$, this is the correct order of $\mathbb{E}\,Z_k(\mathcal{T}_n)$; in fact, $n^{-1/2} Z_{\lfloor x\sqrt{n}\rfloor}(\mathcal{T}_n)$ converges in distribution for every fixed $x \ge 0$, and as a function of x, see [10, 11] (assuming a finite exponential moment) and [21] (the general case, by probabilistic methods).

For small k, on the other hand, $Z_k(\mathcal{T}_n)$ is smaller and it was proven in [16, Theorem 1.13] that

$$\mathbb{E}\,Z_k(\mathcal{T}_n) = O(k), \tag{1.8}$$

uniformly for all $k \ge 1$ and $n \ge 1$. This is the best possible estimate, since for any fixed k,

$$\mathbb{E}\,Z_k(\mathcal{T}_n) \to 1 + k\sigma^2, \qquad \text{as } n \to \infty, \tag{1.9}$$

see Meir and Moon [26] and Janson [16, 17]. (It is shown in [17] that the sequence $\mathbb{E}\,Z_k(\mathcal{T}_n)$ is not always monotone in n, so (1.8) is not a consequence of (1.9).)

Furthermore, for large k, (1.8) is again not sharp. Indeed, if $k \gg \sqrt{n}$, then typically $H(\mathcal{T}_n) < k$ and thus $Z_k(\mathcal{T}_n) = 0$. In fact, as $k \to \infty$, $\mathbb{E}\,Z_k(\mathcal{T}_n)$ decreases exponentially, as is shown by the next theorem, which combines the three phases ($k \ll \sqrt{n}$, $k \asymp n$, $k \gg \sqrt{n}$) in a unified statement. (Drmota and Gittenberger [11] gave the weaker bound $C_3 n^{1/2} e^{-c_3 k/\sqrt{n}}$, assuming an exponential moment on ξ.)

Theorem 1.5. *Suppose that $\mathbb{E}\,\xi = 1$ and $0 < \operatorname{Var}\xi < \infty$. For all $n, k \ge 1$,*

$$\mathbb{E}\,Z_k(\mathcal{T}_n) \le C_4 k e^{-c_4 k^2/n} \tag{1.10}$$

and also

$$\mathbb{E}\,Z_k(\mathcal{T}_n) \le C_5 n^{1/2} e^{-c_5 k^2/n} \tag{1.11}$$

(which is weaker for $k = o(\sqrt{n})$ but equivalent for larger n).

Turning to higher moments of $Z_k(\mathcal{T}_n)$, we first note that for small k there is no result corresponding to (1.10) without assuming higher moments of ξ. In fact, already for $k = 1$, it is easy to see that for any $m \ge 1$,

$$\mathbb{P}(Z_1(\mathcal{T}_n) = m) \to m\,\mathbb{P}(\xi = m)$$

as $n \to \infty$, see [19] and Remark 3.1. It follows by Fatou's lemma, that if $\mathbb{E}\,\xi^{r+1} = \infty$, for some $r > 1$, then $\mathbb{E}\,Z_k(\mathcal{T}_n)^r \to \infty$. The same holds for $\mathbb{E}\,Z_k(\mathcal{T}_n)^r$ for every fixed $k \ge 1$.

Conversely, it was proven in [16, Theorem 1.13] that if $\mathbb{E}\,\xi^{r+1} < \infty$ for an integer $r \geq 1$, then $\mathbb{E}\,Z_k(\mathcal{T}_n)^r = O(k^r)$ uniformly in $k \geq 1$ and $n \geq 1$. (The restriction to integer r is for technical reasons in the proof; we conjecture that the result holds for any real $r \geq 1$.)

On the other hand, the estimate (1.11) extends to higher moments without assuming any moment condition on ξ beyond our standing $0 < \operatorname{Var}\xi < \infty$, i.e., $\mathbb{E}\,\xi^2 < \infty$ and ξ is not constant.

Theorem 1.6. *Suppose that $\mathbb{E}\,\xi = 1$ and $0 < \operatorname{Var}\xi < \infty$. For any $r < \infty$,*

$$\mathbb{E}\big(Z_k(\mathcal{T}_n)/\sqrt{n}\big)^r \leq C_6(r)e^{-c_6k^2/n} \tag{1.12}$$

for all $k, n \geq 1$.

Furthermore,

$$\mathbb{P}\big(Z_k(\mathcal{T}_n) > x\big) \leq C_7e^{-c_6k^2/n-c_7x^2/n} \tag{1.13}$$

for all $x \geq 0$ and $n \geq 1$.

1.1. **Remarks on the limit law.** We say that T is theta distributed if it has distribution function

$$\mathbb{P}(T \leq x) = \sum_{j=-\infty}^{\infty} \left(1 - 2j^2x^2\right) e^{-j^2x^2} = \frac{4\pi^{5/2}}{x^3} \sum_{j=1}^{\infty} j^2 e^{-\pi^2j^2/x^2}, \; x > 0.$$

The appearance of T as the limit law of the height of random conditional Galton–Watson trees was noted in [28, 4, 6, 20, 26, 14]. Furthermore, the maximum of Brownian excursion of duration one is distributed as $T/\sqrt{2}$ (see, e.g., [3]). In (1.3), $W \overset{\mathrm{d}}{=} T/\sqrt{2}$ and $H \overset{\mathrm{d}}{=} T\sqrt{2}$. It takes a moment to verify that for $x \geq 1$,

$$\mathbb{P}(T \geq x) \geq 2e^{-x^2}, \tag{1.14}$$

and for $x \leq 1$,

$$\mathbb{P}(T \leq x) \geq 40\, e^{-\pi^2/x^2}. \tag{1.15}$$

The bound of Theorem 1.1, combined with the limit result (1.3) then shows that

$$c_1 \leq \frac{2}{\sigma^2}.$$

Similarly, the bound of Theorem 1.2, combined with the limit result (1.3) then shows that

$$c_2 \leq \frac{\sigma^2}{2}.$$

It would be nice if c_1 and c_2 could be be made more explicit. In any case, the sub-Gaussian tail behaviour of the bounds in Theorems 1.1 and 1.2 is optimal, modulo a constant factor (depending on ξ).

We also have the trivial observation that

$$W(\mathcal{T}_n)H(\mathcal{T}_n) \geq n - 1.$$

Thus, Theorems 1.1 and 1.2 yield the following left-tail upper bounds:

$$\mathbb{P}(W(\mathcal{T}_n) \le x) \le \mathbb{P}\left(H(\mathcal{T}_n) \ge \frac{n-1}{x}\right) \le C_2 \exp\left(-\frac{c_2(n-2)}{x^2}\right)$$

and

$$\mathbb{P}(H(\mathcal{T}_n) \le x) \le \mathbb{P}\left(W(\mathcal{T}_n) \ge \frac{n-1}{x}\right) \le C_1 \exp\left(-\frac{c_1(n-2)}{x^2}\right).$$

In view of (1.3) and the remark (1.15) about the theta distribution, these bounds are optimal up to the constant factors c_1 and c_2.

2. Preliminaries

The *span* of ξ, denoted $\operatorname{span}(\xi)$, is the largest integer d such that ξ/d a.s. is an integer. Note that $\mathbb{P}(|\mathcal{T}| = n) > 0$, so $\mathcal{T}_n$ exists, if and only if $n \equiv 1$ modulo $\operatorname{span}(\xi)$, except possibly for some small n.

We let ξ_i denote i.i.d. copies of the random variable ξ, and let S_n be the partial sums of $\xi_1, \xi_2, \dots,$

$$S_n := \sum_{i=1}^{n} \xi_i. \tag{2.1}$$

By a classic formula, see e.g. Dwass [12], Kolchin [22, Lemma 2.1.3, p. 105] or Pitman [27], for $n \ge 1$,

$$\mathbb{P}\big(|\mathcal{T}| = n\big) = \frac{1}{n}\,\mathbb{P}\big(S_n = n-1\big), \tag{2.2}$$

and, more generally, for $m, n \ge 1$ and independent copies $\mathcal{T}_1, \dots, \mathcal{T}_m$ of $\mathcal{T}$,

$$\mathbb{P}\Big(\sum_{i=1}^{m} |\mathcal{T}_i| = n\Big) = \frac{m}{n}\,\mathbb{P}\big(S_n = n-m\big). \tag{2.3}$$

Together with the local central limit theorem, (2.2) implies [22, Lemma 2.1.4, p. 105], with $d := \operatorname{span}(\xi)$ (recall that we only consider n such that $n \equiv 1 \pmod{d}$),

$$\mathbb{P}\big(|\mathcal{T}| = n\big) \sim \frac{d}{\sqrt{2\pi}\sigma}\, n^{-3/2}. \tag{2.4}$$

We will use a one-sided tail bound for S_n, which we take from Janson [16], that only requires our (weak) conditions. Note that, apart from the values of the constants, the bound in (2.5) is exactly as the limit given by the local central limit theorem when it applies; hence, at least for m not too large, it is of the best possible kind.

Lemma 2.1 ([16, Lemma 2.1])**.** *Suppose that ξ_i are i.i.d., non-negative and integer-valued random variables, with $\mathbb{E}\,\xi_i = 1$ and $\operatorname{Var}\xi_i < \infty$, and let $S_n := \sum_{i=1}^{n} \xi_i$. Then, for all $n \ge 1$ and $m \ge 0$,*

$$\mathbb{P}(S_n = n-m) \le \frac{C_7}{\sqrt{n}} e^{-c_7 m^2/n}. \tag{2.5}$$

Remark 2.2. We can write the probability in (2.5) as $\mathbb{P}(\sum_{i=1}^{n}(1-\xi_i)=m)$. The point is that even without any assumptions on the tail of ξ_i beyond finite variance, the variables $1-\xi_i$ are bounded above, which is enough for strong tail bounds for $m \geq 0$. (There is no similar bound for $m < 0$ under our weak conditions.) Cf. the related tail bound $\mathbb{P}(S_n \leq n-m) \leq C_8 e^{-c_8 m^2/n}$, which follows by (2.6) below.

We will use the following version of Bernstein's inequality, which is valid for variables with a one-sided bound, see e.g. [15, (2.9)–(2.13)] and [25, Theorem 2.7].

Lemma 2.3. *Let $X_1, X_2, \dots, X_n$ be independent random variables such that $X_i - \mathbb{E}\, X_i \leq b$ for every i. Then, with $V := \sum_{i=1} \operatorname{Var}(X_i)$,*

$$\mathbb{P}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}\, X_i) \geq t\right) \leq \exp\left(-\frac{t^2}{2V + 2bt/3}\right). \tag{2.6}$$

3. A size-biased Galton–Watson tree

Let $\hat{\xi}$ be a random variable with the *size-biased* distribution

$$\mathbb{P}(\hat{\xi} = m) = m\,\mathbb{P}(\xi = m). \tag{3.1}$$

(Note that this is a probability distribution on $\{1, 2, \dots\}$ since $\mathbb{E}\,\xi = 1$, and that $\hat{\xi} \geq 1$.)

Let, for $k \geq 1$, $\widehat{\mathcal{T}}^{(k)}$ be the modified Galton–Watson tree defined as follows: There are two types of nodes: *normal* and *mutant*. Normal nodes have offspring (outdegree) according to independent copies of ξ, while mutant nodes have offspring according to independent copies of $\hat{\xi}$. Moreover, all children of a normal node are normal, while for each mutant node, one of its children is selected uniformly at random and called its *heir*; the heir is mutant if it has depth less than k but normal if the depth is at least k, and all other children are normal. (Alternatively, we can call the mutants *kings*, with a reproductive behaviour different from the common people. At time k, a republic is introduced, and everybody becomes equal.)

There are thus exactly k mutant nodes, which together with the heir v^* of the last mutant node form a path from the root to some node v^* at depth k. We call this path the *spine* of $\widehat{\mathcal{T}}^{(k)}$.

Remark 3.1. This construction with $k = \infty$ was introduced by Lyons, Pemantle and Peres [24], and is called the *size-biased Galton–Watson tree*; in this case the spine is infinite so the tree is infinite. The underlying size-biased Galton–Watson process is the same as the *Q-process* studied in [2, Section I.14]. For any fixed k, the first k generations of $\mathcal{T}_n$ converge in distribution to the first generations of $\widehat{\mathcal{T}}^{(\infty)}$.

Our $\widehat{\mathcal{T}}^{(k)}$ is a truncated version of this, which grows like a normal Galton–Watson tree after generation k; thus $\widehat{\mathcal{T}}^{(k)}$ is a.s. finite.

An equivalent construction is to start with the spine, and attach independent copies of $\mathcal{T}$ to it; the number of such trees attached to each node in the spine except the last one (the top node) has distribution $\hat{\xi}-1$, but the number attached to the top node is ξ.

The probability that a given mutant node has m children and that a given one of them is selected as heir is, by (3.1),

$$\frac{1}{m}\,\mathbb{P}(\hat{\xi}=m)=\mathbb{P}(\xi=m),\qquad m\geq 1.$$

It follows that for any rooted tree T, and any path γ in T from the root to a node at depth k, letting $d_1, d_2, \dots$ denote the outdegrees of the nodes in T, taken in breadth-first order, say,

$$\mathbb{P}(\widehat{\mathcal{T}}^{(k)}=T \text{ with } \gamma \text{ as spine})=\prod_v \mathbb{P}(\xi=d_v)=\mathbb{P}(\mathcal{T}=T). \tag{3.2}$$

Since the possible spines in T are in one-to-one correspondence with the nodes at depth k, the number of them is $Z_k(T)$, and thus

$$\mathbb{P}(\widehat{\mathcal{T}}^{(k)}=T)=Z_k(T)\,\mathbb{P}(\mathcal{T}=T). \tag{3.3}$$

In other words, $\widehat{\mathcal{T}}^{(k)}$ has the distribution of $\mathcal{T}$ biased by Z_k, the size of generation k. In particular, this yields, summing (3.3) over all trees T of size $|T|=n$,

$$\mathbb{P}\big(|\widehat{\mathcal{T}}^{(k)}|=n\big)=\mathbb{E}\big(Z_k(\mathcal{T});\,|\mathcal{T}|=n\big)$$

and thus

$$\mathbb{E}\,Z_k(\mathcal{T}_n)=\frac{\mathbb{E}\big(Z_k(\mathcal{T});\,|\mathcal{T}|=n\big)}{\mathbb{P}(|\mathcal{T}|=n)}=\frac{\mathbb{P}\big(|\widehat{\mathcal{T}}^{(k)}|=n\big)}{\mathbb{P}(|\mathcal{T}|=n)}. \tag{3.4}$$

4. Proofs

Proof of Theorem 1.1. Consider the breadth first search of the Galton–Watson tree. As is well known, this search keeps a queue of Q_i nodes with $Q_0=1$ and the recursion $Q_i=Q_{i-1}-1+\xi_i$, with ξ_i i.i.d. copies of ξ as above; hence $Q_j=1+\widetilde{S}_j$, where $\widetilde{S}_j:=\sum_{i=1}^{j}(\xi_i-1)=S_j-j$. The breadth first search stops, and the tree is completely explored, when Q_j becomes 0; in order for the tree to have size n we thus have $Q_j>0$ for $0\leq j<n$ and $Q_n=0$; equivalently, $\widetilde{S}_j\geq 0$ for $j<n$ and $\widetilde{S}_n=-1$.

When the breadth first search just has completed exploring the nodes at level $k-1$, the queue consists of exactly the nodes at level k. Hence each Z_k is some Q_j, and

$$W:=\max_{k\geq 0} Z_k\leq \max_{j\geq 0} Q_j.$$

As a result, for the conditioned Galton–Watson tree $\mathcal{T}_n$,

$$\begin{aligned}\mathbb{P}(W\geq x+1)&\leq \mathbb{P}\big(\max_j Q_j\geq x+1\big)\\&=\mathbb{P}\big(\max_j \widetilde{S}_j\geq x\mid \widetilde{S}_j\geq 0,\ j<n,\ \text{and } \widetilde{S}_n=-1\big).\end{aligned} \tag{4.1}$$

We get rid of the conditioning on $\widetilde{S}_j \geq 0$ $(j < n)$ by the standard rotation argument: for each (deterministic) sequence $x_1, \ldots, x_n$ of integers ≥ -1 with sum $\sum_{i=1}^n x_i = -1$, there is exactly one rotation $x_i^{(t)} := x_{i+t}$ with $t \in \{0, \ldots, n-1\}$ and indices taken modulo n, such that the partial sums $S_j^{(t)} := \sum_{i=1}^j x_i^{(t)} \geq 0$ for $1 \leq j < n$. Hence, we can obtain $(\widetilde{S}_j)_{j=1}^n$ with the conditional distribution given $\widetilde{S}_j \geq 0$, $j < n$, and $\widetilde{S}_n = -1$, as required in (4.1), by conditioning $(\widetilde{S}_j)_{j=1}^n$ on $\widetilde{S}_n = -1$ and then taking the unique correct rotation. The rotation may change $\max_j \widetilde{S}_j$, but we have

$$\max_{j \leq n} \widetilde{S}_j = \max_{j \leq n} \widetilde{S}_j - \min_{j \leq n} \widetilde{S}_j + 1,$$

and the latter quantity is changed by at most 1 by a rotation of $\widetilde{\xi}_i := \xi_i - 1$, $i = 1, \ldots, n$. Hence, the rotation argument shows that

$$\begin{aligned} \mathbb{P}\Big(\max_{j \leq n} \widetilde{S}_j \geq x \mid \widetilde{S}_j \geq 0,\ j < n, \text{ and } \widetilde{S}_n = -1\Big) \\ \leq \mathbb{P}\Big(\max_{j \leq n} \widetilde{S}_j - \min_{j \leq n} \widetilde{S}_j \geq x \mid \widetilde{S}_n = -1\Big). \end{aligned}$$

By (4.1) we thus have

$$\begin{aligned} \mathbb{P}(\max_j Q_j \geq 2x + 2) &\leq \mathbb{P}\big(\max_{j \leq n} \widetilde{S}_j - \min_{j \leq n} \widetilde{S}_j \geq 2x + 1 \mid \widetilde{S}_n = -1\big) \\ &\leq \mathbb{P}\big(\max_{j \leq n} \widetilde{S}_j \geq x \mid \widetilde{S}_n = -1\big) + \mathbb{P}\big(\min_{j \leq n} \widetilde{S}_j \leq -x - 1 \mid \widetilde{S}_n = -1\big). \end{aligned}$$

Furthermore, the reflection $\xi_i \leftrightarrow \xi_{n+1-i}$, which takes $\widetilde{S}_j \leftrightarrow \widetilde{S}_n - \widetilde{S}_{n-j}$, shows that the last probabilities are the same, and we thus have

$$\mathbb{P}(\max_j Q_j \geq 2x + 2) \leq 2\,\mathbb{P}\big(\max_{j \leq n} \widetilde{S}_j \geq x \mid \widetilde{S}_n = -1\big). \tag{4.2}$$

Fix $x > 0$ and let τ be the stopping time $\min\{j \geq 0 : \widetilde{S}_j \geq x\}$. Then (4.2) can be written

$$\begin{aligned} \mathbb{P}(\max_j Q_j \geq 2x + 2) &\leq 2\,\mathbb{P}\big(\tau < n \mid \widetilde{S}_n = -1\big) \\ &= \frac{2\,\mathbb{P}\big(\widetilde{S}_n = -1 \mid \tau < n\big) \cdot \mathbb{P}(\tau < n)}{\mathbb{P}(\widetilde{S}_n = -1)}. \end{aligned} \tag{4.3}$$

By definition, $\widetilde{S}_\tau \geq x$. Further, for any $t < n$ and $y \geq x$, by Lemma 2.1,

$$\begin{aligned} \mathbb{P}\big(\widetilde{S}_n = -1 \mid \tau = t \text{ and } \widetilde{S}_\tau = y\big) &= \mathbb{P}(\widetilde{S}_n - \widetilde{S}_t = -y - 1) \\ &= \mathbb{P}\big(\widetilde{S}_{n-t} = -(y+1)\big) \leq C_7 n^{-1/2} e^{-c_7 (y+1)^2/(n-t)} \leq C_7 n^{-1/2} e^{-c_7 x^2/n}. \end{aligned}$$

Consequently,

$$\mathbb{P}\big(\widetilde{S}_n = -1 \mid \tau < n\big) \leq C_7 n^{-1/2} e^{-c_7 x^2/n},$$

and (4.3) yields

$$\mathbb{P}(\max_j Q_j \geq 2x+2) \leq \frac{C_9 n^{-1/2} e^{-c_7 x^2/n}}{\mathbb{P}(\widetilde{S}_n = -1)} \leq C_{10} e^{-c_9 x^2/n}, \tag{4.4}$$

since $\mathbb{P}(\widetilde{S}_n = -1) \geq c_{10} n^{-1/2}$ by the standard local central limit theorem. Finally, since $\mathbb{P}(W \geq 2x+2) \leq \mathbb{P}(\max_j Q_j \geq 2x+2)$, the proof is complete. □

Proof of Theorem 1.2. By choosing C_2 sufficiently large we may assume that $h \geq \sqrt{n}$. We may also assume that h is an integer. Our proof of (1.4) is based on the following observation: if v is a node of $\mathcal{T}_n$ with "large" height then *either* there are many edges leaving the path from the root to v, *or* many of the ancestors of v have exactly one child. In the first case, we will be forced to consider whether the majority of edges leaving the root-to-v path lead to nodes which are lexicographically before, or after, v. To do so, we use *lexicographic* and *reverse-lexicographic* depth-first search (DFS) of $\mathcal{T}_n$.

To define lexicographic DFS of $\mathcal{T}_n$, think of $\mathcal{T}_n$ as a plane tree (i.e. as embedded in the Ulam–Harris tree $\mathcal{U}$) and list the nodes of $\mathcal{T}_n$ in lexicographic order as $v_0, v_1, \ldots, v_{n-1}$. We then let $Q^d_0 = 1$ and $Q^d_i = Q^d_{i-1} - 1 + \xi_{v_{i-1}}$, where ξ_{v_i} is the number of children of v_i in $\mathcal{T}_n$. (This is sometimes called the Lukasiewicz path of $\mathcal{T}_n$; see, e.g., [23].) The *reverse-lexicographic* depth-first search of $\mathcal{T}_n$ is the sequence $Q^r_0, \ldots, Q^r_{|\mathcal{T}_n|}$ obtained by performing a lexicographic depth-first search on the mirror image of $\mathcal{T}_n$ (so if the root $\varnothing$ has children $1, \ldots, k$ in $\mathcal{T}_n$, then k is the first rather than last child visited, and so on). We remark that the lexicographic and reverse-lexicographic depth-first search both are identical in distribution to the breadth-first search of $\mathcal{T}_n$.

Now let $p_1 = \mathbb{P}(\xi = 1)$ and let $q_1 = 1 - p_1$. If v is a node of $\mathcal{T}_n$ with $h(v) = h$, then, writing j (resp. k) for the index of v in lexicographic (resp. reverse-lexicographic) order, either $\max(Q^d_j, Q^r_k) \geq (q_1/3)h$, or else at least $(p_1 + q_1/3)h$ of the ancestors of v have exactly one child. Let $\mathcal{S}$ be the set of trees T with $|T| = n$, such that T contains a node v possessing $(p_1 + q_1/3)h(v)$ ancestors with exactly one child and for which $h(v) = h$. Then let $\mathcal{E} := \{\mathcal{T}_n \in \mathcal{S}\} = \bigcup_{T \in \mathcal{S}} \{\mathcal{T}_n = T\}$.

Since Q^d and Q^r have the same distribution as Q, we then have

$$\begin{aligned} \mathbb{P}(H(\mathcal{T}_n) \geq h) &\leq \mathbb{P}(\max_j Q^d_j \geq (q_1/3)h) + \mathbb{P}(\max_k Q^r_k \geq (q_1/3)h) + \mathbb{P}(\mathcal{E}) \\ &= 2\,\mathbb{P}(\max_i Q_i \geq (q_1/3)h) + \mathbb{P}(\mathcal{E}) \\ &\leq C_{11} e^{-c_{11} h^2/n} + \mathbb{P}(\mathcal{E}), \end{aligned} \tag{4.5}$$

the latter inequality holding due to (4.4).

Next, for each tree $T \in \mathcal{S}$, fix a path γ_T from the root of T to a node v with $h(v) = h$ and with at least $(p_1 + q_1/3)h$ ancestors with exactly one child (such a node exists by the definition of $\mathcal{S}$). Then by (3.2),

$$\begin{aligned}\mathbb{P}(\mathcal{T}\in\mathcal{S}) &= \sum_{T\in\mathcal{S}}\mathbb{P}(\mathcal{T}=T)\\ &= \sum_{T\in\mathcal{S}}\mathbb{P}(\hat{T}^{(h)}=T \text{ with } \gamma_T \text{ as spine})\\ &= \mathbb{P}\left(\bigcup_{T\in\mathcal{S}}\{\hat{T}^{(h)}=T \text{ with } \gamma_T \text{ as spine}\}\right)\\ &\le \mathbb{P}\left(\sum_{i=0}^{h-1}\mathbf{1}_{\hat{\xi}_i=1}\ge (p_1+q_1/3)h\right). \end{aligned}\tag{4.6}$$

The $\mathbf{1}_{\hat{\xi}_i=1}$ are Bernoulli(p_1), so by Lemma 2.3,

$$\begin{aligned}\mathbb{P}\left(\sum_{i=0}^{h-1}\mathbf{1}_{\hat{\xi}_i=1}\ge (p_1+q_1/3)h\right) &\le \exp\left(-\frac{(q_1h/3)^2}{2p_1q_1h+2q_1^2h/9}\right)\\ &= \exp\left(-\frac{h}{18p_1/q_1+2}\right). \end{aligned}\tag{4.7}$$

It follows by (2.4) and (4.6)–(4.7) that

$$\begin{aligned}\mathbb{P}(\mathcal{E}) = \frac{\mathbb{P}(\mathcal{T}\in\mathcal{S})}{\mathbb{P}(|\mathcal{T}|=n)} &\le C_{12}n^{3/2}\exp\left(-\frac{h}{18p_1/q_1+2}\right)\\ &\le C_{13}e^{-c_{12}h^2/n}\end{aligned}$$

for all $h\ge\sqrt{n}$. Together with (4.5) we have thus proved

$$\mathbb{P}(H(\mathcal{T}_n)\ge h)\le C_{11}e^{-c_{11}h^2/n}+C_{13}e^{-c_{12}h^2/n},$$

which establishes (1.4). □

Proof of Theorem 1.5. Note first that the case $k>n$ is trivial, since $H(\mathcal{T}_n)\le n$ and $Z_k(\mathcal{T}_n)=0$ for $k>n$. Further, if $k\le\sqrt{n}$, then the result follows from (1.8). Hence it suffices to consider $\sqrt{n}\le k\le n$.

Consider the random tree $\widehat{\mathcal{T}}^{(k)}$ constructed in Section 3. By the alternative construction described there, we can regard the tree as the k mutant nodes (the spine except its top node) together with a random number M of attached independent copies of $\mathcal{T}$. Hence,

$$|\widehat{\mathcal{T}}^{(k)}| \overset{\mathrm{d}}{=} k+\sum_{i=1}^{M}|\mathcal{T}_i|, \tag{4.8}$$

where $\mathcal{T}_i$ are independent copies of $\mathcal{T}$, independent also of M. The number M is the total number of normal children (including the top node) of the k mutants, and thus

$$M \overset{\mathrm{d}}{=} \sum_{i=1}^{k}(\hat{\xi}_i-1)+1, \tag{4.9}$$

where $\hat{\xi}_i$ are i.i.d. with the distribution (3.1).

Thus, for $m > 0$ and $n > k$, using (4.8), (2.3) and Lemma 2.1,

$$\mathbb{P}\big(|\widehat{\mathcal{T}}^{(k)}| = n \mid M = m\big) = \mathbb{P}\Big(\sum_{i=1}^{m} |\mathcal{T}_i| = n-k\Big) = \frac{m}{n-k}\,\mathbb{P}\big(S_{n-k} = n-k-m\big)$$
$$\leq C_7 \frac{m}{(n-k)^{3/2}} e^{-c_7 m^2/(n-k)}. \tag{4.10}$$

The summands $\hat{\xi}_i - 1$ in (4.9) have mean $\mathbb{E}(\hat{\xi}-1) = \mathbb{E}\,\xi^2 - 1 = \sigma^2 > 0$. We truncate them and define $\hat{\xi}'_i := \min(\hat{\xi}_i, K)$, where K is chosen so large that $\mathbb{E}\,\hat{\xi}'_i > 1 + \sigma^2/2$. We apply Bernstein's inequality (2.6) to $-\hat{\xi}'_i$, and obtain, since $\operatorname{Var}(\hat{\xi}'_i) < \infty$ and thus $V = O(n)$,

$$\mathbb{P}\big(M \leq k\sigma^2/4\big) \leq \mathbb{P}\left(\sum_{i=1}^{k}(\hat{\xi}_i - 1) \leq k\sigma^2/4\right) \leq \mathbb{P}\left(\sum_{i=1}^{k}(\hat{\xi}'_i - 1) \leq k\sigma^2/4\right)$$
$$\leq \mathbb{P}\left(\sum_{i=1}^{k}(\hat{\xi}'_i - \mathbb{E}\,\hat{\xi}'_i) \leq -k\sigma^2/4\right) \leq e^{-c_{13}k}. \tag{4.11}$$

Note that $|\widehat{\mathcal{T}}^{(k)}| \geq M + k$ by (4.8), so if $M = m > k\sigma^2/2$, we only have to consider $n \geq m + k > (1+\sigma^2/2)k$, and for such n, $n - k \geq c_{14}n$. Hence, for $m > k\sigma^2/2$, (4.10) yields

$$\mathbb{P}\big(|\widehat{\mathcal{T}}^{(k)}| = n \mid M = m\big) \leq C_{14}\frac{m}{n^{3/2}}e^{-c_7 m^2/n} \leq C_{15}\frac{1}{n}e^{-c_{15}m^2/n}. \tag{4.12}$$

If $\sqrt{n} \leq k \leq n$, (4.11) and (4.12) yield

$$\mathbb{P}\big(|\widehat{\mathcal{T}}^{(k)}| = n\big) \leq e^{-c_{13}k} + \max_{m \geq k\sigma^2/2} C_{15}\frac{1}{n}e^{-c_{15}m^2/n} \leq C_{16}\frac{1}{n}e^{-c_{16}k^2/n}. \tag{4.13}$$

Since $\mathbb{P}(|\mathcal{T}| = n) \geq c_{17}n^{-3/2}$ by (2.4), (3.4) and (4.13) yield, if $\sqrt{n} \leq k \leq n$,

$$\mathbb{E}\,Z_k(\mathcal{T}_n) \leq C_{17}n^{1/2}e^{-c_{16}k^2/n} \leq C_{18}ke^{-c_{16}k^2/n}, \tag{4.14}$$

which completes the proof. (We remarked above that it suffices to consider such k.) □

Proof of Theorem 1.6. First, by Theorem 1.1,

$$\mathbb{P}(Z_k(\mathcal{T}_n) > x) \leq \mathbb{P}\big(W(\mathcal{T}_n) > x\big) \leq C_1 e^{-c_1 x^2/n}.$$

Further, since $Z_k(\mathcal{T}_n) > 0$ implies $H(\mathcal{T}_n) \geq k$, Theorem 1.2 implies that

$$\mathbb{P}(Z_k(\mathcal{T}_n) > x) \leq \mathbb{P}\big(H(\mathcal{T}_n) \geq k\big) \leq C_2 e^{-c_2 k^2/n}.$$

Taking the geometric mean of these bounds we obtain (1.13). Further, (1.13) implies, for any $r > 0$, with $\tilde{Z} := Z_k(\mathcal{T}_n)/\sqrt{n}$,

$$\mathbb{E}\,\tilde{Z}^r = r\int_0^\infty x^{r-1}\,\mathbb{P}(\tilde{Z} > x)\,\mathrm{d}x \leq rC_7\,e^{-c_6k^2/n}\int_0^\infty x^{r-1}e^{-c_7x^2}\,\mathrm{d}x$$
$$= C_{19}(r)e^{-c_6k^2/n}.$$

□

Acknowledgements

Part of this work was completed while all three authors were visiting McGill's Bellairs Research Institute, Barbados, and another part while S.J. was visiting the Centre de Recherches Mathématiques in Montréal.

References

[1] D. Aldous, The continuum random tree II: an overview. *Stochastic Analysis (Proceedings of the Durham Symposium on Stochastic Analysis, Durham, 1990)*, London Mathematical Society Lecture Note Series 167, Cambridge University Press, Cambridge, 1991, pp. 23–70.

[2] K. B. Athreya & P. E. Ney, *Branching Processes.* Springer-Verlag, Berlin, 1972.

[3] P. Biane, J. Pitman & M. Yor, Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bulletin of the American Mathematical Society* **38** (2001), 435–465.

[4] N. G. de Bruijn, D. E. Knuth & S. O. Rice, The average height of planted plane trees, *Graph Theory and Computing*, ed. R.-C. Read, Academic Press, New York, 1972, pp. 5–22.

[5] P. Chassaing, J.-F. Marckert & M. Yor, The height and width of simple trees. *Mathematics and Computer Science (Versailles, 2000)*, Trends Math., Birkhäuser, Basel, 2000, pp. 17–30.

[6] K. L. Chung, Maxima in Brownian excursions. *Bulletin of the American Mathematical Society* **81** (1975), 742–744.

[7] L. Devroye, Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic Methods for Algorithmic Discrete Mathematics*, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer-Verlag, Berlin, 1998, pp. 249–314.

[8] C. Donati-Martin, Some remarks about the identity in law for the Bessel bridge $\int_0^1 \frac{ds}{r(s)} \overset{\text{(law)}}{=} 2\sup_{s\le 1} r(s)$. *Studia Scientiarum Mathematicarum Hungarica* **37** (2001), no. 1–2, 131–144.

[9] M. Drmota, *Random Trees*, Springer, Wien, 2009.

[10] M. Drmota & B. Gittenberger, On the profile of random trees. *Random Structures and Algorithms* **10** (1997), no. 4, 421–451.

[11] M. Drmota & B. Gittenberger, The width of Galton–Watson trees conditioned by the size. *Discrete Mathematics and Theoretical Computer Science* **6** (2004), 387–400.

[12] M. Dwass, The total progeny in a branching process and a related random walk. *Journal of Applied Probability* **6** (1969), 682–686.

[13] P. Flajolet, Z. Gao, A. Odlyzko & B. Richmond, The distribution of heights of binary trees and other simple trees. *Combinatorics, Probability and Computing* **2** (1993), no. 2, 145-156.

[14] P. Flajolet & A. Odlyzko, The average height of binary trees and other simple trees. *Journal of Computer and System Sciences* **25** (1982), 171–213.

[15] W. Hoeffding, Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58** (1963), 13–30.

[16] S. Janson, Random cutting and records in deterministic and random trees. *Random Structures and Algorithms* **29** (2006), no. 2, 139–179.

[17] S. Janson, Conditioned Galton–Watson trees do not grow. *Proceedings, Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities (Nancy, 2006)*, DMTCS Proceedings AG, 2006, pp. 331–334.

[18] S. Janson, On the asymptotic joint distribution of height and width in random trees, *Studia Scientiarum Mathematicarum Hungarica* **45** (2008), no. 4, 451–467.

[19] D. P. Kennedy, The Galton–Watson process conditioned on the total progeny. *Journal of Applied Probability* **12** (1975), 800–806.

[20] D. P. Kennedy, The distribution of the maximum Brownian excursion. *Journal of Applied Probability* **13** (1976), 371–376.

[21] G. Kersting, On the height profile of a conditioned Galton-Watson tree. Preprint, 1998.
`http://ismi.math.uni-frankfurt.de/kersting/research/profile.ps`

[22] V. F. Kolchin, *Random Mappings*. Optimization Software, New York, 1986.

[23] J. F. Le Gall, Random trees and applications. *Probability Surveys* **2** (2005), 245–311.

[24] R. Lyons, R. Pemantle & Y. Peres, Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Annals of Probability* **23** (1995), no. 3, 1125–1138.

[25] C. McDiarmid, Concentration. *Probabilistic Methods for Algorithmic Discrete Mathematics*, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer-Verlag, Berlin, 1998, pp. 195–248.

[26] A. Meir & J. W. Moon, On the altitude of nodes in random trees. *Canadian Journal of Mathematics* **30** (1978), 997–1015.

[27] J. Pitman, Enumerations of trees and forests related to branching processes and random walks. *Microsurveys in Discrete Probability (Princeton, NJ, 1997)*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 41, Amer. Math. Soc., Providence, RI, 1998, pp. 163–180.

[28] A. Rényi & G. Szekeres, On the height of trees. *Journal of the Australian Mathematical Society* **7** (1967), 497–507.

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montréal, Québec, H3A 2K6, Canada
E-mail address: louigi@math.mcgill.ca
URL: http://www.math.mcgill.ca/~louigi/

School of Computer Science, McGill University, 3480 University Street, Montréal, Québec, H3A 2A7, Canada
E-mail address: luc@cs.mcgill.ca
URL: http://cg.scs.carleton.ca/~luc/

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
E-mail address: svante.janson@math.uu.se
URL: http://www.math.uu.se/~svante/