ON GENERAL SUBTREES OF A CONDITIONED GALTON–WATSON TREE

SVANTE JANSON

Abstract. We show that the number of copies of a given rooted tree in a conditioned Galton–Watson tree satisfies a law of large numbers under a minimal moment condition on the offspring distribution.

1. Introduction

Let $\mathcal{T}_n$ be a random conditioned Galton–Watson tree with n nodes, defined by an offspring distribution ξ with mean $\mathbb{E}\,\xi = 1$, and let $\mathbf{t}$ be a fixed ordered rooted tree. We are interested in the number of copies of $\mathbf{t}$ as a (general) subtree of $\mathcal{T}_n$, which we denote by $N_{\mathbf{t}}(\mathcal{T}_n)$. For details of these and other definitions, see Section 2. Note that we consider subtrees in a general sense. (Thus, e.g., not just fringe trees; for them, see similar results in [9].)

The purpose of the present paper is to show the following law of large numbers under minimal moment assumptions. Let $n_{\mathbf{t}}(T)$ be the number of rooted copies of $\mathbf{t}$ in a tree T, i.e., copies with the root at the root of T. Further, let $\Delta(\mathbf{t})$ be the maximum outdegree in $\mathbf{t}$.

Theorem 1.1. *Let $\mathbf{t}$ be a fixed ordered tree, and let $\mathcal{T}_n$ be a conditioned Galton–Watson tree defined by an offspring distribution ξ with $\mathbb{E}\,\xi = 1$ and $\mathbb{E}\,\xi^{\Delta(\mathbf{t})} < \infty$. Also, let $\mathcal{T}$ be a Galton–Watson tree with the same offspring distribution. Then, as $n \to \infty$,*

$$N_{\mathbf{t}}(\mathcal{T}_n)/n \xrightarrow{L^1} \mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}), \tag{1.1}$$

where the limit is finite and given explicitly by (3.2) *below.*

Equivalently,

$$N_{\mathbf{t}}(\mathcal{T}_n)/n \xrightarrow{\mathrm{p}} \mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}), \tag{1.2}$$

and

$$\mathbb{E}\, N_{\mathbf{t}}(\mathcal{T}_n)/n \to \mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}). \tag{1.3}$$

The fact that (1.1) is equivalent to (1.2)–(1.3) is an instance of the general fact that for any random variables, convergence in L^1 is equivalent to convergence in probability together with convergence of the means of the absolute values (i.e., in this case, with non-negative variables, the means); see e.g. [6, Theorem 5.5.4]. We nevertheless state both versions for convenience.

Chyzak, Drmota, Klausner and Kok [1] (see also [2, Section 3.3]) considered patterns in random trees; their patterns differ from the subgraph counts above in that some external vertices are added to $\mathbf{t}$, and that one only considers copies

---

Date: 9 November, 2020.
Supported by the Knut and Alice Wallenberg Foundation.

of $\mathbf{t}$ in a tree T such that each internal vertex in the copy has the same degree in T as in $\mathbf{t}$ (counting also edges to external vertices); equivalently, each vertex in $\mathbf{t}$ is equipped with a number, and one considers only copies of $\mathbf{t}$ where the vertex degrees match these numbers. (Another difference is that [1] consider unrooted trees, but the proof proceeds by first considering rooted [planted] trees. Furthermore, only uniformly random labelled trees are considered in [1], but the proofs extend to suitable more general conditioned Galton–Watson trees, as remarked in [1] and shown explicitly in [11; 12].) It was shown in Chyzak, Drmota, Klausner and Kok [1] that the number of occurences of such a pattern is asymptotically normal, with asymptotic mean and variance both of the order n (except that the variance might be smaller in at least one exceptional degenerate case), which of cource entails a law of large numbers. Moreover, [1] discuss briefly generalizations, including subtrees without further degree conditions as in the present paper; they expect asymptotic normality to hold in this case too, but it seems that their method, which is based on setting up and analyzing a system of functional equations for generating functions, in general would require extensions to infinite systems, which as far as we know has not been pursued. (See [4] for a related problem.) See further Section 5.

Our method is probabilistic, and quite different from the analysis of generating functions in [1].

2. Notation

All trees are rooted and ordered. The root of a tree T is denoted $o = o_T$. The size $|T|$ of a tree T is defined as the number of vertices in T.

The *degree* $d(v)$ of a vertex $v \in T$ always means the outdegree, i.e., the number of children of v. The *degree sequence* of T is the sequence of all degrees $d(v)$, $v \in T$, for definiteness in depth first order. Let $\Delta(T) := \max_{v\in T} d(v)$ be the maximum (out)degree in T.

A (general) *subtree* T' of a tree T is a non-empty connected subgraph of T; we regard a subtree as a rooted tree in the obvious way, with the root being the vertex in T' that is closest to the root in T. Note that for any vertex $v \in T'$, its set of children in T' is a subset of its set of children in T; the order of the children of v in T' is (by definition) the same as their relative order in T.

If $v \in T$, the *fringe subtree* T^v is the subtree of T consisting of v and all its descendants; this is thus a subtree with root v.

If $\mathbf{t}$ and T are ordered rooted tree, let $N_{\mathbf{t}}(T)$ be the number of (general) subtrees of T that are isomorphic to $\mathbf{t}$ (as ordered trees), and let $n_{\mathbf{t}}(T)$ be the number of such subtrees that furthermore have root o_T. Then $n_{\mathbf{t}}(T^v)$ is the number of subtrees with root v isomorphic to $\mathbf{t}$, and thus

$$N_{\mathbf{t}}(T) = \sum_{v\in T} n_{\mathbf{t}}(T^v). \tag{2.1}$$

In other words, $N_{\mathbf{t}}(T)$ is an additive functional with toll function $n_{\mathbf{t}}(T)$, see e.g. [9].

Let $\mathcal{T}$ be a random Galton–Watson tree defined by an offspring distribution $(p_i)_0^\infty$, and let $\mathcal{T}_n$ be the conditioned Galton–Watson tree defined as $\mathcal{T}$ conditioned on $|\mathcal{T}| = n$ (tacitly considering only n such that $\mathbb{P}\bigl(|\mathcal{T}| = n\bigr) > 0$); see e.g. [8] for a survey. We let ξ be a random variable with the distribution $(p_i)_0^\infty$; we call both $(p_i)_0^\infty$ and (with a minor abuse) ξ the *offspring distribution.* We will only consider offspring distributions with $\mathbb{E}\,\xi = 1$ (i.e., ξ is *critical*). (We often repeat this for emphasis.) Let $\sigma^2 := \operatorname{Var}\xi \leqslant \infty$; we tacitly assume $\sigma^2 > 0$, but do not require $\sigma^2 < \infty$ unless we say so.

C and c denote unspecified constants that may vary from one occurrence to the next. They may depend on parameters such as the offspring distribution or the fixed tree $\mathbf{t}$, but they never depend on n.

Convergence in probability and distribution is denoted $\overset{\mathrm{p}}{\longrightarrow}$ and $\overset{\mathrm{d}}{\longrightarrow}$, respectively. Unspecified limits are as $n \to \infty$.

3. Proof

We begin by finding the expectation of $n_{\mathbf{t}}$ for both unconditioned and conditioned Galton–Watson trees. Let

$$S_n := \sum_{i=1}^{n} \xi_i, \tag{3.1}$$

where $\xi_1, \xi_2, \dots$ are i.i.d. copies of ξ.

Lemma 3.1. *Let* $\mathbf{t}$ *be a fixed ordered tree with degree sequence* $d_1, \dots, d_k$*, where thus* $k = |T|$.

(i) *Then*

$$\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}) = \prod_{i=1}^{k} \mathbb{E}\binom{\xi}{d_i} = \prod_{i=1}^{k} \sum_{m_i=d_i}^{\infty} p_{m_i}\binom{m_i}{d_i}. \tag{3.2}$$

(ii) *If* $n > k$*, then, with* $m := \sum_{i=1}^{k} m_i$,

$$\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) = \frac{n}{n-k} \sum_{m_1,\dots m_k \geqslant 0} \prod_{i=1}^{k} p_{m_i}\binom{m_i}{d_i} \cdot \frac{(m-k+1)\,\mathbb{P}(S_{n-k} = n-m-1)}{\mathbb{P}(S_n = n-1)}. \tag{3.3}$$

Proof. (i): We try to construct a copy t' of $\mathbf{t}$ in $\mathcal{T}$, with the given root o. Let m_1 be the root degree of $\mathcal{T}$. Then there are $\binom{m_1}{d_1}$ ways to choose the d_1 children of the root that belong to t'. Fix one of these choices, say $v_{11}, \dots, v_{1d_1}$.

Next, let m_2 be the number of children of v_{11} in $\mathcal{T}$. Given m_2, there are $\binom{m_2}{d_2}$ ways to choose the d_2 children of v_{11} that belong to t'. Fix one of these choices.

Continuing in the same way, taking the vertices of t' in depth first order, we find for every sequence $m_1, \dots, m_k$ of non-negative integers, a total of $\prod_1^k \binom{m_i}{d_i}$ choices, and each of these gives a tree $t' \cong t$ provided the selected vertices in $\mathcal{T}$ have degrees $m_1, \dots, m_k$, which occurs with probability $\prod_{i=1}^k p_{m_i}$. Hence,

$$\begin{aligned}\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}) &= \sum_{m_1,\dots m_k \geqslant 0} \prod_{i=1}^{k} p_{m_i} \prod_{i=1}^{k}\binom{m_i}{d_i} = \sum_{m_1,\dots m_k \geqslant 0} \prod_{i=1}^{k}\Bigl(p_{m_i}\binom{m_i}{d_i}\Bigr) \\ &= \prod_{i=1}^{k} \sum_{m_i=0}^{\infty} p_{m_i}\binom{m_i}{d_i},\end{aligned} \tag{3.4}$$

and (3.2) follows.

(ii): Consider again $\mathcal{T}$. We have just shown that each sequence $m_1, \dots, m_k$ gives $\prod_{i=1}^k \binom{m_i}{d_i}$ choices of possible subtrees $t' \cong t$ in $\mathcal{T}$, where the vertices of t'

are supposed to have degrees $m_1, \dots m_k$ in $\mathcal{T}$. This gives a total of $m = \sum_{i=1}^k m_i$ children, of which $k-1$ are the non-root vertices in t', and thus $m-(k-1)$ are unaccounted children. Then, $|\mathcal{T}| = n$ if and only if these $m-k+1$ children and their descendants yield exactly $n-k$ vertices.

Condition on $m_1, \dots, m_k$ and one of the corresponding choices of t'. The probability that the $m-k+1$ children above and their descendants are $n-k$ vertices is the probability that a Galton–Watson process (with offspring distribution ξ) started witk $m-k+1$ individuals has total progeny $n-k$, which by the Otter–Dwass formula [5] (see also [17] and the further references there) is given by

$$\frac{m-k+1}{n-k}\,\mathbb{P}\bigl(S_{n-k} = n-k-(m-k+1)\bigr). \tag{3.5}$$

Multiplying with $\prod_{i=1}^k p_{m_i}$, the probability that the vertices in t' have the right degrees in $\mathcal{T}$, and summing over all possibilities, we obtain

$$\begin{aligned}\mathbb{E}\bigl[n_{\mathbf{t}}(\mathcal{T}_n)\bigr]\,\mathbb{P}\bigl(|\mathcal{T}| = n\bigr) &= \mathbb{E}\bigl[n_{\mathbf{t}}(\mathcal{T}) \mid |\mathcal{T}| = n\bigr]\,\mathbb{P}\bigl(|\mathcal{T}| = n\bigr) = \mathbb{E}\bigl[n_{\mathbf{t}}(\mathcal{T})\mathbf{1}\{|\mathcal{T}| = n\}\bigr] \\ &= \sum_{m_1,\dots m_k \geqslant 0} \prod_{i=1}^k p_{m_i}\binom{m_i}{d_i} \cdot \frac{m-k+1}{n-k}\,\mathbb{P}(S_{n-k} = n-m-1). \end{aligned} \tag{3.6}$$

By the Otter–Dwass formula again (this time the original case in [15]),

$$\mathbb{P}\bigl(|\mathcal{T}| = n\bigr) = \frac{1}{n} P\bigl(S_n = n-1\bigr) \tag{3.7}$$

and (3.3) follows. (Cf. [8, Lemma 15.9] for a related result.) $\square$

We need estimates of the probabilities $\mathbb{P}\bigl(S_n = n-m\bigr)$. The estimate (3.8) below is standard; we expect that also (3.9) is known, but we have not found a reference, so we give a proof. (It is related to more difficult estimates in e.g. [16] assuming more moments, see Remark 3.3 below.)

Lemma 3.2. *Suppose that $\mathbb{E}\,\xi = 1$ and $\mathbb{E}\,\xi^2 < \infty$. Then, uniformly for all $n \geqslant 1$ and $m \in \mathbb{Z}$,*

$$\mathbb{P}\bigl(S_n = n-m\bigr) \leqslant Cn^{-1/2}, \tag{3.8}$$

$$\mathbb{P}\bigl(S_n = n-m\bigr) \leqslant C|m|^{-1}. \tag{3.9}$$

Proof. (3.8): This is well-known. In fact, the classical local limit theorem, see e.g. [16, Theorem VII.1], gives the much more precise result that, uniformly in $m \in \mathbb{Z}$ as $n \to \infty$,

$$\mathbb{P}\bigl(S_n = n-m\bigr) = \frac{h}{\sigma\sqrt{n}}\Bigl(\frac{1}{\sqrt{2\pi}}e^{-m^2/2\sigma^2 n} + o(1)\Bigr). \tag{3.10}$$

where h is the span of the offspring distribution. (Provided $h|(n-m)$; otherwise the probability is 0.)

(3.9): Let $\varphi(t) := \mathbb{E}\,e^{\mathrm{i}t(\xi-1)}$ be the characteristic function of $\xi - 1 = \xi - \mathbb{E}\,\xi$; note that $\varphi(t)$ is twice differentiable because $\mathbb{E}\,\xi^2 < \infty$. Then, by Fourier inversion,

$$\mathbb{P}\bigl(S_n = n-m\bigr) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\mathrm{i}mt}\varphi(t)^n\,\mathrm{d}t. \tag{3.11}$$

Hence, using an integration by parts,

$$2\pi\mathrm{i}m\,\mathbb{P}\bigl(S_n = n-m\bigr) = \int_{-\pi}^{\pi}\Bigl(\frac{\mathrm{d}}{\mathrm{d}t}e^{\mathrm{i}mt}\Bigr)\varphi(t)^n\,\mathrm{d}t = -\int_{-\pi}^{\pi} e^{\mathrm{i}mt}\frac{\mathrm{d}}{\mathrm{d}t}\bigl(\varphi(t)^n\bigr)\,\mathrm{d}t \tag{3.12}$$

and thus

$$|m|\,\mathbb{P}\bigl(S_n = n-m\bigr) \leqslant \int_{-\pi}^{\pi}\Bigl|\frac{\mathrm{d}}{\mathrm{d}t}\bigl(\varphi(t)^n\bigr)\Bigr|\,\mathrm{d}t = n\int_{-\pi}^{\pi}|\varphi'(t)||\varphi(t)|^{n-1}\,\mathrm{d}t. \tag{3.13}$$

The assumptions yield $\varphi'(0) = \mathbb{E}(\xi-1) = 0$ and $\sup|\varphi''(t)| = |\varphi''(0)| = \operatorname{Var}\xi = C < \infty$, and thus

$$|\varphi'(t)| \leqslant Ct. \tag{3.14}$$

Assume for simplicity that the span of ξ is 1 (the general case is similar, with standard modifications). Then, as is well-known, it is easy to see that there exist $c > 0$ such that

$$|\varphi(t)| \leqslant e^{-ct^2}, \qquad |t| \leqslant \pi. \tag{3.15}$$

Using (3.14) and (3.15) in (3.13) we obtain

$$|m|\,\mathbb{P}\bigl(S_n = n-m\bigr) \leqslant nC\int_{-\pi}^{\pi}|t|e^{-c(n-1)t^2}\,\mathrm{d}t \leqslant Cn\int_0^{\infty}te^{-cnt^2}\,\mathrm{d}t = C, \tag{3.16}$$

which proves (3.9). □

Remark 3.3. In the same way, taking two derivatives inside (3.11), one obtains

$$\mathbb{P}\bigl(S_n = n-m\bigr) \leqslant Cn^{1/2}m^{-2}, \tag{3.17}$$

which is stronger for large m; note that (3.8) and (3.17) imply (3.9). Furthermore, even stronger estimates hold if we assume more moments; see [16, Theorem VII.16] for a precise asymptotic estimate assuming $\mathbb{E}\,\xi^k < \infty$ for some $k \geqslant 3$. In fact, [16, Theorem VII.16] holds for $k = 2$ too, which can be seen by refining the argument above; this is an asymptotic estimate that is more precise than (3.17) (and implies it). □

Lemma 3.4. *Let* $\mathbf{t}$ *be a fixed ordered tree and suppose that* $\mathbb{E}\,\xi = 1$, $\mathbb{E}\,\xi^2 < \infty$ *and* $\mathbb{E}\,\xi^{\Delta(\mathbf{t})} < \infty$. *Then* $\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n) = o\bigl(n^{1/2}\bigr)$.

Proof. Let again the degree sequence of $\mathbf{t}$ be $d_1,\dots,d_k$. For a vector $\mathbf{m} = (m_1,\dots,m_k) \in \mathbb{Z}^k_{\geqslant 0}$, let

$$a_{\mathbf{m}} := \prod_{i=1}^{k} p_{m_i}\binom{m_i}{d_i}. \tag{3.18}$$

Then, (3.2)–(3.3) and the assumption $\mathbb{E}\,\xi^{\Delta}(\mathbf{t}) < \infty$ yield

$$\sum_{\mathbf{m}} a_{\mathbf{m}} = \mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}) < \infty \tag{3.19}$$

and for $n > k$, with as above $m := \sum_i m_i =: |\mathbf{m}|$ (and $C = 1$, actually),

$$\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n) \leqslant C\sum_{\mathbf{m}} a_{\mathbf{m}}\cdot\frac{m\,\mathbb{P}(S_{n-k} = n-m-1)}{\mathbb{P}(S_n = n-1)}. \tag{3.20}$$

Denote the summand in (3.20) by $b_{\mathbf{m},n}$. By the local limit theorem (3.10), as is well-known,

$$\mathbb{P}(S_n = n-1) \sim cn^{-1/2}, \tag{3.21}$$

and thus

$$b_{\mathbf{m},n}/n^{1/2} \leqslant Cma_{\mathbf{m}}\,\mathbb{P}(S_{n-k} = n-m-1). \tag{3.22}$$

Hence, (3.8) implies that for every fixed $\mathbf{m}$, as $n\to\infty$,

$$b_{\mathbf{m},n}/n^{1/2} \leqslant Cma_{\mathbf{m}}n^{-1/2} \to 0. \tag{3.23}$$

Furthermore, (3.22) and (3.9) yield

$$b_{\mathbf{m},n}/n^{1/2} \leqslant Ca_{\mathbf{m}}, \tag{3.24}$$

which is summable by (3.19). Consequently, dominated convergence shows that

$$n^{-1/2}\sum_{\mathbf{m}} b_{\mathbf{m},n} = \sum_{\mathbf{m}} b_{\mathbf{m},n}/n^{1/2} \to 0, \tag{3.25}$$

which together with (3.20) yields the result $n^{-1/2}\,\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n)\to 0$. □

We will see in Example 4.4 below, that the estimate $o(n^{1/2})$ in Lemma 3.4 is best possible in general. However, if we assume another moment on ξ, we can improve the estimate to $O(1)$, and furthermore show that $\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n)$ converges. We next show this, although it is not required for our main result.

Lemma 3.5. *Let* $\mathbf{t}$ *be a fixed tree with degree sequence* $d_1,\dots,d_k$*, and suppose that* $\mathbb{E}\,\xi=1$*. Then, as* $n\to\infty$*,*

$$\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n) \to \sum_{i=1}^{k}(d_i+1)\,\mathbb{E}\binom{\xi}{d_i+1}\prod_{j\neq i}\mathbb{E}\binom{\xi}{d_j}. \tag{3.26}$$

In particular, $\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n)=O(1)$ *if* $\mathbb{E}\,\xi^{\Delta(\mathbf{t})+1}<\infty$*, while* $\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n)\to\infty$ *if* $\mathbb{E}\,\xi^{\Delta(\mathbf{t})+1}=\infty$*.*

Proof. Define again $a_{\mathbf{m}}$ by (3.18), and denote the summand in (3.3) by $b'_{\mathbf{m},n}$, where as above $\mathbf{m}=(m_1,\dots,m_k)\in\mathbb{Z}^k_{\geqslant 0}$. It follows from the local limit theorem (3.10) that for every fixed $\mathbf{m}$, as $n\to\infty$,

$$\frac{\mathbb{P}(S_{n-k}=n-m-1)}{\mathbb{P}(S_n=n-1)} = \frac{h(2\pi\sigma^2(n-k))^{-1/2}\bigl(1+o(1)\bigr)}{h(2\pi\sigma^2 n)^{-1/2}\bigl(1+o(1)\bigr)} \to 1. \tag{3.27}$$

(This holds also if the span $h>1$, assuming as we may that all $p_{m_i}>0$, so $h|m$.) Hence,

$$b'_{\mathbf{m},n} \to a_{\mathbf{m}}(m-k+1). \tag{3.28}$$

Furthermore, by (3.8) and (3.21),

$$\frac{\mathbb{P}(S_{n-k}=n-m-1)}{\mathbb{P}(S_n=n-1)} \leqslant \frac{Cn^{-1/2}}{cn^{-1/2}} = C, \tag{3.29}$$

and thus

$$b'_{\mathbf{m},n} \leqslant Ca_{\mathbf{m}}(m-k+1). \tag{3.30}$$

Consequently, if $\sum_{\mathbf{m}} a_{\mathbf{m}}(m-k+1)<\infty$, then

$$\sum_{\mathbf{m}} b'_{\mathbf{m},n} \to \sum_{\mathbf{m}} a_{\mathbf{m}}(m-k+1) \tag{3.31}$$

by (3.28), (3.30) and dominated convergence. On the other hand, if $\sum_{\mathbf{m}} a_{\mathbf{m}}(m-k+1)=\infty$, then $\sum_{\mathbf{m}} b'_{\mathbf{m}}\to\infty$ by (3.28) and Fatou's lemma, and thus (3.31) holds in this case too. Recalling (3.3), this shows that in any case,

$$\mathbb{E}\,n_{\mathbf{t}}(\mathcal{T}_n) \to \sum_{\mathbf{m}} a_{\mathbf{m}}(m-k+1), \tag{3.32}$$

and it remains only to evaluate the limit.

Since $\mathbf{t}$ is a tree, we have $\sum_{i=1}^k d_i = k-1$, and thus $m-k+1 = \sum_{i=1}^k (m_i - d_i)$. Recalling the definition (3.18) of $a_{\mathbf{m}}$, we thus have

$$\begin{aligned}\sum_{\mathbf{m}} a_{\mathbf{m}}(m-k+1) &= \sum_{\mathbf{m}} \sum_{i=1}^{k} (m_i - d_i) p_{m_i} \binom{m_i}{d_i} \prod_{j\neq i} p_{m_j} \binom{m_j}{d_j} \\ &= \sum_{i=1}^{k} \sum_{m_i=0}^{\infty} p_{m_i}(m_i - d_i) \binom{m_i}{d_i} \prod_{j\neq i} \sum_{m_j=0}^{\infty} p_{m_j} \binom{m_j}{d_j}, \end{aligned} \tag{3.33}$$

which equals the right-hand side of (3.26) because $(m_i - d_i)\binom{m_i}{d_i} = (d_i+1)\binom{m_i}{d_i+1}$. This completes the proof by (3.32). ◻

Remark 3.6. Assume only $\mathbb{E}\,\xi = 1$. If $\widehat{T}$ is the infinite size-biased Galton–Watson tree defined by Kesten [10], see also [8, Section 5], then $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{T}$ in a local topology (i.e., close to the root), see [8, Theorem 7.1], and it follows that

$$n_{\mathbf{t}}(\mathcal{T}_n) \xrightarrow{\mathrm{d}} n_{\mathbf{t}}(\widehat{T}). \tag{3.34}$$

It is not difficult to see that $\mathbb{E}\, n_{\mathbf{t}}(\widehat{T})$ equals the right-hand side of (3.26), which thus says that $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) \to \mathbb{E}\, n_{\mathbf{t}}(\widehat{T})$. (This could presumably be used to give an alternative proof of Lemma 3.5, but we prefer the direct proof above.)

In particular, if $\mathbb{E}\,\xi^{\Delta(\mathbf{t})+1} = \infty$, then $\mathbb{E}\, n_{\mathbf{t}}(\widehat{T}) = \infty$, and thus (3.34) and Fatou's lemma yield $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) \to \infty$. Hence, the last sentence in Lemma 3.5 holds also without the assumption $\mathbb{E}\,\xi^2 < \infty$. ◻

We proceed to the proof of Theorem 1.1. The case $\Delta(\mathbf{t}) \leqslant 1$ is special, since we then do not assume $\mathbb{E}\,\xi^2 < \infty$, but on the other hand this case is simple and rather trivial, so we discuss it separately in the following example.

Example 3.7. Consider the case $\Delta(\mathbf{t}) \leqslant 1$. This means that $\mathbf{t}$ is a path P_k with $k \geqslant 1$ vertices, and thus length $k-1$. A copy of $\mathbf{t}$ in a tree T is thus a path consisting of k vertices $v_1, \dots, v_k$ such that v_{i+1} is a child of v_i; such a path is determined by its endpoint v_k, and every vertex of depth (= distance from the root) at least $k-1$ is the endpoint of a copy of $\mathbf{t}$. Hence, if $\nu_i(T)$ is the number of vertices in T of depth i, then

$$N_{\mathsf{P}_k}(T) = \sum_{i\geqslant k-1} \nu_i(T) = |T| - \sum_{i=0}^{k-2} \nu_i(T). \tag{3.35}$$

In particular, $N_{\mathsf{P}_1}(\mathcal{T}_n) = n$ and $N_{\mathsf{P}_2}(\mathcal{T}_n) = n-1$ are deterministic; these are trivially just the numbers of vertices and edges.

Moreover, as said in Remark 3.6, assuming $\mathbb{E}\,\xi = 1$, the random tree $\mathcal{T}_n$ converges locally in distribution as $n \to \infty$, see [8, Theorem 7.1]; in particular each $\nu_i(\mathcal{T}_n)$ converges in distribution (to $\nu_i(\widehat{T})$) and thus $\nu_i(\mathcal{T}_n) = O_{\mathrm{p}}(1)$ (i.e., is bounded in probability). Hence, for every $k \geqslant 1$, (3.35) implies

$$N_{\mathsf{P}_k}(\mathcal{T}_n) = n + O_{\mathrm{p}}(1). \tag{3.36}$$

In particular, $N_{\mathsf{P}_k}(\mathcal{T}_n)$ is more strongly concentrated than the dispersion of order $n^{1/2}$ typically seen in similar statistics, see e.g. Example 4.2 and Section 5. ◻

Proof of Theorem 1.1. Suppose first $\Delta(\mathbf{t}) \leqslant 1$. Then $\mathbf{t} = \mathsf{P}_k$ for some $k \geqslant 1$ and Example 3.7 shows that (3.36) holds, and thus $N_{\mathsf{P}_k}(\mathcal{T}_n)/n \xrightarrow{\mathrm{p}} 1$. Furthermore, (3.2) yields

$$\mathbb{E}\, n_{\mathsf{P}_k}(\mathcal{T}) = (\mathbb{E}\,\xi)^{k-1} = 1, \tag{3.37}$$

and thus (1.2) holds. Moreover, $N_{\mathsf{P}_k}(\mathcal{T}_n)/n \leqslant 1$ by (3.35), and thus dominated convergence applies to (1.2) and yields (1.3) and (1.1), see e.g. [6, Theorems 5.5.4 and 5.5.5].

In the remainder of the proof we may thus assume $\Delta(\mathbf{t}) \geqslant 2$, and thus, in particular, $\mathbb{E}\,\xi^2 < \infty$. (The arguments below use $\mathbb{E}\,\xi^2 < \infty$, but apply to any $\Delta(\mathbf{t})$.)

Lemma 3.1(i) and the assumption $\mathbb{E}\,\xi^{\Delta(\mathbf{t})} < \infty$ show that $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}) < \infty$, and Lemma 3.4 shows $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) = o\big(n^{1/2}\big)$. Hence (1.2) and (1.3) follow by [9, Remark 5.3]. However, since only a sketch of the proof is given in that remark, let us add some details.

First, (1.3) follows by the argument in the proof of [9, Theorem 1.5(i)], adding the factor $n^{1/2}$ at some places.

Next, define for $M > 0$ the truncation $\nu_{\mathbf{t}}^M(T) := n_{\mathbf{t}}(T) \wedge M$ and let $N_{\mathbf{t}}^M(T) := \sum_{v\in T} \nu_{\mathbf{t}}^M(T^v)$ be the corresponding additive functional, cf. (2.1). Let $\varepsilon > 0$. Since $\nu_{\mathbf{t}}^M(\mathcal{T}) \nearrow n_{\mathbf{t}}(\mathcal{T})$ as $M \to \infty$, we can by monotone convergence, and $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}) < \infty$, choose M such that

$$\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}) - \mathbb{E}\,\nu_{\mathbf{t}}^M(\mathcal{T}) < \varepsilon^2. \tag{3.38}$$

We have proved (1.3), and similarly $\mathbb{E}\, N_{\mathbf{t}}^M(\mathcal{T}_n)/n \to \mathbb{E}\,\nu_{\mathbf{t}}^M(\mathcal{T})$ by [9, Theorem 1.3], since $\nu_{\mathbf{t}}^M$ is bounded. Hence, (3.38) implies that for all sufficiently large n,

$$\mathbb{E}\big|N_{\mathbf{t}}(\mathcal{T}_n)/n - N_{\mathbf{t}}^M(\mathcal{T}_n)/n\big| = \mathbb{E}\, N_{\mathbf{t}}(\mathcal{T}_n)/n - \mathbb{E}\, N_{\mathbf{t}}^M(\mathcal{T}_n)/n < \varepsilon^2. \tag{3.39}$$

Furthermore, [9, Theorem 1.3] also yields $N_{\mathbf{t}}^M(\mathcal{T}_n)/n \xrightarrow{\mathrm{p}} \mathbb{E}\,\nu_{\mathbf{t}}^M(\mathcal{T})$. Consequently, using also (3.38) again, (3.39) and Markov's inequality, if n is large,

$$\begin{aligned}
&\mathbb{P}\big(\big|N_{\mathbf{t}}(\mathcal{T}_n)/n - \mathbb{E}\, n_{\mathbf{t}}(\mathcal{T})\big| > 3\varepsilon\big) \\
&\leqslant \mathbb{P}\big(\big|N_{\mathbf{t}}(\mathcal{T}_n)/n - N_{\mathbf{t}}^M(\mathcal{T}_n)/n\big| > \varepsilon\big) + \mathbb{P}\big(\big|N_{\mathbf{t}}^M(\mathcal{T}_n)/n - \mathbb{E}\,\nu_{\mathbf{t}}^M(\mathcal{T})\big| > \varepsilon\big) \\
&\leqslant 2\varepsilon.
\end{aligned} \tag{3.40}$$

Hence, (1.2) holds.

Finally, as said earlier, (1.2) and (1.3) are together equivalent to the L^1 convergence (1.1). □

4. Examples

We give some simple but illuminating examples. Recall also Example 3.7.

Example 4.1. Let $t = t_{q,r}$ consist of two paths with $q+1$ and $r+1$ vertices, joined at the root; here $q, r \geqslant 1$. We have $k = 1 + q + r$ and $d_1 = 2$ while $d_i = 1$ for $i > 1$; thus $\Delta(\mathbf{t}) = 2$. Since $\mathbb{E}\,\xi = 1$, (3.2) yields

$$\mathbb{E}\, n_{t_{q,r}}(\mathcal{T}) = \mathbb{E}\binom{\xi}{2} = \frac{\mathbb{E}\,\xi^2 - 1}{2} = \frac{\sigma^2}{2}. \tag{4.1}$$

Hence, Theorem 1.1 yields, for any $q, r \geqslant 1$,

$$N_{t_{q,r}}(\mathcal{T}_n)/n \xrightarrow{L^1} \sigma^2/2. \tag{4.2}$$

□

Example 4.2. Consider the special case $q = r = 1$ of Example 4.1. Then $t_{1,1}$ is a cherry, i.e., a root with two children. If a vertex v in a tree T has degree $d(v)$, then the number of cherries rooted at v is $\binom{d(v)}{2}$, and thus

$$N_{t_{1,1}}(T) = \sum_{v \in T} \binom{d(v)}{2} = \sum_{r=1}^{\infty} \binom{r}{2} X_r(T), \tag{4.3}$$

where $X_r(T)$ is the number of vertices of degree r in T.

It is known that $X_r(\mathcal{T}_n)/n \xrightarrow{\text{p}} p_r$, see e.g. [8, Theorem 7.11]. Hence, (4.2) (with $q = r = 1$) is what we would get by dividing (4.3) by n and taking the limit inside the sum; if the degree distribution is bounded, the sum is finite so this is rigorous and (4.2) (still with $q = r = 1$) follows from (4.3).

In this case we can say much more than (4.2). It was proved in [13], see also [3], that $X_r(\mathcal{T}_n)$ is asymptotically normal, with

$$\frac{X_r(\mathcal{T}_n) - np_r}{\sqrt{n}} \xrightarrow{\text{d}} N\bigl(0, \gamma_r^2\bigr) \tag{4.4}$$

for some explicit γ_r^2. This was extended to joint convergence for all r in [7], provided $\mathbb{E}\,\xi^3 < \infty$. Hence, at least if ξ is bounded, it follows from (4.3) that $N_{t_{1,1}}(\mathcal{T}_n)$ is asymptotically normal, with

$$\frac{N_{t_{1,1}}(\mathcal{T}_n) - n\sigma^2/2}{\sqrt{n}} \xrightarrow{\text{d}} N\bigl(0, \gamma^2\bigr) \tag{4.5}$$

for some explicit $\gamma^2 \geqslant 0$. There are degenerate cases where $\gamma^2 = 0$. For example, for full binary trees ($\mathbb{P}(\xi = 2) = \mathbb{P}(\xi = 0) = \frac{1}{2}$), all degrees are 0 or 2, and then each $X_r(T)$ is a deterministic function of $|T|$; hence (4.3) shows that $N_{t_{1,1}}(\mathcal{T}_n)$ is deterministic. More generally, the same happens for full m-ary trees, with $\xi \in \{0, m\}$ a.s., for any $m \geqslant 2$. But it can be seen from the covariances given in [7] that $\gamma^2 > 0$ in all other cases with bounded ξ. See further Section 5. □

Example 4.3. Let $\ell \geqslant 1$, and let $\varpi_\ell(T)$ be the number of (undirected) paths of length ℓ in T. For definiteness, we count undirected paths, so this equals the number of unordered pairs (v, w) of vertices of distance ℓ. There are two cases:

(i) v is an ancestor of w, or conversely; the number of such pairs is $N_{\mathsf{P}_\ell}(T)$.
(ii) Neither v nor w is an ancestor of the other. Then v and w are the two leaves in a copy of $t_{q,r}$ with $q, r \geqslant 1$ and $q + r = \ell$. For given q and r, the number of such pairs equals $N_{t_{q,r}}(T)$

Consequently,

$$\varpi_\ell(T) = n_{\mathsf{P}_\ell}(T) + \sum_{q=1}^{\ell-1} N_{t_{q,\ell-q}}(T). \tag{4.6}$$

Hence, Examples 3.7 and 4.1 yield

$$\varpi_\ell(\mathcal{T}_n)/n \xrightarrow{L^1} 1 + (\ell - 1)\frac{\sigma^2}{2}. \tag{4.7}$$

For example, taking $\xi \sim \mathrm{Po}(1)$ we obtain (forgetting the ordering) a uniformly random unordered labelled tree; we have $\sigma^2 = 1$ and thus (4.7) yields

$$\varpi_\ell(\mathcal{T}_n) \xrightarrow{L^1} (\ell + 1)/2. \tag{4.8}$$

Similarly, taking $\xi \sim \mathrm{Ge}(1/2)$ we obtain a uniformly random ordered tree; we have $\sigma^2 = 2$ and thus (4.7) then yields

$$\varpi_\ell(\mathcal{T}_n) \xrightarrow{L^1} \ell. \tag{4.9}$$

Taking $\xi \sim \mathrm{Bi}(2, 1/2)$ we obtain a uniformly random binary tree; we have $\sigma^2 = 1/2$ and thus (4.7) now yields

$$\varpi_\ell(\mathcal{T}_n) \xrightarrow{L^1} (\ell + 3)/4. \tag{4.10}$$

□

The following example shows that the estimate $o(n^{1/2})$ in Lemma 3.4 is best possible.

Example 4.4. For simplicity, let the tree $\mathbf{t}$ be a star, where the root has degree $\Delta \geqslant 2$ and its children are leaves with degree 0. (The argument is easily modified to any tree $\mathbf{t}$ with $\Delta(\mathbf{t}) \geqslant 2$.) Thus $k := |t| = \Delta + 1$. Assume that the span of ξ is 1.

The local limit theorem (3.10) implies that if n is large and $m \leqslant n^{1/2}$, then

$$\mathbb{P}(S_{n-k} = n - m - 1) \geqslant cn^{-1/2}, \tag{4.11}$$

and thus, using (3.21),

$$\mathbb{P}(S_{n-k} = n - m - 1)/\,\mathbb{P}(S_n = n - 1) \geqslant c. \tag{4.12}$$

Hence, by (3.3) and considering there only terms with $m_2 = \dots = m_k = 0$,

$$\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) \geqslant c \sum_{\Delta < m_1 \leqslant n^{1/2}} p_{m_1} \binom{m_1}{\Delta} m_1 \geqslant c \sum_{\Delta < m \leqslant n^{1/2}} p_m m^{\Delta+1}. \tag{4.13}$$

If $\varepsilon > 0$, and we let $p_m = m^{-\Delta-1-\varepsilon}$ for large m, then $\mathbb{E}\,\xi^\Delta < \infty$, and (4.13) yields, for large n,

$$\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) \geqslant c \sum_{\Delta < m \leqslant n^{1/2}} m^{-\varepsilon} \geqslant cn^{(1-\varepsilon)/2}. \tag{4.14}$$

Hence, for any $\varepsilon > 0$, $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n)$ can grow faster than $n^{1/2-\varepsilon}$.

Similarly, we can find an offspring distribution $(p_m)_0^\infty$ satifying the conditions such that $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) = n^{1/2-o(1)}$; we omit the details. Moreover, for any given sequence $\delta(n) \searrow 0$, we can find $(p_m)_0^\infty$ such that $\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_n) \geqslant \delta(n) n^{1/2}$, at least for a subsequence. To see this, take an increasing sequence $(m_j)_1^\infty$ with $\sum_{j=1}^\infty j\delta(m_j^2) < 1$. Let $p_{m_j} := j\delta(m_j^2) m_j^{-\Delta}$, and $p_m = 0$ for all other $m \geqslant 2$, choosing p_0 and p_1 such that $\sum_i p_i = \sum_i ip_i = 1$. Also, let $n_j := m_j^2$. Then (4.13) implies that, for large j,

$$\mathbb{E}\, n_{\mathbf{t}}(\mathcal{T}_{n_j}) \geqslant cp_{m_j} m_j^{\Delta+1} = cjm_j\delta(m_j^2) \geqslant m_j\delta(m_j^2) = n_j^{1/2}\delta(n_j). \tag{4.15}$$

□

5. Asymptotic normality?

We showed in Example 4.2 that if ξ is bounded, then $N_{t_{1,1}}(\mathcal{T}_n)$ is asymptotically normal in the sense that (4.5) holds (although $\gamma^2 = 0$ is possible). In fact, this holds for any fixed tree $\mathbf{t}$.

Proposition 5.1. *Assume that ξ is bounded. Then, for any fixed tree $\mathbf{t}$,*

$$\frac{N_{\mathbf{t}}(\mathcal{T}_n) - n\mu_{\mathbf{t}}}{\sqrt{n}} \overset{\mathrm{d}}{\longrightarrow} N\bigl(0, \gamma_{\mathbf{t}}^2\bigr), \tag{5.1}$$

for $\mu_{\mathbf{t}} := \mathbb{E}\, n_{\mathbf{t}}(\mathcal{T})$ and some $\gamma_{\mathbf{t}}^2 \geqslant 0$.

Proof. This follows from the result by Chyzak, Drmota, Klausner and Kok [1] on patterns discussed in Section 1 (extended to conditioned Galton–Watson trees [1; 11; 12]); the assumption on ξ means that vertex degrees are bounded by some constant, and thus there is a finite number of patterns that correspond to subtrees isomorphic to $\mathbf{t}$; hence $N_{\mathbf{t}}(\mathcal{T}_n)$ is a linear combination of pattern counts, and the result follows from the joint asymptotic normality of the latter. (See also [14] for a special case.)

Alternatively, this is an application of [9, Theorem 1.13]: the functional $n_{\mathbf{t}}$ is local (as defined in [9]) and for trees with degrees bounded by some constant K, $n_{\mathbf{t}}$ is bounded. Hence (5.1) follows from [9, Theorem 1.13]. ☐

We conjecture that this behaviour is typical, and that Proposition 5.1 holds for every ξ with $\mathbb{E}\,\xi = 1$ that satisfies a suitable moment condition. However, it seems that substantial additional work would be required to show this. As said in the introduction, this was briefly discussed in [1], but it seems that the method there requires extensions to infinite systems of functional equations. Similarly, the application of [9, Theorem 1.13] requires $n_{\mathbf{t}}(\mathcal{T}_n)$ to be bounded, which is not the case when ξ is unbounded. It is possible that this may be overcome by truncations and some variance estimates, but again more work is needed. (The extension in [18] applies to the case when $\mathbf{t}$ is a star with root degree Δ (including Example 4.2 with $\Delta = 2$) and $\mathbb{E}\,\xi^{2\Delta+1} < \infty$; this might suggest further extensions.) This problem is thus left for future research.

Note also that there are degenerate cases when the asymptotic variance in (5.1) $\gamma_{\mathbf{t}}^2 = 0$; see Examples 3.7 and 4.2. (Then (5.1) does not give asymptotic normality; only a concentration result.) However, we conjecture that this is an exception, occuring only in a few special cases.

Acknowledgement. I thank Stephan Wagner for helpful comments.

References

[1] Frédéric Chyzak, Michael Drmota, Thomas Klausner & Gerard Kok: The distribution of patterns in random trees. *Combin. Probab. Comput.* **17** (2008), no. 1, 21–59. MR 2376422

[2] Michael Drmota: *Random Trees. An interplay between Combinatorics and Probability.* SpringerWienNewYork, 2009. MR 2484382

[3] Michael Drmota & Bernhard Gittenberger: The distribution of nodes of given degree in random trees. *J. Graph Theory* **31** (1999), no. 3, 227–253. MR 1688949

[4] Michael Drmota, Lander Ramos & Juanjo Rué: Subgraph statistics in subcritical graph classes. *Random Structures Algorithms* **51** (2017), no. 4, 631–673. MR 3718593

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

[6] Allan Gut: *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013. MR 2977961

[7] Svante Janson: Moment convergence in conditional limit theorems. *J. Appl. Probab.* **38** (2001), no. 2, 421–437. MR 1834751

[8] Svante Janson: Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probability Surveys* **9** (2012), 103–252. MR 2908619

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

[10] Harry Kesten: Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.* **22** (1986), no. 4, 425–487. MR 0871905

[11] Gerard Kok: Pattern distribution in various types of random trees. *2005 International Conference on Analysis of Algorithms*, *Discrete Math. Theor. Comput. Sci. Proc.* **AD** (2005), 223–230. MR 2193121

[12] Gerard J. P. Kok: *The distribution of patterns in random trees.* Thesis, Institut für Diskrete Mathematik und Geometrie, TU Wien, Austria, 2005.

[13] Valentin F. Kolchin: *Random Mappings.* Nauka, Moscow, 1984 (Russian). English transl.: Optimization Software, New York, 1986. MR 0865130

[14] Xueliang Li & Yiyang Li: The asymptotic behavior of the Estrada index for trees. *Bull. Malays. Math. Sci. Soc. (2)* **36** (2013), no. 1, 97–106. MR 2989291

[15] Richard Otter: The multiplicative process. *Ann. Math. Statistics* **20** (1949), 206–224. MR 0030716

[16] Valentin V. Petrov: *Sums of Independent Random Variables.* Springer-Verlag, Berlin, 1975. MR 0388499

[17] Jim 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. MR 1630413

[18] Dimbinaina Ralaivaosaona, Matas Šileikis & Stephan Wagner: A central limit theorem for almost local additive tree functionals. *Algorithmica* **82** (2020), no. 3, 642–679. MR 4058420

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: `svante.janson@math.uu.se`

URL: `http://www2.math.uu.se/∼svante/`