CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Distribution of tree parameters by martingale approach

Mikhail Isaev[1] *, Angus Southwell[1] and Maksim Zhukovskii[2,3,4]

[1]School of Mathematics, Monash University, Clayton, VIC, Australia, [2]Laboratory of Combinatorial and Geometric Structures, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russian Federation, [3]Caucasus Mathematical Center, Adyghe State University, Maykop, Republic of Adygea, Russian Federation and [4]The Russian Presidential Academy of National Economy and Public Administration, Moscow, Russian Federation
*Corresponding author. Email: mikhail.isaev@monash.edu

Abstract

For a uniform random labelled tree, we find the limiting distribution of tree parameters which are stable (in some sense) with respect to local perturbations of the tree structure. The proof is based on the martingale central limit theorem and the Aldous–Broder algorithm. In particular, our general result implies the asymptotic normality of the number of occurrences of any given small pattern and the asymptotic log-normality of the number of automorphisms.

1. Introduction

The distribution of various random variables associated with trees is widely studied in the literature. Typically, the tree parameters that behave additively exhibit normal distribution, which was observed by Drmota [7, Chapter 3], Janson [16], and Wagner [27]. For example, the number of leaves or, more generally, the number of vertices of a given degree satisfies a central limit theorem (CLT) for many random models: labelled trees, unlabelled trees, plane trees, forests; see Drmota and Gitteberger [8] and references therein for more details.

The classical limit theorems of probability theory are impractical for random trees due to the dependency of adjacencies. Instead, one employs more elaborate tools such as the analysis of generating functions [2], the conditional limit theorems [12], and Hwang's quasi-power theorem [13]. These methods are particularly efficient for parameters that admit a recurrence relation, which is often the case for trees.

The martingale CLT [4] is a powerful tool that has been extensively used to study random structures. Nevertheless, it is surprisingly overlooked in the context of the distribution of tree parameters and the vast majority of known results rely on the methods mentioned in the paragraph above. We are aware of only a few applications of the martingale CLT: Smythe [25] and Mahmoud [18] analysed growth of leaves in the random trees related to urn models; Móri [22] examined the max degree for Barabási–Albert random trees; Fen and Hu [9] considered the

CrossMark

Figure 1. S_i^{jk} removes ij from a tree and adds ik (dashed).

Zagreb index for random recursive trees; Sulzbach [26] studied the path length in a random model encapsulating binary search trees, recursive trees and plane-oriented recursive trees.

We prove a CLT for an arbitrary tree parameter using the martingale approach. Unlike other methods, the parameter is not required to be of a specific form or to satisfy a recurrence relation. Our only assumption is that the parameter is stable with respect to small perturbations in the sense that precisely specified below. We also bound the rate of convergence to the normal distribution. In this paper, we restrict our attention to unrooted labelled trees even though martingales appear naturally in many other random settings. This is sufficient to demonstrate the power of the new approach and cover several important applications that go beyond the toolkit of existing methods.

Let $\mathcal{T}_n$ be the set of trees whose vertices are labelled by $[n] := \{1, \ldots, n\}$ and T be a uniform random element of $\mathcal{T}_n$. By Cayley's formula, we have $|\mathcal{T}_n| = n^{n-2}$. For a tree $T \in \mathcal{T}_n$ and two vertices $i, j \in [n]$, let $d_T(i, j)$ denote the distance between i and j that is the number of edges in the unique path from i to j in T. For $A, B \subseteq [n]$, let

$$d_T(A, B) := \min_{u \in A, v \in B} d_T(u, v).$$

Throughout the paper, we identify graphs and their edge sets. Consider an operation defined S_i^{jk} as follows. If $ij \in T$ and $ik \notin T$, let $S_i^{jk} T$ be the graph obtained from T by deleting the edge ij and inserting the edge ik; see Figure 1 below.

Observe that $S_i^{jk} T$ is a tree if and only if the path from j to k in T does not contain the vertex i. We refer the operation S_i^{jk} as a *tree perturbation*.

Let $\mathbb{R}^+$ denote the set of non-negative real numbers. For $\alpha \in \mathbb{R}^+$, we say a tree parameter $F : \mathcal{T}_n \to \mathbb{R}$ is α-*Lipschitz* if

$$|F(T) - F\left(S_i^{jk} T\right)| \leqslant \alpha.$$

for all $T \in \mathcal{T}_n$ and triples (i, j, k) that $S_i^{jk} T$ is a tree. We also require that the effects on the parameter F of sufficiently distant perturbations S_i^{jk} and S_a^{bc} *superpose*; that is

$$F\left(S_i^{jk} S_a^{bc} T\right) - F(T) = \left(F\left(S_i^{jk} T\right) - F(T)\right) + \left(F\left(S_a^{bc} T\right) - F(T)\right).$$

For $\rho \in \mathbb{R}^+$, we say F is ρ-*superposable* if the above equation holds for all $T \in \mathcal{T}_n$ and triples (i, j, k), (a, b, c) such that $S_i^{jk} T$, $S_a^{bc} T$, $S_i^{jk} S_a^{bc} T$ are trees and $d_T(\{j, k\}, \{b, c\}) \geqslant \rho$. Note that the sets $\{j, k\}$ and $\{b, c\}$ are at the same distance in all four trees T, $S_i^{jk} T$, $S_a^{bc} T$, and $S_i^{jk} S_a^{bc} T$. Thus, $d_T(\{j, k\}, \{b, c\})$ is an appropriate measure for the distance between the two tree perturbations S_i^{jk} and S_a^{bc}.

For a random variable X let

$$\delta_K[X] := \sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(X - \mathbb{E}[X] \leqslant t(\text{Var}[X])^{1/2}\right) - \Phi(t) \right|,$$

where $\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-x^2/2} dx$. In other words, $\delta_K[X]$ is the Kolmogorov distance between the scaled random variable X and the standard normal distribution. We say $X = X_n$ is *asymptotically normal* if $\delta_K[X] \to 0$ as $n \to \infty$.

In the following theorem, F, α, and ρ stand for sequences parametrised by a positive integer n that is $(F, \alpha, \rho) = (F_n, \alpha_n, \rho_n)$. We omit the subscripts for notation simplicity. All asymptotics in the paper refer to $n \to \infty$ and the notations $o(\,\cdot\,)$, $O(\,\cdot\,)$, $\Theta(\,\cdot\,)$ have the standard meaning.

Theorem 1.1. *Let a tree parameter $F : \mathcal{T}_n \to \mathbb{R}$ be α-Lipschitz and ρ-superposable for some $\alpha > 0$ and $\rho \geqslant 1$. Assume also that, for a fixed constant $\varepsilon > 0$,*

$$\frac{n\alpha^3}{(\text{Var}\,[F(T)]\,)^{3/2}} + \frac{n^{1/4}\alpha\rho}{(\text{Var}\,[F(T)]\,)^{1/2}} = O(n^{-\varepsilon}).$$

Then, $F(T)$ is asymptotically normal. Moreover, $\delta_K[F(T)] = O(n^{-\varepsilon'})$ for any $\varepsilon' \in (0, \varepsilon)$.

To clarify the assumptions Theorem 1.1, we consider a simple application to the aforementioned parameter $L(T)$, the number of leaves in a tree T. The distribution of $L(T)$ was derived for the first time by Kolchin [17], using generating functions and the connection to the Galton–Watson branching process. Theorem 1.1 immediately leads to the following result:

Corollary 1.2. *$L(T)$ is asymptotically normal and $\delta_K[L(T)] = O(n^{-1/4+\epsilon})$ for any $\epsilon > 0$.*

Proof. For any tree $T \in \mathcal{T}_n$ and a triple (i, j, k) that $S_i^{jk}T$ is a tree, the numbers of leaves of T and $S_i^{jk}T$ differ by at most one. Thus, L is α-Lipchitz on $\mathcal{T}_n$ with $\alpha = 1$.

Next, observe that if T, $S_i^{jk}T$, $S_a^{bc}T$, and $S_i^{jk}S_a^{bc}T$ are trees and $\{j, k\} \cap \{b, c\} = \emptyset$, then

$$L(T) - L\left(S_i^{jk}T\right) - L\left(S_a^{bc}T\right) + L\left(S_i^{jk}S_a^{bc}T\right) = 0.$$

Indeed, the trees T, $S_i^{jk}T$, $S_a^{bc}T$, $S_i^{jk}S_a^{bc}T$ have the same sets of leaves except possibly vertices $\{j, k, b, c\}$. However, any vertex from $\{j, k, b, c\}$ contributes to the same number of negative and positive terms in the left-hand side of the above. This implies that L is ρ-superposable with $\rho = 1$.

It is well known that $\text{Var}[L(T)] = (1 + o(1))n/e$; see, for example, [21, Theorem 7.7]. Then, all the assumptions of Theorem 1.1 are satisfied with $\alpha = \rho = 1$ and $\varepsilon = 1/4$. This completes the proof. $\square$

Remark 1.3. The rates of convergence $\delta_K[F(T)] = O(n^{-1/4+\epsilon})$ are typical in applications of Theorem 1.1 because, for many examples, $\text{Var}[F(T)]$ is linear and α, ρ are bounded by some power of $\log n$. Wagner [29] pointed out that Hwang's quasi-power theorem [13] leads to a better estimate $\delta_K[L(T)] = O(n^{-1/2+\epsilon})$ for the number of leaves. This matches the rates of convergence in the classical Berry–Esseen result (for a sum of i.i.d. variables) and, thus, is likely optimal. It remains an open question whether the bound $\delta_K[F(T)] = O(n^{-1/2+\epsilon})$ always hold for an arbitrary α-Lipschitz and ρ-superposable tree parameter F (assuming the variance is linear and α and ρ are not too large).

The asymptotic normality of the number of vertices in T with a given degree is proved identically to Corollary 1.2. However, for many other applications, a tree parameter F might behave badly on a small set of trees. Then, Theorem 1.1 does not work directly since α and ρ are too large. For example, a single perturbation S_i^{jk} can destroy a lot of paths on three vertices in a tree with large degrees. To overcome this difficulty, one can apply Theorem 1.1 to a parameter $\tilde{F}$, which is related to F, but ignores the vertices with degrees larger $\log n$. This trick does not change the limiting distribution because the trees with large degrees are rare: Moon [21, formula (7.3)] showed that, for any $d \in [n]$,

$$\mathbb{P}(T \text{ has a vertex with degree} > d) \leqslant n/d! \tag{1}$$

Similarly, one can restrict attention to the trees for which the neighbourhoods of vertices do not grow very fast. Let

$$\beta(T) = \max_{i,d\in[n]} \frac{|\{j\in[n] : d_T(i,j)=d\}|}{d}. \tag{2}$$

In this paper, we prove the following result, which might be of independent interest.

Theorem 1.4. $\mathbb{P}\left(\beta(\boldsymbol{T}) \geqslant \log^4 n\right) \leqslant e^{-\omega(\log n)}$.

Remark 1.5. The distribution of the height profiles in branching processes is a well-studied topic. In particular, the number of vertices in $\boldsymbol{T}$ at distance at most d from a given vertex was already considered by Kolchin [17]. However, we could not find a suitable large deviation bound for $\beta(\boldsymbol{T})$ in the literature. In fact, the constant 4 in the exponent of the logarithm in the bound above is not optimal, but sufficient for our purposes.

The structure of the paper is as follows. In Section 2, we analyse the number of occurrences of an arbitrary tree pattern. For various interpretations of the notion "occurrence," the asymptotic normality in this problem was established by Chysak, Drmota, Klausner, Kok [5] and Janson [16]. Applying Theorem 1.1, we not only confirm these results but also allow much more general types of occurrences. In particular, we prove the asymptotical normality for the number of induced subgraphs isomorphic to a given tree of fixed size and for the number of paths of length up to $n^{1/8-\varepsilon}$. Both of these applications go beyond the setup of [5, 16]. In Section 3, we derive the distribution of the number of automorphisms of $\boldsymbol{T}$ and confirm the conjecture by Yu [30]. To our knowledge, this application of Theorem 1.1 is also not covered by any of the previous results.

We prove Theorem 1.1 in Section 5, using a martingale construction based on the Aldous–Broder algorithm [1] for generating random labelled spanning trees of a given graph. Section 4 contains the necessary background on the theory of martingales. We also use martingales to prove Theorem 1.4 in Section 6. This proof is independent of Section 5 and, in fact, Theorem 1.4 is one of the ingredients that we need for our main result, Theorem 1.1. We also use Theorem 1.4 in the application to long induced paths to bound the number of the paths affected by one perturbation; see Theorem 2.9.

Tedious technical calculations of the variance for the pattern and automorphism counts are given in Appendices A and B.

2. Pattern counts

In this section, we apply Theorem 1.1 to analyse the distribution of the number of occurrences of a tree pattern H as an induced subtree in uniform random labelled tree $\boldsymbol{T}$. To our knowledge, the strongest results for this problem were obtained by Chysak et al. [5] and Janson [16].

Chysak et al. [5] consider occurrences of a pattern H as an induced subgraph of a tree T with the additional restriction that the *internal* vertices in the pattern match the degrees the corresponding vertices in T. That is, the other edges of T can only be adjacent to leaves of H. For example, the tree T on Figure 2 contains only three paths on three vertices in this sense, namely $T[\{1, 5, 8\}]$, $T[\{1, 3, 6\}]$, and $T[\{3, 6, 13\}]$. In particular, the induced path on vertices 1, 2, 7 is not counted since the internal vertex 2 is adjacent to 4. The result by Chysak, Drmota, Klausner, Kok is given below.

Theorem 2.1 ([5, Theorem 1]). *Let H to be a given finite tree. Then the limiting distribution of the number of occurrences of H (in the sense described above) in* $\boldsymbol{T}$ *is asymptotically normal with mean and variance asymptotically equivalent to* μn *and* $\sigma^2 n$, *where* $\mu > 0$ *and* $\sigma^2 \geqslant 0$ *depend on the pattern H and can be computed explicitly and algorithmically and can be represented as polynomials (with rational coefficients) in* $1/e$.
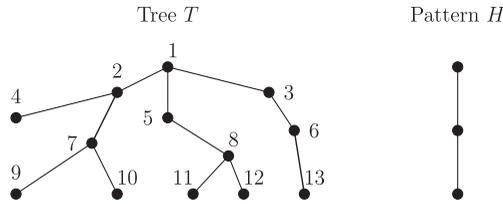
Figure 2. A labelled tree *T* and a pattern *H*.

Janson [16] considers the subtree counts $\eta_H(T)$ defined differently. Set vertex 1 to be a root of $T \in \mathcal{T}_n$. For any other vertex v, let T_v be the subtree consisting of v an all its descendants. Such subtrees are called *fringe subtrees*. The parameter $\eta_H(T)$ equals the number of fringe subtrees isomorphic to H (with a root). For example, the tree T on Figure 2 contains only one path with three vertices (rooted at end vertex), namely $T[\{3, 6, 13\}]$. In particular, the induced paths $T[\{1, 5, 8\}]$ and $T[\{1, 3, 6\}]$ are not counted since they are not fringe subtrees. Janson [16] proved the following result about joint asymptotic normality for several such subtree counts.

Theorem 2.2 ([16, Corollary 1.8]) *Let $\boldsymbol{T}_n^{\mathrm{GW}}$ be a conditioned Galton–Watson tree of order n with offspring distribution ξ, where $\mathbb{E}[\xi] = 1$ and $0 < \sigma^2 := \mathrm{Var}[\xi] < \infty$. Then, the subtree counts $\eta_H(\boldsymbol{T}_n^{\mathrm{GW}})$ (for all H from a given set of patterns) are asymptotically jointly normal.*

Janson [16, Corollary 1.8] also gives expressions for the covariances of the limiting distribution in terms of the distribution of the corresponding unconditioned Galton–Watson tree. To relate this model to uniform random labelled tree $\boldsymbol{T}$, one need to take the conditioned Galton–Watson tree of order n with the Poisson offspring distribution.

We consider a more general type of tree counts which encapsulates both counts from above. In fact, it was suggested by Chysak et al. [5]: "*...we could also consider pattern-matching problems for patterns in which some degrees of certain possibly external "filled" nodes must match exactly while the degrees of the other, possibly internal "empty" nodes might be different. But then the situation is more involved.*" Then, in [5, Section 5.3] they explain that having an internal "empty" node leads to serious complications in their approach.

We define our tree parameter formally. Let H be a tree with ℓ vertices $v_1, \ldots, v_\ell$. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_\ell) \in \{0, 1\}^\ell$. We say the pattern $(H, \boldsymbol{\theta})$ *occurs* in a tree $T \in \mathcal{T}_n$ if there exists a pair of sets (U, W) such that $W \subset U \subset [n]$ and

- the induced subgraph $T[U]$ is isomorphic to H,
- the set W corresponds to all vertices v_i with $\theta_i = 1$ ("empty" nodes),
- there is no edge in T between $U - W$ and $[n] - U$.

Denote by $N_{H,\boldsymbol{\theta}}(T)$ the number of occurrences of the pattern $(H, \boldsymbol{\theta})$ in T that is the number of different pairs (U, W) satisfying the above. It equals the number of ways to choose suitable identities for $v_1, \ldots, v_\ell$ in $[n]$ divided by $|\mathrm{AUT}\,(H, \boldsymbol{\theta})|$, the number of automorphisms of H that preserve $\boldsymbol{\theta}$. In particular, if $\theta_i = 1$ for all $i \in [\ell]$ then $N_{H,\boldsymbol{\theta}}(T)$ is the number of induced subgraphs in T isomorphic H. If $\theta_i = 1$ whenever i is a leaf of H, then $N_{H,\boldsymbol{\theta}}(T)$ is the tree count considered in Theorem 2.1. If $\theta_i = 1$ for exactly one vertex $i \in [\ell]$ which is a leaf in H, then $N_{H,\boldsymbol{\theta}}(T)$ counts fringe subtrees.

In Section 2.2, we prove that $N_{H,\boldsymbol{\theta}}(\boldsymbol{T})$ is asymptotically normal for any fixed H and $\boldsymbol{\theta} \in \{0, 1\}^\ell$ with at least one non-zero component (where ℓ is the number of vertices in H). Note that if $\theta_i = 0$ for all $i \in [\ell]$ and $n > \ell$, then $N_{H,\boldsymbol{\theta}}(\boldsymbol{T}) = 0$ since at least one vertex corresponding to H must be adjacent to other vertices in T. Our approach also works for growing patterns. We demonstrate it for the case when H is a path.

2.1. Moments calculation

To apply Theorem 1.1, we need a lower bound for $\mathrm{Var}(N_{H,\boldsymbol{\theta}}(T))$. One can compute the moments of $N_{H,\boldsymbol{\theta}}(T)$ using the following formula for the number of trees containing a given spanning forest. Lemma 2.3 is a straightforward generalisation of [21, Theorem 6.1] with almost identical proof, which we include here for the sake of completeness.

Lemma 2.3. *Let $S = H_1 \sqcup \ldots \sqcup H_k$ be a forest on $[n]$ and B_i be non-empty subsets (not necessarily proper) of $V(H_i)$ for all $i \in [k]$. Then, the number of trees T on $[n]$ containing all edges of H such that $\deg_T(v) = \deg_S(v)$ for every v outside $B_1 \cup \ldots \cup B_k$ equals $b_1 \cdots b_k (b_1 + \cdots + b_k)^{k-2}$, where b_i is the number of vertices in B_i.*

Proof. Any desired tree T corresponds to a tree T_H on k vertices labelled by $H_1, \ldots, H_k$ for which the vertices H_i and H_j are adjacent if and only if there is an edge between H_i and H_j in T. If $d_1, \ldots, d_k$ are degrees of T_H, then the number of trees T corresponding to T_H equals $b_1^{d_1} \ldots b_k^{d_k}$ since we can only use vertices from $B_1 \cup \ldots \cup B_k$ for edges of T. From [21, Theorem 3.1], we know that the number of trees on k vertices with degrees $d_1, \ldots, d_k$ is $\binom{k-2}{d_1-1,\ldots,d_k-1}$. Thus, the total number of such trees T is

$$\sum_{(d_1,\ldots,d_k)} b_1^{d_1} \ldots b_k^{d_k} \binom{k-2}{d_1-1,\ldots,d_k-1} = b_1 \ldots b_k (b_1 + \cdots + b_k)^{k-2},$$

where the sum is over all positive integers sequences that $d_1 + \cdots + d_k = 2k - 2$. $\square$

For an ℓ-tuple $\boldsymbol{u} = (u_1, \ldots, u_\ell) \in [n]^\ell$ with distinct coordinates, let $\mathbb{1}_{\boldsymbol{u}}(T)$ be the indicator of the event that a pattern $(H, \boldsymbol{\theta})$ occurs in T with $u_1, \ldots, u_\ell$ corresponding to the vertices of H. Let $s := \sum_{i=1}^{\ell} \theta_i$. Applying Lemma 2.3 to a forest consisting of one nontrivial component isomorphic to H and dividing by $|\mathcal{T}_n| = n^{n-2}$, we find that

$$\mathbb{E}\left[\mathbb{1}_{\boldsymbol{u}}(T)\right] = \frac{s(n-\ell+s)^{n-\ell-1}}{n^{n-2}} = \frac{se^{-\ell+s+O(\ell^2/n)}}{n^{\ell-1}}. \tag{3}$$

Summing over all choices for $\boldsymbol{u}$ and dividing by $|\mathrm{AUT}(H, \boldsymbol{\theta})|$ to adjust overcounting, we get

$$\mathbb{E}\left[N_{H,\boldsymbol{\theta}}(T)\right] = \frac{1}{|\mathrm{AUT}(H, \boldsymbol{\theta})|} \sum_{\boldsymbol{u}} \mathbb{E}\left[\mathbb{1}_{\boldsymbol{u}}(T)\right] = n\frac{se^{-\ell+s+O(\ell^2/n)}}{\mathrm{AUT}(H, \boldsymbol{\theta})}.$$

In particular, this formula agrees with Theorem 2.1 that μ is a polynomial with rational coefficients in $1/e$. Similarly, for the variance, we have

$$\mathrm{Var}\left[N_{H,\boldsymbol{\theta}}(T)\right] = \frac{1}{|\mathrm{AUT}(H, \boldsymbol{\theta})|^2} \sum_{\boldsymbol{u},\boldsymbol{u}'} \mathrm{Cov}(\mathbb{1}_{\boldsymbol{u}}(T), \mathbb{1}_{\boldsymbol{u}'}(T)), \tag{4}$$

where the sum over all ℓ-tuples $\boldsymbol{u}, \boldsymbol{u}' \in [n]^\ell$ with distinct coordinates. Then, we can also use Lemma 2.3 (with one or two nontrivial components) to compute $\mathrm{Cov}(\mathbb{1}_{\boldsymbol{u}}(T), \mathbb{1}_{\boldsymbol{u}'}(T))$. However, this computation is much more involved: one needs to consider all possible ways the pattern $(H, \boldsymbol{\theta})$ intersects with itself. Nevertheless, for a fixed pattern, it is not difficult to see that $\mathbb{E}\left[\mathbb{1}_{\boldsymbol{u}}(T)\right]$ and $\mathbb{E}\left[\mathbb{1}_{\boldsymbol{u}}(T)\mathbb{1}_{\boldsymbol{u}'}(T)\right]$ are polynomials with integer coefficients in $1/e$ divided by some power of n. This

observation is already sufficient to establish the bound $\text{Var}\left[N_{H,\boldsymbol{\theta}}(T)\right] = \Omega(n)$ for the case when $\sum_{i=1}^{\ell} \theta_i < \ell$.

Lemma 2.4. *Let $(H, \boldsymbol{\theta})$ be a fixed pattern, ℓ be the number of vertices in tree H, and $s := \sum_{i=1}^{\ell} \theta_i$. Then, there exist a polynomial $p_{H,\boldsymbol{\theta}}$ of degree at most $2\ell - 2s$ with integer coefficients that*

$$\text{Var}\left[N_{H,\boldsymbol{\theta}}(T)\right] = n \, \frac{p_{H,\boldsymbol{\theta}}(1/e)}{|\text{AUT}(H, \boldsymbol{\theta})|^2} + O(1).$$

Moreover, if $s < \ell$ then $p_{H,\boldsymbol{\theta}}(1/e) > 0$.

Proof. Consider any ℓ-tuples $\boldsymbol{u}, \boldsymbol{u}' \in [n]^{\ell}$ with distinct coordinates. If the coordinates of $\boldsymbol{u}$ and $\boldsymbol{u}'$ form disjoint sets, then applying Lemma 2.3 to a forest consisting of two nontrivial component isomorphic to H, we find that

$$\mathbb{E}\left[\mathbb{1}_{\boldsymbol{u}}(T)\mathbb{1}_{\boldsymbol{u}'}(T)\right] = \frac{s^2 \, (n - 2\ell + 2s)^{n-2\ell}}{n^{n-2}}.$$

Using (3), we get that

$$\text{Cov}(\mathbb{1}_{\boldsymbol{u}}(T), \mathbb{1}_{\boldsymbol{u}'}(T)) = \frac{s^2}{n^{2\ell-2}} \left(e^{-2\ell + 2s - \frac{(2\ell-2s)^2}{2n} + O(n^{-2})} - e^{2\left(-\ell+s - \frac{(\ell-s)^2}{2n}\right) + O(n^{-2})} \right)$$

$$= -\frac{s^2(\ell - s)^2 e^{-2\ell + 2s}}{n^{2\ell-1}} + O(n^{-2\ell}).$$

Then, the contribution of such $\boldsymbol{u}, \boldsymbol{u}'$ to the sum $\sum_{\boldsymbol{u},\boldsymbol{u}'} \text{Cov}(\mathbb{1}_{\boldsymbol{u}}(T), \mathbb{1}_{\boldsymbol{u}'}(T))$ in (4) equals

$$-ns^2(\ell - s)^2 e^{-2\ell + 2s} + O(1).$$

Next, we proceed to the case when the sets formed by the coordinates of $\boldsymbol{u}$ and $\boldsymbol{u}'$ intersect. Let a be the size of the union of these two sets and

$$b := |\{u_i : \theta_i = 1\} \cap \{u'_i : \theta_i = 1\}|.$$

Note that $\ell - s \leqslant a - b \leqslant 2\ell - 2s$. Then, using Lemma 2.3 (and also (3)), we find that

$$\text{Cov}(\mathbb{1}_{\boldsymbol{u}}(T), \mathbb{1}_{\boldsymbol{u}'}(T)) = \frac{b(n - a + b)^{n - a - 1}}{n^{n-2}} - \left(\frac{s(n - \ell + s)^{n - \ell - 1}}{n^{n-2}} \right)^2$$

$$= \frac{1 + O(n^{-1})}{n^{a-1}} \cdot \begin{cases} be^{-a+b}, & \text{if } a \leqslant 2\ell - 2, \\ be^{-a+b} - s^2 e^{-2\ell + 2s}, & \text{if } a = 2\ell - 1. \end{cases}$$

We say a pair $(\boldsymbol{u}, \boldsymbol{u}')$ is equivalent to $(\boldsymbol{w}, \boldsymbol{w}')$ if there is a permutation σ of the set $[n]$ that $w_i = \sigma(u_i)$ and $w'_i = \sigma(u'_i)$ for all $i \in [\ell]$. Note that the number of pairs equivalent to $(\boldsymbol{u}, \boldsymbol{u}')$ is exactly $(n)_a$. Then, the contribution of the equivalence class to the sum $\sum_{\boldsymbol{u},\boldsymbol{u}'} \text{Cov}(\mathbb{1}_{\boldsymbol{u}}(T), \mathbb{1}_{\boldsymbol{u}'}(T))$ in (4) is $nbe^{-a+b} + O(1)$ or $nbe^{-a+b} - s^2 e^{-2\ell+2s} + O(1)$. Summing over all equivalence classes, we complete the proof of the first part.

For the second part, observe in the above that $a - b = \ell - s$ if and only if the sets of coordinates of $\boldsymbol{u}$ and $\boldsymbol{u}'$ coincide and $\{u_i : \theta_i = 1\} = \{u'_i : \theta_i = 1\}$. In particular, we have $a < 2\ell - 1$ so $\text{Cov}(\mathbb{1}_{\boldsymbol{u}}(T), \mathbb{1}_{\boldsymbol{u}'}(T)) > 0$. Then, the coefficient corresponding to $x^{-\ell+s}$ in $p_{H,\theta}(x)$ is strictly positive so the polynomial $p_{H,\theta}(x)$ is not trivial. Since the number $1/e$ is transcendental, we conclude that $p_{H,\theta}(1/e)$ is not zero. Also, $p_{H,\theta}(1/e)$ cannot be negative since $\text{Var}\left[N_{H,\theta}(T)\right] \geqslant 0$ so it can only be positive. This completes the proof. $\qquad\square$

For a tree $T \in \mathcal{T}_n$, let $N_H(T) := N_{H,\theta}(T)$ if $\theta_i = 1$ for all $i \in [\ell]$ that is $N_H(T)$ is the number of induced subgraphs of T isomorphic to H. Unfortunately, the lemma above cannot guarantee that $\mathrm{Var}\,[N_H(T)] = \Omega(n)$. In this case, the polynomial $p_{H,\theta}$ is a non-negative constant, but an additional argument is required to show that it is not zero.

Lemma 2.5. *For any fixed tree H with degrees $h_1, \ldots, h_\ell$, we have*

$$\mathrm{Var}\,[N_H(T)] \geqslant \frac{n}{|\mathrm{AUT}\,(H)\,|^2} \sum_{j \geqslant 2} c_j^2\, j! + O(1),$$

where $c_j = \sum_{i=1}^{\ell} \left(\binom{h_i}{j} + (\ell - 1)\binom{h_i - 1}{j} \right)$. In particular, $c_2 > 0$ if $\ell \geqslant 3$.

The proof of Lemma 2.5 is given in Appendix A of the ArXiv version [15] of the current paper. The key idea of this proof is to estimate the variance of the conditional expectation value of $N_H(T)$ given the degree sequence of T.

Remark 2.6. There is a different way to show $\mathrm{Var}\left[N_{H,\theta}(T)\right] = \Omega(n)$ for any fixed H and θ (including the case $\theta_i = 1$ for all $i \in [\ell]$). First, one establishes that $\mathbb{P}(N_{H,\theta}(T) = x_n) = o(1)$ for any sequence x_n. Reducing/incrementing the number of fringe copies of H in a clever way shows that $\mathbb{P}(N_{H,\theta}(T) = x_n)$ is not much larger than $\mathbb{P}(N_{H,\theta}(T) = x_n - k) + \mathbb{P}(N_{H,\theta}(T) = x_n + k)$ for all k from a sufficiently large set. This implies that $\mathrm{Var}\left[N_{H,\theta}(T)\right] \to \infty$. Therefore, $p_{H,\theta} > 0$ so $\mathrm{Var}\left[N_{H,\theta}(T)\right] = \Omega(n)$. In fact, the proof of Lemma 2.5 given in [15, Appendix A] is more technically involved than this idea, but it extends better to growing substructures.

Using formula (4), we also obtain a precise estimate of $\mathrm{Var}\,[N_H(T)]$ for the case when H is a path. With slight abuse of notations, let $P_\ell(T) := N_{P_\ell}(T)$ that is the number of paths on ℓ vertices in a tree $T \in \mathcal{T}_n$.

Lemma 2.7. *Let $\ell > 2$ and $\ell = O(n^{1/2})$, then*

$$\mathrm{Var}\,[P_\ell(T)] = \left(1 + O\left(\frac{\ell^2}{n}\right)\right) n\, \frac{\ell(\ell - 1)^2(\ell - 2)}{24}.$$

Proof. For the induced path counts formula (4) simplifies as follows:

$$\mathrm{Var}\,[P_\ell(T)] = \frac{1}{4} \sum_{u,u'} \mathrm{Cov}(\mathbb{1}_u(T), \mathbb{1}_{u'}(T)).$$

For $i \in [\ell]$, let Σ_i be the set of pairs (u, u') that the sets formed by its coordinates have exactly i elements in common. From (3), we have that $\mathbb{E}\,[\mathbb{1}_u(T)] = \ell n^{1-\ell}$. Using Lemma 2.3, we get $\mathbb{E}\left[\mathbb{1}_u(T)\mathbb{1}_{u'}(T)\right] = \ell^2 n^{2-2\ell}$ for $(u, u') \in \Sigma_0$, so

$$\sum_{(u,u') \in \Sigma_0} \mathrm{Cov}(\mathbb{1}_u(T), \mathbb{1}_{u'}(T)) = 0.$$

Applying Lemma 2.3, it is a routine to check that

$$\sum_{(u,u') \in \Sigma_1} \mathrm{Cov}(\mathbb{1}_u(T), \mathbb{1}_{u'}(T)) = |\Sigma_1| \left((2\ell - 1)n^{2-2\ell} - \ell^2 n^{2-2\ell} \right)$$

$$= -\frac{(n)_{2\ell - 1}}{n^{2\ell - 2}} \ell^2(\ell - 1)^2 = -\left(1 + O\left(\frac{\ell^2}{n}\right)\right) n\ell^2(\ell - 1)^2.$$

Similarly, for $2 \leqslant i \leqslant \ell$, we get

$$\sum_{(u,u') \in \Sigma_i} \mathrm{Cov}(\mathbb{1}_u(T), \mathbb{1}_{u'}(T)) = |\Sigma_i| \left((2\ell - i)n^{1 - 2\ell + i} - \ell^2 n^{2 - 2\ell} \right)$$

$$= \left(1 + O\left(\frac{\ell}{n}\right) \right) \frac{2(n)_{2\ell - i}}{n^{2\ell - i - 1}} (\ell - i + 1)^2 (2\ell - i)$$

$$= \left(1 + O\left(\frac{\ell^2}{n}\right) \right) 2n(\ell - i + 1)^2 (2\ell - i).$$

Summing the above bounds for $\Sigma_0, \ldots, \Sigma_\ell$ and using

$$-\ell^2 (\ell - 1)^2 + 2 \sum_{i=2}^{\ell} (\ell - i + 1)^2 (2\ell - i) = \frac{\ell(\ell - 1)^2 (\ell - 2)}{6},$$

we get the stated formula for $\mathrm{Var} P_\ell(T)$. □

2.2. Asymptotic normality of pattern counts

Here we apply Theorem 1.1 to derive the limiting distribution of the pattern counts $N_{H,\theta}(T)$. In fact, all applications of Theorem 1.1 typically have short proofs leaving the lower bound for the variance to be the most technically involved part.

Theorem 2.8. *Let H be a tree on ℓ vertices and $\theta \in \{0, 1\}^\ell$ be a non-zero vector. Then $N_{H,\theta}(T)$ is asymptotically normal and $\delta_K \left[N_{H,\theta}(T) \right] = O(n^{-1/4 + \varepsilon})$ for any $\varepsilon > 0$.*

Proof. For a tree $T \in \mathcal{T}_n$, let $F(T)$ be the number of occurrences of (H, θ) in the induced subforest of T for the set of vertices with degrees at most $\log n$ in T.

Removing one edge from T can only destroy at most $\log^\ell n$ patterns (H, θ) counted in $F(T)$. Thus, F is α-Lipschitz with $\alpha = 2 \log^\ell n$. If two perturbations S_i^{jk} and S_a^{bc} are at distance at least 3ℓ in T, then every pattern (H, θ) counted in $F(S_i^{jk} S_a^{bc} T) - F(T)$ (with positive or negative sign) is present in exactly one of the terms $F(S_i^{jk} T) - F(T)$ and $F(S_a^{bc} T) - F(T)$ (with the same sign). Thus, F is ρ-superposable with $\rho = 3\ell$.

From (1), we know that

$$\mathbb{P}(F(T) \neq N_{H,\theta}(T)) = e^{-\omega(\log n)}.$$

Since the values of these random variables are not bigger than n^ℓ, we get

$$\mathbb{E}[F(T)] = \mathbb{E}[N_{H,\theta}(T)] + e^{-\omega(\log n)},$$

$$\mathrm{Var}[F(T)] = \mathrm{Var}[N_{H,\theta}(T)] + e^{-\omega(\log n)}.$$

Combining Lemmas 2.4 and 2.5, we get that $\mathrm{Var}[F(T)] = \Omega(n)$. Applying Theorem 1.1, we complete the proof. □

In the next result, we allow the pattern to grow, but restricted to the case when H is a path and all θ_i equal 1 (all vertices are "empty").

Theorem 2.9. *Let $\ell = O(n^{1/8 - \delta})$ for some fixed $\delta \in (0, 1/8)$. Then $P_\ell(T)$ is asymptotically normal and $\delta_K [P_\ell(T)] = O(n^{-\varepsilon'})$ for any $\varepsilon' \in (0, 2\delta)$.*

Proof. For a tree $T \in \mathcal{T}_n$, let

$$V_{\mathrm{good}}(T) := \left\{ i \in [n] : \text{for all } d \in [n], \text{ we have } |\{j \in [n] : d_T(i, j) = d\}| \leqslant d \log^4 n \right\}.$$

Define $F(T)$ to be the number of induced paths on ℓ vertices in the forest $T[V_{\mathrm{good}}(T)]$.

The number of ℓ-paths counted in $F(T)$ containing any fixed edge is at most

$$\log^8 n \sum_{i=1}^{\ell-2} i(\ell - i - 1) \leqslant \frac{1}{2}\ell^3 \log^8 n.$$

Arguing similarly to the proof of Theorem 2.8, we conclude that F is α-Lipschitz with $\alpha = \ell^3 \log^8 n$ and ρ-superposable with $\rho = 3\ell$. From Theorem 1.4, we also get

$$\mathbb{P}(F(\boldsymbol{T}) \neq P_\ell(\boldsymbol{T})) = e^{-\omega(\log n)}.$$

Next, for a tree $T \in \mathcal{T}_n$, observe that $F(T) \leqslant P_\ell(T) \leqslant n^2$ since any path in T is uniquely determined by the choice of its end vertices. The rest of the argument is identical to the proof of Theorem 2.8. □

3. Number of automorphisms

An *automorphism* of a graph G is a bijection $\sigma : V(G) \to V(G)$ such that the edge set of G is preserved under σ. Bona and Flajolet [3] studied this parameter for random unlabelled rooted non-plane trees and random phylogenetic trees (rooted non-plane binary trees with labelled leaves). They showed that in both cases the distribution is *asymptotically lognormal*; that is, the logarithm of the number of automorphisms in a random tree is asymptotically normal. McKeon [20] proved asymptotic formulas for the number automorphisms in related random models of unlabelled locally restricted trees.

In her PhD thesis, Yu [30] determined the asymptotics of $\mathbb{E}\left[\log |\mathrm{AUT}\,(\boldsymbol{T})|\right]$ for uniform random labelled tree $\boldsymbol{T}$. She also made the following conjecture:

Conjecture 3.1. [30] The distribution of $|\mathrm{AUT}\,(\boldsymbol{T})|$ is asymptotically lognormal.

In this section, we prove this conjecture. Unfortunately, we cannot immediately apply Theorem 1.1 to derive the distribution of the number of automorphisms since the logarithm of this parameter is not ρ-superposable for a sufficiently small ρ. This happens because some trees have automorphisms affected by both perturbations S_i^{jk} and S_a^{bc} even if $d_T(\{j, k\}, \{b, c\})$ is large. Instead, we start by looking at $\mathrm{AUT}_r\,(T)$, the subgroup of $\mathrm{AUT}\,(T)$ consisting of automorphisms $\sigma \in \mathrm{AUT}\,(T)$ such that $\sigma(r) = r$, where r is some fixed vertex from $[n]$. In other words, $\mathrm{AUT}_r\,(T)$ is the number of rooted automorphisms of a tree T with root r, or equivalently the stabilizer of r.

The parameter $\mathrm{AUT}_r\,(\boldsymbol{T})$ is easier to work with while also remaining asymptotically very similar to $\mathrm{AUT}\,(\boldsymbol{T})$. The ease of analysis comes from the product representation of $|\mathrm{AUT}_r\,(T)|$ given by Yu [30, Corollary 2.1.3].

$$|\mathrm{AUT}_r\,(T)| = \prod_{i \in [n]} \prod_B N_i(B, T, r)! \tag{5}$$

The product over B represents a product over isomorphism classes of rooted unlabelled trees. Define a *branch* of T at v to be a subtree rooted at an immediate descendent (with respect to r) of v. That is the branch is a fringe subtree of T at this descendent. The term $N_i(B, T, r)$ denotes the number of branches isomorphic to B at vertex i. Factorisation (5) also follows from the result of Stacey and Holton that says every rooted automorphism is a product of branch transpositions [24, Lemma 2.4].

We give an example of (5) in Figure 3 for a tree on 9 vertices. There are only three types of branches in this tree with respect to the root $r = 1$, namely B_1, B_2, and B_3. Vertex 1 has two branches isomorphic to B_2, and thus $N_1(B_2, T, r)! = 2! = 2$. It also has one branch isomorphic to B_1, and thus $N_1(B_1, T, r)! = 1$. Vertex 2 has three branches isomorphic to B_3, and
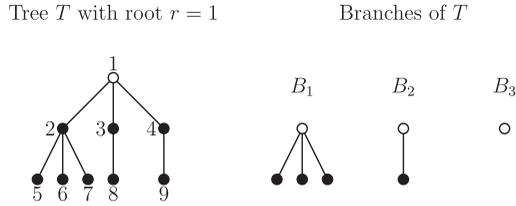
Figure 3. A labelled tree on the left and its (rooted, unlabelled) branches on the right.

thus $N_2(B_3, T, r)! = 3! = 6$. Vertices 3 and 4 each have one branch isomorphic to B_3, and thus $N_3(B_3, T, r)! = N_4(B_3, T, r)! = 1$. Applying (5) shows that $|\text{AUT}_r(T)| = 3! \cdot 2! = 12$.

To define our tree parameter $F(T)$, we look at a subgroup of $\text{AUT}_r(T)$ based on small automorphisms. We define a *small branch* to be a branch with at most $4 \log n$ vertices, any branch that is not small is *large*. A *small automorphism* is an automorphism where any vertex that is the root of a large branch is fixed. For a given tree T, let $\text{AUT}_{\text{small}} \subseteq \text{AUT}_r(T)$ be the set of small automorphisms.

Lemma 3.2. $\text{AUT}_{\text{small}}$ *is a subgroup of* $\text{AUT}_r(T)$.

Proof. Observe that any automorphism in $\text{AUT}_{\text{small}}$ must also have an inverse in $\text{AUT}_{\text{small}}$, since they move the same vertices. Furthermore, to prove closure under composition, suppose that $a, b \in \text{AUT}_{\text{small}}$ but $ab \notin \text{AUT}_{\text{small}}$. Let B be a large branch that is mapped by ab onto B'. Then all of the vertices in B are moved by either a or b. Since $a \in \text{AUT}_{\text{small}}$, there are some vertices in B not moved by a; denote this set by X. Since B is connected, there exists an edge between X and $V(B) \backslash X$ in the edge set of B. Thus, there exists an edge between aX and $aV(B)$ in T; however, this creates a cycle and thus a contradiction. Thus, ab must also only move small branches, and thus $ab \in \text{AUT}_{\text{small}}$. Thus, $\text{AUT}_{\text{small}}$ is a subgroup. $\square$

The parameter $F(T)$ is obtained by writing $|\text{AUT}_{\text{small}}|$ in the same product representation as $|\text{AUT}_r(T)|$ and taking the logarithm:

$$F(T) := \log|\text{AUT}_{\text{small}}| = \sum_{i \in [n]} \sum_{B \in \mathcal{B}_{\text{small}}} \log(N_i(B, T, r)!). \tag{6}$$

Here $\mathcal{B}_{\text{small}}$ is the set of small branches.

Remark 3.3. In fact, the parameter F defined above belongs to a larger class of additive functionals considered by Janson [16] and Wagner [27]. They established a general CLT for this type of parameters. [16, Theorem 1.3] and [27, Theorem 2] do not cover the number of automorphisms in T because $\mathbb{E}\left[\left(\sum_B \log(N_i(B, T, r)!)\right)^2\right]$ is not vanishing. In fact, it is bounded below by the second moment of the number of leaves attached to a given vertex which tends to a positive constant; see also the estimates given in [15, Appendix B].

Next, we show that $F(T)$ satisfies assumptions of Theorem 1.1 while also being very close to $\log|\text{AUT}_r(T)|$.

Lemma 3.4. *Let* $\alpha = 3 \log n$ *and* $\rho = 10 \log n$. *Then* $F(T)$ *as defined in* (6) *is* α-*Lipschitz and* ρ-*superposable.*

Proof. To prove the Lipschitz property, we show that for any two trees T and T' differing by a perturbation S_i^{jk}, the order of $\text{AUT}_{\text{small}}$ for each tree can differ by at most a factor of n^3. Any automorphism of T fixing $\{i, j, k\}$ is an automorphism of T', since all other edges remained static so their orbits are unaffected. Let G_{ijk} be the subgroup of $\text{AUT}_{\text{small}}$ that fixes $\{i, j, k\}$. Then the

cosets of this subgroup are defined by where they send each of these vertices. Since there are at most n such options for each element in the set, we get at most n^3 cosets. By Lagrange's theorem, we get that

$$\text{AUT}_{\text{small}}(T) \geqslant \frac{|\text{AUT}_{\text{small}}(T')|}{n^3}$$

and vice versa by swapping the roles of T and T'. Taking the logarithm of both sides gives the desired bound.

Next, we show that F is ρ-superposable. Suppose $d = d_T \left(\{j, k\}, \{b, c\} \right) > 10 \log n$. Then suppose an automorphism $\sigma \in \text{AUT}_{\text{small}}(T)$ is created or destroyed by S_i^{jk}. Then σ must not fix $\{i, j, k\}$. Any path between one of $\{j, k\}$ and one of $\{b, c\}$ must be longer than $10 \log n$. Therefore, any parent vertex in the tree is strictly more than $5 \log n$ distance from at least one vertex in each pair. So σ must fix $\{a, b, c\}$ and all lower branches, since each branch moved by the automorphism is at most $4 \log n$. So S_a^{bc} cannot affect the presence or absence of σ in $\text{AUT}(T)$. Similarly, any automorphism created or destroyed by S_a^{bc} cannot be affected by S_i^{jk}. Thus,

$$F \left(S_i^{jk} S_a^{bc} T \right) - F(T) = \left(F \left(S_i^{jk} T \right) - F(T) \right) + \left(F \left(S_a^{bc} T \right) - F(T) \right).$$

This completes the proof. $\qquad\square$

In the next lemma, we derive bounds needed to compare $\text{AUT}(T)$ and $F(T)$.

Lemma 3.5. *The following statements hold.*

   a. $\left| \log |\text{AUT}_r(T)| - \log |\text{AUT}(T)| \right| \leqslant \log n$ *for all* $T \in \mathcal{T}_n$,

   b. $\mathbb{P} \left(F(T) \neq \log |\text{AUT}_r(T)| \right) = O \left(\frac{1}{n^3} \right),$

   c. $\mathbb{E} \left[\log |\text{AUT}(T)| \right] - \mathbb{E} [F(T)] = O(\log n),$

   d. $\text{Var} \left[\log |\text{AUT}(T)| \right] - \text{Var} [F(T)] = O \left(\sqrt{n \log n} \right).$

Proof. Each automorphism in $\text{AUT}_r(T)$ is an automorphism in $\text{AUT}(T)$. The group $\text{AUT}(T)$ operates on $[n]$ such that $\text{AUT}_r(T)$ is the stabilizer of r. Hence,

$$|\text{AUT}_r(T)| \leqslant |\text{AUT}(T)| = |\text{Orbit}(r)| \times |\text{AUT}_r(T)| \leqslant n |\text{AUT}_r(T)|.$$

Thus, we get (a). Parts (b) follows almost immediately from results by Yu [30, Corollary 2.2.2]. To show part (c), we use parts (a) and (b) and observe $F(T) \leqslant |\text{AUT}_r(T)| \leqslant \log n! \leqslant n \log n$ to get that

$$\mathbb{E} \left[\log |\text{AUT}(T)| - F(T) \right] < \max_T |\log |\text{AUT}(T)| - \log |\text{AUT}_r(T)||$$

$$+ \mathbb{P} \left(F(T) \neq \log |\text{AUT}_r(T)| \right) n \log n$$

$$\leqslant \log n + O \left(\frac{\log n}{n^2} \right) = O \left(\log n \right).$$

Finally, we proceed to part (d). Let $W = F(T) - \log |\text{AUT}_r(T)|$ and $Z = \log |\text{AUT}_r(T)| - \log |\text{AUT}(T)|$. From Lemma 3.5(a,b,c), we get that

$$|\text{Var}[W] + \text{Cov}(Z, W)| \leqslant \mathbb{P} \left(F(T) \neq \log |\text{AUT}_r(T)| \right) 2n^2 \log^2 n = O \left(\frac{\log^2 n}{n} \right)$$

$$\text{Var}[Z] \leqslant \mathbb{E} Z^2 \leqslant \log^2 n,$$

$$|\text{Cov} \left(F(T), W + Z \right)| \leqslant (\text{Var}[F(T)] \, \text{Var}[W + Z])^{1/2} = O \left(\sqrt{n \log n} \right).$$

Then, we have

$$\mathrm{Var}\left[\log|\mathrm{Aut}\,(\boldsymbol{T})|\right] = \mathrm{Var}\left[F(\boldsymbol{T}) - [W + Z]\right] = \mathrm{Var}\left[F(\boldsymbol{T})\right] + O\left(\sqrt{n\log n}\right).$$

$\square$

The final ingredient needed to apply Theorem 1.1 is a bound on the variance of $F(\boldsymbol{T})$, given in the lemma below.

Lemma 3.6. *For sufficiently large n, we have* $\mathrm{Var}\left[F(\boldsymbol{T})\right] \geqslant 0.002\,n$.

The proof of Lemma 3.6 is lengthy and quite technical. We include it in Appendix B of the ArXiv version [15] of the current paper.

Now, we are ready to prove the following result.

Theorem 3.7. *Conjecture 3.1 is true. Furthermore,* $\delta_K\left[\log|\mathrm{Aut}\,(\boldsymbol{T})|\right] = O\left(n^{-\frac{1}{4}+\epsilon}\right)$ *and* $\delta_K\left[\log|\mathrm{Aut}_r\,(\boldsymbol{T})|\right] = O\left(n^{-\frac{1}{4}+\epsilon}\right)$ *for any $\epsilon > 0$.*

Proof of Theorem 3.7. Combining Lemmas 3.4, 3.5, and 3.6, we get that the parameter F defined in (6) satisfies all the assumptions of Theorem 1.1 and $\delta_k[F(\boldsymbol{T})] = O(n^{-1/4+\epsilon})$ for any $\epsilon > 0$. Using Lemma 3.5 and recalling that $F(T) \leqslant \log|\mathrm{Aut}_r\,(\boldsymbol{T})| \leqslant \log\mathrm{Aut}\,(\boldsymbol{T})$, we get that $\log|\mathrm{Aut}_r\,(\boldsymbol{T})|$ and $\log|\mathrm{Aut}\,(\boldsymbol{T})|$ have the same limiting distribution as F(T) (with the same bound for the Kolmogorov distance). $\square$

Remark 3.8. Recently, Stufler and Wagner [28] have also announced progress in showing that the distributions of $|\mathrm{Aut}\,(T)|$ and $|\mathrm{Aut}_r\,(T)|$ are asymptotically lognormal; however, it has not yet appear in any published or arXiv paper. Their method is based on the analysis of the generating function and is different from our approach. Stufler and Wagner gave much more accurate values for the mean and variance in their talk [28], specifically $\mathbb{E}\left[\log|\mathrm{Aut}\,(T)|\right] \approx 0.052290n$ and $\mathrm{Var}\left[\log|\mathrm{Aut}\,(T)|\right] = 0.039498n$.

4. Tools from the theory of martingales

Let $\mathcal{P} = (\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A sequence $\mathcal{F}_0, \ldots, \mathcal{F}_n$ of sub-σ-fields of $\mathcal{F}$ is a *filtration* if $\mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_n$. A sequence $Y_0, \ldots, Y_n$ of random variables on $\mathcal{P}$ is a *martingale with respect to* $\mathcal{F}_0, \ldots, \mathcal{F}_n$ if

  i. Y_i is $\mathcal{F}_i$-measurable and $|Y_i|$ has finite expectation, for $0 \leq i \leq n$;
  ii. $\mathbb{E}\left[Y_i \mid \mathcal{F}_i\right] = Y_{i-1}$ for $1 \leq i \leq n$.

In the following we will always assume that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and so $Y_0 = \mathbb{E}[Y_n]$.

In this section, we state some general results on concentration and limiting distribution for martingales. In fact, we only need these results for discrete uniform probability spaces, where the concept of martingale reduces to average values over increasing set systems. In this case, Ω is a finite set and each σ-field $\mathcal{F}_i$ is generated by unions of blocks of a partition of Ω. Following McDiarmid [19], for $i = 0, \ldots, n$ we define the *conditional range* of a random variable X on $\mathcal{P}$ as

$$\mathrm{ran}[X \mid \mathcal{F}_i] := \sup\,[X \mid \mathcal{F}_i] + \sup\,[-X \mid \mathcal{F}_i]. \tag{7}$$

Here, $\sup\,[X \mid \mathcal{F}_i]$ is the $\mathcal{F}_i$-measurable random variable which takes the value at $\omega \in \Omega$ equal to the maximum value of X over the block of $\mathcal{F}_i$ containing ω (and similarly for $-X$). More generally, "supremum" can be replaced by "essential supremum". For more information about conditional range and diameter, see, for example, [14, Section 2.1] and references therein. We will use that the conditional range is a seminorm and, in particular, it is subadditive.

Our first tool is the following result of McDiarmid [19]. Further in this section, the notation $\mathrm{ran}_i[\,\cdot\,]$ stands for $\mathrm{ran}[\cdot \mid \mathcal{F}_i]$.

Theorem 4.1 ([19, Theorem 3.14]) *Let* $Y_0, Y_1, \ldots, Y_n$ *be a real-valued martingale with respect to the filtration* $\{\emptyset, \Omega\} = \mathcal{F}_0, \mathcal{F}_1 \ldots, \mathcal{F}_n$. *Denote*

$$R^2 := \sum_{i=1}^{n} (\mathrm{ran}_{i-1}[Y_i])^2 .$$

Then, for any $r, t > 0$

$$\mathbb{P}\left(|Y_n - Y_0| \geqslant t\right) \leqslant 2\exp\left(-2t^2/r^2\right) + 2\mathbb{P}\left(R^2 > r^2\right).$$

The normalized quadratic variation of a martingale sequence $\mathbf{Y} = (Y_0, \ldots, Y_n)$ is defined by

$$Q[\mathbf{Y}] := \frac{1}{\mathrm{Var}\,[Y_n]} \sum_{i=1}^{n} (Y_i - Y_{i-1})^2.$$

Observe that

$$\mathbb{E}\left[(Y_i - Y_{i-1})^2\right] = \mathbb{E}\left[\mathrm{Var}[Y_i \mid F_{i-1}]\right] = \mathbb{E}\left[\mathbb{E}\left[Y_i^2 - Y_{i-1}^2 \mid F_{i-1}\right]\right] = \mathbb{E}\left[Y_i^2 - Y_{i-1}^2\right]. \qquad (8)$$

Thus,

$$\mathbb{E}Q[\mathbf{Y}] = \frac{1}{\mathrm{Var}\,[Y_n]} \sum_{i=1}^{n} \left(\mathbb{E}\left[Y_i^2\right] - \mathbb{E}\left[Y_{i-1}^2\right]\right) = 1.$$

A classical result by Brown [4] states that if the increments $Y_i - Y_{i-1}$ have finite variances, $Q[\mathbf{Y}] \xrightarrow{\text{prob.}} 1$ as $n \to \infty$ and a certain Lindeberg-type condition is satisfied then the limiting distribution of Y_n is normal, i.e. $\delta_{\mathrm{K}}[Y_n] \to 0$. For a more restricted class of martingales with bounded differences, these conditions can be slightly simplified and will be sufficient for our purposes. Our second tool is the following result of Mourrat [23] which gives an explicit bound on the rate of convergence in the CLT under a strengthened condition that the normalized quadratic variation $Q[\mathbf{Y}]$ converges to 1 in L^p.

Theorem 4.2 ([23, Theorem 1.5.]) *Let* $p \in [1, +\infty)$ *and* $\gamma \in (0, +\infty)$. *There exists a constant* $C_{p,\gamma} > 0$ *such that, for any real martingale sequence* $\mathbf{Y} = (Y_0, \ldots, Y_n)$ *satisfying* $|Y_i - Y_{i-1}| \leqslant \gamma$ *for all* $i = 1, \ldots, n$,

$$\delta_{\mathrm{K}}[Y_n] \leqslant C_{p,\gamma} \left(\frac{n \log n}{(\mathrm{Var}\,[Y_n])^{3/2}} + \left(\mathbb{E}\left[|Q[\mathbf{Y}] - 1|^p\right] + (\mathrm{Var}\,[Y_n])^{-p}\right)^{1/(2p+1)} \right).$$

One way to bound the term $\mathbb{E}\left[|Q[\mathbf{Y}] - 1|^p\right]$ in the above is by applying Theorem 4.1 to the martingale for $Q[\mathbf{Y}]$ with respect to the same filtration, as which gives the following lemma.

Lemma 4.3. *Let* $Y_0, \ldots, Y_n$ *be a real-valued martingale with respect to the filtration* $\{\emptyset, \Omega\} = \mathcal{F}_0, \ldots, \mathcal{F}_n$. *For* $\hat{q} > 0$, *let* $\mathcal{A}_{\hat{q}}$ *denote the event*

$$\sum_{i=1}^{n} \left(\mathrm{ran}_{i-1}\left[\mathrm{Var}\,[Y_n \mid \mathcal{F}_i]\right] + (\mathrm{ran}_{i-1}[Y_i])^2\right)^2 > \left(\hat{q}\,\mathrm{Var}\,[Y_n]\right)^2 .$$

Then, for any $p \in [1, +\infty)$, *we have*

$$\mathbb{E}\left[|Q[\mathbf{Y}] - 1|^p\right] \leqslant c_p\,\hat{q}^p + 2\mathbb{P}\left(\mathcal{A}_{\hat{q}}\right)\sup|Q[\mathbf{Y}] - 1|^p,$$

where $c_p = 2p \int_0^{+\infty} e^{-2x^2} x^{p-1} dx$.

Proof. By definition, we have that $|Y_i - Y_{i-1}| \leqslant \mathrm{ran}_{i-1}[Y_i]$ for all $i \in [n]$. Therefore,

$$\mathrm{ran}_{i-1}\left[(Y_i - Y_{i-1})^2\right] \leqslant (\mathrm{ran}_{i-1}[Y_i])^2.$$

Observe also $\mathrm{ran}_{i-1}\left[(Y_j - Y_{j-1})^2\right] = 0$ for any $j < i$. Then, using (8) and the subadditivity of the conditional range, we get that

$$\mathrm{ran}_{i-1}\left[\mathbb{E}\left[Q[\boldsymbol{Y}] \mid \mathcal{F}_i\right]\right] = \frac{1}{\mathrm{Var}\left[Y_n\right]}\mathrm{ran}_{i-1}\left[\sum_{j=i}^{n} \mathbb{E}\left[(Y_j - Y_{j-1})^2 \mid \mathcal{F}_i\right]\right]$$

$$= \frac{\mathrm{ran}_{i-1}\left[\mathrm{Var}\left[Y_n \mid \mathcal{F}_i\right] + (Y_i - Y_{i-1})^2\right]}{\mathrm{Var}\left[Y_n\right]}$$

$$\leqslant \frac{\mathrm{ran}_{i-1}\left[\mathrm{Var}\left[Y_n \mid \mathcal{F}_i\right]\right] + (\mathrm{ran}_{i-1}[Y_i])^2}{\mathrm{Var}\left[Y_n\right]}.$$

Applying Theorem 4.1 to the martingale $\{\mathbb{E}\left[Q[\boldsymbol{Y}] \mid \mathcal{F}_i\right]\}_{i=0,\dots,n}$, we find that

$$\mathbb{P}\left(|Q[\boldsymbol{Y}] - 1| \geqslant t\right) \leqslant 2\exp\left(-2t^2/\hat{q}^2\right) + 2\mathbb{P}\left(\mathcal{A}_{\hat{q}}\right).$$

Substituting this bound into

$$\mathbb{E}\left[|Q(\boldsymbol{Y}) - 1|^p\right] = \int_0^{t_{\max}} \mathbb{P}\left(|Q(\boldsymbol{Y}) - 1| \geqslant t\right) pt^{p-1}dt$$

and changing the variable $t = \hat{q}x$, we complete the proof. Here, $t_{\max} = \sup|Q(\boldsymbol{Y}) - 1|$. $\qquad\square$

Using the formulas for $\mathbb{E}\left[(Y_j - Y_{j-1})^2 \mid \mathcal{F}_i\right]$ similar to (8), we find that

$$\mathrm{Var}\left[Y_n \mid \mathcal{F}_i\right] = \sum_{j=i+1}^{n} \mathbb{E}\left[(Y_j - Y_{j-1})^2 \mid \mathcal{F}_i\right]. \tag{9}$$

Then, by the subadditivity of the conditional range, we get the next bound, which will be useful in applying Lemma 4.3.

$$\mathrm{ran}_{i-1}\left[\mathrm{Var}\left[Y_n \mid \mathcal{F}_i\right)\right] \leqslant \sum_{j=i+1}^{n} \mathrm{ran}_{i-1}\mathbb{E}\left[(Y_j - Y_{j-1})^2 \mid \mathcal{F}_i\right]. \tag{10}$$

The Doob martingale construction is another important tool in our argument. Suppose $\boldsymbol{X} = (X_1, \dots, X_n)$ is a random vector on $\mathcal{P}$ taking values in S and $f : S \to \mathbb{R}$ is such that $f(\boldsymbol{X})$ has bounded expectation. Consider the filtration $\mathcal{F}_0, \dots \mathcal{F}_n$ defined by $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$ which is the σ-field generated by random variables $X_1, \dots X_i$. Then, the Doob martingale $\boldsymbol{Y}^{\mathrm{Doob}} = \boldsymbol{Y}^{\mathrm{Doob}}(f, \boldsymbol{X})$ is defined by, for all $i = 0, \dots, n$,

$$Y_i^{\mathrm{Doob}} := \mathbb{E}\left[f(X_1, \dots, X_n) \mid \mathcal{F}_i\right].$$

In case of finite S, the random variables Y_i^{Doob}, $\mathrm{Var}\left[Y_n^{\mathrm{Doob}} \mid \mathcal{F}_j\right]$ and $\mathrm{ran}_i[Y_n^{\mathrm{Doob}}]$ can be seen as functions $f_i, v_i, r_i : S \to \mathbb{R}$ of the random vector $\boldsymbol{X}$ defined as follows: for $\boldsymbol{x} \in S$,

$$f_i(\boldsymbol{x}) := \mathbb{E}\left[f(\boldsymbol{X}) \mid X_1 = x_1, \dots, X_i = x_i\right] = \mathbb{E}\left[f(x_1, \dots, x_i, X_{i+1} \dots, X_n)\right],$$

$$v_i(\boldsymbol{x}) := \mathrm{Var}\left[f(\boldsymbol{X}) \mid X_1 = x_1, \dots, X_i = x_i\right] = \mathrm{Var}\left[f(x_1, \dots, x_i, X_{i+1} \dots, X_n)\right],$$

$$r_i(\boldsymbol{x}) := \mathrm{ran}\left[f(\boldsymbol{X}) \mid X_1 = x_1, \dots, X_i = x_i\right]$$

$$= \max_{\boldsymbol{y}} f(x_1, \dots, x_i, y_{i+1} \dots, y_n) - \min_{\boldsymbol{y}} f(x_1, \dots, x_i, y_{i+1} \dots, y_n), \tag{11}$$

where $x_1, \ldots, x_i$ are fixed and $X_{i+1}, \ldots, X_n$ are random and both max and min are over $\boldsymbol{y} \in S$ such that $y_j = x_j$ for $j = 1, \ldots, i$. If, in addition, random variables $X_1, \ldots, X_n$ are independent then

$$|Y_i^{\mathrm{Doob}} - Y_{i-1}^{\mathrm{Doob}}| \leqslant \mathrm{ran}_{i-1}\left[Y_i^{\mathrm{Doob}}\right] \leqslant \max_{\boldsymbol{x}, \boldsymbol{x}'} |f(\boldsymbol{x}) - f(\boldsymbol{x}')|, \qquad (12)$$

where the maximum is over $\boldsymbol{x}, \boldsymbol{x}' \in S$ that differ only in the ith coordinate.

In particular, the Doob martingale process is applicable for functions of random permutations since we can represent them as vectors. Let S_n be the set of permutations of $[n]$. We write $\omega = (\omega_1, \ldots, \omega_n) \in S_n$ if ω maps j to ω_j. The product of two permutuations $\omega, \sigma \in S_n$ is defined by

$$\omega \circ \sigma := (\omega_{\sigma_1}, \ldots, \omega_{\sigma_n})$$

which corresponds to the composition of ω and σ if we treat them as functions on $[n]$. For a function $f : S_n \to \mathbb{R}$ and $1 \leqslant i \neq j \leqslant n - 1$, define

$$\alpha_i[f] := \sum_{a=i+1}^{n} \frac{\max_{\omega \in S_n} |f(\omega) - f(\omega \circ (ia))|}{n - i},$$

$$\Delta_{ij}[f] := \sum_{a=i+1}^{n} \sum_{b=j+1}^{n} \frac{\max_{\omega \in S_n} |f(\omega) - f(\omega \circ (ia)) - f(\omega \circ (jb)) + f(\omega \circ (jb) \circ (ia))|}{(n - i)(n - j)}.$$

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a uniform random element of S_n and $\boldsymbol{Y}^{\mathrm{Doob}}(f, \boldsymbol{X})$ be the Doob martingale sequence for $f(\boldsymbol{X})$. Note that $Y_n^{\mathrm{Doob}} = Y_{n-1}^{\mathrm{Doob}} = f(\boldsymbol{X})$ since the first $n - 1$ coordinates X_i determine the permutation $\boldsymbol{X}$ uniquely.

Lemma 4.4. *If $Y^{\mathrm{Doob}} = Y^{\mathrm{Doob}}(f, \boldsymbol{X})$ where $f : S_n \to \mathbb{R}$ and $\boldsymbol{X}$ is a uniform random element of S_n, then*

a. $|Y_i^{\mathrm{Doob}} - Y_{i-1}^{\mathrm{Doob}}| \leqslant \mathrm{ran}_{i-1}\left[Y_i^{\mathrm{Doob}}\right] \leqslant \alpha_i[f]$, *for all $1 \leqslant i \leqslant n - 1$.*

b. $\mathrm{ran}_{i-1}\left[\mathbb{E}\left[(Y_j^{\mathrm{Doob}} - Y_{j-1}^{\mathrm{Doob}})^2 \mid \mathcal{F}_i\right]\right] \leqslant 2\alpha_j[f]\Delta_{ij}[f]$, *for all $1 \leqslant i < j \leqslant n - 1$.*

Proof. To show the first inequality in part (a), we observe that

$$-\sup(-Y_i^{\mathrm{Doob}} \mid \mathcal{F}_{i-1}) \leqslant Y_{i-1}^{\mathrm{Doob}} \leqslant \sup(Y_i^{\mathrm{Doob}} \mid \mathcal{F}_{i-1}),$$

by definition. The other bounds is a special case of [11, Lemma 2.1.] for real-valued random variables, where the conditional range is the same as the conditional diameter. $\square$

5. Martingales for tree parameters

To prove Theorem 1.1, we use the martingale based on the Aldous–Broder algorithm, which generates a random spanning tree of a given graph G. Here is a quick summary: (1) consider the random walk starting from any vertex; (2) every time we traverse an edge which takes us to a vertex we have not yet explored, add this edge to the tree; (3) stop when we visited all vertices. The resulting random graph has uniform distribution over the set of spanning trees of G, for more details see [1]. If G is the complete graph K_n, $n \geqslant 2$, this construction can be rephrased as the following two-stage procedure [1, Algorithm 2]:

I. For $1 \leqslant i \leqslant n - 1$ connect vertex $i + 1$ to vertex $V_i = \min\{i, U_i\}$, where $\boldsymbol{U} = (U_1, \ldots, U_{n-1})$ is uniformly distributed on $[n]^{n-1}$.

II. Relabel vertices $1, \ldots, n$ as $X_1, \ldots, X_n$, where $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a uniform random permutation from S_n.

Let $T(\boldsymbol{u})$ is the tree produced at stage I given that $\boldsymbol{U} = \boldsymbol{u}$. For a permutation $\omega \in S_n$ and a tree $T \in \mathcal{T}_n$, let

$$T^\omega := \text{ the tree obtained from } T \text{ by relabelling according to } \omega.$$

From [1] we know that $T(\boldsymbol{U})^{\boldsymbol{X}}$ has uniform distribution on the set $\mathcal{T}_n$. Now, a tree parameter $F : \mathcal{T}_n \to \mathbb{R}$ can be seen as a function with domain $[n]^{n-1} \times S_n$. Consider the functions $\hat{F} : \mathcal{T}_n \to \mathbb{R}$ and $F_T : S_n \to \mathbb{R}$ defined by

$$\hat{F}(T) := \mathbb{E}\left[F(T^{\boldsymbol{X}})\right], \qquad F_T(\omega) := F(T^\omega). \tag{13}$$

Let $\boldsymbol{Y} = (Y_0, \dots, Y_{n-1})$ and $\boldsymbol{Z}(T) = (Z_0(T), \dots, Z_{n-1}(T))$ be the Doob martingale sequences for $\hat{F}(T(\boldsymbol{U}))$ and $F_T(\boldsymbol{X})$, respectively: for $i = 0, \dots, n-1$,

$$Y_i := \mathbb{E}\left[\hat{F}(T(\boldsymbol{U})) \mid \mathcal{F}_i\right] \qquad \text{and} \qquad Z_i(T) := \mathbb{E}\left[F_T(\boldsymbol{X}) \mid \mathcal{G}_i\right], \tag{14}$$

where the filtrations are $\mathcal{F}_i = \sigma(U_1, \dots, U_i)$ and $\mathcal{G}_i = \sigma(X_1, \dots, X_i)$. We construct the martingale for $F(\boldsymbol{T})$ by combining the above two sequences together. Further in this section, we will use the following notations for conditional statistics of a random variable W with respect to $\mathcal{F}_i$ and $\mathcal{G}_i$:

$$\mathbb{E}_{\mathcal{F}_i}[W] := \mathbb{E}\left[W \mid \mathcal{F}_i\right], \qquad \mathbb{E}_{\mathcal{G}_i}[W] := \mathbb{E}[W \mid \mathcal{G}_i],$$

$$\mathrm{Var}_{\mathcal{F}_i}[W] := \mathrm{Var}\left[W \mid \mathcal{F}_i\right], \qquad \mathrm{Var}_{\mathcal{G}_i}[W] := \mathrm{Var}[W \mid \mathcal{G}_i],$$

$$\sup_{\mathcal{F}_i}[W] := \sup[W \mid \mathcal{F}_i], \qquad \sup_{\mathcal{G}_i}[W] := \sup[W \mid \mathcal{G}_i],$$

$$\mathrm{ran}_{\mathcal{F}_i}[W] := \mathrm{ran}[W \mid \mathcal{F}_i], \qquad \mathrm{ran}_{\mathcal{G}_i}[W] := \mathrm{ran}[W \mid \mathcal{G}_i].$$

5.1. Properties of F_T and $\hat{F}$

First, we study properties of functions F_T and $\hat{F}$ from (13) given that the parameter F is α-Lipschitz and ρ-superposable.

Lemma 5.1. *Let a tree parameter $F : \mathcal{T}_n \to \mathbb{R}$ be α-Lipschitz and ρ-superposable for some $\alpha \geqslant 0$ and $\rho \geqslant 1$, then*

  a. *$\hat{F}$ is α-Lipschitz and ρ-superposable.*

Furthermore, the following holds for all trees $T \in \mathcal{T}_n$ and permutations $\omega \in S_n$.

  b. *If (ia) is a transposition from S_n, then*

$$|F_T(\omega) - F_T(\omega \circ (ia))| \leqslant \alpha(\deg_T(i) + \deg_T(a)),$$

  where $\deg_T(i)$, $\deg_T(a)$ are degrees of i, a in the tree T.

  c. *Let $T' = \mathrm{S}_q^{rs} T$ be a tree for some triple (q, r, s). If (ia) is a transposition from S_n that $d_T(\{i, a\}, \{r, s\}) \geqslant \rho + 1$, then*

$$F_T(\omega) - F_T(\omega \circ (ia)) - F_{T'}(\omega) + F_{T'}(\omega \circ (ia)) = 0.$$

  d. *If (ia), (jb) are transpositions from S_n such that $d_T(\{i, a\}, \{j, b\}) \geqslant \rho + 2$, then*

$$F_T(\omega) - F_T(\omega \circ (ia)) - F_T(\omega \circ (jb)) + F_T(\omega \circ (jb) \circ (ia)) = 0.$$

Proof. For any permutation $\omega = (\omega_1, \dots, \omega_n) \in S_n$ define the function $F_\omega : \mathcal{T}_n \to \mathbb{R}$ by $F_\omega(T) := F(T^\omega)$. If $\mathrm{S}_i^{jk} T$ is a tree then $(\mathrm{S}_i^{jk} T)^\omega = \mathrm{S}_{\omega_i}^{\omega_j \omega_k} T^\omega$. Relabelling also does not change the distances, that is, $d_T(a, b) = d_{T^\omega}(\omega_a, \omega_b)$ for all $a, b \in [n]$. Thus, F_ω is α-Lipschitz and ρ-superposable. Averaging over all ω proves part (a).

For part (b), we show that the tree $T^{(ia)}$ can be obtained from T by performing at most $\deg_T(i) + \deg_T(a)$ tree perturbations S_x^{yz}. We denote the set of these perturbations by $\mathcal{P}_T^{ia}$. Let u and v be the vertices on the path from i to a in T adjacent to i and a, respectively. Consider $\deg_T(i) - 1$ perturbations S_x^{ia} for all vertices $x \neq u$ adjacent to i and $\deg_T(a) - 1$ perturbations S_x^{ai} for all for all vertices $x \neq v$ adjacent to a. If $d_T(a, i) \leqslant 2$ then performing these $\deg_T(i) + \deg_T(a) - 2$ perturbations in any order turns T into $T^{(ia)}$. Otherwise, all vertices i, a, u, v are distinct and we need two more perturbations S_i^{uv} and S_a^{vu} to obtain $T^{(ia)}$. This defines the set $\mathcal{P}_T^{ia}$. Now, since F is α-Lipschitz, the value of the function changes by at most α after each perturbation so

$$\left| F(T) - F\left(T^{(ia)} \right) \right| \leqslant \alpha(\deg_T(i) + \deg_T(a)).$$

The above holds for any $T \in \mathcal{T}_n$. Substituting T^ω and observing $\deg_{T^\omega}(\omega_i) = \deg_T(i)$, we prove part (b).

Before proving parts (c) and (d), we outline some important properties of the set $\mathcal{P}_T^{ia}$ of the tree perturbations that turn T into $T^{(ia)}$.

    i. The perturbations of $\mathcal{P}_T^{ia}$ can be performed in any order, that is, all intermediate graphs are trees.

    ii. $d_T(\{x, y, z\}, \{i, a\}) = 0$ for any $S_x^{yz} \in \mathcal{P}_T^{ia}$, that is, $\{x, y, z\} \cap \{i, a\} \neq \emptyset$.

    iii. the distance from any $w \in [n]$ to $\{i, a\}$ is unchanged by perturbations S_x^{ia} or S_x^{ai}.

    iv. the distance from any $w \in [n]$ to $\{i, a\}$ can increase after performing one of the perturbations S_i^{uv} or S_a^{vu} but then it decreases back to the initial value after performing the second (so it never gets smaller than the initial distance $d_T(w, \{i, a\})$).

For (c), observe first that $d_T(\{i, a\}, \{r, s\}) \geqslant 2$ implies that i and a are adjacent to the same sets of vertices in T and T'. Consider first the case when both u and v belong to the path from i to a in the tree T'. For example, this is always the case when the path from i to a is not affected by removing the edge qr. Then, by definition, $\mathcal{P}_T^{ia} = \mathcal{P}_{T'}^{ia}$ that is we can use the same sets of perturbations to change labels i and a in both trees. We order them arbitrary to form a sequence $(S_1, \ldots, S_k)$. Note also that for any perturbation $S_x^{yz} \in \mathcal{P}_T^{ia}$ we have $d_T(\{y, z\}, \{r, s\}) \geqslant \rho$ due to the property (ii) and $d_T(\{i, a\}, \{r, s\}) \geqslant \rho + 1$. Since F is ρ-superposable and using properties (iii) and (iv), we get that

$$F(S_t \cdots S_1 T) - F(S_t \cdots S_1 T') - F(S_{t+1} \cdots S_1 T) + F(S_{t+1} \cdots S_1 T') = 0.$$

Summing up these equalities for all $t = 0, \ldots k - 1$, we get that

$$F(T) - F(T') - F\left(T^{(ia)} \right) + F\left(T'^{(ia)} \right) = 0. \tag{15}$$

We still need to consider the case when removing qr changes the path from i to a such that u or v do not lie on the path anymore. In this case, one have to be slightly more careful with the order of perturbations $(S_1, \ldots, S_k)$ to avoid the appearance of cycles in $S_t \cdots S_1 T'$. Without loss of generality, we may assume that $d_T(i, q) < d_T(i, r)$ (otherwise, swap the roles of i and a). Let v' be the vertex adjacent to a that lies on the path from i to a in T'. In notations of part (b), we define $S_1 = S_i^{uv}$ and $S_2 = S_{v'}^{ai}$, then put the remaining perturbations in any order. A sequence $(S_1, \ldots, S_k)$ defined in this way ensures that all intermediate steps from T' to $T'^{(ia)}$ are trees. Repeating the same argument as above, we prove (15). To complete the proof of part (c), we just need to substitute T by T^ω similarly to part (b).

Finally, we prove (d) by repeatedly using part (c) for a sequence of perturbations $S_q^{rs} \in \mathcal{P}_T^{jb}$ that turn T into $T^{(jb)}$. We can apply part (c) for all intermediate trees T' because the assumption $d_T(\{i, a\}, \{j, b\}) \geqslant \rho + 2$ together with properties (ii), (iii), (iv) implies that $d_{T'}(\{i, a\}, \{r, s\}) \geqslant \rho + 1$.

5.2. Martingale properties

Here, we establish the properties of martingales Y and $Z(T)$ from (14) needed to apply the results of Section 4. For a tree $T \in \mathcal{T}_n$ and $A, B \subset [n]$, define

$$\mathbb{1}_T^\rho(A, B) := \begin{cases} 1, & \text{if } d_T(A, B) < \rho, \\ 0, & \text{otherwise.} \end{cases}$$

We will repeatedly use the fact that for any $T \in \mathcal{T}_n$ and $i \in [n]$, we have

$$\sum_{j=1}^{n} \mathbb{1}_T^\rho(\{i\}, \{j\}) \leqslant \rho^2 \beta(T), \tag{16}$$

where $\beta(T)$ is the parameter defined in (2). In the following, for simplicity of notations, we write $\mathbb{1}_T^\rho(i, B)$, or $\mathbb{1}_T^\rho(A, j)$, or $\mathbb{1}_T^\rho(i, j)$ when A, or B, or both are one-element sets. Let $\mathcal{T}_n^d \subset \mathcal{T}_n$ be the set of trees with degrees at most d. We denote by $a \wedge b$ the minimum of two real numbers a, b.

Lemma 5.2. *Let $F : \mathcal{T}_n \to \mathbb{R}$ be α-Lipschitz and ρ-superposable for some $\alpha \geqslant 0$ and $\rho \geqslant 1$. Then, the following holds for all $i \in [n-1]$, $d \in \mathbb{R}^+$ and $T \in \mathcal{T}_n^d$.*

a. $|Y_i - Y_{i-1}| \leqslant \operatorname{ran}_{\mathcal{F}_{i-1}}[Y_i] \leqslant \alpha$.

b. $\operatorname{ran}_{\mathcal{F}_{i-1}}\left[\operatorname{Var}_{\mathcal{F}_i}[Y_{n-1}]\right] \leqslant 32\alpha^2\rho^2 \sup_{\mathcal{F}_{i-1}}\left[\mathbb{E}_{\mathcal{F}_i}[\beta(T(\boldsymbol{U}))]\right].$

c. $|Z_i(T) - Z_{i-1}(T)| \leqslant \operatorname{ran}_{\mathcal{G}_{i-1}}[Z_i(T)] \leqslant \max_{\omega,(ia)\in S_n} |F(T^\omega) - F(T^{\omega\circ(ia)})| \leqslant 2\alpha d.$

d. $\operatorname{ran}_{\mathcal{G}_{i-1}}\left[\operatorname{Var}_{\mathcal{G}_i}[Z_{n-1}(T)]\right] \leqslant 64\alpha^2 d^2 (\rho+2)^2 \beta(T) \log n.$

e. *Let $V(\boldsymbol{u}) := \operatorname{Var}[Z_{n-1}(T(\boldsymbol{u}))] = \operatorname{Var}\left[F_{T(\boldsymbol{u})}(X)\right]$. Then, $0 \leqslant V(\boldsymbol{U}) \leqslant 4\alpha^2 n^2$ and*

$$\operatorname{ran}_{\mathcal{F}_{i-1}}\left[\mathbb{E}_{\mathcal{F}_i}\left[V(\boldsymbol{U})\mathbb{1}_{T(\boldsymbol{U})\in\mathcal{T}_n^d}\right]\right]$$
$$\leqslant \alpha^2 \sup_{\mathcal{F}_{i-1}}\left[\mathbb{E}_{\mathcal{F}_i}\left[4n^2 \mathbb{1}_{T(\boldsymbol{U})\notin\mathcal{T}_n^d} + 8d^2(\rho+1)^2\beta(T(\boldsymbol{U}))\right]\right].$$

Proof. Using bound (12), we find that

$$|Y_i - Y_{i-1}| \leqslant \operatorname{ran}_{\mathcal{F}_{i-1}}[Y_i] \leqslant \max |\hat{F}(T(\boldsymbol{u})) - \hat{F}(T(\boldsymbol{u}'))|,$$

where $\boldsymbol{u}, \boldsymbol{u}' \in [n]^{n-1}$ differ in ith coordinate. Observe that

$$T(\boldsymbol{u}') = \mathrm{S}_{i+1}^{i\wedge u_i\, i\wedge u_i'} T(\boldsymbol{u}). \tag{17}$$

From Lemma 5.1(a), we know that $\hat{F}(T)$ is α-Lipschitz. Part (a) follows.

As explained in (11), we have $Y_i = f_i(\boldsymbol{U})$, where

$$f_i(\boldsymbol{u}) = \mathbb{E}\left[\hat{F}(T(\boldsymbol{U})) \mid u_{\leqslant i}\right]$$

and $\mathbb{E}(\cdot \mid u_{\leqslant i})$ stands for $\mathbb{E}(\cdot \mid U_1 = u_1, \ldots, U_i = u_i)$. Let $0 \leqslant i < j \leqslant n-1$. Using formula (17), we find that

$$f_j(\boldsymbol{u}) - f_{j-1}(\boldsymbol{u}) = \frac{1}{n}\sum_{u=1}^{n} \mathbb{E}\left[\hat{F}(T(\boldsymbol{U})) - \hat{F}\left(\mathrm{S}_{j+1}^{j\wedge u_j\, j\wedge u} T(\boldsymbol{U})\right) \mid u_{\leqslant j}\right].$$

Consider $\boldsymbol{u}' \in [n]^{n-1}$ that differs from $\boldsymbol{u}$ only in ith coordinate. Then, we have

$$
\begin{aligned}
f_j(\boldsymbol{u}) - f_{j-1}(\boldsymbol{u}) - f_j(\boldsymbol{u}') + f_{j-1}(\boldsymbol{u}') &= \frac{1}{n} \sum_{u=1}^{n} \mathbb{E}\bigg[\hat{F}(T(\boldsymbol{U})) - \hat{F}\left(S_{j+1}^{j \wedge u_j \, j \wedge u} T(\boldsymbol{U}) \right) \\
&\quad - \hat{F}\left(S_{i+1}^{i \wedge u_i \, i \wedge u_i'} T(\boldsymbol{U}) \right) + \hat{F}\left(S_{j+1}^{j \wedge u_j \, j \wedge u} S_{i+1}^{i \wedge u_i \, i \wedge u_i'} T(\boldsymbol{U}) \right) \mid u_{\leqslant j} \bigg]
\end{aligned}
$$

From part (a), we have $0 \leqslant |f_j - f_{j-1}| \leqslant \alpha$. Observe also that if $U_1 = u_1, \ldots, U_{j-1} = u_{j-1}$ and $v \in [i]$ then

$$
d_{T(\boldsymbol{U})}(v, \{i \wedge u_i, i \wedge u_i'\}) = d_{T(\boldsymbol{u})}(v, \{i \wedge u_i, i \wedge u_i'\}).
$$

That is, the distance between v and $\{i \wedge u_i, i \wedge u_i'\}$ is completely determined by $u_1, \ldots, u_{j-1}$ and v. From Lemma 5.1(a), we know that $\hat{F}(T)$ is ρ-superposable. Thus, we find that

$$
\begin{aligned}
|(f_j(\boldsymbol{u}) - f_{j-1}(\boldsymbol{u}))^2 - (f_j(\boldsymbol{u}') - f_{j-1}(\boldsymbol{u}'))^2| &\leqslant 2\alpha |f_j(\boldsymbol{u}) - f_{j-1}(\boldsymbol{u}) - f_j(\boldsymbol{u}') + f_{j-1}(\boldsymbol{u}')| \\
&\leqslant \frac{4\alpha^2}{n} \sum_{u=1}^{n} \mathbb{1}_{T(\boldsymbol{u})}^{\rho}(\{j \wedge u_j, j \wedge u\}, \{i \wedge u_i, i \wedge u_i'\}))
\end{aligned}
$$

Using (16), we can bound

$$
\begin{aligned}
\frac{1}{n} \sum_{u=1}^{n} \mathbb{E}&\left[\mathbb{1}_{T(\boldsymbol{u})}^{\rho}(\{j \wedge u_j, j \wedge u\}, \{i \wedge u_i, i \wedge u_i'\}) \mid u_{\leqslant j-1} \right] \\
&= \frac{1}{n^2} \sum_{u=1}^{n} \sum_{u_j=1}^{n} \mathbb{1}_{T(\boldsymbol{u})}^{\rho}(\{j \wedge u_j, j \wedge u\}, \{i \wedge u_i, i \wedge u_i'\}) \\
&\leqslant 2 \cdot \mathbb{1}_{T(\boldsymbol{u})}^{\rho}(j, \{i \wedge u_i, i \wedge u_i'\}) + \frac{2}{n} \sum_{k=1}^{j-1} \mathbb{1}_{T(\boldsymbol{u})}^{\rho}(k, \{i \wedge u_i, i \wedge u_i'\}) \\
&\leqslant 2 \cdot \mathbb{1}_{T(\boldsymbol{u})}^{\rho}(j, i \wedge u_i) + 2 \cdot \mathbb{1}_{T(\boldsymbol{u})}^{\rho}(j, i \wedge u_i') + \frac{4}{n} \rho^2 \beta(T(\boldsymbol{u})).
\end{aligned}
$$

Similarly to (11), let $\operatorname{ran}_{\mathcal{F}_{i-1}} \left[\operatorname{Var}_{\mathcal{F}_i}[Y_{n-1}] \right] = r(U_1, \ldots, U_{i-1})$. Using (9), (12) and taking the conditional expectation given $U_1 = u_1, \ldots, U_{i-1} = u_{i-1}$ for the bounds above, we obtain that

$$
\begin{aligned}
r(u_1, \ldots, u_{i-1}) &= \max_{u_i, u_i' \in [n]} \Bigg| \sum_{j=i+1}^{n} \mathbb{E}\left[(f_j(\boldsymbol{U}) - f_{j-1}(\boldsymbol{U}))^2 \mid u_{\leqslant i-1}, U_i = u_i \right] \\
&\qquad\qquad - \sum_{j=i+1}^{n} \mathbb{E}\left[(f_j(\boldsymbol{U}) - f_{j-1}(\boldsymbol{U}))^2 \mid u_{\leqslant i-1}, U_i = u_i' \right] \Bigg| \\
&\leqslant \frac{16\alpha^2}{n} \max_{u \in [n]} \sum_{j=i+1}^{n} \mathbb{E}\left[\mathbb{1}_{T(\boldsymbol{u})}^{\rho}(j, i \wedge u) + \rho^2 \beta(T(\boldsymbol{U})) \mid u_{\leqslant i-1}, U_i = u \right] \\
&\leqslant 32\alpha^2 \rho^2 \max_{u_i \in [n]} \mathbb{E}(\beta(T(\boldsymbol{U})) \mid u_{\leqslant i}).
\end{aligned}
$$

This completes the proof part (b).

Part (c) immediately follows from Lemma 4.4(a) and Lemma 5.1(b). Indeed,

$$
\alpha_i[F_T] \leqslant \max_{\omega, (ia) \in S_n} |F_T(\omega) - F_T(\omega \circ (ia))| \leqslant 2\alpha d. \tag{18}
$$

For (d), recall from (10) that

$$\text{ran}_{\mathcal{G}_{i-1}} \left[\text{Var}_{\mathcal{G}_i}[Z_{n-1}(T)] \right] \leqslant \sum_{j=i+1}^{n-1} \text{ran}_{\mathcal{G}_{i-1}} \left[\mathbb{E}_{\mathcal{G}_i} \left[(Z_j(T) - Z_{j-1}(T))^2 \right] \right]. \tag{19}$$

We will apply Lemma 4.4(b) to estimate the right-hand side of (19). From Lemma 5.1(d) and the bound (18), we get that

$$|F_T(\omega) - F_T(\omega \circ (ia)) - F_T(\omega \circ (jb)) + F_T(\omega \circ (jb) \circ (ia))| \leqslant 4\alpha d \, \mathbb{1}_T^{\rho+2}(\{i, a\}, \{j, b\}).$$

Bounding

$$\mathbb{1}_T^{\rho+2}(\{i, a\}, \{j, b\}) \leqslant \mathbb{1}_T^{\rho+2}(i, j) + \mathbb{1}_T^{\rho+2}(i, b) + \mathbb{1}_T^{\rho+2}(a, j) + \mathbb{1}_T^{\rho+2}(a, b)$$

and using (16), we find that, for $1 \leqslant i < j \leqslant n - 1$,

$$\Delta_{ij}[F_T] \leqslant 4\alpha d \sum_{a=i+1}^{n} \sum_{b=j+1}^{n} \frac{\mathbb{1}_T^{\rho+2}(\{i, a\}, \{j, b\})}{(n-i)(n-j)} \leqslant 4\alpha d \left(\mathbb{1}_T^{\rho+2}(i, j) + \frac{3(\rho+2)^2 \beta(T)}{n-j} \right).$$

Combining (16), (18), Lemma 4.4(b) and the inequality

$$1 + 3 \sum_{k=1}^{n-1} k^{-1} \leqslant 4 + 3 \log n \leqslant 4 \log n,$$

we obtain that

$$\text{ran}_{\mathcal{G}_{i-1}} \left[\text{Var}_{\mathcal{G}_i}[Z_{n-1}(T)] \right] \leqslant \sum_{j=i+1}^{n-1} 16\alpha^2 d^2 \left(\mathbb{1}_T^{\rho+2}(i, j) + \frac{3(\rho+2)^2 \beta(T)}{n-j} \right)$$

$$\leqslant 64\alpha^2 d^2 (\rho+2)^2 \beta(T) \log n.$$

Finally, we proceed to part (e). Since F is α-Lipschitz, we have $|F(T) - F(T')| \leqslant 2\alpha n$ for any two trees $T, T' \in \mathcal{T}_n$. Indeed, applying at most n perturbations of type S_x^{y1}, where x is a leaf, we can turn any tree into a star centered at vertex 1. Thus, we can bound

$$0 \leqslant V(\boldsymbol{u}) \leqslant 4\alpha^2 n^2.$$

Then, for any $\mathcal{A} \subset [n]^{n-1}$ and $u_1, \ldots, u_{i-1} \in [n]$,

$$\text{ran}\left[\mathbb{E}_{\mathcal{F}_i} \left[V(\boldsymbol{U}) \mathbb{1}_{\boldsymbol{U} \in \mathcal{A}} \right] \mid u_{\leqslant i-1} \right]$$

$$= \max_{u \in [n]} \mathbb{E} \left[V(\boldsymbol{U}) \mathbb{1}_{\boldsymbol{U} \in \mathcal{A}} \mid u_{\leqslant i-1}, U_i = u \right] - \min_{u \in [n]} \mathbb{E} \left[V(\boldsymbol{U}) \mathbb{1}_{\boldsymbol{U} \in \mathcal{A}} \mid u_{\leqslant i-1}, U_i = u \right]$$

$$\leqslant 4\alpha^2 n^2 \max_{u_i \in [n]} \mathbb{P}(\boldsymbol{U} \notin \mathcal{A} \mid u_{\leqslant i}) + \max_{u_i, u \in [n]} \mathbb{E} \left[(V(\boldsymbol{U}) - V(\boldsymbol{U}')) \mathbb{1}_{\boldsymbol{U}, \boldsymbol{U}' \in \mathcal{A}} \mid \boldsymbol{u}_{\leqslant i}, U_i' = u \right]$$

where $\boldsymbol{U}'$ differs from $\boldsymbol{U}$ in ith coordinate only. For the following we put $\mathcal{A} = \{\boldsymbol{u} \in [n]^{n-1} : T(\boldsymbol{u}) \in \mathcal{T}_n^d\}$. It remains to bound $V(\boldsymbol{U}) - V(\boldsymbol{U}')$ when $T(\boldsymbol{U}), T(\boldsymbol{U}') \in \mathcal{T}_n^d$.

Consider any $\boldsymbol{u}, \boldsymbol{u}' \in [n]^{n-1}$ that differ in ith coordinate only and $T(\boldsymbol{u}), T(\boldsymbol{u}') \in \mathcal{T}_n^d$. If $T(\boldsymbol{u}) = T(\boldsymbol{u}')$, then $V(\boldsymbol{u}) = V(\boldsymbol{u}')$. Otherwise, recalling (17), we can find some relabelling $\sigma \in S_n$ that the trees $T = T(\boldsymbol{u})^\sigma$, $T' = T(\boldsymbol{u}')^\sigma$ satisfy $T' = S_3^{12} T$ and

$$0 = d_T(1, \{1, 2\}) \leqslant \cdots \leqslant d_T(n, \{1, 2\}).$$

Note that $\mathrm{Var}[F_T(X)] = V(u)$ and $\mathrm{Var}[F_{T'}(X)] = V(u')$. Using Lemma 5.1(c) and (18), we find that, for any $1 \leqslant i < a \leqslant n$,

$$|F_T(\omega) - F_T(\omega \circ (ia)) - F_{T'}(\omega) + F_{T'}(\omega \circ (ia))| \leqslant 4\alpha d\, \mathbb{1}_T^{\rho+1}(\{i, a\}, \{1, 2\})$$

$$\leqslant 4\alpha d\, \mathbb{1}_T^{\rho+1}(i, \{1, 2\}).$$

Applying Lemma 4.4(a) to the function $F_T - F_{T'}$, we obtain

$$|Z_i(T) - Z_{i-1}(T) - Z_i(T') + Z_{i-1}(T')| \leqslant \alpha_i(F_T - F_{T'}) \leqslant 4\alpha d\, \mathbb{1}_T^{\rho+1}(i, \{1, 2\}).$$

We have already proved in part (b) that $|Z_i(T) - Z_{i-1}(T)| \leqslant 2\alpha d$. Using (8) and (16), we bound

$$V(u) - V(u') = \mathrm{Var}\left[Z_{n-1}(T)\right] - \mathrm{Var}\left[Z_{n-1}(T')\right]$$

$$= \sum_{i=1}^{n-1} \mathbb{E}\left[(Z_i(T) - Z_{i-1}(T))^2 - (Z_i(T') - Z_{i-1}(T'))^2\right]$$

$$\leqslant \sum_{i=1}^{n-1} 4\alpha^2 d^2 \mathbb{1}_T^{\rho+1}(i, \{1, 2\}) \leqslant 8\alpha^2 d^2 (\rho + 1)^2 \beta(T).$$

Part (e) follows.　$\square$

5.3. Proof of Theorem 1.1

Before proving of Theorem 1.1, we need one more lemma. Let

$$\mathcal{U}_{\mathrm{small}} := \{u \in [n]^{n-1} : T(u) \in \mathcal{T}_n^{\log n} \text{ and } \beta(T(u)) \leqslant \log^4 n\},$$

$$\mathcal{U}_{\mathrm{big}} := \{u \in [n]^{n-1} : T(u) \in \mathcal{T}_n^{2\log n} \text{ and } \beta(T(u)) \leqslant 2\log^4 n\}.$$

Lemma 5.3. *The following asymptotics bounds hold for any* $u \in \mathcal{U}_{\mathrm{small}}$, $u \in [n]$:

$$\mathbb{P}(U \notin \mathcal{U}_{\mathrm{small}}) = e^{-\omega(\log n)}, \qquad \mathbb{P}(U \notin \mathcal{U}_{\mathrm{big}} \mid u_{\leqslant i-1}, U_i = u) = e^{-\omega(\log n)}.$$

Proof. The first bound follows immediately from (1) and Theorem 1.4. For the second, observe that, for any $u'_1, \ldots, u'_{i-1} \in [n]$,

$$\mathbb{P}\left(U \in \mathcal{U}_{\mathrm{small}} \mid U_1 = u'_1, \ldots, U_{i-1} = u'_{i-1}\right) \leqslant \mathbb{P}\left(U \in \mathcal{U}_{\mathrm{big}} \mid u_{\leqslant i-1}\right).$$

Indeed, let U, U' are such that $U_j = u_j$ and $U'_j = u'_j$ for $j \in [i-1]$ and $U_j = U'_j$ for $j \geqslant i$. Then, $T(U) \subset T(U') \cup T(u)$ because the edges corresponding from $i-1$ steps of the Aldous–Broder algorithm for $T(U)$ lie in $T(u)$, while the remaining edges are covered by $T(U'))$. We know that $u \in \mathcal{U}_{\mathrm{small}}$. Therefore, if $U' \in \mathcal{U}_{\mathrm{small}}$, then $U \in \mathcal{U}_{\mathrm{big}}$.

Next, averaging over all $u'_1, \ldots, u'_{i-1} \in [n]$, we conclude that

$$\mathbb{P}\left(U \notin \mathcal{U}_{\mathrm{big}} \mid u_{\leqslant i-1}\right) \leqslant \mathbb{P}\left(U \notin \mathcal{U}_{\mathrm{small}}\right).$$

Note that, for any $u \in [n]$,

$$\mathbb{P}(U \notin \mathcal{U}_{\mathrm{big}} \mid u_{\leqslant i-1}, U_i = u) = \frac{\mathbb{P}(U \notin \mathcal{U}_{\mathrm{big}}, U_i = u \mid u_{\leqslant i-1})}{\mathbb{P}(U_i = u \mid u_{\leqslant i-1})} \leqslant n\mathbb{P}\left(U \notin \mathcal{U}_{\mathrm{small}}\right).$$

Recalling $\mathbb{P}\left(U \notin \mathcal{U}_{\mathrm{small}}\right) = e^{-\omega(\log n)}$, we complete the proof.

Now we are ready to prove Theorem 1.1, our main result. Let Y and $Z(T)$ be the martingales from (14). Consider the sequence $W = (W_0, \ldots, W_{2n-2})$ defined by

$$
W_i := \begin{cases} Y_i, & \text{if } i = 0, \ldots, n-1, \\ Z_{i-n+1}(T(U)), & \text{if } i \geqslant n \text{ and } T(U) \in \mathcal{T}_n^{\log n}, \\ Y_{n-1}, & \text{if } i \geqslant n \text{ and } T(U) \notin \mathcal{T}_n^{\log n}. \end{cases}
$$

Note that W is a martingale with the respect to the filtration $\mathcal{F}_0', \ldots, \mathcal{F}_{2n-2}'$, where $\mathcal{F}_i' = \mathcal{F}_i$, for $i \leqslant n-1$ and $\mathcal{F}_i' = \mathcal{F}_{n-1} \times \mathcal{G}_{i-n+1}$, for $i \geqslant n$. Using (1), (8), and Lemma 5.2(e), we get that

$$
\mathrm{Var}\,[W_{2n-2}] = \mathrm{Var}\,[Y_{n-1}] + \mathbb{E}\left[V(U)\mathbb{1}_{T(U) \in \mathcal{T}_n^{\log n}}\right]
$$

$$
= \mathrm{Var}\,[Y_{n-1}] + \mathbb{E}\,[V(U)] - 4\alpha^2 n^2 e^{-\omega(\log n)} = \mathrm{Var}\,[F(T)] - \alpha^2 e^{-\omega(\log n)}.
$$

Then, by assumptions of Theorem 1.1, we get $\mathrm{Var}\,[W_{2n-2}] = \left(1 + e^{-\omega(\log n)}\right) \mathrm{Var}\,[F(T)]$ and

$$
\alpha^2 = O\left(n^{-2/3 - 2\varepsilon/3}\right) \mathrm{Var}\,[W_{2n-2}], \qquad \alpha^2 \rho^2 = O\left(n^{-1/2 - 2\varepsilon}\right) \mathrm{Var}\,[W_{2n-2}]. \tag{20}
$$

Using Lemma 5.2(a,c), we obtain that, for all $i \in [2n-2]$,

$$
W_i - W_{i-1} = O(\alpha \log n). \tag{21}
$$

Let $u \in [n]^{n-1} \in U_{\text{small}}$. Combining Lemmas 5.2(b,d,e) and 5.3 and observing $\beta(T) \leqslant n^2$ for all $T \in \mathcal{T}_n$, we get that, for all $i \in [n-1]$

$$
\mathrm{ran}\left[\mathrm{Var}_{\mathcal{F}_i}[Y_{n-1}] \mid u_{\leqslant i-1}\right] = O\left(\alpha^2 \rho^2 \log^4 n\right)
$$

$$
\mathrm{ran}_{\mathcal{G}_{i-1}}[Z_i(T(u))] = O\left(\alpha^2 \rho^2 \log^7 n\right)
$$

$$
\mathrm{ran}\left[\mathbb{E}_{\mathcal{F}_i}[V(U)] \mid u_{\leqslant i-1}\right] = O\left(\alpha^2 \rho^2 \log^6 n\right)
$$

Note that, in the case of the event $U \in \mathcal{U}_{\text{small}}$, we have $W_i = Z_i(T(U))$ and

$$
\mathrm{Var}[W_{2n-2} \mid \mathcal{F}_i] = \mathrm{Var}_{\mathcal{F}_i}[Y_{n-1}] + \mathbb{E}_{\mathcal{F}_i}[V(U)].
$$

Then, we obtain that if $U \in \mathcal{U}_{\text{small}}$ then, for all $i \in [2n-2]$,

$$
\mathrm{ran}\left[\mathrm{Var}[W_{2n-2} \mid \mathcal{F}_i'] \mid \mathcal{F}_{i-1}'\right] = O(\alpha^2 \rho^2 \log^7 n).
$$

Using (20), we conclude that, with probability $1 - e^{-\omega(\log n)}$,

$$
\sum_{i=1}^{2n-2} \left(\mathrm{ran}[\mathrm{Var}[W_{2n-2} \mid \mathcal{F}_i'] \mid \mathcal{F}_{i-1}'] + \left(\mathrm{ran}\left[W_i \mid \mathcal{F}_{i-1}'\right]\right)^2\right)^2
$$

$$
= O(\alpha^4 \rho^4 n \log^{14} n) = O(n^{-4\varepsilon} \log^{14} n)\left(\mathrm{Var}[W_{2n-2}]\right)^2.
$$

Let $\tilde{\varepsilon} \in (0, \varepsilon)$. Setting $\hat{q} = n^{-2\tilde{\varepsilon}}$ and applying Lemma 4.3, we get that, for any $p \in [1, +\infty)$,

$$
\mathbb{E}\left[|Q[W] - 1|^p\right] = O\left(n^{-2p\tilde{\varepsilon}} + \sup |Q[W] - 1|^p e^{-\omega(\log n)}\right).
$$

Using (21) and (20), we can bound

$$
Q[W] = \frac{1}{\mathrm{Var}[W_{2n-2}]} \sum_{i=1}^{2n-2} (W_i - W_{i-1})^2 = O\left(n^{1/3}\right).
$$

Applying Theorem 4.2 to the scaled martingale sequence $W/(\alpha \log n)$, we get that

$$\delta_K[W_{2n-2}] = O\left(\left(\frac{\alpha^2 \log^2 n}{\mathrm{Var}[W_{2n-2}]}\right)^{3/2} n \log n + \left(n^{-2\tilde{p}\varepsilon} + e^{-\omega(\log n)} n^{p/3}\right)^{1/(2p+1)}\right)$$

$$= O\left(n^{-\varepsilon} \log^4 n + n^{-2p\tilde{\varepsilon}/(2p+1)}\right) = O(n^{-2p\tilde{\varepsilon}/(2p+1)}).$$

We can make $2p\tilde{\varepsilon}/(2p+1) \geqslant \varepsilon'$ for any $\varepsilon' \in (0, \varepsilon)$ by taking $\tilde{\varepsilon}$ to be sufficiently close to ε and p to be sufficiently large. Recalling that $W_{2n-2} = F(T(U)^X)$ with probability $1 - e^{\omega(\log n)}$ (that is for the event $T(U) \in \mathcal{T}_n^{\log n}$) and $\mathrm{Var}[W_{2n-2}] = \left(1 + e^{-\omega(\log n)}\right) \mathrm{Var}[F(T)]$, the required bound for $\delta_K[F(T)]$ follows.

Remark 5.4. The proof of Theorem 1.1 can be significantly simplified under additional assumption that the tree parameter F is symmetric. Namely, we would not need the martingale sequence $Z(T)$, the bounds of Section 5.1, and we would only use parts (a), (b) from Lemma 5.2. In fact, a symmetric version of Theorem 1.1 would be sufficient to cover all applications given in Sections 2 and 3. Our decision to consider arbitrary tree parameters serves two purposes. First, the result is significantly stronger. Second, the analysis of martingales based on functions with dependent random variables is essential for extensions to more sophisticated tree models.

Remark 5.5. Combining Lemma 5.2(a,c) and Theorem 4.1 one can easily derive fast decreasing bounds for the tail of the distribution of $F(T)$, provided a tree parameter F is α-Lipschitz. Cooper, McGrae and Zito [6, Section 4] used a different martingale construction for trees to establish the concentration of $F(T)$ around its expectation; however, they needed more restrictive assumptions about the tree parameter F.

6. The balls in random trees are not too large

In this section, we prove Theorem 1.4 using martingales. For a tree $T \in \mathcal{T}_n$, let $\Gamma_T^k(v)$ be the set of all vertices at distance exactly k from v. Theorem 1.4 follows immediately from Lemmas 6.2 and 6.4 (stated below) by summing over all $|\Gamma_T^k(v)|$ for $k = 1, \ldots, d$ and using the union bound over all vertices $v \in [n]$.

Let $a > b$ be positive integers. Let A be an arbitrary set of a vertices from $[n]$, and B be its subset on b vertices. Consider event $\mathcal{E}_{A,B}$ that A induces a tree and vertices of $A \backslash B$ have neighbours only in A. For $T \in \mathcal{T}_n$, let $\xi_{A,B}(T)$ be the number of neighbours of B in T outside A. Below, we denote the random variable $\xi_{A,B}(T)$ simply $\xi_{A,B}$.

Lemma 6.1. *The conditional distribution of $\xi_{A,B} - 1$ subject to $\mathcal{E}_{A,B}$ is binomial with parameters $(n - a - 1, \frac{b}{n-a+b})$.*

Proof. Let T_0 be a tree on A. Consider event $\mathcal{E}_{A,B,T_0}$ that A induces exactly the given subtree T_0 and vertices of $A \backslash B$ have neighbours only in A. By Lemma 2.3,

$$\left|\mathcal{E}_{A,B,T_0}\right| = b(n - a + b)^{n-a-1}.$$

Let $k \in \mathbb{N}$. By Lemma 2.3,

$$\left|\{\xi_{A,B} = k\} \cap \mathcal{E}_{A,B,T_0}\right| = b^k \binom{n-a-1}{k-1} (n-a)^{n-a-k}.$$

Therefore,

$$\mathbb{P}(\xi_{A,B} = k \mid \mathcal{E}_{A,B}) = \frac{\sum_T \mathbb{P}(\xi_{A,B} = k \mid \mathcal{E}_{A,B,T}) \mathbb{P}(\mathcal{E}_{A,B,T})}{\sum_T \mathbb{P}(\mathcal{E}_{A,B,T})}$$

$$= \mathbb{P}(\xi_{A,B} = k \mid \mathcal{E}_{A,B,T_0})$$

$$= \left(\frac{b}{n-a+b} \right)^{k-1} \left(1 - \frac{b}{n-a+b} \right)^{n-a-k} \binom{n-a-1}{k-1},$$

which is the required distribution. □

Fix a vertex $v \in [n]$. Define the sequence of random variables $X_0, \ldots, X_n$ by

$$X_0 := 1, \qquad \text{and} \qquad X_k := |\Gamma_T^k(v)| \text{ for all } k \in [n].$$

From Lemma 6.1, we have $X_1 - 1 \sim \mathrm{Bin}(n-2, \frac{1}{n})$. Notice that, for $k \geq 1$, the vertices of $\Gamma_T^{k+1}(v)$ are adjacent only to the vertices of $\Gamma_T^k(v)$ in $\bigsqcup_{j \leq k+1} \Gamma_T^j(v)$. Let $(x_1, \ldots, x_k)$ be a sequence of positive integers such that $1 + x_1 + \ldots + x_k \leq n$. By Lemma 6.1, if $x_1 + \ldots + x_k \leq n-3$, then the conditional distribution of $X_{k+1} - 1$ subject to $(X_1 = x_1, \ldots, X_k = x_k)$ is binomial with parameters $n - x_1 - \ldots - x_k - 2$ and $\frac{x_k}{n - x_1 - \ldots - x_{k-1} - 1}$. If $x_1 + \ldots + x_k = n-2$, then $X_{k+1} = 1$. Finally, if $x_1 + \ldots + x_k = n-1$, then $X_{k+1} = 0$.

Lemma 6.2. *There exists a sequence $X_0 = X_0', X_1', \ldots, X_n'$ such that*

- *$X_k' \geq X_k$,*
- *for $k \geq 0$, the distribution of $X_{k+1}' - 1$ subject to $X_j = x_j, X_j' = x_j', j \in [k]$, is*

$$\begin{cases} \mathrm{Bin}\left(n - \sum_{j=0}^{k-1} x_j, \frac{x_k'}{n - \sum_{j=0}^{k-1} x_j} \right), & \text{if } n - \sum_{j=0}^{k-1} x_j \geq x_k', \\ x_k' \text{ with probability } 1, & \text{otherwise.} \end{cases}$$

Proof. It is straightforward since, for every k, we preserve the denominator of the second parameter of the binomial distribution but make the first one larger. □

Note that $(X_k' - k)_{k \in [n]}$ is a martingale sequence. Unfortunately, we can not apply Theorem 4.1 directly because every X_k' ranges in a large interval (mostly for small k). Instead, we cut the tails of these random variables and construct a new martingale. To do that we need the following property of binomial distributions.

Lemma 6.3. *Let N and $a \leq N$ be positive integers, $\xi \sim \mathrm{Bin}(N, \frac{a}{N})$. Then, for every $b \in \mathbb{N}$, there exists an interval $\mathcal{I} = \mathcal{I}(N, a, b) \subset [a - b, a + b]$ such that*

- *$\mathbb{P}(\xi \notin \mathcal{I}) \leq N^2 \mathbb{P}(\xi \notin [a - b, a + b])$,*
- *$\exists c \in [a - b, a + b]$ such that the function $f : \mathbb{R} \to \mathbb{R}$ defined by*

$$f(x) := \begin{cases} x, & x \in \mathcal{I} \\ c, & x \notin \mathcal{I} \end{cases}$$

satisfies $\mathbb{E}\left[f(\xi) \right] = a$.

Proof. For $a = N/2$, we get the result by setting $\mathcal{I} = [a - b, a + b]$ and $c = a$. For the following, without loss of the generality, we may assume $a < N/2$ since the proof for $a > N/2$ is symmetric.

Let us consider the set $\mathcal{S}$ of all integers s such that

$$\mathbb{E}[\xi \mathbb{1}_{\{\xi \in [a-s, a+b]\}}] \geqslant a\mathbb{P}(\xi \in [a-s, a+b]). \tag{22}$$

It is clear that $0 \in \mathcal{S}$. However, for every $x \in \mathbb{N}$, $\mathbb{P}(\xi = a - x) > \mathbb{P}(\xi = a + x)$. Indeed,

$$\frac{\mathbb{P}(\xi = a - x)}{\mathbb{P}(\xi = a + x)} = \frac{\left(1 + \frac{x}{a}\right)\left(1 + \frac{x-1}{a}\right) \cdots \left(1 - \frac{x-1}{a}\right)}{\left(1 + \frac{x}{N-a}\right)\left(1 + \frac{x-1}{N-a}\right) \cdots \left(1 - \frac{x-1}{N-a}\right)} > 1.$$

Therefore, $b \notin \mathcal{S}$. Let s^* be the maximum integer from $\mathcal{S}$. Then, $s^* \in [1, b-1]$ and

$$\mathbb{E}[\xi \mathbb{1}_{\{\xi \in [a-s^*-1, a+b]\}}] < a\mathbb{P}(\xi \in [a-s^*-1, a+b]). \tag{23}$$

Let us prove that $\mathcal{I} = [a - s^*, a + b]$ is the desired interval. From (23), we get

$$\begin{aligned}
&\mathbb{E}[(a - s^* - 1)\mathbb{1}_{\{\xi \notin \mathcal{I}\}} + \xi \mathbb{1}_{\{\xi \in \mathcal{I}\}}] \\
&= \mathbb{E}[(a - s^* - 1)]\mathbb{1}_{\{\xi \notin [a-s^*-1, a+b]\}} + \xi \mathbb{1}_{\{\xi \in [a-s^*-1, a+b]\}}] \\
&< (a - s^* - 1)\mathbb{P}(\xi \notin [a-s^*-1, a+b]) + a\mathbb{P}(\xi \in [a-s^*-1, a+b]) < a.
\end{aligned}$$

Moreover, since (22) holds for $s = s^*$,

$$\mathbb{E}[a\mathbb{1}_{\{\xi \notin \mathcal{I}\}} + \xi \mathbb{1}_{\{\xi \in \mathcal{I}\}}] \geqslant a\mathbb{P}(\xi \notin \mathcal{I}) + a\mathbb{P}(\xi \in \mathcal{I}) = a.$$

Therefore, there exists $c \in (a - s^* - 1, a]$ such that $\mathbb{E}[cI(\xi \notin \mathcal{I}) + \xi I(\xi \in \mathcal{I})] = a$.

It remains to estimate $\mathbb{P}(\xi \notin \mathcal{I})$ from above. Notice that, from (23),

$$a\mathbb{P}(\xi \in [a-s^*-1, a+b]) + (a - s^*)\mathbb{P}(\xi < a - s^* - 1) + N\mathbb{P}(\xi > a + b) > a.$$

Therefore, $s^*\mathbb{P}(\xi < a - s^* - 1) < N\mathbb{P}(\xi > a + b)$. Since $2a\mathbb{P}(\xi = a - s^* - 2) > \mathbb{P}(\xi = a - s^* - 1)$, we get

$$\mathbb{P}(\xi < a - s^*) < (2a + 1)\mathbb{P}(\xi < a - s^* - 1) \leqslant N^2\mathbb{P}(\xi > a + b),$$

and this immediately implies that $\mathbb{P}(\xi \notin \mathcal{I}) \leqslant N^2\mathbb{P}(\xi \notin [a - b, a + b])$. $\qquad \square$

Now, we are ready to construct a martingale sequence that coincides with $X'_k - k$ with probability very close to 1, but is more suitable for applying Theorem 4.1. For every $k \geqslant 2$, consider the event

$$\mathcal{B}_k := \left\{ n - \sum_{j=0}^{k-2} X_j \geqslant X'_{k-1} \right\}.$$

For $\omega \in \mathcal{B}_k$, denote

$$\mathcal{I}_k := \mathcal{I}\left(n - \sum_{j=0}^{k-2} X_j, X'_{k-1}, \sqrt{X'_{k-1} \log n} \right),$$

$$f_k := f\left(n - \sum_{j=0}^{k-2} X_j, X'_{k-1}, \sqrt{X'_{k-1} \log n} \right).$$

Let

$$\mathcal{E}_k := \mathcal{B}_k \cap \left(\bigcap_{j=1}^{k} \{X'_j - 1 \in \mathcal{I}_j\} \right).$$

Define the sequence $(Y_k)_{k \in [n]}$ as follows. Let $Y_0 := X'_0 = 1$. For $k \geqslant 1$, set

$$Y_k := [f_k(X'_k - 1) - (k - 1)]\mathbb{1}_{\mathcal{E}_k} + Y_{k-1}\mathbb{1}_{\overline{\mathcal{E}_k}}.$$

Using Lemmas 6.2 and 6.3, we find that $(Y_0, Y_1, \ldots, Y_n)$ is a martingale sequence with respect to the filtration $\mathcal{F}_i = \sigma(X_j, X'_j : 0 \leqslant j \leqslant i)$ for all $i \in \{0, 1, \ldots, n\}$.

Lemma 6.4. *Let $c > 0$ be a fixed constant. Then, the following bounds hold:*

 a. $\mathbb{P}\left(\exists k \in [n]: Y_k > k \log^4 n\right) \leqslant e^{-\omega(\log n)}$,
 b. $\mathbb{P}\left(\exists k \in [n]: Y_k \neq X'_k - k\right) \leqslant e^{-\omega(\log n)}$.

Proof. For (a), we apply Theorem 4.1. First, we estimate the conditional ranges. From Lemma 6.3, we get that, for all $k \in [n]$

$$\mathrm{ran}_k[Y_{k+1}] \leqslant 2\sqrt{X'_k} \log n \, \mathbb{1}_{\mathcal{E}_k} = 2\sqrt{Y_k + k} \log n \, \mathbb{1}_{\mathcal{E}_k}.$$

We prove by induction on k that $\mathsf{P}(Y_k > k \log^4 n) \leqslant \exp[-c \log^2 n]$, where $c > 0$ does not depend on k and n. For $k = 1$, we have $\mathbb{P}(Y_1 > \log^4 n) \leqslant \mathbb{P}(Y_1 > \log n) = 0$.

Assume that $\mathbb{P}(Y_j > j \log^4 n - j) \leqslant \exp[-\log^2 n(1 + o(1))]$ for all $j \leqslant k$. Then, with a probability at least $1 - n \exp[-\log^2 n(1 + o(1))] = 1 - \exp[-\log^2 n(1 + o(1))]$,

$$\sum_{j=1}^{k+1} (\mathrm{ran}_{j-1}[Y_j])^2 \leqslant 4 \log^2 n \sum_{j=0}^{k} (Y_j + j) \leqslant 2k^2 \log^6 n.$$

Therefore, by Theorem 4.1,

$$\mathbb{P}\left(Y_{k+1} > (k+1) \log^4 n - (k+1)\right)$$

$$\leqslant 2 \exp\left[-\frac{(k+1)^2}{k^2} \log^2 n(1 + o(1))\right] + 2 \exp\left[-\log^2 n(1 + o(1))\right]$$

$$= \exp\left[-\log^2 n(1 + o(1))\right].$$

This proves (a).

For (b), observe that, by the definition of Y_k,

$$\mathbb{P}(\exists k \quad Y_k \neq X'_k) = \mathbb{P}\left(\bigcup_k \mathcal{B}_k \backslash \mathcal{E}_k\right) \leqslant \sum_{k=1}^{n} \mathbb{P}\left(X'_k - 1 \notin \mathcal{I}_k \mid \mathcal{B}_k\right).$$

Each term in the sum above is $e^{-\omega(\log n)}$ by Lemma 6.5 and the difinition of X'_k given in Lemma 6.2. Part (b) follows.

Lemma 6.5. *For n large enough and all positive integers $a \leqslant N$, a random variable $\xi \sim \mathrm{Bin}(N, a/N)$ satisfies the following:*

$$\mathbb{P}(|\xi - a| > \sqrt{a} \log n) \leqslant \exp\left(-\frac{1}{5} \log n \log \log n\right).$$

Proof. By the Chernoff bounds,

$$\mathbb{P}(\xi \geqslant a + \sqrt{a} \log n) \leqslant \exp\left[\sqrt{a} \log n - (a + \sqrt{a} \log n) \ln\left(1 + \frac{\log n}{\sqrt{a}}\right)\right],$$

$$\mathbb{P}(\xi \leqslant a - \sqrt{a} \log n) \leqslant \exp\left[-\frac{1}{2} \log^2 n\right]$$

It is straightforward to check that the stated bound holds for all possible values of a. $\qquad\square$

References

[1] Aldous, D. J. (1990) The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.* **3**(4) 450–465.

[2] Bender, E. A. and Richmond, L. B. (1983) Central and local limit theorems applied to asymptotic enumeration II: multivariate generating functions. *J. Combin. Theory Ser. A* **34** 255–265.

[3] Bóna, M. and Flajolet, P. (2009) Isomorphism and symmetries in random phylogenetic trees. *Appl. Prob. Trust* **46** 1005–1019.

[4] Brown, B. M. (1971) Martingale central limit theorems. *Ann. Math. Stat.* **42**(1) 59–66.

[5] Chysak, F., Drmota, M., Klausner, T. and Kok, G. (2008) The distribution of patterns in random trees. *Comb. Prob. Comput.* **17**(1) 21–59.

[6] Cooper, C., McGrae, A. R. A. and Zito, M. (2009) Martingales on trees and the empire chromatic number of random trees. In *Fundamentals of Computation Theory 2009* (M. Kutyłowski, M. Gębala and W. Charatonik, eds), Vol. 5699 of *LNCS*, Springer, Heidelberg, pp. 74–83.

[7] Drmota, M. (2009) *Random Trees*. Springer, Vienna.

[8] Drmota, M. and Gittenberger, B. (1999) The distribution of nodes of given degree in random trees. *J. Graph Theory* **31** 227–253.

[9] Feng, Q. and Hu, Z. (2011) On the Zagreb index of random recursive trees. *J. Appl. Probab.* **48** 1189–1196.

[10] Greenhill, C., Isaev, M., Kwan, M. and McKay, B. D. (2017) The average number of spanning trees in sparse graphs with given degrees. *Eur. J. Comb.* **63** 6–25.

[11] Greenhill, C., Isaev, M. and McKay, B. D. (2018) Subgraph counts for dense random graphs with specified degrees. Preprint, submitted in 2018. arxiv:1801.09813 [math.CO]

[12] Holst, L. (1981) Some conditional limit theorems in exponential families. *Ann. Prob.* **9**(5) 818–830.

[13] Hwang, H.-K. (1998) On convergence rates in the central limit theorems for combinatorial structures. *Eur. J. Comb.* **19** 329–343.

[14] Isaev, M. and McKay, B. D. (2018) Complex martingales and asymptotic enumeration. *Random Struct. Algorithms* **52**(4) 617–661.

[15] Isaev, M., Southwell, A. and Zhukovskii, M. Distribution of tree parameters by martingale approach, e-preprint arXiv:1912.09838.

[16] Janson, S. (2016) Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. *Random Struct. Algorithms* **48**(1) 57–101.

[17] Kolchin, V. F. (1977) Branching processes, random trees, and a generalized scheme of arrangements of particles. *Mat. Zametki* **21**(5) 691–705.

[18] Mahmoud, H. (2004) Pólya urn models and connections to random trees: a review. *J. Iranian Stat. Soc.* **2** 53–114.

[19] McDiarmid, C. (1998) Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*, *Algorithms and Combinatorics* **16** 195–248.

[20] McKeon, K. A. (1996) The expected number of symmetries in locally restricted trees II. *Discrete Appl. Math.* **66**(3) 245–253,

[21] Moon, J. W. (1970) Counting labelled trees. In *Canadian Mathematical Monographs*, Vol. **1**, Canadian Mathematical Congress, Montreal.

[22] Móri, T. F. (2005) The maximum degree of the Barabási–Albert random tree. *Comb. Prob. Comput.* **14**(3) 339–348.

[23] Mourrat, J.-C. (2013) On the rate of convergence in the martingale central limit theorem. *Bernoulli* **19**(2) 633–645.

[24] Stacey, K. C. and Holton, D. A. (1977) On the automorphism group of a tree. *J. Comb. Theory Ser. B* **22**(1) 10–25.

[25] Smythe, R. T. (1996) Central limit theorems for urn models. *Stochastic Process. Appl.* **65**(1) 115–137.

[26] Sulzbach, H. (2017) On martingale tail sums for the path length in random trees. *Random Struct. Algorithms* **50**(3) 493–508.

[27] Wagner, S. (2015) Central Limit theorems for additive tree parameters with small toll functions. *Comb. Prob. Comput.* **24** 329–353.

[28] Wagner, S. and Stufler, B. (2016) The number of automorphisms of random trees, a talk at BIRS Workshop, October 2016.

[29] Wagner, S. (2019) personal communication.

[30] Yu, L. (2012) Automorphisms of random trees, PhD Thesis, *Drexel University*.