Asymptotic Normality of Almost Local Functionals in Conditioned Galton-Watson Trees

Dimbinaina Ralaivaosaona[1]
Department of Mathematical Sciences, Stellenbosch University, Stellenbosch 7602, South Africa
naina@sun.ac.za

Matas Šileikis[2]
The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic. With institutional support RVO:67985807.
matas.sileikis@gmail.com

Stephan Wagner[3]
Department of Mathematical Sciences, Stellenbosch University, Stellenbosch 7602, South Africa
swagner@sun.ac.za

Abstract

An additive functional of a rooted tree is a functional that can be calculated recursively as the sum of the values of the functional over the branches, plus a certain toll function. Janson recently proved a central limit theorem for additive functionals of conditioned Galton-Watson trees under the assumption that the toll function is local, i.e. only depends on a fixed neighbourhood of the root. We extend his result to functionals that are almost local, thus covering a wider range of functionals. Our main result is illustrated by two explicit examples: the (logarithm of) the number of matchings, and a functional stemming from a tree reduction process that was studied by Hackl, Heuberger, Kropf, and Prodinger.

2012 ACM Subject Classification Mathematics of computing → Trees

Keywords and phrases Galton-Watson trees, central limit theorem, additive functional

Digital Object Identifier 10.4230/LIPIcs.AofA.2018.33

1 Introduction

A functional F that associates a value $F(T)$ with every rooted tree is said to be *additive* if it satisfies a recursion of the form

$$F(T) = \sum_{i=1}^{k} F(T_i) + f(T), \tag{1}$$

where $T_1, T_2, \ldots, T_k$ are the branches of T and f is a so-called “toll function”, another function that assigns a value to every rooted tree. If T only consists of the root (so that $k = 0$), we interpret the empty sum as 0 and set $F(T) = f(T)$. Of course, every functional F is additive in this sense (for a suitable choice of f), so the usefulness of the concept depends on what is known about the toll function f.

[1] The first author was partially supported by the Division for Research Development (DRD) of Stellenbosch University.
[2] The second author was supported by the Czech Science Foundation, grant number GJ16-07822Y.
[3] The third author is supported by the National Research Foundation of South Africa under grant number 96236.

29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2018).
Editors: James Allen Fill and Mark Daniel Ward; Article No. 33; pp. 33:1–33:17
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

An important special case is the number of occurrences of a prescribed “fringe subtree”. A fringe subtree is an induced subtree of a rooted tree that consists of one of the nodes and all its descendants. Now fix a rooted tree S. We say that S occurs on the fringe of T if there is a fringe subtree of T that is isomorphic to S. The number of occurrences of S as a fringe subtree in T (i.e., the number of nodes v of T for which the fringe subtree rooted at v is isomorphic to S) is an additive functional, which we shall denote by $F_S(T)$. Indeed, one has

$$F_S(T) = \sum_{i=1}^{k} F(T_i) + f_S(T),$$

where

$$f_S(T) = \begin{cases} 1 & S \text{ is isomorphic to } T, \\ 0 & \text{otherwise.} \end{cases}$$

This is because an occurrence of S in T is either an occurrence in one of the branches, or comprises the entire tree T. Every additive functional can be expressed as a linear combination of these elementary functionals: it is easy to see (e.g. by induction) that a functional satisfying (1) can be expressed as

$$F(T) = \sum_{S} f(S) F_S(T).$$

Functionals of the form F_S are known to be asymptotically normally distributed in different classes of trees, notably simply generated trees/Galton-Watson trees [6, 14], which will also be the topic of this paper, and classes of increasing trees [4, 10]. In view of this and several other important examples of additive functionals that satisfy a central limit theorem, general schemes have been devised that yield a central limit theorem under different technical assumptions. This includes work on simply generated trees/Galton-Watson trees [6, 14] (labelled trees, plane trees and d-ary trees are well-known special cases) as well as Pólya trees [14] and increasing trees [10, 14] (specifically recursive trees, d-ary increasing trees and generalised plane-oriented recursive trees). It is worth mentioning, however, that there are also many instances of additive functionals that are not normally distributed in the limit, since the toll functions can be quite arbitrary. A well-known example is the case of the path length, i.e. the sum of the distances of all nodes to the root. It satisfies (1) with toll function

$$f(T) = |T| - 1,$$

and, when suitably normalised, its limiting distribution for simply generated trees is the Airy distribution (see [11]).

Previous results [4, 6, 10, 14], while giving rather general conditions on the toll function that imply normality, are unfortunately still insufficient to cover all possible examples one might be interested in. This paper is essentially an extension of Janson’s work [6] on local functionals. By weakening the conditions he makes on the toll functions, we arrive at a new general central limit theorem that can be applied to a variety of examples that were not previously covered. Two such examples are presented in detail in this extended abstract: one is concerned with the number of matchings of a tree, the other settles an open problem from a paper of Hackl, Heuberger, Kropf and Prodinger [3] on tree reductions.

A *local* functional (as considered in Janson’s paper [6]) is a functional for which the value of the toll function can be determined from the knowledge of a fixed neighbourhood of the root. A typical example is the number of nodes with a given outdegree: the corresponding toll function (whose value is either 0 or 1) is completely determined by the root degree. We relax this condition somewhat (to what we call "almost local functionals") in our main theorem. Intuitively speaking, functionals that satisfy our conditions have toll functions that can be approximated well from knowledge of a neighbourhood of the root, with the approximation getting better the wider the neighbourhood is chosen.

The model of random trees that we consider here are *conditioned Galton-Watson trees*: these are determined by an offspring distribution ξ, which we will assume to be normalised to satisfy $\mathbb{E}\xi = 1$. We also assume that $\mathrm{Var}\xi$ is finite and nonzero (to avoid a degenerate case). The Galton-Watson process starts from a single node, the root. At time t, all nodes at level t (distance t from the root) generate a number of children according to the offspring distribution ξ. The numbers of children of different nodes on the same level are mutually independent. The outcome of this process, which ends when all nodes at level t generate 0 children, is a random tree $\mathcal{T}$ (almost surely finite). By conditioning the process to "die out" when the total number of nodes is n, we obtain a conditioned Galton-Watson tree, which will be denoted by $\mathcal{T}_n$.

Conditioned Galton-Watson trees are known to be essentially equivalent to so-called *simply generated trees* [2, Section 3.1.4]. Classical examples include rooted labelled trees (corresponding to a Poisson distribution for ξ), plane trees (corresponding to a geometric distribution for ξ) and binary trees (with a distribution whose support is $\{0, 2\}$).

We conclude the introduction with some more notation: for a tree T, we let $T^{(M)}$ be its restriction to the first M levels, i.e. all nodes whose distance to the root is at most M. A local functional as defined above is thus a functional for which the value of $f(T)$ is determined by $T^{(M)}$ for some fixed M (the "cut-off"). The conditioned Galton-Watson tree $\mathcal{T}_n$ is known to converge in the local topology induced by these restrictions to the (infinite) *size-biased* Galton-Watson tree $\hat{\mathcal{T}}$ as defined by Kesten [8]: one has

$$\mathbb{P}(\hat{\mathcal{T}}^{(M)} = T) = w_M(T)\mathbb{P}(\mathcal{T}^{(M)} = T)$$

for all trees T, where $w_M(T)$ is the number of nodes of depth M in T.

For a rooted tree T (possibly infinite), we let $\deg(T)$ denote the degree of the root of T. Finally, it will be convenient for us to use the Vinogradov notation $\ll$ interchangeably with the O-notation, i.e. $f(n) \ll g(n)$ and $f(n) = O(g(n))$ both mean that $|f(n)| \le Kg(n)$ for a fixed positive constant K and all sufficiently large n.

2 The general theorem

Let us now formulate our main result, which is a central limit theorem for additive functionals under suitable technical conditions on the toll function f.

▶ **Theorem 1.** *Let $\mathcal{T}_n$ be a conditioned Galton-Watson tree of order n with offspring distribution ξ, where ξ satisfies $\mathbb{E}\xi = 1$ and $0 < \sigma^2 := \mathrm{Var}\xi < \infty$. Assume further that $\mathbb{E}\xi^{2\alpha+1} < \infty$ for some integer $\alpha \ge 0$. Consider a functional f of finite rooted ordered trees with the property that there is an absolute constant $C_0 > 0$ such that*

$$|f(T)| \le C_0 \deg(T)^{\alpha}. \tag{2}$$

Furthermore, let $(p_M)_{M\ge1}$ be a sequence of positive real numbers with $p_M \to 0$, and assume that f satisfies the following:

- *for every $M \in \{1, 2, \dots\}$,*

$$\mathbb{E}\left|f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)})\,|\,\hat{\mathcal{T}}^{(M)}\right)\right| \le p_M \tag{3}$$

uniformly in N, with $N \ge M$,

- *there is a sequence of positive integers* $(M_n)_{n\geq 1}$ *such that for large enough* n,

$$\mathbb{E}\left|f(\mathcal{T}_n) - f\left(\mathcal{T}_n^{(M_n)}\right)\right| \leq p_{M_n}. \tag{4}$$

If $a_n := n^{-1/2}(n^{\max\{\alpha,1\}}p_{M_n} + M_n^2)$ *satisfies*

$$a_n \to 0, \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{a_n}{n} < \infty, \tag{5}$$

then

$$\frac{F(\mathcal{T}_n) - n\mu}{\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \gamma^2) \tag{6}$$

where $\mu = \mathbb{E}f(\mathcal{T})$, *and* $0 \leq \gamma < \infty$.

▶ Remark. The proof of the above theorem is a generalisation of Janson's proof of his theorem for bounded and local functionals in [6]. By slightly weakening the condition on the offspring distribution ξ, we are able to reduce the boundedness condition to (2). However, the main difficulty to overcome is the fact that our toll function is no longer local. To give a simple example, an essential part of the proof is the existence of the expectation $\mathbb{E}f(\hat{\mathcal{T}})$. When f is local with a cut-off M, then $f(\hat{\mathcal{T}}) := f(\hat{\mathcal{T}}^{(M)})$. So, $\mathbb{E}f(\hat{\mathcal{T}})$ is simply defined to be $\mathbb{E}f(\hat{\mathcal{T}}^{(M)})$. In our case, where f is not necessarily local, we can define

$$\mathbb{E}f(\hat{\mathcal{T}}) := \lim_{M\to\infty} \mathbb{E}f(\hat{\mathcal{T}}^{(M)}), \tag{7}$$

which may not exist in general. However, if f satisfies (3), then we can show that $\mathbb{E}f(\hat{\mathcal{T}})$ exists. Indeed,

$$\begin{aligned}|\mathbb{E}f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(N)})| &= \left|\mathbb{E}\left(f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)}) \,|\, \hat{\mathcal{T}}^{(M)}\right)\right)\right| \\ &\leq \mathbb{E}\left|f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)}) \,|\, \hat{\mathcal{T}}^{(M)}\right)\right| \leq p_M,\end{aligned}$$

which tends to zero as $M \to \infty$, uniformly for $N \geq M$. In other words, $(\mathbb{E}f(\hat{\mathcal{T}}^{(M)}))_{M\geq 1}$ is a Cauchy sequence, so the limit (7) exists.

Throughout the rest of the paper, the offspring distribution ξ is assumed to satisfy $\mathbb{E}\xi = 1$, $\mathbb{P}(\xi = 0) > 0$, $0 < \sigma^2 := \mathrm{Var}\xi < \infty$, and $\mathbb{E}\xi^{2\alpha+1} < \infty$ for some fixed integer $\alpha \geq 0$. The distribution of the number of nodes at level k, w_k, for the three random trees $\mathcal{T}$, $\hat{\mathcal{T}}$, and $\mathcal{T}_n$ will play an important role in our proof. This parameter has been studied in [5], and in particular, the following results were proved there: for every positive integer $r \leq \max\{2\alpha, 1\}$, we have

$$\mathbb{E}\left(w_k(\mathcal{T})^r\right) = O(k^{r-1}),\ \mathbb{E}(w_k(\hat{\mathcal{T}})^r) = O(k^r),\ \text{and}\ \mathbb{E}\left(w_k(\mathcal{T}_n)^r\right) = O(k^r), \tag{8}$$

where the constants in the O-terms depend on the offspring distribution ξ only. Moreover, for a rooted tree T, we know that $|T^{(M)}| = \sum_{k=0}^{M} w_k(T)$. Hence, we can deduce from the estimates in (8), for $r = 1$, that

$$\mathbb{E}|\mathcal{T}^{(M)}| = O(M),\ \mathbb{E}|\hat{\mathcal{T}}^{(M)}| = O(M^2),\ \text{and}\ \mathbb{E}|{\mathcal{T}_n}^{(M)}| = O(M^2). \tag{9}$$

In fact, it can be shown that $\mathbb{E}|\mathcal{T}^{(M)}| = M + 1$. We are also going to make extensive use of the higher moments of the root degree. By definition, the distribution of $\deg(\mathcal{T})$ is ξ, so we know the higher moments of $\deg(\mathcal{T})$. On the other hand, note that $\deg(T) = w_1(T)$. So, as particular cases of the estimates in (8), we have

$$\mathbb{E}(\deg(\hat{\mathcal{T}})^r) < \infty \text{ and } \mathbb{E}\left(\deg(\mathcal{T}_n)^r\right) = O(1), \tag{10}$$

for every positive integer $r \leq \max\{2\alpha, 1\}$, where the implied constant in the second estimate is independent of n.

3 Mean and variance

We first look at the expectation $\mathbb{E}f(\mathcal{T}_n)$. As it is also the case in [6], one of the key observations in the proof of Theorem 1 is the fact that $\mathbb{E}f(\mathcal{T}_n)$ is asymptotically equal to $\mathbb{E}f(\hat{\mathcal{T}})$ (which is finite, cf. Remark 2) with an explicit bound on the error term. This is made precise in the following lemma:

▶ **Lemma 2.** *If f satisfies the conditions of Theorem 1, then*

$$\mathbb{E}f(\mathcal{T}_n) = \mathbb{E}f(\hat{\mathcal{T}}) + O(p_{M_n} + n^{-1/2} M_n^2). \tag{11}$$

Proof (sketch). We let M_n be defined as in Theorem 1, but write $M = M_n$ for easy reading. Notice first that

$$\begin{aligned}|\mathbb{E}f(\mathcal{T}_n) - \mathbb{E}f(\hat{\mathcal{T}})| \\ \leq |\mathbb{E}f(\mathcal{T}_n) - \mathbb{E}f({\mathcal{T}_n}^{(M)})| + |\mathbb{E}f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}f(\hat{\mathcal{T}})| + |\mathbb{E}f({\mathcal{T}_n}^{(M)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})|.\end{aligned} \tag{12}$$

The first term on the right side is at most p_M by assumption (4). The second term is also bounded above by p_M in view of (3), using the same argument as in Remark 2: we have

$$\begin{aligned}|\mathbb{E}f(\hat{\mathcal{T}}^{(N)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})| &= \left|\mathbb{E}\left(f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)})|\hat{\mathcal{T}}^{(M)}\right)\right)\right| \\ &\leq \mathbb{E}\left|f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)})|\hat{\mathcal{T}}^{(M)}\right)\right| \leq p_M,\end{aligned}$$

uniformly for $N \geq M$. Therefore,

$$|\mathbb{E}f(\hat{\mathcal{T}}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})| = \lim_{N\to\infty} |\mathbb{E}f(\hat{\mathcal{T}}^{(N)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})| \leq p_M.$$

The estimate of the term $|\mathbb{E}f({\mathcal{T}_n}^{(M)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})|$ is rather technical and therefore given in the appendix. It can be shown, using the bound (2), that

$$|\mathbb{E}f({\mathcal{T}_n}^{(M)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})| = O\left(n^{-1/2} M^2 \mathbb{E}(\deg(\hat{\mathcal{T}})^{\alpha+1}) + n^{-1} M^2 \mathbb{E}(\deg(\mathcal{T}_n)^{\alpha+1})\right). \tag{13}$$

In view of (10), the moment $\mathbb{E}(\deg(\hat{\mathcal{T}})^{\alpha+1})$ is finite and $\mathbb{E}(\deg(\mathcal{T}_n)^{\alpha+1})$ is $O(1)$. Therefore, we conclude that

$$|\mathbb{E}f(\mathcal{T}_n) - \mathbb{E}f(\hat{\mathcal{T}})| \ll p_M + n^{-1/2} M^2 = p_{M_n} + n^{-1/2} {M_n}^2,$$

which is equivalent to the statement in the lemma. ◀

Lemma 2 is already enough to show that $\mathbb{E}F(\mathcal{T}_n) = \mu n + o(\sqrt{n})$, where $\mu = \mathbb{E}f(\mathcal{T})$, by simply applying Part (i) of [6, Theorem 1.5] to the shifted toll function $f(T) - \mathbb{E}f(\hat{\mathcal{T}})$. Next, we estimate the variance of $F(\mathcal{T}_n)$.

▶ **Lemma 3.** *Assume that f satisfies the conditions of Theorem 1. Moreover, set $a_k = k^{-1/2}(k^{\max\{\alpha,1\}}\, p_{M_k} + M_k^2)$ (as in Theorem 1) and $\mu_k = \mathbb{E}f(\mathcal{T}_k)$. Moreover, set $N = \min\{|T| : f(T) \neq 0\}$. Then we have*

$$n^{-1/2}\mathrm{Var}\left(F(\mathcal{T}_n)\right)^{1/2} \ll \left(\sup_{k\geq N} a_k + \sum_{k=N}^{\infty} \frac{a_k}{k}\right)^{1/2} + \sup_{k\geq N}|\mu_k| + \sum_{k=N}^{\infty}\frac{|\mu_k|}{k}. \tag{14}$$

Proof (sketch). We follow the proof of [6, Theorem 6.12]. We start with a decomposition $f(T) = f'(T) + f''(T)$, where $f'(T) = f(T) - \mu_{|T|}$ and $f''(T) = \mu_{|T|}$. In view of Minkowski's inequality $\mathrm{Var}(X+Y)^{1/2} \leq \mathrm{Var}(X)^{1/2} + \mathrm{Var}(Y)^{1/2}$, it suffices to check that (14) holds for the following cases:

(i) if $f(T) = \mu_{|T|}$, that is, f depends on $|T|$ only,

(ii) when $\mathbb{E}f(\mathcal{T}_k) = 0$ for every k.

Case (i) works precisely as in [6, Theorem 6.7] and gives a bound

$$\mathrm{Var}\left(F(\mathcal{T}_n)\right)^{1/2} \ll n^{1/2}\left(\sup_{k\geq N}|\mu_k| + \sum_{k=N}^{\infty}\frac{|\mu_k|}{k}\right). \tag{15}$$

The contribution from $k < N$ is zero, since $\mu_k = 0$ for $k < N$. So we only consider Case (ii), where $\mathbb{E}f(\mathcal{T}_k) = 0$ for every k. By [6, (6.28)], we have

$$\frac{1}{n}\mathrm{Var}\left(F(\mathcal{T}_n)\right) \leq 2\sum_{k=N}^{n}\frac{\mathbb{P}\left(S_{n-k} = n-k\right)}{\mathbb{P}\left(S_n = n-1\right)}\pi_k\mathbb{E}(f(\mathcal{T}_k)F(\mathcal{T}_k)), \tag{16}$$

where $\pi_k = \mathbb{P}\left(|\mathcal{T}| = k\right)$, and S_k is the sum of k independent copies of ξ. From [6, Lemma 5.2], we know that

$$\frac{\mathbb{P}\left(S_{n-k} = n-k\right)}{\mathbb{P}\left(S_n = n-1\right)} \ll \frac{n^{1/2}}{(n-k+1)^{1/2}},$$

uniformly for $1 \leq k \leq n$. Recalling that $\pi_k = O(k^{-3/2})$, which can also be found in [6], we obtain

$$\frac{1}{n}\mathrm{Var}\left(F(\mathcal{T}_n)\right) \ll \sum_{k=N}^{n}\frac{n^{1/2}}{(n-k+1)^{1/2}\,k^{3/2}}\mathbb{E}(f(\mathcal{T}_k)F(\mathcal{T}_k)). \tag{17}$$

So it remains to estimate $\mathbb{E}(f(\mathcal{T}_k)F(\mathcal{T}_k))$. It can be shown (see appendix) that

$$\mathbb{E}(f(\mathcal{T}_k)F(\mathcal{T}_k)) \ll k^{\max\{\alpha,\,1\}}\, p_{M_k} + \mathbb{E}(\deg(\mathcal{T}_k)^{2\alpha}) + M_k^2\,\mathbb{E}(\deg(\mathcal{T}_k)^{\alpha+1}). \tag{18}$$

Once again, by means of the second estimate in (10), $\mathbb{E}(\deg(\mathcal{T}_k)^{2\alpha})$ and $\mathbb{E}(\deg(\mathcal{T}_k)^{\alpha+1})$ are both bounded above by constants. Thus, we have

$$\mathbb{E}(f(\mathcal{T}_k)F(\mathcal{T}_k)) \ll k^{\max\{\alpha,\,1\}}\, p_{M_k} + M_k^2 = k^{1/2}a_k, \tag{19}$$

where a_k is defined as in Theorem 1. Applying (19) to (17), we get

$$\frac{1}{n}\mathrm{Var}F(\mathcal{T}_n) \ll \sum_{k=N}^{n}\frac{n^{1/2}\,a_k}{(n-k+1)^{1/2}\,k} \ll \sum_{k=N}^{n/2}\frac{a_k}{k} + \sup_{k\geq n/2}a_k \sum_{n/2\leq k\leq n}\frac{1}{(n-k+1)^{1/2}\,n^{1/2}} \tag{20}$$

Noting that the last sum on the right side is $O(1)$, the result follows by applying Minkowski's inequality to combine the results from the two cases. ◀

4 Central limit theorem

We use a truncation argument as in the proof of [6, Theorem 1.5]. This is formulated in the following lemma:

▶ **Lemma 4.** *Let $(X_n)_{n\geq 1}$ and $(W_{N,n})_{N,n\geq 1}$ be sequences of centred random variables. If we have*

- $W_{N,n} \xrightarrow{d}_n W_N$*, and* $W_N \xrightarrow{d}_N W$*, for some random variables* $W, W_1, W_2, \dots$
- $\mathrm{Var}(X_n - W_{N,n}) = O(\sigma_N^2)$ *uniformly in* n*, and* $\sigma_N^2 \to_N 0$*,*

then $X_n \xrightarrow{d}_n W$*.*

This lemma is a simple consequence of [7, Theorem 4.28] or [1, Theorem 4.2].

Proof of Theorem 1. We may assume, without loss of generality, that $\mathbb{E}f(\hat{\mathcal{T}}) = 0$, by subtracting $\mathbb{E}f(\hat{\mathcal{T}})$ from f if it is not zero, because shifting f by a constant will only add a deterministic term in $F(\mathcal{T}_n)$. For each k, let μ_k denote the expectation $\mathbb{E}f(\mathcal{T}_k)$. By Lemma 2, we have

$$|\mu_k| = |\mathbb{E}f(\mathcal{T}_k)| \ll p_{M_k} + k^{-1/2}M_k^2 \leq a_k. \tag{21}$$

For a positive integer N, let $f^{(N)}$ be the truncated functional defined by $f^{(N)}(T) = f(T)\,\mathbb{I}_{\{|T|<N\}}$ and $F^{(N)}$ be the additive functional associated to the toll function $f^{(N)}$. It is important to notice that $f^{(N)}$ is local, for any fixed N. So, if f satisfies the conditions of Theorem 1, then $f^{(N)}$ also satisfies the conditions of Theorem 1. Note further that $\mathbb{E}f^{(N)}(\mathcal{T}_k) = \mu_k$ if $k < N$, and zero otherwise. Hence, we have $|\mathbb{E}f^{(N)}(\mathcal{T}_k)| \leq |\mu_k|$ for every positive integer N. Let

$$W_{N,n} := \frac{F^{(N)}(\mathcal{T}_n) - \mathbb{E}F^{(N)}(\mathcal{T}_n)}{\sqrt{n}}, \text{ and } X_n := \frac{F(\mathcal{T}_n) - \mathbb{E}F(\mathcal{T}_n)}{\sqrt{n}}.$$

Since $f^{(N)}$ has finite support, by [6, Theorem 1.5], we have

$$W_{N,n} \xrightarrow{d}_n \mathcal{N}(0, \gamma_N^2),$$

where

$$\begin{aligned}\gamma_N^2 &= \lim_{n\to\infty} n^{-1}\mathrm{Var}(F^{(N)}(\mathcal{T}_n)) \\ &= 2\mathbb{E}\left(f^{(N)}(\mathcal{T})\,(F^{(N)}(\mathcal{T}) - |\mathcal{T}|\mu^{(N)})\right) - \mathrm{Var} f^{(N)}(\mathcal{T}) - \frac{(\mu^{(N)})^2}{\sigma^2},\end{aligned}$$

and $\mu^{(N)} = \mathbb{E}f^{(N)}(\mathcal{T})$.

Next we need to show that $\lim_{N\to\infty}\gamma_N$ exists. To that end, we take an arbitrary integer $M \geq N$. We have

$$\gamma_M - \gamma_N = \lim_{n\to\infty} n^{-1/2}\left(\mathrm{Var}(F^{(M)}(\mathcal{T}_n))^{1/2} - \mathrm{Var}(F^{(N)}(\mathcal{T}_n))^{1/2}\right)$$

If we apply Minkowski's inequality to the random variables $F^{(M)}(\mathcal{T}_n) - F^{(N)}(\mathcal{T}_n)$ and $F^{(N)}(\mathcal{T}_n)$, we obtain

$$\mathrm{Var}(F^{(M)}(\mathcal{T}_n))^{1/2} \leq \mathrm{Var}\Big(F^{(M)}(\mathcal{T}_n) - F^{(N)}(\mathcal{T}_n)\Big)^{1/2} + \mathrm{Var}(F^{(N)}(\mathcal{T}_n))^{1/2}.$$

Consequently,

$$|\gamma_M - \gamma_N| = \lim_{n\to\infty} n^{-1/2} |\mathrm{Var}(F^{(M)}(\mathcal{T}_n))^{1/2} - \mathrm{Var}(F^{(N)}(\mathcal{T}_n))^{1/2}|$$
$$\le \limsup_{n\to\infty} n^{-1/2} \mathrm{Var}\Big(F^{(M)}(\mathcal{T}_n) - F^{(N)}(\mathcal{T}_n)\Big)^{1/2}.$$

The toll function associated to the functional $F^{(M)} - F^{(N)}$ is $f^{(M)} - f^{(N)}$, which is zero for all trees of order smaller than N. Hence, the idea of Lemma 3 can be used to estimate the variance $\mathrm{Var}(F^{(M)}(\mathcal{T}_n) - F^{(N)}(\mathcal{T}_n))^{1/2}$, and we obtain

$$|\gamma_M - \gamma_N| \ll \left(\sup_{k\ge N} a_k + \sum_{k=N}^{\infty} \frac{a_k}{k}\right)^{1/2} + \sup_{k\ge N} |\mu_k| + \sum_{k=N}^{\infty} \frac{|\mu_k|}{k}$$
$$\ll \left(\sup_{k\ge N} a_k + \sum_{k=N}^{\infty} \frac{a_k}{k}\right)^{1/2} + \sup_{k\ge N} a_k + \sum_{k=N}^{\infty} \frac{a_k}{k}.$$

The last line follows from (21). By the condition (5) of Theorem 1, we also deduce that $|\gamma_M - \gamma_N| \to_N 0$ uniformly for $M \ge N$. Hence, the sequence $(\gamma_N)_N$ is a Cauchy sequence, which implies that $\gamma := \lim_{N\to\infty} \gamma_N$ exists.

Similarly, we have

$$\mathrm{Var}(X_n - W_{N,n})^{1/2} = n^{-1/2} \mathrm{Var}(F(\mathcal{T}_n) - F^{(N)}(\mathcal{T}_n))^{1/2}$$
$$\ll \left(\sup_{k\ge N} a_k + \sum_{k=N}^{\infty} \frac{a_k}{k}\right)^{1/2} + \sup_{k\ge N} a_k + \sum_{k=N}^{\infty} \frac{a_k}{k},$$

which tends to zero as $N \to \infty$ uniformly in n, so Lemma 4 applies and the proof of Theorem 1 is complete. ◀

5 Examples

In this section, we give two representative applications of our main theorem (further examples will be provided in the full version). The absolute values of the toll functions in both examples are not bounded by positive constants, but they are both bounded above by the root degree. Hence, we need α to be at least 1, i.e. $\mathbb{E}\xi^3 < \infty$.

5.1 The number of matchings

The number of matchings in random trees has been studied previously, and means and variances have been determined for different classes of trees [9, 12, 13]. However, in order to obtain a limiting distribution, one has to consider the logarithm of this quantity. For a rooted tree T, let $m(T)$ be the total number of matchings of T and $m_0(T)$ be the number of matchings of T that do not cover the root (by this, we mean matchings that do not contain an edge incident to the root). It is easy to see that these parameters can be determined recursively in the following way:

$$m_0(T) = \prod_i m(T_i), \tag{22}$$

$$m(T) = m_0(T) + \sum_i m_0(T_i) \prod_{j\ne i} m(T_j). \tag{23}$$

Defining an additive functional $F(T) := \log m(T)$, we observe from (22) that the associated toll function is

$$f(T) = F(T) - \sum_i F(T_i) = \log m(T) - \sum_i \log m(T_i) = -\log\left(\frac{m_0(T)}{m(T)}\right).$$

It is convenient to define $\rho(T) = \frac{m_0(T)}{m(T)}$, which is the probability that a random matching does not cover the root, when all matchings are equally likely. By (22) and (23), $\rho(T)$ also satisfies a recursion

$$\rho(T) = \frac{1}{1 + \sum_i \rho(T_i)}. \tag{24}$$

It follows immediately that $0 \le f(T) \le \log(1 + \deg(T))$. Hence, the condition (2) of Theorem 1 is satisfied by f with $\alpha = 1$. Next, we measure the difference between $f(T)$ and $f(T^{(M)})$. Define the exact bounds on ρ given the first M levels:

$$\rho^M_{\min}(T) := \inf\{\rho(S) : S^{(M)} = T^{(M)}\}, \qquad \rho^M_{\max}(T) := \sup\{\rho(S) : S^{(M)} = T^{(M)}\}.$$

The functions $\rho^M_{\min}(T)$ and $\rho^M_{\max}(T)$, $M = 0, 1, 2, \ldots$ can also be determined recursively from the root branches $T_1, T_2, \ldots$ by observing $\rho^0_{\min}(T) = 0$ and $\rho^0_{\max}(T) = 1$ for any T, and for any $M \ge 1$, we have

$$\rho^M_{\max}(T) = \frac{1}{1 + \sum_i \rho^{M-1}_{\min}(T_i)} \quad \text{and} \quad \rho^M_{\min}(T) = \frac{1}{1 + \sum_i \rho^{M-1}_{\max}(T_i)}. \tag{25}$$

Since $\rho(T), \rho(T^{(M)}) \in [\rho^M_{\min}(T), \rho^M_{\max}(T)]$, we obtain

$$\frac{\rho^M_{\min}(T)}{\rho^M_{\max}(T)} \le \frac{\rho(T)}{\rho(T^{(M)})} \le \frac{\rho^M_{\max}(T)}{\rho^M_{\min}(T)}. \tag{26}$$

Writing $\tau^M(T) := \log(\rho^M_{\max}(T)/\rho^M_{\min}(T)) \ge 0$, (26) gives us

$$|f(T) - f(T^{(M)})| \le \tau^M(T). \tag{27}$$

Using (25), we get

$$\tau^M(T) = -\log\left(\frac{1 + \sum_i \rho^{M-1}_{\min}(T_i)}{1 + \sum_i \rho^{M-1}_{\max}(T_i)}\right) = -\log\left(\frac{1 + \sum_i \rho^{M-1}_{\max}(T_i)\exp(-\tau^{M-1}(T_i))}{1 + \sum_i \rho^{M-1}_{\max}(T_i)}\right).$$

Since the term inside the logarithm on the right side can be regarded as an expectation (of the expression $\exp(-\tau^{M-1}(T_i))$), applying Jensen's inequality to the convex function $-\log x$ yields

$$\begin{aligned}\tau^M(T) &\le \frac{1}{1 + \sum_i \rho^{M-1}_{\max}(T_i)} \sum_i \rho^{M-1}_{\max}(T_i)\tau^{M-1}(T_i) \\ &\le \frac{\max_i \rho^{M-1}_{\max}(T_i)}{1 + \max_i \rho^{M-1}_{\max}(T_i)} \sum_i \tau^{M-1}(T_i) \le \frac{1}{2}\sum_i \tau^{M-1}(T_i).\end{aligned} \tag{28}$$

From (25) it is clear that $\rho^1_{\max}(T) = 1$ and $\rho^1_{\min}(T) = (1 + \deg(T))^{-1}$ for any T. Therefore

$$\tau^1(T) = \log(1 + \deg(T)) \le \deg(T). \tag{29}$$

Let $v_1, v_2, \ldots, v_{w_{M-1}(T)}$ be the nodes at level $M-1$ of T. By iterating (28) $M-1$ times and applying (29), we arrive at the bound

$$\tau^M(T) \le 2^{-(M-1)} \sum_{i=1}^{w_{M-1}(T)} \tau^1(T_{v_i}) \le 2^{-(M-1)} \sum_{i=1}^{w_{M-1}(T)} \deg(T_{v_i}) \le 2^{-(M-1)} w_M(T). \quad (30)$$

Combining (27) and (30), we obtain

$$|f(T) - f(T^{(M)})| \le 2^{-M+1} w_M(T). \quad (31)$$

This is essentially enough to show that the remaining conditions of Theorem 1 are satisfied by our toll function. Let us first check (3). Note that for any $N \ge M$, we have

$$\mathbb{E}\left| f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)}) \,|\, \hat{\mathcal{T}}^{(M)} \right) \right| \le \mathbb{E}\left(\mathbb{E}\left(|f(\hat{\mathcal{T}}^{(M)}) - f(\hat{\mathcal{T}}^{(N)})| \,\middle|\, \hat{\mathcal{T}}^{(M)} \right) \right).$$

Using (31), we deduce that for any $N \ge M$,

$$\mathbb{E}\left(|f(\hat{\mathcal{T}}^{(M)}) - f(\hat{\mathcal{T}}^{(N)})| \,\middle|\, \hat{\mathcal{T}}^{(M)} \right) \le 2^{-M+1} \mathbb{E}\left(w_M(\hat{\mathcal{T}}^{(N)}) \,\middle|\, \hat{\mathcal{T}}^{(M)} \right).$$

By taking the expectations, and using $w_M(\hat{\mathcal{T}}^{(N)}) = w_M(\hat{\mathcal{T}})$ as well as the estimate $\mathbb{E} w_M(\hat{\mathcal{T}}) = O(M)$ (see (8)), we get

$$\mathbb{E}\left| f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)}) \,|\, \hat{\mathcal{T}}^{(M)} \right) \right| \ll M\, 2^{-M}. \quad (32)$$

To check (4) we use (31) and $\mathbb{E} w_M(\mathcal{T}_n) = O(M)$ (see (8)) and get

$$\mathbb{E}|f(\mathcal{T}_n) - f({\mathcal{T}_n}^{(M)})| = \mathbb{E}\left(\mathbb{E}\left(|f(\mathcal{T}_n) - f({\mathcal{T}_n}^{(M)})| \,\middle|\, {\mathcal{T}_n}^{(M)} \right) \right) \le \mathbb{E}\left(2^{-M+1} w_M(\mathcal{T}_n) \right) \ll M 2^{-M}, \quad (33)$$

where the implied constant is independent of n. To sum up, (32) and (33) show that the conditions of Theorem 1 are satisfied for the choice $p_M := C_1 M 2^{-M}$ and $M_n := \lfloor C_2 \log n \rfloor$ with sufficiently large positive constants C_1 and C_2.

5.2 Tree reductions

An old leaf is a leaf that is the leftmost child of its parent node, and an old path is a maximal path with the property that its lower endpoint is an old leaf, and its internal nodes are all nodes of outdegree 1 that are leftmost children of their parents. As in [3], consider the process of reducing a tree by cutting off all old paths from the tree at each step. This process is called old path-reduction. For a given positive integer r, and for a tree T, let $X_r(T)$ be the number of nodes in the reduced tree after the first r steps of the old path-reduction process. The authors of [3] proved estimates for the mean and variance of $X_r(\mathcal{T}_n)$ for the special case where $\mathcal{T}_n$ is the random plane (=ordered) tree on n nodes, but they did not derive a limiting distribution. Theorem 1 can be applied to show asymptotic normality for this case. However, we do not need to restrict ourselves to plane trees.

We let

$$F_r(T) = |T| - X_r(T),$$

which corresponds to the number of deleted nodes after r steps in T. The functional F_r is additive with toll function f_r, where

$$f_r(T) = \sum_j \eta_T(T_j)$$

and the sum is over all branches T_j, with

$$\eta_T(T_j) = \begin{cases} 1 & \text{if the root of } T_j \text{ is deleted within the first } r \text{ steps,} \\ 0 & \text{otherwise.} \end{cases}$$

We can immediately see that

$$0 \leq f_r(T) \leq \deg(T).$$

Next, we show that f_r is almost local. For a tree T, let T^* be the planted tree where the root of T is connected to a new node, which becomes the root of T^*. Let $\kappa = \min\{k \geq 2 : \mathbb{P}(\xi = k) > 0\}$ (this must exist under our current assumptions on ξ), and let T_0 be the complete κ-ary tree of depth r. It is clear that $F_r(T_0^*) \neq 1$, i.e. T_0^* is not reduced to the root in r steps, and

$$\mathbb{P}(\mathcal{T} = T_0) > 0. \tag{34}$$

For each positive integer M, let $\mathcal{B}_M$ be the set of all trees T (not necessarily finite) of height at least $M-1$ such that $F_r((T^{(M-1)})^*) = 1$ (i.e. the tree $T^{(M-1)}$ vanishes after the first r steps of the reduction). It is important to notice here that a rooted tree T is not reduced to a single node after the first r steps of the reduction if the fixed tree T_0 appears as a *subtree* of T (by *subtree*, we mean a subtree of the form $T_v^{(k)}$ for some integer $k \geq 0$ and some node v of T). This observation is key in the proof of the next lemma, which can be found in the appendix.

▶ **Lemma 5.** *There is a positive constant $c < 1$, that depends only on ξ and r, such that*

$$\mathbb{P}(\mathcal{T} \in \mathcal{B}_M) \ll c^M \text{ and } \mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M\right) \ll c^M.$$

For a finite tree T, the only possibility for which $f_r(T^{(M)}) \neq f_r(T)$ is when there is a root branch T_j of T such that $T_j^{(M-1)}$ vanishes after the first r steps of the reduction of $T^{(M)}$, but T_j does not vanish after the first r steps of the reduction of T. This means that if $f_r(T^{(M)}) \neq f_r(T)$, then T must have a branch in $\mathcal{B}_M$. Therefore, we have

$$\mathbb{P}\left(f_r(\mathcal{T}^{(M)}) \neq f_r(\mathcal{T})\right) \leq \sum_{k=1}^{\infty} k\mathbb{P}(\xi = k)\,\mathbb{P}(\mathcal{T} \in \mathcal{B}_M) \ll c^M.$$

The estimate on the right follows from Lemma 5. As an immediate consequence of this, we have

$$\mathbb{P}\left(f_r({\mathcal{T}_n}^{(M)}) \neq f_r(\mathcal{T}_n)\right) \leq \frac{\mathbb{P}\left(f_r(\mathcal{T}^{(M)}) \neq f_r(\mathcal{T})\right)}{\mathbb{P}(|\mathcal{T}| = n)} \ll n^{3/2}\, c^M.$$

Hence,

$$\mathbb{E}|f_r({\mathcal{T}_n}^{(M)}) - f_r(\mathcal{T}_n)| \ll n^{3/2} c^M \max_{|T|=n} |f_r(T^{(M)}) - f_r(T)| \ll n^{5/2} c^M. \tag{35}$$

Let $\mathcal{E}_M$ be the event $\bigcup_{N>M}\{f_r(\hat{\mathcal{T}}^{(M)}) \neq \mathbb{E}(f_r(\hat{\mathcal{T}}^{(N)}) \,|\, \hat{\mathcal{T}}^{(M)})\}$. Then, for any $N \geq M$, we have

$$\left|f_r(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}(f_r(\hat{\mathcal{T}}^{(N)}) \,|\, \hat{\mathcal{T}}^{(M)})\right| \ll \deg(\hat{\mathcal{T}}^{(M)}) I_{\mathcal{E}_M}.$$

For $\hat{\mathcal{T}}$ to be in $\mathcal{E}_M$, $\hat{\mathcal{T}}$ must have a root branch in $\mathcal{B}_M$. Therefore,

$$\mathbb{E}\Big|f_r(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}(f_r(\hat{\mathcal{T}}^{(N)}) \mid \hat{\mathcal{T}}^{(M)})\Big| \ll \sum_{k=1}^{\infty} k\mathbb{P}\left(\deg(\hat{\mathcal{T}}) = k\right)\left((k-1)\mathbb{P}\left(\mathcal{T} \in \mathcal{B}_M\right) + \mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M\right)\right). \quad (36)$$

In view of Lemma 5, we have

$$\mathbb{E}\Big|f_r(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}(f_r(\hat{\mathcal{T}}^{(N)}) \mid \hat{\mathcal{T}}^{(M)})\Big| \ll c^M \sum_{k=1}^{\infty} k^2\mathbb{P}\left(\deg(\hat{\mathcal{T}}) = k\right) \ll c^M, \quad (37)$$

since $\mathbb{E}(\deg(\hat{\mathcal{T}})^2) < \infty$ if $\mathbb{E}\xi^3 < \infty$. The estimates (35) and (37) confirm that f_r is indeed almost local where, for example, $p_M = c_2^M$, for some c_2 with $c < c_2 < 1$, and $M_n = \lfloor (\log n)^2 \rfloor$.

▶ Remark. We only made very little use of the actual definition of the old path-reduction. To be precise, we only used it when we argued that our constructed T_0 does not vanish after the first r steps of the reduction and that all trees that contain T_0 as a subtree will not be reduced after the first r steps. This means that the same proof will work for any tree reduction with a similar property. This includes all tree reductions considered in [3].

References

1 Patrick Billingsley. *Convergence of probability measures.* John Wiley & Sons, Inc., New York-London-Sydney, 1968.

2 Michael Drmota. *Random trees.* SpringerWienNewYork, Vienna, 2009.

3 Benjamin Hackl, Clemens Heuberger, Sara Kropf, and Helmut Prodinger. Fringe Analysis of Plane Trees Related to Cutting and Pruning. *Aequ. Math.*, 92(2):311–353, 2018.

4 Cecilia Holmgren, Svante Janson, and Matas Šileikis. Multivariate normal limit laws for the numbers of fringe subtrees in m-ary search trees and preferential attachment trees. *Electron. J. Combin.*, 24(2):Paper 2.51, 49, 2017.

5 Svante Janson. Random cutting and records in deterministic and random trees. *Random Structures Algorithms*, 29(2):139–179, 2006.

6 Svante Janson. Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Structures Algorithms*, 48(1):57–101, 2016.

7 Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

8 Harry Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(4):425–487, 1986.

9 Martin Klazar. Twelve countings with rooted plane trees. *European J. Combin.*, 18(2):195–210, 1997.

10 Dimbinaina Ralaivaosaona and Stephan Wagner. Additive functionals of d-ary increasing trees. *arXiv 1605.03918*, 2016.

11 Lajos Takács. Conditional limit theorems for branching processes. *J. Appl. Math. Stochastic Anal.*, 4(4):263–292, 1991.

12 Stephan Wagner. Correlation of graph-theoretical indices. *SIAM J. Discrete Math.*, 21(1):33–46, 2007.

13 Stephan Wagner. On the number of matchings of a tree. *European J. Combin.*, 28(4):1322–1330, 2007.

14 Stephan Wagner. Central limit theorems for additive tree parameters with small toll functions. *Combin. Probab. Comput.*, 24(1):329–353, 2015.

A Appendix

Proof of Estimate (13)**.** From the proof of [6, Lemma 5.9] (see (5.42) there), we have, for T with $|T| \le n/2$, that

$$\mathbb{P}\left(\mathcal{T}_n{}^{(M)} = T\right) = \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right)\left(1 + O\left(\frac{|T|}{n^{1/2}}\right)\right). \tag{38}$$

Using (38), we infer

$$\begin{aligned}
|\mathbb{E}f(\mathcal{T}_n{}^{(M)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})| \\
&= \left|\sum_T f(T)\mathbb{P}\left(\mathcal{T}_n{}^{(M)} = T\right) - \sum_T f(T)\mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right)\right| \\
&\le \sum_{|T|\le n/2} |f(T)| \left|\mathbb{P}\left(\mathcal{T}_n{}^{(M)} = T\right) - \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right)\right| + \\
&\quad \sum_{|T|>n/2} |f(T)| \left(\mathbb{P}\left(\mathcal{T}_n{}^{(M)} = T\right) + \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right)\right) \\
&\ll \sum_T \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right) \frac{\deg(T)^\alpha |T|}{n^{1/2}} \\
&\quad + \sum_{|T|>n/2} \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right)\deg(T)^\alpha + \sum_{|T|>n/2} \mathbb{P}\left(\mathcal{T}_n{}^{(M)} = T\right)\deg(T)^\alpha.
\end{aligned}$$

We can now estimate each of the three terms in the last two lines separately. First, we have

$$\begin{aligned}
\sum_T \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right) \frac{\deg(T)^\alpha |T|}{n^{1/2}} &= n^{-1/2}\mathbb{E}(\deg(\hat{\mathcal{T}}^{(M)})^\alpha |\hat{\mathcal{T}}^{(M)}|) \\
&= n^{-1/2}\mathbb{E}\left(\deg(\hat{\mathcal{T}}^{(M)})^\alpha\, \mathbb{E}\left(|\hat{\mathcal{T}}^{(M)}| \,\middle|\, \deg(\hat{\mathcal{T}}^{(M)})\right)\right).
\end{aligned}$$

Conditioning on $\deg(\hat{\mathcal{T}}^{(M)})$ (which is the same as $\deg(\hat{\mathcal{T}})$ for $M \ge 1$), $\hat{\mathcal{T}}$ consists of a root, a copy of $\hat{\mathcal{T}}$ and $\deg(\hat{\mathcal{T}}) - 1$ independent copies of $\mathcal{T}$. Thus, by the estimates in (9), we have

$$\mathbb{E}\left(|\hat{\mathcal{T}}^{(M)}| \,\middle|\, \deg(\hat{\mathcal{T}}^{(M)})\right) = O(M^2 \deg(\hat{\mathcal{T}}^{(M)})).$$

Therefore,

$$\mathbb{E}\left(\deg(\hat{\mathcal{T}}^{(M)})^\alpha\, \mathbb{E}\left(|\hat{\mathcal{T}}^{(M)}| \,\middle|\, \deg(\hat{\mathcal{T}}^{(M)})\right)\right) \ll M^2\mathbb{E}(\deg(\hat{\mathcal{T}}^{(M)})^{\alpha+1}),$$

which yields

$$\sum_T \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right) \frac{\deg(T)^\alpha\, |T|}{n^{1/2}} \ll n^{-1/2}M^2\mathbb{E}(\deg(\hat{\mathcal{T}}^{(M)})^{\alpha+1}). \tag{39}$$

For the second term, we have

$$\begin{aligned}
\sum_{|T|>n/2} \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right)\deg(T)^\alpha \\
&= \sum_{k\ge1} k^\alpha\, \mathbb{P}\left(|\hat{\mathcal{T}}^{(M)}| > n/2 \ \text{and} \ \deg(\hat{\mathcal{T}}) = k\right) \\
&= \sum_{k\ge1} k^\alpha\, \mathbb{P}\left(\deg(\hat{\mathcal{T}}) = k\right)\mathbb{P}\left(|\hat{\mathcal{T}}^{(M)}| > n/2 \,\middle|\, \deg(\hat{\mathcal{T}}) = k\right).
\end{aligned}$$

By Markov's inequality and a similar argument as before, we obtain

$$\mathbb{P}\left(|\hat{\mathcal{T}}^{(M)}| > n/2 \,\middle|\, \deg(\hat{\mathcal{T}}) = k\right) \le \frac{2\mathbb{E}(|\hat{\mathcal{T}}^{(M)}| \mid \deg(\hat{\mathcal{T}}) = k)}{n} \ll \frac{kM^2}{n}.$$

Thus,

$$\sum_{|T|>n/2} \mathbb{P}\left(\hat{\mathcal{T}}^{(M)} = T\right) \deg(T)^\alpha \ll n^{-1}M^2\mathbb{E}(\deg(\hat{\mathcal{T}})^{\alpha+1}). \tag{40}$$

Finally, for the last term, we proceed in a similar fashion:

$$\begin{aligned}\sum_{|T|>n/2} &\mathbb{P}\left({\mathcal{T}_n}^{(M)} = T\right) \deg(T)^\alpha \\ &\le \sum_{k\ge1} k^\alpha\, \mathbb{P}\left(|{\mathcal{T}_n}^{(M)}| > n/2 \text{ and } \deg(\mathcal{T}_n) = k\right) \\ &\le \sum_{k\ge1} k^\alpha\, \mathbb{P}\left(\deg(\mathcal{T}_n) = k\right) \mathbb{P}\left(|{\mathcal{T}_n}^{(M)}| > n/2 \,\middle|\, \deg(\mathcal{T}_n) = k\right).\end{aligned}$$

If $\mathcal{T}_{n,1}, \mathcal{T}_{n,2}, \dots, \mathcal{T}_{n,k}$ are the branches of $\mathcal{T}_n$, given that $\deg(\mathcal{T}_n) = k$, then, conditioning on their sizes $n_1, n_2, \dots, n_k$, they are k independent conditioned Galton-Watson trees $\mathcal{T}_{n_1}, \mathcal{T}_{n_2}, \dots, \mathcal{T}_{n_k}$. On the other hand, we have

$$|{\mathcal{T}_n}^{(M)}| = 1 + \sum_{i=1}^{k} |\mathcal{T}_{n,i}^{(M-1)}|.$$

Thus,

$$\begin{aligned}\mathbb{E}\left(|\mathcal{T}_n^{(M)}| \mid \deg(\mathcal{T}_n) = k\right) &= \mathbb{E}\left(\mathbb{E}\left(|\mathcal{T}_n^{(M)}| \,\middle|\, n_1, n_2, \cdots, n_k\right)\right) \\ &= 1 + \sum_{i=1}^{k} \mathbb{E}\left(\mathbb{E}\left(|\mathcal{T}_{n_i}^{(M-1)}| \,\middle|\, n_1, n_2, \cdots, n_k\right)\right) \ll kM^2,\end{aligned}$$

which again follows from the last estimate in (9). Now, Markov's inequality yields

$$\mathbb{P}\left(|{\mathcal{T}_n}^{(M)}| > n/2 \,\middle|\, \deg(\mathcal{T}_n) = k\right) \ll n^{-1}kM^2.$$

Therefore, making use of (10) once again, we have

$$\begin{aligned}\sum_{|T|>n/2} &\mathbb{P}\left({\mathcal{T}_n}^{(M)} = T\right) \deg(T)^\alpha \\ &\ll n^{-1}M^2 \sum_{k\ge1} k^{\alpha+1}\, \mathbb{P}\left(\deg(\mathcal{T}_n) = k\right) = n^{-1}M^2\mathbb{E}(\deg(\mathcal{T}_n)^{\alpha+1}).\end{aligned} \tag{41}$$

Combining the estimates (39), (40), and (41), we finally arrive at the estimate

$$|\mathbb{E}f({\mathcal{T}_n}^{(M)}) - \mathbb{E}f(\hat{\mathcal{T}}^{(M)})| \ll n^{-1/2}M^2\mathbb{E}(\deg(\hat{\mathcal{T}})^{\alpha+1}) + n^{-1}M^2\mathbb{E}(\deg(\mathcal{T}_n)^{\alpha+1}), \tag{42}$$

which is exactly as we claimed in (13). ◀

Proof of Estimate (18)**.** We decompose $F(\mathcal{T}_k)$ according to the depth $d(v)$ of the nodes:

$$F(\mathcal{T}_k) = \sum_{v\in\mathcal{T}_k} f(\mathcal{T}_{k,v}) = \sum_{d(v)<M} f(\mathcal{T}_{k,v}) + \sum_{d(v)\ge M} f(\mathcal{T}_{k,v}) =: S_1 + S_2. \tag{43}$$

We first observe that

$$\begin{aligned}\mathbb{E}|f(\mathcal{T}_k)S_1| &\ll \mathbb{E}\Big(\deg(\mathcal{T}_k)^\alpha \sum_{d(v)<M} \deg(\mathcal{T}_{k,v})^\alpha\Big)\\ &= \mathbb{E}\Big(\deg(\mathcal{T}_k)^\alpha\, \mathbb{E}\Big(\sum_{d(v)<M} \deg(\mathcal{T}_{k,v})^\alpha \,\Big|\, \deg(\mathcal{T}_k)\Big)\Big).\end{aligned}$$

Next, for any positive integer $m \le M$, we have

$$\mathbb{E}\Big(\sum_{d(v)<m} \deg(\mathcal{T}_{k,v})^\alpha \,\Big|\, \mathcal{T}_k^{(m-1)}\Big) = \sum_{d(v)<m-1} \deg(\mathcal{T}_{k,v})^\alpha + O\left(w_{m-1}(\mathcal{T}_k)\right).$$

This is because the $w_{m-1}(\mathcal{T}_k)$ fringe subtrees with roots at level $m-1$, conditioned on their sizes, are conditioned Galton-Watson trees and thus by (10) the root degrees are $O(1)$. Taking the expectation conditioned on $\deg(\mathcal{T}_k)$, again by the same argument, and by the estimate $\mathbb{E}w_{m-1}(\mathcal{T}_k) = O(m)$ as in (8), we have

$$\mathbb{E}\Big(\sum_{d(v)<m} \deg(\mathcal{T}_{k,v})^\alpha \,\Big|\, \deg(\mathcal{T}_k)\Big) = \mathbb{E}\Big(\sum_{d(v)<m-1} \deg(\mathcal{T}_{k,v})^\alpha \,\Big|\, \deg(\mathcal{T}_k)\Big) + O(m\deg(\mathcal{T}_k)).$$

Thus, iterating from M, we obtain

$$\mathbb{E}\Big(\sum_{d(v)<M} \deg(\mathcal{T}_{k,v})^\alpha \,\Big|\, \deg(\mathcal{T}_k)\Big) \ll \deg(\mathcal{T}_k)^\alpha + M^2\deg(\mathcal{T}_k).$$

Therefore,

$$\mathbb{E}|f(\mathcal{T}_k)S_1| \ll \mathbb{E}(\deg(\mathcal{T}_k)^{2\alpha}) + M^2\mathbb{E}(\deg(\mathcal{T}_k)^{\alpha+1}). \tag{44}$$

For S_2, we condition on $\mathcal{T}_k{}^{(M)}$ and the sizes of the fringe subtrees $\mathcal{T}_{k,v_i}$, $i = 1, \ldots, w_M(\mathcal{T}_k)$, induced by nodes at level M. Conditionally, each $\mathcal{T}_{k,v_i}$ is distributed as $\mathcal{T}_{n_i}$. From the assumption that for every n we have $\mathbb{E}f(\mathcal{T}_n) = 0$ it follows (see [6, (6.25)]) that $\mathbb{E}F(T_{k,v_i}) = 0$ and therefore

$$\mathbb{E}\left(S_2 \,|\, \mathcal{T}_k{}^{(M)}\right) = 0. \tag{45}$$

Let us define $\tilde{f}_M(\mathcal{T}_k) := \mathbb{E}(f(\mathcal{T}_k) \,|\, \mathcal{T}_k{}^{(M)})$. Note that

$$\mathbb{E}(\tilde{f}_M(\mathcal{T}_k)S_2) = \mathbb{E}\Big(\mathbb{E}\left(\tilde{f}_M(\mathcal{T}_k)S_2 \,|\, \mathcal{T}_k{}^{(M)}\right)\Big) = \mathbb{E}\left(\tilde{f}_M(\mathcal{T}_k)\mathbb{E}\left(S_2 \,|\, \mathcal{T}_k{}^{(M)}\right)\right) = 0.$$

Hence,

$$|\mathbb{E}(f(\mathcal{T}_k)S_2)| = |\mathbb{E}(S_2(f(\mathcal{T}_k) - \tilde{f}_M(\mathcal{T}_k)))| \le \max|S_2|\,\mathbb{E}|f(\mathcal{T}_k) - \tilde{f}_M(\mathcal{T}_k)|.$$

It is important to notice here that the expectation in the last term remains unchanged if $f(T)$ is shifted by $\mu_{|T|}$ (this is the reason why we can assume that $\mathbb{E}f(\mathcal{T}_k) = 0$ and f still satisfies the conditions of Theorem 1). By the triangle inequality and the definition of $\tilde{f}_M(\mathcal{T}_k)$, we have

$$\begin{aligned}|f(\mathcal{T}_k) - \tilde{f}_M(\mathcal{T}_k)| &\le |f(\mathcal{T}_k) - f(\mathcal{T}_k{}^{(M)})| + |f(\mathcal{T}_k{}^{(M)}) - \tilde{f}_M(\mathcal{T}_k)|\\ &\le |f(\mathcal{T}_k) - f(\mathcal{T}_k{}^{(M)})| + \mathbb{E}(|f(\mathcal{T}_k{}^{(M)}) - f(\mathcal{T}_k)| \,|\, \mathcal{T}_k{}^{(M)}).\end{aligned}$$

Taking the expectation again, and using our condition (4), we obtain

$$\mathbb{E}|f(\mathcal{T}_k) - \tilde{f}_M(\mathcal{T}_k)| \le 2p_M.$$

Here, we are assuming that $M = M_k$. On the other hand, we have

$$|S_2| \le \sum_{d(v)\ge M} |f(\mathcal{T}_{k,v})| \le \sum_{v\in\mathcal{T}_k} \deg(\mathcal{T}_{k,v})^\alpha.$$

Since α is a nonnegative integer, the last term is bounded above by $(\sum_{v\in\mathcal{T}_k} \deg(\mathcal{T}_{k,v}))^\alpha$ (which is equal to $(k-1)^\alpha$) except for $\alpha = 0$. Hence, we get

$$\max |S_2| \le k^{\max\{\alpha,1\}}.$$

Therefore,

$$\mathbb{E}(f(\mathcal{T}_k)F(\mathcal{T}_k)) \ll k^{\max\{\alpha,1\}} p_M + \mathbb{E}(\deg(\mathcal{T}_k)^{2\alpha}) + M^2\mathbb{E}(\deg(\mathcal{T}_k)^{\alpha+1}),$$

as claimed. ◀

Proof of Lemma 5. We start with the first estimate. We notice that for $\mathcal{T}$ to be in $\mathcal{B}_M$, $\mathcal{T}^{(r)}$ must not be equal to T_0. So

$$\mathbb{P}(\mathcal{T} \in \mathcal{B}_M) = \sum_{T\ne T_0} \mathbb{P}\left(\mathcal{T}^{(r)} = T\right)\mathbb{P}\left(\mathcal{T} \in \mathcal{B}_M \,|\, \mathcal{T}^{(r)} = T\right).$$

Conditioning on the event $\{\mathcal{T}^{(r)} = T\}$, the rest of $\mathcal{T}$ is a forest consisting of $w_r(T)$ independent copies of $\mathcal{T}$. Hence, by the union bound, we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{T} \in \mathcal{B}_M) &\le \sum_{T\ne T_0} \mathbb{P}\left(\mathcal{T}^{(r)} = T\right) w_r(T)\mathbb{P}(\mathcal{T} \in \mathcal{B}_{M-r})\\
&= \mathbb{P}(\mathcal{T} \in \mathcal{B}_{M-r}) \sum_{T\ne T_0} \mathbb{P}\left(\mathcal{T}^{(r)} = T\right) w_r(T)\\
&\le \mathbb{P}(\mathcal{T} \in \mathcal{B}_{M-r}) \sum_{T\ne T_0} \mathbb{P}(\mathcal{T} = T)\, w_r(T).
\end{aligned}$$

If we let $q = \sum_{T\ne T_0} \mathbb{P}(\mathcal{T} = T)\, w_r(T)$, then

$$\mathbb{P}(\mathcal{T} \in \mathcal{B}_M) \le q\,\mathbb{P}(\mathcal{T} \in \mathcal{B}_{M-r}). \tag{46}$$

On the other hand, we know that

$$q + w_r(T_0)\mathbb{P}(\mathcal{T} = T_0) = \sum_{T} \mathbb{P}(\mathcal{T} = T)\, w_r(T) = \mathbb{E}w_r(\mathcal{T}) = 1.$$

In view of (34), we deduce that $q < 1$. Therefore, iterating (46) yields

$$\mathbb{P}(\mathcal{T} \in \mathcal{B}_M) \le q^{\lfloor M/r\rfloor} \le c_1^M, \tag{47}$$

where $c_1 := q^{1/r} < 1$, proving the first estimate.

For the second estimate, we also begin in a similar fashion, i.e. we have

$$\mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M\right) = \sum_{T\ne T_0} \mathbb{P}\left(\hat{\mathcal{T}}^{(r)} = T\right)\mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M \,|\, \hat{\mathcal{T}}^{(r)} = T\right).$$

Here, when conditioning on the event $\{\hat{\mathcal{T}}^{(r)} = T\}$, the rest of $\hat{\mathcal{T}}$ is a forest consisting of $w_r(T) - 1$ independent copies of $\mathcal{T}$ and a copy of $\hat{\mathcal{T}}$. Thus,

$$\mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M\right) \leq \sum_{T \neq T_0} \mathbb{P}\left(\hat{\mathcal{T}}^{(r)} = T\right) \left((w_r(T) - 1)\mathbb{P}\left(\mathcal{T} \in \mathcal{B}_{M-r}\right) + \mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_{M-r}\right)\right).$$

Using (47), letting $q_2 = \sum_{T \neq T_0} \mathbb{P}\left(\hat{\mathcal{T}}^{(r)} = T\right)(w_r(T) - 1)$ (which is finite since it is bounded above by $\mathbb{E} w_r(\hat{\mathcal{T}}) < \infty$), and noting that $q \geq \sum_{T \neq T_0} \mathbb{P}\left(\hat{\mathcal{T}}^{(r)} = T\right)$ by the definition of $\hat{\mathcal{T}}$, we obtain

$$\mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M\right) \ll q_2\, c_1^{M-r} + q\, \mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_{M-r}\right).$$

Iterating this, we have

$$\mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M\right) \ll q^{\lfloor M/r \rfloor} + q_2 \sum_{j=1}^{\lfloor M/r \rfloor} q^{j-1}\, c_1^{M-jr}.$$

Since we set $q = c_1^r$, the latter estimate becomes

$$\mathbb{P}\left(\hat{\mathcal{T}} \in \mathcal{B}_M\right) \ll M\, c_1^M.$$

The proof is completed by choosing any constant $c > 0$ such that $c_1 < c < 1$. ◀