THE TOP EIGENVALUE OF UNIFORMLY RANDOM TREES

LOUIGI ADDARIO-BERRY, GÁBOR LUGOSI, AND ROBERTO IMBUZEIRO OLIVEIRA

ABSTRACT. Let $\mathbf{T}_n$ be a uniformly random tree with vertex set $[n] = \{1, \ldots, n\}$, let $\Delta_{\mathbf{T}_n}$ be the largest vertex degree in $\mathbf{T}_n$, and let $\lambda_1(\mathbf{T}_n), \ldots, \lambda_n(\mathbf{T}_n)$ be the eigenvalues of its adjacency matrix, arranged in decreasing order. We prove that $|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| \to 0$ in expectation as $n \to \infty$, and additionally prove probability tail bounds for $|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}|$. Writing a_n for any median of $\Delta_{\mathbf{T}_n}$, we also prove that $|\lambda_k(\mathbf{T}_n) - \sqrt{a_n}| \to 0$ in expectation, uniformly over $1 \le k \le e^{\log^\beta(n)}$, for any fixed $\beta \in (0, 1/2)$.

The proof is based on the trace method and thus on counting closed walks in a random tree. To this end, we develop novel combinatorial tools for encoding walks in trees that we expect will find other applications. In order to apply these tools, we show that uniformly random trees — after appropriate "surgery" — satisfy, with high probability, the properties required for the combinatorial bounds to be effective.

1. Introduction

In this work, we derive precise information about the largest eigenvalue of the adjacency matrix of large uniformly random trees. For a finite graph $G = (V, E)$, the adjacency matrix of G is the $V \times V$ symmetric matrix $A = A(G)$ whose (u, v) entry is 1 if $\{u, v\} \in E$ and is 0 otherwise. Write Δ_G for the largest degree of any vertex in G, and write $\lambda_1(G) \ge \lambda_2(G) \ge \cdots \ge \lambda_n(G)$ for the ordered sequence of eigenvalues of the adjacency matrix of G, where $n = |V|$. Also, for positive integers n write $[n] := \{1, \ldots, n\}$.

Theorem 1.1. *For $n \ge 1$, let $\mathbf{T}_n$ be a uniformly random tree with vertex set $[n]$. Then*

$$\mathbf{E}|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| \to 0$$

as $n \to \infty$. Moreover, there exists a universal constant $L > 0$ such that, as $n \to \infty$,

$$\mathbf{P}\left\{|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| > L \left(\frac{\log^2 \Delta_{\mathbf{T}_n}}{\sqrt{\Delta_{\mathbf{T}_n}}}\right)^{1/15}\right\} \le n^{-0.8+o(1)}.$$

It is natural to ask whether the terms $\Delta_{\mathbf{T}_n}$ in the above theorem can be replaced by deterministic values. For the error bars on $|\lambda_1 - \sqrt{\Delta_{\mathbf{T}_n}}|$, they can. It follows from results of Moon [19, Theorem 2 and Lemma 4] that $\Delta_{\mathbf{T}_n} \cdot \left(\frac{\log n}{\log \log n}\right)^{-1} \to 1$ in probability and that $\mathbf{P}\left\{\Delta_{\mathbf{T}_n} < 0.99 \frac{\log n}{\log \log n}\right\} < -\exp(-(1 + o(1))n^{0.01})$, and so the probability bound in the above theorem also implies that there is a universal constant $L > 0$ such that

$$\mathbf{P}\left\{|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| > L \left(\frac{(\log \log n)^{2.5}}{\sqrt{\log n}}\right)^{1/15}\right\} \le n^{-0.8+o(1)}.$$

On the other hand, the upper tail of $\Delta_{\mathbf{T}_n}$ is heavy enough that the probability bound in the theorem would not be true if $|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}|$ were replaced by, e.g., $\left|\lambda_1(\mathbf{T}_n) - \sqrt{\frac{\log n}{\log \log n}}\right|$, or indeed by any other fixed sequence. However, the expectation bound *does* hold with $\sqrt{\Delta_{\mathbf{T}_n}}$

replaced by its median; moreover, we are able to prove the same expectation bound not just for $\lambda_1(\mathbf{T}_n)$ but for $\lambda_k(\mathbf{T}_n)$, for k quite large.

Theorem 1.2. *For $n \geq 1$, let $\mathbf{T}_n$ be a uniformly random tree with vertex set $[n]$. Let a_n be any median of $\Delta_{\mathbf{T}_n}$. Fix any constant $\beta \in (0, 1/2)$ and let $k(n) = \lceil \exp((\log n)^\beta) \rceil$. Then*

$$\mathbf{E}|\lambda_1(\mathbf{T}_n) - \sqrt{a_n}| \to 0$$

and

$$\mathbf{E}|\lambda_{k(n)}(\mathbf{T}_n) - \sqrt{a_n}| \to 0.$$

Together with Theorem 1.1, it follows that also

$$\mathbf{E}|\lambda_{k(n)}(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| \to 0,$$

by the triangle inequality. Moreover, since $\lambda_1 \geq \lambda_i \geq \lambda_{k(n)}$ for all $i \in [k(n)]$, the same estimates hold for $\mathbf{E}|\lambda_{i(n)}(\mathbf{T}_n) - \sqrt{a_n}|$ and for $\mathbf{E}|\lambda_{i(n)}(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}|$, for any sequence $(i(n), n \geq 1)$ with $1 \leq i(n) \leq k(n)$. We conjecture that for any $\beta \in (1/2, 1]$ there is $c = c(\beta) > 0$ such that, with $\ell(n) = \lceil \exp((\log n)^\beta) \rceil$, then $\mathbf{P}\{\lambda_{\ell(n)}(\mathbf{T}_n) \leq \sqrt{a_n} - c\} \to 1$; this would imply that the range of $k(n)$ covered by Theorem 1.2 is essentially optimal.

We expect that the methodology we develop here can also be used for controlling the top eigenvalues of other random tree models. Informally, in order to be applicable, it requires maximum-degree vertices to be relatively few in number and to be well-separated from one-another in the tree. (We provide a fairly detailed overview of our proof technique in Section 1.2.) However, for the moment we do not see a way to formulate a more precise set of requirements which would allow us to treat many models simultaneously; thus far, model-specific conditions and calculations seem unavoidable to us.

1.1. **Related work.** It is known in some generality that the bulk spectral properties of sparse random graphs are determined locally. In particular, suppose that $(G_n, n \geq 1)$ is a sequence of finite random rooted graphs, where G_n has n vertices, and write

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(G_n)}$$

for the empirical spectral measure of G_n. If the sequence $(G_n, n \geq 1)$ converges in distribution in the local weak sense to a locally finite random rooted graph G, then μ_n converges weakly to a limiting measure $\mu = \mu_G$; see [9], where this is proved under a uniform integrability condition on the root degree, and [1, 8], which handle the general case. More precise results about the bulk of the spectrum are available when the local weak limit is a (deterministic or random) tree [6, 7, 12, 23].

For the edge of the spectrum, fewer general results are available, and indeed one expects more model-dependence. In the setting of the Erdős-Rényi random graph $G(n, p)$, Krivelevich and Frieze showed that for any sequence $(p_n, n \geq 1)$ of probabilities, $\lambda_1(G(n, p_n)) = (1 + o(1)) \max(np_n, \sqrt{\Delta_{G(n, p_n)}})$ in probability as $n \to \infty$, where $\Delta_{G(n, p_n)}$ denotes the maximum degree of the graph. The threshold $np_n \approx \sqrt{\Delta_{G(n, p_n)}}$ between the regime where np_n is larger and that where $\sqrt{\Delta_{G(n, p_n)}})$ is larger occurs around $p_n^* = (\log n)/\log(4/e)$. Alt, Ducatez and Knowles [3–5] have shown that p_n^* is the boundary between the localized and delocalized phases for the spectral edge of sparse Erdős-Rényi random graphs, in the sense that for $\log \log n \ll p_n < (1 - o(1))p_n^*$ the eigenvectors corresponding to eigenvalues near the edge of the spectrum are concentrated on a small number of vertices, whereas for $p_n > (1 + o(1))p_n^*$ this is not the case. (We are being slightly informal here, and refer the reader to the papers cited above for more precise statements.) To see how the value $\sqrt{\Delta_{G(n, p_n)}}$ arises, note that for

any finite graph $G = (V, E)$, if v is a vertex of G with degree Δ_G, then the vector $x : V \to \mathbb{R}$ given by

$$x_w = \begin{cases} \sqrt{\Delta_G} & \text{if } w = v \\ 1 & \text{if } \{v, w\} \in E \\ 0 & \text{otherwise} \end{cases}$$

satisfies that $\langle x, Ax \rangle \geq \sqrt{\Delta_G}|x|^2$, so by Rayleigh's formula x is a witness that $\lambda_1(G) \geq \sqrt{\Delta_G}$. Moreover, x is very localized, taking nonzero values only on a single vertex and its neighbourhood; so this suggests that if $\lambda_1(G) \approx \sqrt{\Delta_G}$ then perhaps the extremal eigenvectors are also localized.

Very recently, Hiesmayr and McKenzie [16] have built on the techniques of [5] to describe the extremal eigenvalues and eigenvectors of $G_{n,p}$ in the regime $\log^{-1/15} n \leq np \leq \log^{1/40} n$, showing that $|\lambda_1(G_{n,p}) - \sqrt{\Delta_{G_{n,p}}}| \to 0$ in probability, and, more strongly, that with high probability each of the top $e^{\log^{1/8} n}$ eivenvalues of $G_{n,p}$ is essentially determined by the graph structure of the metric ball of radius two around a vertex of near-maximal degree. Hiesmayr and McKenzie in fact prove quantitative bounds, and also establish localization of the corresponding eigenvectors. The case $np = \Theta(1)$ of their result in particular answers a question of Guionnet [15, Section 3.3].

The above results for the spectral edge of Erdős-Rényi random graphs have some similarity to ours; and indeed, the local structure of an Erdős-Rényi random graph is similar to that of the random trees whose study we undertake (both are locally described by a Poisson branching process). We expect that eigenvector localization near the spectral edge also occurs in many random tree models, including those considered in this paper, but our proof technique - which we now sketch - only gives access to the extremal eigenvalues, and not to the associated eigenvectors.

1.2. **Proof overview.** This section both presents our approach to proving Theorems 1.1 and 1.2 and introduces some concepts and several pieces of notation which are required in the sequel.

The bulk of the work is in proving the upper bound of the theorem. Our proof of the upper bound is based on the trace method, which we now recall. Fix a graph $G = (V, E)$ with $|V| = n$ and write $A = A(G)$. Then $\{\lambda_i(A^k), i \in [n]\} = \{\lambda_i(A)^k, i \in [n]\}$ for all $k \geq 1$, so

$$\begin{aligned} \lambda_1(G) &= \lim_{k \to \infty} \Big(\sum_{i \in [n]} \lambda_i(A)^{2k} \Big)^{1/(2k)} \\ &= \lim_{k \to \infty} \Big(\sum_{i \in [n]} \lambda_i(A^{2k}) \Big)^{1/(2k)} \\ &= \lim_{k \to \infty} \Big(\sum_{v \in V} (A^{2k})_{vv} \Big)^{1/(2k)} \\ &= \lim_{k \to \infty} \Big(\max_{v \in V} (A^{2k})_{vv} \Big)^{1/(2k)}, \end{aligned}$$

where for the third equality we have used the trace formula. Now, $(A^{2k})_{vv}$ is nothing but the number of closed walks of length $2k$ from v to v in G, by this we mean a sequence of vertices $(w_0, w_1, \ldots, w_{2k})$ with $w_0 = w_{2k} = v_i$ such that $\{w_{i-1}, w_i\} \in E$ for each $i \in [2k]$. The preceding displayed identity thus yields an equivalence between estimating $\lambda_1(G)$ and estimating the number of long closed walks in G. Writing $\mathcal{N}_k(v, G)$ for the set of closed walks of length $2k$ from v to v in G and $N_k(v, G) = |\mathcal{N}_k(v, G)|$ and $N_k(G) = \max_{v \in V} N_k(v, G)$, we

may rewrite the above identity for $\lambda_1(G)$ as

$$\lambda_1(G) = \lim_{k \to \infty} \left(\max_{v \in V} N_k(v, G) \right)^{1/(2k)} = \lim_{k \to \infty} N_k(G)^{1/(2k)} . \tag{1}$$

We now turn to the setting of trees. For a tree $T = (V, E)$ and a walk $w = (w_0, w_1, \ldots, w_j)$ in T, we write $D(w) = (d_0(w), \ldots, d_j(w)) := (\text{dist}_T(w_0, w_i), 0 \le i \le j)$ for the sequence of graph distances to w_0 in T, along the walk w. If w is a closed walk then $j = 2k$ is even and $D(w)$ is a *Dyck path* — a non-negative lattice path with $d_0(w) = d_{2k}(w) = 0$ and $|d_i(w) - d_{i-1}(w)| = 1$ for all $i \in [2k]$. We will usually be considering $D(w)$ for closed walks w, and have chosen the notation accordingly. We may then write

$$N_k(v, T) = \sum_d \{w \in \mathcal{N}_k(v, T) : D(w) = d\} ,$$

where the sum is over Dyck paths d of length $2k$.

An easy way to bound this expression from above is to note that, whatever d may be,

$$\{w \in \mathcal{N}_k(v, T) : D(w) = d\} \le \Delta_T^k .$$

This holds since in any Dyck path D of length $2k$ there are m increasing steps. An increasing step in $D(w)$ corresponds to a step of w which increases the distance to v in T; and, from any vertex u of T there are at most Δ_T choices for such a step. The decreasing steps yield no additional choices, since T is a tree; every vertex u of T aside from v has a unique neighbour u' with $\text{dist}_T(u', v) = \text{dist}_T(u, v) - 1$. Since the number of Dyck paths of length $2k$ is $\frac{1}{k+1}\binom{2k}{k}$, we deduce that

$$N_k(T) = \max_{v \in V} N_k(v, T) \le \frac{1}{k+1}\binom{2k}{k}\Delta^k \le (4\Delta)^k . \tag{2}$$

Using (1), this yields the well-known fact that $\lambda_1(T) \le \lim_{k \to \infty}(\max_v N_k(v, T))^{1/2k} \le 2\sqrt{\Delta_T}$. This bound is tight for general trees for large Δ_T; see, e.g., Godsil [14], Stevanović [24] who show that for any tree, the largest eigenvalue is at most $2\sqrt{\Delta_T - 1}$.[1]

We next illustrate the basic idea of how we improve this bound under additional conditions on the tree, but first introduce one required piece of notation: for a tree $T = (V, E)$ and $r \in \mathbb{N}$, let $\Delta_T^{(r)} = \max_{v \in V} |\{w \in V : \text{dist}_T(v, w) = r\}|$ be the largest size of an r'th neighbourhood in T.

Note that any Dyck path $d = d_0 d_1 \ldots d_{2k}$ either begins with the string 010 or with the string 012. (We shall often write $d_0 d_1 \ldots d_{2k}$ instead of $(d_0, \ldots, d_{2k})$ and 010 instead of $(0, 1, 0)$, *et cetera*, for succinctness.) It is straightforward to see that

$$|\{w \in \mathcal{N}_m(v, T) : d_0(w)d_1(w)d_2(w) = 010\}| \le \Delta_T N_{k-1}(v, T) \le \Delta_T N_{k-1}(T),$$

since a walk counted in the above set is simply the concatenation of a walk of length 2 and a walk of length $2(k - 1)$. A walk $w \in \mathcal{N}_k(v, T)$ with $d_0(w)d_1(w)d_2(w) = 012$ may also be decomposed: after w_2 must appear a closed walk from w_2 to w_2 which avoids w_1, followed by a closed walk from w_1 to w_1 which avoids w_0, and finally a closed walk from w_0 to w_0; each of these may be a walk of length zero. Since there are at most $\Delta_T^{(2)}$ choices for w_2, it follows that

$$|\{w \in \mathcal{N}_k(v, T) : d_0(w)d_1(w)d_2(w) = 012\}| \le \Delta_T^{(2)} \cdot \sum_{\substack{i,j,k \ge 0 \\ i+j+k=k-2}} N_i(T)N_j(T)N_k(T) .$$

---

[1] In fact, we may easily recover the bound $2\sqrt{\Delta_T - 1}$ from this argument. To do so, simply note that in (1) we can replace $\max_{v \in V} N_k(v, T)$ by $N_k(w, T)$ for *any* fixed vertex w. If we fix a vertex w of degree less than Δ_T, then the number of *increasing* steps from any vertex is at most $\Delta_T - 1$.

Combining these bounds, we obtain that

$$N_k(T) \le \Delta_T N_{k-1}(T) + \Delta_T^{(2)} \cdot \sum_{\substack{i,j,k \ge 0 \\ i+j+k=k-2}} N_i(T) N_j(T) N_k(T)$$

We may then apply this case analysis recursively to bound each of the terms $N_{k-1}(T)$, $N_i(T)$, $N_j(T)$ and $N_k(T)$, and hope that something useful comes out of it.

The set of words $\{010, 012\}$ is a *prefix code* for Dyck paths; any Dyck path has a unique prefix in this set. For example, another prefix code is $\{010, 0121, 0123\}$. We may generalize the above case analysis and recursive approach to any set of prefix codes. Our bounds end up depending crucially on two quantities. First, we require control on the greatest number of ways that each given code word can appear as a prefix of a Dyck path of a walk in the graph (for example, our bound on this for the word 012 was $\Delta_T^{(2)}$). Second, when controlling

$$|\{w \in \mathcal{N}_k(v,T) : D(w) = d\}|$$

for a given Dyck path d, our bounds depend on the number of times each code word appears when recursively decomposing d until no further decomposition is possible.

In order to explain more precisely, a little combinatorics is unavoidable. Say that a sequence $c = (c_i, 0 \le i \le j)$ is a *meander* if $c_0 = 0$, $c_i \ge 0$ for all $0 \le i \le j$, and $|c_i - c_{i-1}| = 1$ for all $i \in [j]$; it is an *excursion* if additionally $c_j = 0$. The *final value* of c is $f(c) := c_j$; its *length* is $\text{len}(c) = j$. A finite set C of meanders is a *prefix code* (or just *code*, for short) if for any Dyck path d of positive length there is a unique meander $c \in C$ such that c is a prefix of d. The unique code C with $01 \in C$ is $C = \{01\}$; we call this code *trivial*. Observe that if C is nontrivial then $010 \in C$.

For a prefix code $C = (c^{(1)}, \ldots, c^{(a)})$, to any Dyck path $d = (d_0, d_1, \ldots, d_{2k})$ we associate the vector

$$\vec{t}(C, d) = (t_b(C, d), 1 \le b \le a),$$

where $t_b(C, d)$ is the number of times the code word $c^{(b)}$ is used when decomposing d using the code C. Formally, let $p = p(d, c) \in [a]$ be the unique index for which $c^{(p)}$ is a prefix of d. Let $\eta_0, \ldots, \eta_{f(c^{(p)})+1}$ be the unique increasing sequence of integers with $\eta_0 = \text{len}(c^{(p)})$ and $\eta_{f(c^{(p)})+1} = 2k$ such that for all $0 \le j \le f(c^{(p)})$,

$$d^{(j)} := (d_{\eta_j+i} - (f(c^{(p)}) - j), 0 \le i \le \eta_{j+1} - \eta_j)$$

is a Dyck path. Then for $1 \le b \le a$ inductively define

$$t_b(C, d) = \mathbf{1}_{[b=p]} + \sum_{j=0}^{f(c_p)} t_b(C, d^{(j)}).$$

Note that the length of d is the sum of $\text{len}(c^{(p)})$, $f(c^{(p)})$, and the lengths of $d^{(0)}, \ldots, d^{(f(c^{(p)}))}$, so by induction we have

$$2k = \sum_{b \in [a]} t_b(C, d)(\text{len}(c^{(b)}) + f(c^{(b)})).$$

With a slight abuse of notation, given a closed walk w in a tree T we also define

$$\vec{t}(C, w) = \vec{t}(C, D(w)).$$

For a meander c of length j we let

$$\Delta_T(c) = \max_{v \in V} |\{\text{walks } w = (w_0, \ldots, w_j) \text{ in } T : w_0 = v, D(w) = c\}|.$$

For example, if $c = 012 \ldots h$ then $\Delta_T(c) = \Delta_T^{(h)}$; for another example, if $c = 01212$ then $\Delta_T(c) \le \Delta_T^{(2)} \cdot (\Delta_T - 1)$, since for any $v \in V$, a walk $w = w_0 w_1 w_2 w_3 w_4$ with $w_0 = v$ and with

$D(\mathrm{w}) = 01212$ is uniquely determined by choosing w_2 with $\mathrm{dist}_T(w_0, w_2) = 2$ and $w_4 \neq w_0$ with $\mathrm{dist}_T(w_1, w_4) = 1$.

The following proposition gives an idea of how decomposing walks by prefix codes allows us to bound the number of closed walks in a tree. Given a prefix code $C = (\mathrm{c}^{(1)}, \dots, \mathrm{c}^{(a)})$ and a vector $\vec{t} = (t_1, \dots, t_a)$ of non-negative integers, for a vertex v in a tree T we denote by $N_{C, \vec{t}}(v)$ the number of closed walks w from v to v in T with $t_b(C, \mathrm{w}) = t_b$ for all $1 \leq b \leq a$; the tree T in question will always be clear from context. Note that the length of any such walk is $2k = 2k(\vec{t}) := \sum_{b=1}^{a} t_b(\mathrm{len}(\mathrm{c}^{(b)}) + f(\mathrm{c}^{(b)}))$.

Proposition 1.3. *Fix a tree T, a code $C = \{\mathrm{c}^{(b)}, 1 \leq b \leq a\}$ and a vector of non-negative integers $\vec{t} = (t_1, \dots, t_a)$. Let $q = q(\vec{t}) = \sum_{b=1}^{a} t_b \cdot (1 + f(\mathrm{c}^{(b)}))$ and $k = k(\vec{t}) = \sum_{b=1}^{a} t_b(\mathrm{len}(\mathrm{c}^{(b)}) + f(\mathrm{c}^{(b)}))/2$. Then for any $\Delta > 0$, setting*

$$g_b = \left((1 \vee 2ef(\mathrm{c}^{(b)})) \right) \cdot \frac{\Delta_T(\mathrm{c}^{(b)})}{\Delta^{(\mathrm{len}(\mathrm{c}^{(b)}) + f(\mathrm{c}^{(b)}))/2}} \right)^{1/(1 + f(\mathrm{c}^{(b)}))},$$

for any vertex v of T we have

$$N_{C, \vec{t}}(v) \leq e q^{1/2} \Delta^k \left(\sum_{b=1}^{a} g_b \right)^{q}.$$

The proposition is proved by a careful analysis of the number of ways that a Dyck path can be partitioned by a given code, together with an application of a convexity bound (specifically, the non-negativity of the Kullback-Liebler divergence).

To get a feeling for how one can exploit such a bound, we consider what it yields for a couple of simple codes. First, consider the trivial code $C^* = \{01\}$. Then $\mathrm{len}(\mathrm{c}^{(1)}) = 1$, $f^{(1)} := f(\mathrm{c}^{(1)}) = 1$, and therefore if $\mathrm{len}(\mathrm{w}) = 2k$ then $q(\vec{t}) = 2k = 2t_1$. In this case the vector $\vec{t}$ has only one component t_1 and the only possible value is $\vec{t} = (k)$. Thus, the number $N_k(v, T)$ of closed walks from v to v in T of length $2k$ equals $N_{C, \vec{t}}(v)$ and $q(\vec{t}) = 2k(\vec{t})$. Taking $\Delta = \Delta_T$, then $g_1 = 2e$, and hence the bound of Proposition 1.3 becomes $N_k(v, T) \leq ek^{1/2}(4e^2 \Delta_T)^k$.

Next consider the code $C^* = \{\mathrm{c}^{(1)}, \mathrm{c}^{(2)}, \mathrm{c}^{(3)}\} = \{010, 0121, 0123\}$. For this code we have $f^{(1)} = 0, f^{(2)} = 1, f^{(3)} = 3$ and $\mathrm{len}(\mathrm{c}^{(1)}) = 2, \mathrm{len}(\mathrm{c}^{(2)}) = 3, \mathrm{len}(\mathrm{c}^{(3)}) = 3$. For any closed walk $\mathrm{w} = (w_0, \dots, w_{2k})$ in T, writing $t_i = t_i(C^*, \mathrm{w})$, we have

$$2k = 2t_1 + 4t_2 + 6t_3,$$

and $q = q(\vec{t}(C^*, \mathrm{w})) = t_1 + 2t_2 + 4t_3 = k + t_3$. For any tree T we have $\Delta_T(\mathrm{c}^{(b)}) = \Delta_T^{(b)}$ for $b = 1, 2, 3$, since in each case an embedding of a code word $\mathrm{c}^{(b)}$ is uniquely specified by the choice of its initial vertex and the unique vertex at distance b. We thus have $g_1 = 1$, $g_2 = (2e\Delta^{(2)}/\Delta_T^2)$, and $g_3 = (6e\Delta^{(3)}/\Delta_T^3)^{1/4}$; it follows by Proposition 1.3, again applied with $\Delta = \Delta_T$ that for any vertex v,

$$N_{C, \vec{t}}(v) \leq e q^{1/2} \Delta_T^k \left(1 + \sqrt{\frac{2e\Delta_T^{(2)}}{\Delta_T^2}} + \sqrt[4]{\frac{6e\Delta_T^{(3)}}{\Delta_T^3}} \right)^{q} \leq e(2k)^{1/2} \Delta_T^k \left(1 + \sqrt{\frac{2e\Delta_T^{(2)}}{\Delta_T^2}} + \sqrt[4]{\frac{6e\Delta_T^{(3)}}{\Delta_T^3}} \right)^{2k},$$

where we have used that $q = k + t_3 \leq 2k$. Thus, the number $N_k(v, T)$ of closed walks from v to v in T of length $2k$ may be bounded as

$$N_k(v, T) \leq e(2k)^{1/2} \Delta_T^k \cdot \sum_{t_1 + 2t_2 + 3t_3 = k} \left(1 + \sqrt{\frac{2e\Delta_T^{(2)}}{\Delta_T^2}} + \sqrt[4]{\frac{6e\Delta_T^{(3)}}{\Delta_T^3}} \right)^{2k}$$

$$\leq e(2k)^{1/2} \Delta_T^k k^3 \left(1 + \sqrt{\frac{2e\Delta_T^{(2)}}{\Delta_T^2}} + \sqrt[4]{\frac{6e\Delta_T^{(3)}}{\Delta_T^3}} \right)^{2k}.$$

Taking $2k$'th roots on both sides, letting $k \to \infty$ and applying (1) this yields that

$$\lambda_1(T) \le \Delta_T^{1/2}\left(1 + \sqrt{\frac{2e\Delta_T^{(2)}}{\Delta_T^2}} + \sqrt[4]{\frac{6e\Delta_T^{(3)}}{\Delta_T^3}}\right). \tag{3}$$

This bound is useful if one can control the ratios $\Delta_T^{(2)}/\Delta_T^2$ and $\Delta_T^{(3)}/\Delta_T^3$ in a meaningful manner, which is the case for the random trees considered in this paper. Indeed, it follows from the structural properties shown later[2] that, with high probability, for all vertices v in the random tree $\mathbf{T}_n$, the number of vertices within distance three from v is $O(\Delta_{\mathbf{T}_n} \log^3 \Delta_{\mathbf{T}_n})$, and therefore $(\Delta_{\mathbf{T}_n}^{(2)}/\Delta_{\mathbf{T}_n}^2)^{1/2} = O\left(\Delta_{\mathbf{T}_n}^{-1/2} \log^{3/2} \Delta_{\mathbf{T}_n}\right)$ and $(\Delta_{\mathbf{T}_n}^{(3)}/\Delta_{\mathbf{T}_n}^3)^{1/4} = O\left(\log^{1/4} \Delta_{\mathbf{T}_n}/\Delta_{\mathbf{T}_n}^{1/2}\right)$. This implies that, with high probability, the largest eigenvalue of $\mathbf{T}_n$ is at most

$$\sqrt{\Delta_{\mathbf{T}_n}} + O\left(\log^{3/2} \Delta_{\mathbf{T}_n}\right) .$$

Unfortunately, Proposition 1.3 is not powerful enough to prove Theorem 1.1. Actually, in order to get strong enough bounds on the number of long closed walks in $\mathbf{T}_n$, we end up additionally needing to keep track of the degrees of vertices we visit along the walk. This makes our codes more complicated, since they need to also encode this information. We partition the nodes of a tree T into three types: high-degree nodes, with degree at least $0.95\Delta_T$; medium-degree nodes, with degree between $\Delta_T^{1/5}$ and $0.95\Delta_T$; and low-degree nodes, which are all the rest. We then augment the Dyck path encoding the distance of a closed path from its starting location with a sequence of vertex types along the sequence. For example, an augmented Dyck path may read

$$((0,h),(1,h),(2,l),(1,h),(2,m),(1,h),(0,h)),$$

corresponding to a walk which moved from a high-degree node v to a neighbouring high-degree node w, then made two "zig-zag" steps, visiting a low-degree and a medium-degree neighbour of w in turn, each time returning to w, and finally returned to v.

Recall that a star tree with central vertex of degree Δ has maximum eigenvalue $\sqrt{\Delta}$, and the only possible closed walks on such a star (starting from the central vertex) consist of a sequence of "zigzag" steps from a high-degree vertex to low-degree vertices and back again. Given that we aim to prove that $\mathbf{E}|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| = o(1)$, this suggests that for large k the dominant contribution to $N_k(v, \mathbf{T}_n)$ should consist of zigzag steps from a high-degree vertex to its low or medium-degree neighbours. It turns out to be useful to bound the contribution coming from such steps separately. Thus, we prove bounds on $N_k(v, T)$ by first bounding the number of walks of length $2j \le 2k$ which make no "zigzags from high-degree vertices", then estimating how many ways such walks can be decorated with zigzags to form a walk of length $2k$. (For technical reasons we also need to initially count walks which make no "zigzags of height two", of the form $((0,m/h),(1,l),(2,l),(1,l),(0,m/h))$, and then also control the number of ways such height-two zigzags can be added.) The bound we end up using appears as Proposition 2.6, in Section 2.2. We apply it with a code consisting of 50 words, which appears in Figure 4 in Section 5; the need to track vertex types as well as distance is the reason we need so many code words.

It turns out that to apply Proposition 2.6 to the tree $\mathbf{T}_n$ in order to bound $\mathbf{E}|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}|$, it would suffice to show that $\mathbf{T}_n$ satisfied the following set of structural properties with sufficiently high probability. For a tree $T = (V, E)$ and $v \in V$, write $\deg_T(v)$ for the degree of v in T, and write $C_T(v)$ for the set of all vertices $w \in V$ such that the unique path from v to w does not contain two consecutive vertices with degree $< (\log n)^{1/5}$; we call $C_T(v)$

---

[2]Specifically, the upper bound on third neighbourhood sizes from Corollary 3.7, together with the lower bound on $\Delta_{\mathbf{T}_n}$ coming from Lemma 4.5.

the cluster of v in T. Then the properties we end up requiring of a tree $T = (V, E)$ with n vertices, in order for our combinatorial bounds to be strong enough, are the following.

N1 The maximum degree of T is $\Delta_T \geq 0.99 \, (\log n / \log \log n)$.
N2 For all $x \in V$, $\sum_{y \in C_T(v)} \deg_T(y) \mathbf{1}_{[\deg_T(y) \geq (\log n)^{1/5}]} \leq 3 \log n$.
N3 There are no adjacent nodes that both have degree $\geq 0.9 \log n / \log \log n$.
N4 For all $x \in V$, $C_T(x)$ contains $< 9 \log \log n$ vertices of degree $\geq (\log n)^{1/5}$.

We prove that a random tree $\mathbf{T}_n$ satisfies **N1** − **N3** with high probability by instead analyzing the probability that these properties hold for the family tree of a Poisson(1) branching process conditioned to have total size n, then using the fact that the graph structure of $\mathbf{T}_n$ is equivalent to that of such a conditioned branching process. However, we were unable to thereby establish that $\mathbf{T}_n$ satisfies property **N4** with high probability, and we do not in fact believe it to be true. What saves the proof, and allows us to make the above approach work, is a *rewiring lemma*, which permits us to "clean up" our tree to obtain one with the requisite properties, without decreasing λ_1. This rewiring lemma seems like a potentially valuable general tool, and its proof is quite simple, so we close the sketch of the upper bound by stating it and immediately providing its proof.

Given a finite tree $T = (V, E)$ and distinct vertices $v, w \in V$, the (v, w)-rewiring of T, denoted by $T_{v,w}$, is the graph obtained from T as follows: let $\Gamma(v, T)$ denote the neighborhood of v in T, and $u \in \Gamma(v, T)$ be the unique neighbor of v on the path connecting v to w in T. Then the rewired tree $T_{v,w} = (V, E_{v,w})$ has the same vertex set as T and edge set

$$E_{v,w} := E \cup \{sw \,:\, s \in \Gamma(v, T) - \{u\}\} - \{sv \,:\, s \in \Gamma(v, T) - \{u\}\}.$$

In words, $T_{v,w}$ is obtained from T by replacing each edge sv such that $s \neq u$ by the edge sw. The following fact is immediate.

Fact 1.4. *$T_{v,w}$ is still a tree, and the unique path connecting v to w in $T_{v,w}$ is the same as in T. Moreover, the degrees $\deg_{T_{v,w}}(x)$, for $x \in V$, satisfy:*

$$\deg_{T_{v,w}}(x) = \begin{cases} 1, & x = v; \\ \deg_T(v) + \deg_T(w) - 1, & x = w; \\ \deg_T(x), & x \in V \backslash \{v, w\}. \end{cases} \tag{4}$$

Lemma 1.5 (Rewiring lemma). *With the above notation, $\max\{\lambda_1(T_{v,w}), \lambda_1(T_{w,v})\} \geq \lambda_1(T)$.*

Proof. Assume $V = [n]$ for simplicity and let A denote the adjacency matrix of T. By Perron-Frobenius, its largest eigenvalue is achieved by a vector $x \in \mathbb{R}^n$ with $|x| = 1$ and non-negative entries, so that:

$$\langle x, Ax \rangle = 2 \sum_{\{i,j\} \in E} x_i x_j = \lambda_1(T) \geq 0.$$

Assume without loss of generality that $x_v \leq x_w$, and let B denote the adjacency matrix of $T_{v,w}$. We will show that $\lambda_1(T_{v,w}) \geq \lambda_1(T)$. In fact, since

$$\langle x, Bx \rangle \leq \lambda_1(T_{v,w}) |x|^2 = \lambda_1(T_{v,w}),$$

it suffices to argue that $\langle x, Bx \rangle \geq \langle x, Ax \rangle$. To do this, let $u \in \Gamma_T(v)$ be as in the definition of $T_{v,w}$. Then

$$\langle x, Bx \rangle - \langle x, Ax \rangle = 2 \sum_{s \in \Gamma_T(v) \backslash \{u\}} (x_w - x_v) \, x_s \geq 0,$$

because $x_s \geq 0$ for all $s \in [n]$ and $x_w \geq x_v$ by assumption. □

We conclude the overview by briefly sketching the proof of Theorem 1.2, which is rather straightforward once Theorem 1.1 is at hand. Since $\lambda_k \leq \lambda_1$ for all $k \geq 1$, to prove the theorem it suffices to prove that $\mathbf{E}\lambda_1(\mathbf{T}_n) \leq \sqrt{a_n} + o(1)$ and that $\mathbf{E}\lambda_k(\mathbf{T}_n) \geq \sqrt{a_n} - o(1)$, for a_n as in the theorem. We accomplish this via a straightforward but useful fact about trees: if T is a rooted tree, and T has $2s$ nodes with at least c children, then $\lambda_s(T) \geq \sqrt{c}$. This is proved by

first constructing a set $(x^{(1)}, \ldots, x^{(s)})$ of orthogonal vectors such that $\langle x^{(i)}, Ax^{(i)} \rangle \geq \sqrt{c}|x^{(i)}|^2$ for $1 \leq i \leq s$, then using the Rayleigh quotient principle (see Lemma 5.2, below). The lemma allows us to bound $\mathbf{E}\lambda_k(\mathbf{T}_n)$ from below by showing that with high probability $\mathbf{T}_n$ contains many vertices of near-maximal degree, which we accomplish using the fact that the graph structure of $\mathbf{T}_n$ is the same as that of a Poisson branching process conditioned to have total progeny n. The latter fact also allows us to show that $\mathbf{E}\sqrt{\Delta_n}$ is within $o(1)$ of $\sqrt{a_n}$, which, together with Theorem 1.1, yields the required upper bound on $\mathbf{E}\lambda_1(\mathbf{T}_n)$.

1.3. **Overview of the remainder of the paper**. The rest of the paper is structured as follows. In Section 2 we establish the combinatorial tools that we use for counting walks in trees, culminating in Corollary 2.7 that gives a general upper bound in terms of certain properties of the tree and the code used for encoding walks. In Section 3 we prove some structural properties of random labeled trees that hold with high probability. These properties are required to effectively use the combinatorial bound of Corollary 2.7. As mentioned above, these "typicality" properties are not quite sufficient to derive the desired bound for the largest eigenvalue. In Section 4 we show how random trees can be "safely rewired" using Lemma 1.5 so that they become amenable to using our combinatorial tools, without decreasing their top eigenvalue. Finally, in Section 5 we put all ingredients of the proof together and describe a code that is sufficiently detailed, allowing us to conclude the proofs of Theorems 1.1 and 1.2.

2. **Combinatorial bounds**

2.1. **Decompositions of Dyck paths and the proof of Proposition 1.3**. In this section we elaborate on our method for bounding $N_k(T)$ via decompositions of Dyck paths. Fix a prefix code $C = (c^{(1)}, \ldots, c^{(a)})$ and a Dyck path $\mathrm{d} = (d_0, d_1, \ldots, d_{2k})$.

We begin by defining a partition $P = P(C, \mathrm{d})$ of $[2k]$ which encodes "what codeword each step belongs to"; this is quite similar to how we defined $\vec{t}(C, \mathrm{d})$ in Section 1.2.

First, if there is $0 < i < 2k$ such that $d_i = 0$ then, recursively, Dyck paths $(d_0, \ldots, d_i)$ and $(d_i, \ldots, d_{2k})$ induce partitions P_1 and P_2 of $\{1, \ldots, i\}$ and $\{i+1, \ldots, 2k\}$; in this case we set $P = P_1 \cup P_2$.

We may now assume that $d_i \neq 0$ for all $0 < i < 2k$. Let $\sigma(0)$ be minimal so that $(d_i, 0 \leq i \leq \sigma(0))$ is an element of C, and write $h = d_{\sigma(0)}$.

If $h = 0$ then necessarily $\sigma(0) = 2k$; in this case we set $P = \{[2k]\}$. Otherwise, for $1 \leq i \leq h$, let $\sigma(i) = \min\{m \geq \sigma(i-1) : d_m = h - i\}$, and set

$$P_0 = \big\{\{1, \ldots, \sigma(0) - 1, \sigma(0), \sigma(1), \ldots, \sigma(h)\}\big\}.$$

Note that we necessarily have $\sigma(h) = 2k$. Also, for each $0 \leq i < h$, we have $d_{\sigma(i+1)} = d_{\sigma(i)} - 1$. It follows that if $\sigma(i+1) > \sigma(i) + 1$ then

$$(d_j - (h - i); \ \sigma(i) \leq j \leq \sigma(i+1) - 1)$$

is an excursion (and thus a Dyck path) so, recursively, it induces a partition P_i of $\{\sigma(i) + 1, \ldots, \sigma(i+1) - 1\}$. (For convenience, if $\sigma(i+1) = \sigma(i) + 1$ then we take $P_i = \emptyset$.) Now set $P(C, \mathrm{d}) = \bigcup_{0 \leq i < h} P_i$.

Figure 1 contains a graphical representation of the partition of the Dyck path

$$0121210123232121210$$

by the code $C^* = \{c^{(1)}, c^{(2)}, c^{(3)}\} = \{010, 0121, 0123\}$. The corresponding partition is

$$P(C^*, \mathrm{d}) = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}$$
$$= \big\{\{1, 2, 3, 6\}, \{4, 5\}, \{7, 8, 9, 10, 13, 18\}, \{11, 12\}, \{14, 15\}, \{16, 17\}\big\}. \tag{5}$$

For the trivial code $\{01\}$, the partition would instead be

$$P(\{01\}, \mathrm{d}) = \big\{\{1, 6\}, \{2, 3\}, \{4, 5\}, \{7, 18\}, \{8, 13\}, \{9, 10\}, \{11, 12\}, \{14, 15\}, \{16, 17\}\big\};$$
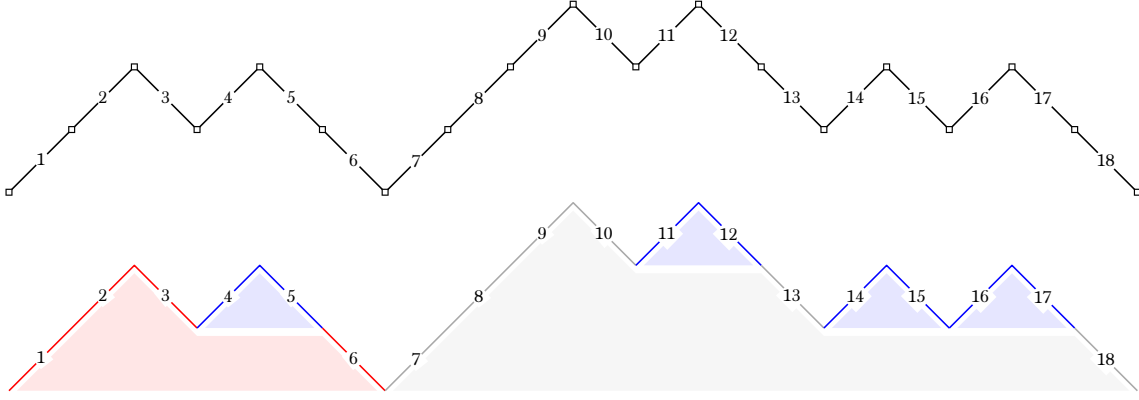
FIGURE 1. Top: a Dyck path of length 18. Bottom: a graphical depiction of the partition induced by the code $C^* = \{c^{(1)}, c^{(2)}, c^{(3)}\} = \{010, 0121, 0123\}$.

For a closed walk w in a tree T, we will also write $P(C, \text{w}) := P(C, D(\text{w}))$. It is immediate that if two walks w, w′ have $D(\text{w}) = D(\text{w}')$, then $P(C, \text{w}) = P(C, \text{w}')$. This implies that for any tree T, any vertex v of T, and any code C, we may rewrite $N_k(v, T)$, the number of walks from v to v of length $2k$, as

$$\sum_{\{\text{partitions } P \text{ of } [2k]\}} \sum_{\{\text{Dyck paths } \text{d}:P(C,\text{d})=P\}} \#\{\text{walks w in } T : w_0 = w_{2k} = v, D(\text{w}) = \text{d}\}.$$

An easy way to bound this expression from above is to note that whatever d may be, the inner summand is at most Δ_T^k; this yields the bound (2) from Section 1.2. To improve this bound, we need to control how much information required to recover $D(\text{w})$ from $P(C, \text{w})$ and to recover w from $D(\text{w})$, and we now turn to this.

Fix a Dyck path d and a code $C = \{c^{(b)}, 1 \le b \le a\}$, and let $P = P(C, \text{d})$. Fix any part π of P, list the elements of π in increasing order as $\pi(1), \ldots, \pi(s)$, and let $\pi(0) = \pi(1) - 1$. Then we write $c(\pi) = c(\pi, C, \text{d})$ for the unique meander $c \in C$ which is a prefix of

$$(d_{\pi(t)} - d_{\pi(0)}, 0 \le t \le s).$$

For example, for the Dyck path in Figure 1, and the partition given in (5), we have

$$(c(\pi_1), c(\pi_2), c(\pi_3), c(\pi_4), c(\pi_5), c(\pi_6)) = (0121, 010, 0123, 010, 010, 010)$$
$$= (c^{(2)}, c^{(1)}, c^{(3)}, c^{(1)}, c^{(1)}, c^{(1)}). \tag{6}$$

Lemma 2.1. *For any code C, any Dyck path* d *is recoverable from the set* $\{(\pi, c(\pi)), \pi \in P(C, \text{w})\}$.

Proof. Write π_1 for the part of $P(C, \text{d})$ containing 1 and $c(\pi) = c_1 c_2 \ldots c_k$. Then listing the elements of π_1 in increasing order as $i_1, \ldots, i_k$, by construction we have $d_{i_j} = c_j$ for $1 \le j \le k$. The result then follows by induction on the length of d, since what remains is to reconstruct $f(c(\pi)) + 1$ sub-paths of d which are themselves shifted Dyck paths of length strictly less than that of d. □

In fact, d can be recovered from somewhat less data than what is used in Lemma 2.1, since a part of π corresponding to a codeword c always begins with len(c) consecutive integers. More precisely, for any part π of $P(C, \text{d})$, if $c(\pi) = c \in C$ then, writing $\pi^{\min} = \min(k : k \in \pi)$, necessarily

$$\{\pi^{\min}, \pi^{\min} + 1, \ldots, \pi^{\min} + \text{len}(c) - 1\} \subset \pi, \tag{7}$$

by the construction of $P(C, \text{d})$. For example, for the Dyck path d in Figure 1, the third part of the partition $P(C^*, \text{d})$ is $\pi_3 = \{7, 8, 9, 10, 13, 18\}$. Given the information that $c(\pi_3) = 0123$, if we additionally know that $\min(k : k \in \pi_3) = 7$ then necessarily also $\{8, 9\} \subset \pi_3$, so we may construct π_3 from the set $\pi_3' = \{7, 10, 13, 18\}$ and the information that $c(\pi_3) = 0123$. Applying

this to all of the parts, we see that π may be reconstructed from the data in (6) together with the sets

$$\{\{1,6\},\{4\},\{7,10,13,18\},\{11\},\{14\},\{16\}\}. \tag{8}$$

The example in Figure 2 may be helpful in understanding the coming paragraph. For a part π of $P(C,\mathrm{d})$, writing

$$\pi' = \pi \setminus \{k \in \pi : \pi^{\min} < k < \pi^{\min} + \mathrm{len}(\mathrm{c}(\pi))\},$$

then in the setting of Lemma 2.1, the inclusion (7) implies that we may recover d from the data $\{(\pi',c(\pi)),\pi \in P(C,\mathrm{d})\}$. However, slightly more is true. We may "compress" the collection $(\pi',\pi \in P(C,\mathrm{w}))$ by removing gaps, to form a partition of a *consecutive* sequence of integers, without sacrificing recoverability, as follows. Write $\Pi' = \bigcup_{\pi \in P(C,\mathrm{w})} \pi'$, and for $\pi \in P(C,\mathrm{w})$, let

$$\hat{\pi} = \{\mathrm{rank}(i,\Pi'),i \in \pi'\},$$

where $\mathrm{rank}(i,\Pi') = \#\{j \in \Pi' : j \le i\}$. For example, if Π' is the collection of sets in (8), then $(\hat{\pi},\pi \in P(C,\mathrm{w}))$ is the partition

$$\{\{1,3\},\{2\},\{4,5,7,10\},\{6\},\{8\},\{9\}\} .$$

Corollary 2.2. *For any code C, any Dyck path* d *is recoverable from the data* $\{(\hat{\pi},\mathrm{c}(\pi)),\pi \in P(C,\mathrm{d})\}$.

Proof. For each $k \in \bigcup_{\pi \in P(C,\mathrm{d})} \hat{\pi}$, let

$$k' = k + \sum_{\{\pi \in P(C,\mathrm{d}):\hat{\pi}^{\min} < k\}} (\mathrm{len}(c(\pi)) - 1).$$

Then for all $\pi \in P(C,\mathrm{d})$, it holds that $\pi' = \{k' : k \in \hat{\pi}\}$, so $\{\pi',\pi \in P(C,\mathrm{d})\}$ is recoverable from $\{(\hat{\pi},\mathrm{c}(\pi)),\pi \in P(C,\mathrm{d})\}$. By the observations preceding the corollary and Lemma 2.1, the result follows. $\square$

This corollary gives us a way of using codes to bound the number of walks in a graph. Given a code $C = \{\mathrm{c}^{(b)},1 \le b \le a\}$, and $b \in [a]$, for a Dyck path d write

$$P_b(C,\mathrm{d}) = \{\pi \in P(C,\mathrm{d}) : \mathrm{c}(\pi) = \mathrm{c}^{(b)}\}$$

for the set of parts of $P(C,\mathrm{w})$ corresponding to the codeword $\mathrm{c}^{(b)}$. Then, for $1 \le b \le a$, let

$$t_b = t_b(C,\mathrm{d}) = |P_b(C,\mathrm{d})|$$

denote the number of times code word $\mathrm{c}^{(b)}$ is used in the partitioning. We then set $\vec{t}(C,\mathrm{d}) = (t_b(C,\mathrm{d}),b \in [a])$; this definition agrees with the one from Section 1.2. If w is a closed walk in a tree T, we also write $P(C,\mathrm{w}) = P(C,D(\mathrm{w}))$ and $\vec{t}(C,\mathrm{w}) = (t_b(C,\mathrm{w}),b \in [a]) := (t_b(C,D(\mathrm{w})),b \in [a])$. This allows us to decompose the set of walks w of a certain length according to number of occurrences t_b of each code word in $P(C,\mathrm{w})$.

In the next proposition we bound the number of closed walks from a vertex to itself with a given "profile" $\vec{t} = (t_1,\ldots,t_a)$. Recall from Section 1.2 that for a vertex v in a tree T, we write $N_{C,\vec{t}}(v)$ for the number of closed walks w from v to v in T with $t_b(C,\mathrm{w}) = t_b$ for all $1 \le b \le a$. Note that the length of any such walks is $2k = 2k(\vec{t}) = \sum_{b=1}^{a} t_b(\mathrm{len}(\mathrm{c}^{(b)}) + f(\mathrm{c}^{(b)}))$. Recall also that for a meander $\mathrm{c} = (c_0,\ldots,c_j)$, we write

$$\Delta_T(\mathrm{c}) = \max_{v \in V} |\{\text{walks } \mathrm{w} = (w_0,\ldots,w_j) \text{ in } T : w_0 = v, D(\mathrm{w}) = \mathrm{c}\}| .$$
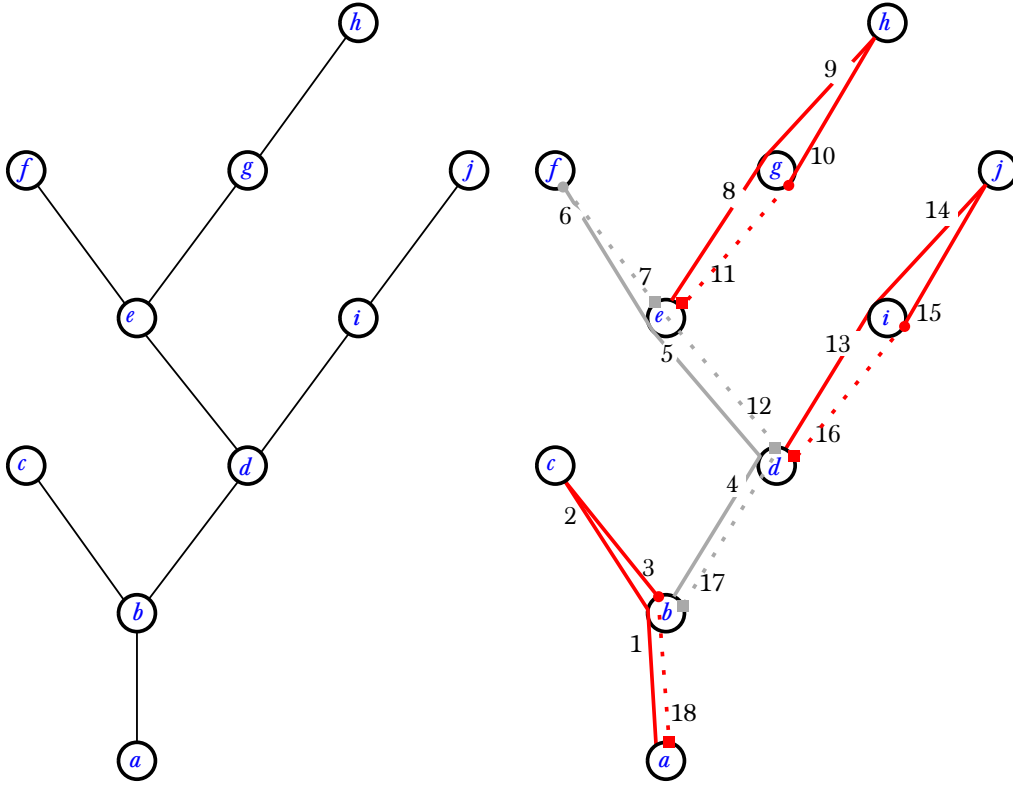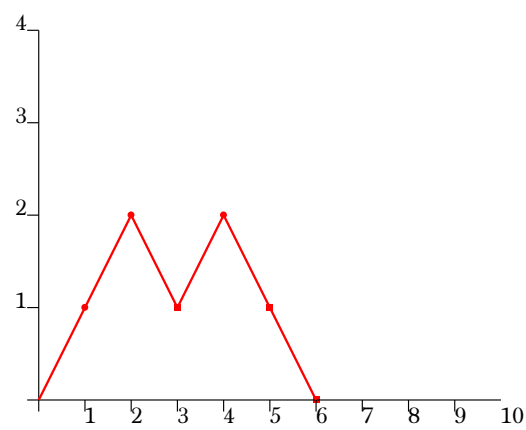
(A) The contour walk w of this tree is *abcbdefeghgedijidba*. The corresponding Dyck path $D(w)$ is 0121234345432343210.

(B) Let $C = \{010, 0121, 0123\}$; then $P(C, w)$ has parts $\{1, 2, 3, 18\}$, $\{4, 5, 6, 7, 12, 17\}$, $\{8, 9, 10, 11\}$ and $\{13, 14, 15, 16\}$.
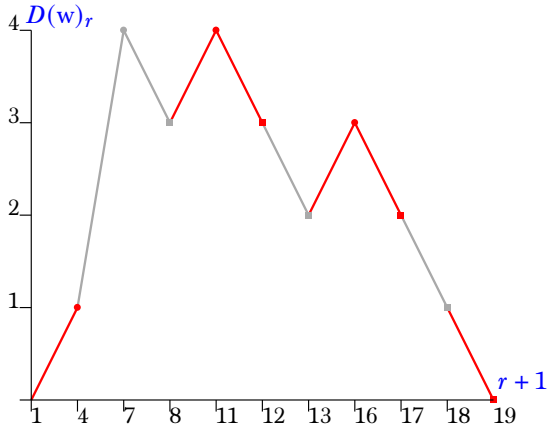
(C) A plot of $(D(w)_r : r+1 \in \Pi' \cup \{\text{len}(w)+1\})$.

(D) The walk S corresponding to codeword 0121.

FIGURE 2. For w and C as above, the collection $\{\pi' : \pi \in P(C, w)\}$ equals $\{\{1, 18\}, \{4, 7, 12, 17\}, \{8, 11\}, \{13, 16\}\}$, and $\{\hat{\pi} : \pi \in P(C, w)\}$ equals $\{\{1, 10\}, \{2, 3, 6, 9\}, \{4, 5\}, \{7, 8\}\}$. With $c^{(2)} = 0121$ and $c^{(3)} = 0123$, then $\Pi_2 = \{1, 4, 5, 7, 8, 10\}$ and $\Pi_3 = \{2, 3, 6, 9\}$.

Proposition 2.3. *Fix a code $C = \{c^{(b)}, 1 \leq b \leq a\}$, and a vector of non-negative integers $\vec{t} = (t_1, \ldots, t_a)$, and let $q(\vec{t}) = \sum_{b=1}^{a} t_b \cdot (1 + f(c^{(b)}))$. Then for any tree T and any vertex v of T,*

$$N_{C, \vec{t}}(v) \leq \binom{q(\vec{t})}{t_b(1 + f(c^{(b)})), 1 \leq b \leq a} \prod_{b=1}^{a} \frac{1}{t_b(1 + f(c^{(b)})) + 1} \binom{t_b(1 + f(c^{(b)})) + 1}{t_b} \cdot \Delta_T(c^{(b)})^{t_b}.$$

(9)

Proof. Fix a closed walk w in T and a codeword $c \in C$. If $\pi \in P(C, \text{w})$ and $c(\pi) = c$ then $|\pi| = \text{len}(c) + f(c)$ and $|\hat{\pi}| = |\pi| - (\text{len}(c) - 1) = 1 + f(c)$. It follows that $\hat{P}(C, \text{w}) := (\hat{\pi}, \pi \in P(C, \text{w}))$ is a partition of $[q(\vec{t})]$. Writing $\Pi_b = \cup_{\pi \in P(C, \text{w}):c(\pi)=c^{(b)}} \hat{\pi}$, then $(\Pi_b, 1 \le b \le a)$ is another, coarser partition of $[q(\vec{t})]$, in which $|\Pi_b| = t_b(1 + f(c^{(b)}))$. Accordingly, the number of possibilities for the partition $(\Pi_b, 1 \le b \le a)$, as w varies over walks with $(t_b(C, \text{w}), 1 \le b \le a) = \vec{t}$, is at most

$$\binom{q(\vec{t})}{t_b(1 + f(c^{(b)})), 1 \le b \le a}. \tag{10}$$

We next focus our attention on the number of words which could give rise to the partition $(\Pi_b, 1 \le b \le a)$. Figure 2 should be useful in understanding the arguments of this paragraph. Consider a given value of b, write $f = f(c^{(b)})$, and define a sequence $S = (S(p), 0 \le p \le |\Pi_b|)$ as follows. List the elements of Π_b in increasing order as $p_1, \ldots, p_{|\Pi_b|}$. Set $S(0) = 0$, $S(1) = f$, and for $1 < i \le |\Pi_b|$ let

$$S(i) = \begin{cases} S(i-1) + f & \text{if } p_i \text{ and } p_{i-1} \text{ are in different parts of } \hat{P}(C, \text{w}) \\ S(i-1) - 1 & \text{if } p_i \text{ and } p_{i-1} \text{ are in the same part of } \hat{P}(C, \text{w}). \end{cases}$$

For any part π of $P(C, \text{w})$ with $c(\pi) = c^{(b)}$, we have $|\hat{\pi}| = |\pi'| = h+1$ by construction; from this it is immediate that $S(|\Pi_b|) = 0$. Moreover, if $\sigma, \rho \in P(C, \text{w})$ are distinct and $\rho_j < \sigma^{\min} < \rho_{j+1}$, where ρ_j and ρ_{j+1} are consecutive elements of ρ, then in fact $\sigma \subset \{\rho_j+1, \ldots, \rho_{j+1}-1\}$; this is an immediate consequence of the definition of $P(C, \text{w})$. This nesting structure implies that there is some part $\pi \in P(C, \text{w})$ with $c(\pi) = c^{(b)}$ such that if $\hat{\pi}^{\min} = p_i$, then

$$\{p_i, p_{i+1}, \ldots, p_{i+f}\} = \hat{\pi}.$$

We then have

$$(S(i), S(i+1), \ldots, S(i+f)) = (S(i-1) + f, S(i-1) + f - 1, \ldots, S(i-1)).$$

It follows by induction that S is non-negative. Moreover, the nesting structure likewise implies that the collection $(\pi \in P(C, \text{w}), c(\pi) = c^{(b)})$ can be recovered from Π_b and S. The number of walks S of length $|\Pi_b| = t_b(1 + f)$ with step sizes in $\{-1, f\}$, starting and finishing at zero and staying weakly above zero is precisely $\frac{1}{t(1+f)+1}\binom{t(1+f)+1}{t}$ - see, e.g., [18, Theorem 10.4.1].

We apply the preceding bound to each set Π_b. Combined with (10) we see that, having constrained w to satisfy that $(t_b(C, \text{w}), 1 \le b \le a) = \vec{t}$, the number of possibilities for the collection

$$\{(\hat{\pi}, c(\pi)), \pi \in P(C, \text{w})\}$$

is at most

$$\binom{q(\vec{t})}{t_b(1 + f(c^{(b)})), 1 \le b \le a} \prod_{b=1}^{a} \cdot \frac{1}{t_b(1 + f(c^{(b)})) + 1} \binom{t_b(1 + f(c^{(b)})) + 1}{t_b}. \tag{11}$$

By Corollary 2.2, it follows that the number of possibilities for the Dyck path $D(\text{w})$ is bounded by the expression in (11).

Finally, to recover w from $D = D(\text{w})$, it suffices to embed the subwalks corresponding to the parts of $P(C, w)$ into G. For a part π of $P(C, w)$, having specified the location of $w_{\pi^{\min}-1}$, by definition there are at most $\Delta_T(c(\pi))$ possibilities for the subwalk $(w_p, p \in \pi)$. It follows that the number of walks with $D(\text{w}) = D$ is at most

$$\prod_{b=1}^{a} \Delta_T(c^{(b)})^{t_b},$$

which completes the proof. $\qquad \square$

We are now prepared to prove Proposition 1.3.

Proof of Proposition 1.3. We bound the multinomial coefficient in (9) with the aid of the in-equalities $(n/e)^n \le n! \le e\sqrt{n}(n/e)^n$. We obtain

$$\binom{q}{t_b(1+f(c^{(b)})),\, 1 \le b \le a} \le e\sqrt{q}\, e^{-q} q^q \prod_{b=1}^{a} \frac{1}{(t_b(1+f(c^{(b)})))!}$$

$$= e\sqrt{q} \prod_{b=1}^{a} \frac{(q/e)^{t_b(1+f(c^{(b)}))}}{(t_b(1+f(c^{(b)})))!}$$

$$\le e\sqrt{q} \prod_{b=1}^{a} \left(\frac{q}{t_b(1+f(c^{(b)}))} \right)^{t_b(1+f(c^{(b)}))}.$$

We also have

$$\frac{1}{t_b(1+f(c^{(b)}))+1} \binom{t_b(1+f(c^{(b)}))+1}{t_b} \le \binom{t_b(1+f(c^{(b)}))}{t_b} \le (1 \vee 2ef(c^{(b)}))^{t_b},$$

the second inequality holding since both sides are 1 when $f(c^{(b)}) = 0$, and when $f(c^{(b)}) \ge 1$ we have

$$\binom{t_b(1+f(c^{(b)}))}{t_b} \le \left(\frac{et_b(1+f(c^{(b)}))}{t_b} \right)^{t_b} = (e(1+f(c^{(b)})))^{t_b} \le (2ef(c^{(b)}))^{t_b}.$$

Using these in the upper bound of Proposition 2.3 yields the bound

$$N_{C,\vec{t}}(v) \le e\sqrt{q} \prod_{b=1}^{a} \left((1 \vee 2ef(c^{(b)})) \Delta_T(c^{(b)}) \right)^{t_b} \left(\frac{q}{t_b(1+f(c^{(b)}))} \right)^{t_b(1+f(c^{(b)}))}$$

$$= e\sqrt{q}\Delta^k \prod_{b=1}^{a} \left(\frac{(1 \vee 2ef(c^{(b)})) \Delta_T(c^{(b)})}{\Delta^{(\mathrm{len}(c^{(b)})+f(c^{(b)}))/2}} \right)^{t_b} \left(\frac{q}{t_b(1+f(c^{(b)}))} \right)^{t_b(1+f(c^{(b)}))},$$

where in the second line we have multiplied and divided by $\Delta^k = \Delta^{\sum_{b=1}^{a} t_b(\mathrm{len}(c^{(b)})+f(c^{(b)}))/2}$. Setting $\rho_b = t_b(1+f(c^{(b)}))/q$ for $1 \le b \le a$, we may rewrite the above bound as

$$e\sqrt{q}\Delta^k \exp\left(q \sum_{b=1}^{a} \rho_b \log\left(\frac{g_b}{\rho_b}\right) \right)$$

Since $\sum_{b=1}^{a} \rho_b = 1$, by the non-negativity of the Kullback-Leibler divergence,

$$\sum_{b=1}^{a} \rho_b \log\left(\frac{g_b}{\rho_b}\right) \le \log\left(\sum_{b=1}^{a} g_b\right),$$

from which the result follows.   □

2.2. **Refined bounds via nonbacktracking walks.** As discussed in the proof overview, in order to strengthen the above bounds, we need to refine our decomposition of walks in trees according to the times when they visit high-degree vertices. This is achieved by classifying vertices into three types according to their degrees ("high," "moderate," and "low" degrees, defined below) and partitioning walks not only by their distances to the starting vertex but also by the sequence of types of the vertices. The prefix codes may be easily extended to encode the type information as well. Part of our strategy will be to separately quantify the contribution to the number of closed walks arising from "backtracking" subwalks, of type xyx or $xyzyx$, where x is a vertex of medium or high degree and y and z are vertices of low degree. (We shall call these *simple* and *double* backtracking steps, respectively.) This will allow us to exploit finer structural properties of uniformly random trees, which we establish in Sections 3 and 4.

Formally, we define an (M,H)-*tree* to be a 4-tuple $T = (V,E,M,H)$, where (V,E) is a tree and $M,H \subset V$ are disjoint sets of vertices; we will eventually take M to be the set of "moderately high degree vertices" and H to be the set of "very high degree vertices." Given a walk w $= (w_i, 0 \le i \le j)$ in an (M,H)-tree, we define a *type sequence* $\tau = (\tau_i(T,\mathrm{w}), 0 \le i \le j)$, by setting

$$\tau_i(T,\mathrm{w}) = \begin{cases} h & \text{if } w_i \in H \\ m & \text{if } w_i \in M \\ \ell & \text{if } w_i \in V \setminus (M \cup H) . \end{cases}$$

(Here h, m, and ℓ stand for "high," "moderate", and "low", indicating the degree type of the vertex visited at step i of the walk.) We let $E_T(\mathrm{w}) = ((\mathrm{dist}_T(w_0, w_i), \tau_i(T,\mathrm{w})), 0 \le i \le j)$, and call $E_T(\mathrm{w})$ the *encoding path* corresponding to the walk w. Note that extracting the first coordinate from each of the pairs in the encoding path yields the Dyck path $D_T(\mathrm{w})$.

Given a walk w in T:

- A *simple backtracking step* is a subwalk (w_i, w_{i+1}, w_{i+2}) of w such that $w_{i+2} = w_i$; $w_i \in M \cup H$; $w_{i+1} \notin M \cup H$; and $\mathrm{dist}_T(w_0, w_{i+1}) > \mathrm{dist}_T(w_0, w_i)$.
- A *double backtracking step* is a subwalk $(w_i, w_{i+1}, w_{i+2}, w_{i+3}, w_{i+4})$ of w such that $w_{i+4} = w_i$; $w_i \in M \cup H$; $w_{i+1}, w_{i+2} \notin M \cup H$; and $\mathrm{dist}_T(w_0, w_{i+2}) > \mathrm{dist}_T(w_0, w_{i+1}) > \mathrm{dist}_T(w_0, w_i)$.

In fact, the condition that $\mathrm{dist}_T(w_0, w_{i+2}) > \mathrm{dist}_T(w_0, w_{i+1})$ may be deduced from the other conditions in the definition of double backtracking steps, but we find it clearer to write it explicitly. An (M,H)-*doubly-nonbacktracking walk* (or (M,H)-*walk* for short) is a walk w in T that contains no simple or double backtracking steps.

An (M,H)-*meander* is a sequence d $= ((d_i, \tau_i), 0 \le i \le j)$ which is the encoding path of some (M,H)-walk; it is an excursion if $d_j = 0$. We define $f(\mathrm{d}) = d_j$ and call $f(\mathrm{d})$ the final value of d. An (M,H)-meander d is an (M,H)-*excursion* if $f(\mathrm{d}) = 0$. A set C of (M,H)-meanders is an (M,H)-*code* if for any (M,H)-excursion d $= ((d_i, \tau_i), 0 \le i \le j)$ of positive length, there is a unique (M,H)-meander c $\in C$ such that c is a prefix of d.

The "trivial" (M,H)-code consists of nine words, each of the form $((0,*),(1,*))$, where each entry of "$*$" may be replaced by an ℓ, m, or h. Additionally, using m/h to denote an entry which can be either an m or an h, the simplest non-trivial (M,H)-code has eleven words and is shown in Figure 3. Using the above shorthand, it can be written as

$$\{((0,\ell),(1,*)), ((0,m/h),(1,m/h)), ((0,m/h),(1,\ell),(2,m/h))\} .$$

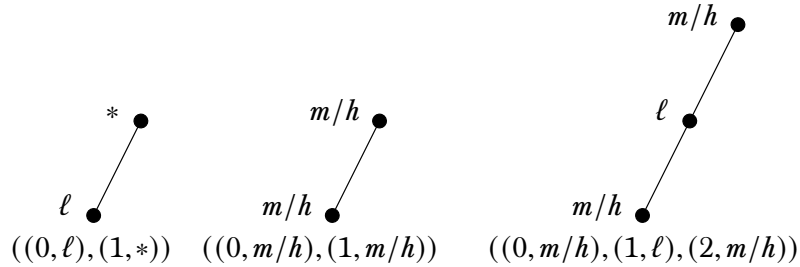FIGURE 3. The simplest non-trivial (M,H)-code.

For an (M,H)-excursion d $= ((d_i, \tau_i), 0 \le i \le j)$ and an (M,H)-code C, the partition $P(C,\mathrm{d})$ is defined just as in Section 2.1. Fix any part π of $P(C,\mathrm{d})$, list the elements of π in increasing order as $\pi(1), \ldots, \pi(s)$, and let $\pi(0) = \pi(1) - 1$. Then there is a unique (M,H)-meander c$(\pi) = \mathrm{c}(\pi, C, \mathrm{d}) \in C$ such that

$$\mathrm{c}(\pi) = ((d_{\pi(r)} - d_{\pi(0)}, \tau_{\pi(r)}), 0 \le r \le s') ,$$

for some $s' \leq s = |\pi|$. For a closed walk w in an (M,H)-tree T, we also write $P(C,\mathrm{w}) = P(C,E(\mathrm{w})))$. Like in Section 2.1, if walks w,w′ have $E_T(\mathrm{w}) = E_T(\mathrm{w}')$, then $P(C,\mathrm{w}) = P(C,\mathrm{w}')$. Thus, for any tree T, any vertex v of T, and any code C, we may rewrite $N_k(v,T)$, the number of walks from v to v of length $2k$, as

$$\sum_{\{\text{partitions } P \text{ of } [2k]\}} \sum_{\{(M,H)-\text{excursions } \mathrm{d}:P(C,\mathrm{d})=P\}} \#\{\text{walks w in } T : w_0 = w_{2k} = v, E(\mathrm{w}) = \mathrm{d}\} .$$

The definitions just preceding Proposition 2.3 also carry over naturally to the setting of (M,H)-trees. For an (M,H)-tree T, a node v of T, and an (M,H)-meander $\mathrm{c} = ((c_i,\tau_i),0 \leq i \leq j)$, let

$$\deg_T(v,\mathrm{c}) = \#\{(M,H)\text{-walks w} = (w_0,\ldots,w_j) \text{ in } T : w_0 = v, E_T(\mathrm{w}) = \mathrm{c}\},$$

and let $\Delta_T(\mathrm{c}) = \max\{\deg_T(v,\mathrm{c}) : v \text{ a node of } T\}$.

Given an (M,H)-code $C = \{\mathrm{c}^{(b)}, 1 \leq b \leq a\}$ and a codeword $\mathrm{c}^{(b)} = ((c_i^{(b)},\tau_i^{(b)}),0 \leq i \leq m) \in C$, for an (M,H)-walk w write

$$P_b(C,\mathrm{w}) = \{\pi \in P(C,\mathrm{w}) : c(\pi) = \mathrm{c}^{(b)}\}$$

for the set of parts of $P(C,\mathrm{w})$ corresponding to the (M,H)-meander $\mathrm{c}^{(b)}$. Then, for $1 \leq b \leq a$, let

$$t_b = t_b(C,\mathrm{w}) = |P_b(C,\mathrm{w})| .$$

Further recycling the notation of Section 2.1, given an (M,H)-code $C = (\mathrm{c}^{(b)}, 1 \leq b \leq a)$, for a vector of non-negative integers $\vec{t} = (t_1,\ldots,t_a)$ and a node v in an (M,H)-tree T, we denote by $N_{C,\vec{t}}(v)$ the number of (M,H)-walks w from v to v in T with $t_b(C,\mathrm{w}) = t_b$ for all $1 \leq b \leq a$. The next proposition bounds the number of such walks from above.

Proposition 2.4. *Fix an (M,H)-tree $T = (V,E,M,H)$ and an (M,H)-code $C = \{\mathrm{c}^{(b)}, 1 \leq b \leq a\}$, and write*

$$\varepsilon = \varepsilon(C,T) := (2a)^2 \max_{b \in [a]} \frac{\Delta_T(\mathrm{c}^{(b)})^{\frac{2}{\mathrm{len}(\mathrm{c}^{(b)})+f(\mathrm{c}^{(b)})}}}{\Delta_T} \tag{12}$$

Then for any vector $\vec{t} = (t_1,\ldots,t_a) \in \mathbb{Z}_{\geq 0}^a$, setting $k(\vec{t}) := \sum_{b \in [a]} t_b(\mathrm{len}(\mathrm{c}^{(b)}) + f(\mathrm{c}^{(b)}))/2$, we have that

$$\max_{v \in T} N_{C,\vec{t}}(v) \leq 2^a (\varepsilon \Delta_T)^{k(\vec{t})}.$$

Crucially, the upper bound established for $N_{C,\vec{t}}(v)$ does not depend on v, and only depends on $\vec{t}$ via $k(\vec{t})$. The form in which we state the bound may seem curious, since we could cancel the factor of Δ_T with the denominator in (12). We wrote the bound in this way because ε will show up again later, in an expression where such a cancellation is not possible. Now, for a given value of $k \in \mathbb{N}$, there are at most a^{2k} choices for the type vector $\vec{t}$ with $k(\vec{t}) = k$. Therefore, letting $N_k^{M,H}(v,T)$ denote the number of closed (M,H)-walks of length $2k$ from v to v in T and setting $N_k^{M,H}(T) = \max_{v \in V} N_k^{M,H}(v,T)$, we have the following corollary.

Corollary 2.5. *For any (M,H)-tree $T = (V,E,M,H)$ and (M,H)-code $C = \{\mathrm{c}^{(b)}, 1 \leq b \leq a\}$, for any integer $k \geq 0$ we have $N_k^{M,H}(T) \leq 2^a (a^2 \varepsilon \Delta_T)^k$, where $\varepsilon = \varepsilon(C,T)$ is as in (12).*

When we later apply this bound, we will need the parameter ε to be much smaller than $\Delta_T^{-1/2}$; it is there that the right choice of code will be crucial.

Proof of Proposition 2.4. Fix $v \in V$ and $\vec{t}$ as above, and define $q = q(\vec{t}) = \sum_{b=1}^{a} t_b \cdot (1 + f(\mathbf{c}^{(b)}))$. Precisely the same proof as for Proposition 2.3 gives the bound

$$N_{C,\vec{t}}(v)$$

$$\leq \binom{q(\vec{t})}{t_b(1 + f(\mathbf{c}^{(b)})), 1 \leq b \leq a} \prod_{b=1}^{a} \frac{1}{t_b(1 + f(\mathbf{c}^{(b)})) + 1} \binom{t_b(1 + f(\mathbf{c}^{(b)})) + 1}{t_b} \cdot \Delta_T(\mathbf{c}^{(b)})^{t_b} .$$

In order to simplify this bound, note that

$$\binom{q(\vec{t})}{t_b(1 + f(\mathbf{c}^{(b)})), 1 \leq b \leq a} \leq a^{q(\vec{t})} \leq a^{2k(\vec{t})} ,$$

and that for each $b \in [a]$,

$$\frac{1}{t_b(1 + f(\mathbf{c}^{(b)})) + 1} \binom{t_b(1 + f(\mathbf{c}^{(b)})) + 1}{t_b} \leq 2^{t_b(1 + f(\mathbf{c}^{(b)})) + 1} ,$$

leading to

$$N_{C,\vec{t}}(v) \leq a^{2k(\vec{t})} 2^{a + 2k(\vec{t})} \prod_{b=1}^{a} \Delta_T(\mathbf{c}^{(b)})^{t_b}$$

$$= \Delta_T^{k(\vec{t})} a^{2k(\vec{t})} 2^{a + 2k(\vec{t})} \prod_{b=1}^{a} \left(\frac{\Delta_T(\mathbf{c}^{(b)})^{\frac{2}{(\mathrm{len}(\mathbf{c}^{(b)}) + f(\mathbf{c}^{(b)}))}}}{\Delta_T} \right)^{t_b(\mathrm{len}(\mathbf{c}^{(b)}) + f(\mathbf{c}^{(b)}))/2} .$$

The product is bounded by

$$\left(\frac{\varepsilon}{(2a)^2} \right)^{\sum_{b=1}^{a} t_b(\mathrm{len}(\mathbf{c}^{(b)}) + f(\mathbf{c}^{(b)}))} = \left(\frac{\varepsilon}{(2a)^2} \right)^{k(\vec{t})} .$$

Plugging this back into the previous display finishes the proof. □

2.3. **Bounding the number of walks**. In this section we establish the main combinatorial bound for $N_k(v, T)$, which we recall is the total number of closed walks of length $2k$ from v to v in T. To this end, first we convert T into an (M, H)-tree by defining H as the set of vertices of degree at least $0.95\Delta_T$ where Δ_T is the maximum degree in T. Moreover, we define M as the set of vertices of degree less than $0.95\Delta_T$ but at least κ for a threshold $\kappa > 0$ specified later. We say that a vertex is of high degree if $v \in H$, it is of moderate degree if $v \in M$ and it is of low degree otherwise. Recall that $\Delta_T^{(2)}$ is the greatest size of any second neighbourhood in T.

Proposition 2.6. *Fix any real number $\kappa > 0$. Let $T = (V, E, M, H)$ be an (M, H)-tree with H the set of vertices of degree at least $0.95\Delta_T$ and M the set of vertices of $V \setminus H$ of degree at least κ. Suppose that there are no adjacent vertices in T of degree at least $0.95\Delta_T$. Fix an (M, H)-code $C = \{\mathbf{c}^{(b)}, 1 \leq b \leq a\}$ and let*

$$\varepsilon = \varepsilon(C, T) = (2a)^2 \max_{b \in [a]} \frac{\Delta_T(\mathbf{c}^{(b)})^{\frac{2}{(\mathrm{len}(\mathbf{c}^{(b)}) + f(\mathbf{c}^{(b)}))}}}{\Delta_T}$$

Then for all $v \in V$,

$$N_k(v, T) \leq 2^a (k+1)^2 \Delta_T^k \left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2} \right)^{2k} \left(1 + e^{20} a^2 \varepsilon \right)^k .$$

The proposition has a direct consequence for the largest eigenvalue of T. Indeed, since

$$\lambda_1(T) = \lim_{k \to \infty} \left(\max_{v \in V} N_k(v, T) \right)^{\frac{1}{2k}} ,$$

we have the following result.

Corollary 2.7. *In the setting of Proposition 2.6, the largest eigenvalue of T satisfies*

$$\lambda_1(T) \leq \sqrt{\Delta_T}\left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2}\right)\sqrt{1 + e^{20}a^2\varepsilon}.$$

To put this into perspective, recall that our goal is to prove an eigenvalue bound of the form $\lambda_1(T) \leq \sqrt{\Delta_T} + o(1)$ for most trees on n vertices. The corollary implies that such a bound follows if both $\Delta_T^{(2)}/\Delta_T$ and $\varepsilon = \varepsilon(C, T)$ are much smaller than $1/\sqrt{\Delta_T}$. While $\Delta_T^{(2)}/\Delta_T$ depends only on the tree, ε depends also on the choices of code C and threshold κ (since κ determines M and H). It will turn out that taking $\kappa = \log^{1/5} n$ is good enough for our purposes, if the right code is used.

Let us now prove the proposition.

Proof of Proposition 2.6. In order to bound $N_k(v, T)$, we partition the set of walks of length $2k$ from v to v into equivalence classes. An equivalence class is determined by

(1) the number of single backtracking steps of type xyx where x has high degree;
(2) the number of single backtracking steps of type xyx where x is of moderate degree;
(3) the number of double backtracking steps of type $xyzyx$ where x is of high degree;
(4) the number of double backtracking steps of type $xyzyx$ where x is of moderate degree;

To formalize this, given a walk $\mathsf{w} = (w_0, \ldots, w_{2k})$ in T, the *reduced (M, H)- walk* is obtained from w by removing all backtracking steps from w. In other words, if $(w_i, \ldots, w_{i+j})$ is a simple or double backtracking step of w, then replace w by the subwalk $(w_0, \ldots, w_i, w_{i+j+1}, \ldots, w_{2k})$, and repeat until no simple or double backtracking steps remain. From the definition of backtracking steps, it is straightforward to verify that the resulting walk w^- does not depend on the order in which the backtracking steps are removed, and that w^- is an (M, H)-walk as defined in the previous subsection.

Writing $\langle \mathsf{w}^- \rangle$ for the equivalence class of w^-, which consists of all walks w of length $2k$ from v to v whose reduced (M, H)-walk is w^-, we then have

$$N_k(v, T) = \sum_{\mathsf{w}^-} |\langle \mathsf{w}^- \rangle|, \tag{13}$$

where sum on the right is over all (M, H)-walks w^- from v to v of (even) length $0 \leq 2j \leq 2k$.

To bound this expression, group (M, H)-walks by their lengths. We know from Corollary 2.5 that $N_j^{M,H}(v, T)$, the number of (M, H)-walks of length $2j$ from v to itself, is at most $2^a a^{2j} \varepsilon^j \Delta_T^j$, so

$$N_k(v, T) \leq \sum_{j=0}^{k} (2^a a^{2j} \varepsilon^j \Delta_T^j)\, Z(j, k), \tag{14}$$

where

$$Z(j, k) := \max\{|\langle \mathsf{w}^- \rangle| \ : \ \mathsf{w}^- \text{ is an } (M, H)\text{-walk of length } 2j\}.$$

We now make the following claim.

Claim 2.8. *For each $0 \leq j \leq k$,*

$$Z(j, k) \leq (k+1)^2 \left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2}\right)^{2k} \binom{k}{j} e^{20j} \Delta_T^{k-j}.$$

The proposition follows from combining this claim with (14) and applying the binomial theorem. The remainder of the proof thus consists of the proof of Claim 2.8. We fix $0 \leq j \leq k$ from now on, and perform the following steps.

Step 1 We find a function $F : \{\ell, m, h\}^{2j+1} \to \mathbb{R}_+$ with the following property. Consider an (M, H)-walk w^- of length $2j$ and let $\tau = (\tau_i(C, \mathrm{w}^-), 0 \le i \le 2j) \in \{\ell, m, h\}^{2j+1}$ be the corresponding type sequence. Then:

$$|\langle \mathrm{w}^- \rangle| \le F(\tau).$$

This reduces the problem of bounding $Z(j, m)$ to a maximization of F over a restricted set that we will define below.

Step 2 We find an explicit $\tau_{\max} \in \{\ell, m, h\}^{2j+1}$ such that $F(\tau_{\max}) \ge F(\tau)$ for any τ that is the type sequence of some w^-. We then compute $F(\tau_{\max})$ explicitly to find an upper bound for $Z(j, k)$.

Step 3 We finish the proof of the claim by bounding $F(\tau_{\max})$.

Let us now present these steps in detail.

Step 1: find a function F. Consider an (M, H)-walk w^- of length $2j$, and let $\tau = (\tau_i, 0 \le i \le j) \in \{\ell, m, h\}^{2j+1}$. Any walk w of length $2k$ whose reduced (M, H)-walk is w^- can be constructed in the following way:

For each $0 \le i \le 2j$ with $\tau_i \in \{m, h\}$, choose $\sigma_i, \delta_i \in \mathbb{Z}_{\ge 0}$; for $\tau_i = \ell$, set $\sigma_i = \delta_i = 0$. For each $0 \le i \le 2j$, we will add σ_i single backtracking steps and δ_i doubly backtracking steps to location w_i^- along the walk. To ensure that the total length of the resulting walk is $2k$, we require that $\sum_{i=0}^{2j}(\sigma_i + 2\delta_i) = k - j$.

At each location $0 \le i \le 2j$, having fixed σ_i and δ_i, the single and double backtracking steps at location i can happen in a total of up to $\binom{\sigma_i + \delta_i}{\sigma_i}$ possible orders. The number of choices for each single backtracking step is the degree of vertex w_i^-, which is at most Δ_T if $\tau_i = h$ and at most $0.95\,\Delta_T$ if $\tau_i = m$. The number of choices for the double backtracking steps is at most the size of the second neighborhood of w_i^-, which we bound by $\Delta_T^{(2)}$. All in all, the number of possible choices for the single and double backtracking steps at position i, for a given choice of σ_i and τ_i, is at most

$$\binom{\sigma_i + \delta_i}{\sigma_i} (\Delta_T^{(2)})^{\delta_i} \, \Delta_T^{\sigma_i} \, (0.95)^{\sigma_i \mathbf{1}_{[\tau_i^- = m]}}.$$

The upshot of the above discussion is this. Given $x = (x_i)_{i=0}^{2j} \in \{\ell, m, h\}^{2j+1}$, define

$$S(x) := \left\{ ((\sigma_i, \delta_i), 0 \le i \le 2j) \in \mathbb{Z}_{\ge 0}^{2(2j+1)} : \begin{array}{c} \sum_{i=0}^{2j}(\sigma_i + 2\delta_i) = (k - j) \\ \text{and} \\ \forall 0 \le i \le 2j, \text{ if } x_i = \ell \text{ then } \sigma_i = \delta_i = 0 \end{array} \right\}$$

and

$$F(x) := \sum_{(\sigma_i, \delta_i)_{i=0}^{2j} \in S(x)} \prod_{i=0}^{2j} \binom{\sigma_i + \delta_i}{\sigma_i} (\Delta_T^{(2)})^{\delta_i} \, \Delta_T^{\sigma_i} \, (0.95)^{\sigma_i \mathbf{1}_{[x_i = m]}}.$$

Then

$$|\langle \mathrm{w}^- \rangle| \le F(\tau).$$

By assumption, T does not contain two adjacent vertices of type h. Therefore, τ cannot contain more than $j + 1$ coordinates equal to h. This allows us to conclude that

$$Z(j, k) \le \max \left\{ F(\tau) \, : \, \tau \in \{\ell, m, h\}^{2j+1}, \sum_{i=0}^{2j} \mathbf{1}_{[\tau_i = h]} \le j + 1 \right\}. \tag{15}$$

Step 2: find the maximizer and compute the maximum.

Inspection reveals that F has the following properties.

(1) for any $x \in \{\ell, m, h\}^{2j+1}$, $F(x)$ is invariant under permutations of the coordinates of x.

(2) Changing some coordinates of x from ℓ to m can only increase $F(x)$. Indeed, such a chance to x can only increase the set $S(x)$ and the combinatorial factors in the sum defining F.

(3) Likewise, changing some coordinates of x from m to h can only increase F.

Recalling (15), we conclude that

$$Z(j,k) \le F(\tau_{\max}) \tag{16}$$

where

$$\tau_{\max} = h^{j+1} m^j.$$

We note in passing that $\tau_{\max}$ is not the type sequence of any (M,H)-walk in T, as it contains consecutive occurences of h. The alternating sequence $hmhmh \ldots mh$, which also maximizes F, could in principle correspond to a (M,H)-walk on some tree. However, working with $\tau_{\max}$ is a bit easier in what follows.

To continue, we find a more convenient expression for $F(\tau_{\max})$. First of all notice that

$$S(\tau_{\max}) := \left\{ (\sigma_i, \delta_i)_{i=0}^{2j} \in \mathbb{Z}_{\ge 0}^{2(2j+1)} : \sum_{i=0}^{2j} (\sigma_i + 2\delta_i) = (k-j) \right\}$$

is the disjoint union of the sets

$$S(s_h, s_m, d) := \left\{ (\sigma_i, \delta_i)_{i=0}^{2j} \in \mathbb{Z}_{\ge 0}^{2\,(2j+1)} : \begin{array}{l} \sum_{i=0}^{j} \sigma_i = s_h, \\ \sum_{i=j+1}^{2j} \sigma_i = s_m, \\ \sum_{i=0}^{2j} \delta_i = d \end{array} \right\}. \tag{17}$$

where (s_h, s_m, d) range over all triples in $\mathbb{Z}_{\ge 0}^3$ with $s_h + s_m + 2d = k - j$. Intuitively, s_h and s_m correspond to the number of single backtracking steps at high and moderate vertices, respectively, that were removed to obtain the (M,H)-walk, whereas d corresponds to double backtracking steps (the intuition is not quite right, because $\tau_{\max}$ does not correspond to a (M,H)-walk, as noted above).

For each element $(\sigma_i, \delta_i)_{i=0}^{2j} \in S(s_h, s_m, d)$, we have

$$\prod_{i=0}^{2j} (\Delta_T^{(2)})^{\delta_i} \Delta_T^{\sigma_i} (0.95)^{\sigma_i \mathbf{1}_{[t_i = m]}} = \Delta_T^{s_h} (0.95\Delta_T)^{s_m} (\Delta_T^{(2)})^d. \tag{18}$$

On the other hand, the sum

$$\sum_{(\sigma_i, \delta_i)_{i=0}^{2j} \in S(s_h, s_m, d)} \prod_{i=0}^{2j} \binom{\sigma_i + \delta_i}{\sigma_i}$$

counts the number of ways one can write a string of the form

$$h\,R_0\,h\,R_2\,\ldots\,h\,R_j\,m\,R_{j+1}\,m\,R_{j+2}\,\ldots\,m\,R_{2j} \tag{19}$$

where

- each R_i is a permutation of $a^{\sigma_i} b^{\delta_i}$, for some choice of $\sigma_i, \delta_i \in \mathbb{Z}_{\ge 0}$ (so there are $\binom{\sigma_i + \delta_i}{\sigma_i}$ choices for R_i once these two numbers is fixed);
- the total number of characters equal to a in the strings $R_0, \ldots, R_j$ is $\sum_{i=0}^{j} \sigma_i = s_h$;
- the total number of characters equal to a in $R_{j+1}, \ldots, R_{2j}$ is $\sum_{i=j+1}^{2j} \sigma_i = s_m$;
- the total number of characters equal to b in $R_0, R_1, \ldots, R_{2j}$ is $\sum_{i=0}^{2j} \delta_i = d$.

We now give an alternative way to count these strings. Notice that such a string has total length $2j + 1 + s_h + s_m + d$, and always begins with 'h'. There are thus

$$\binom{2j + s_h + s_m + d}{d}$$

ways to choose the position of the d's in a string of the above form. Once the d's are removed, one is left with a string of the form

$$h\, a^{\sigma_0}\, h\, a^{\sigma_1}\, \dots\, ha^{\sigma_j}\, ma^{\sigma_{j+1}}\, \dots\, ma^{\sigma_{2j}}$$

with $\sum_{i=0}^{j} \sigma_i = s_h$ and $\sum_{i=j+1}^{2j} \sigma_i = s_m$. There are therefore

$$\binom{j + s_h}{s_h}\binom{j - 1 + s_m}{s_m}$$

total choices for $(\sigma_0, \dots, \sigma_{2j})$. The upshot of this discussion is that

$$\sum_{(\sigma_i, \delta_i)_{i=0}^{2j} \in S(s_h, s_m, d)} \prod_{i=0}^{2j} \binom{\sigma_i + \delta_i}{\sigma_i} = \binom{2j + s_h + s_m + d}{d}\binom{j + s_h}{s_h}\binom{j - 1 + s_m}{s_m}.$$

Combining the above identity with (18) and the definition of $S(s_h, s_m, d)$ in (17), we obtain

$$F(\tau_{\max}) = \sum_{\substack{(s_h, s_m, d) \in \mathbb{Z}_{\geq 0}^3 \\ s_h + s_m + 2d = k - j}} \binom{2j + s_h + s_m + d}{d}\binom{j + s_h}{s_h}\binom{j - 1 + s_m}{s_m} \Delta_T^{s_h}\, (0.95\Delta_T)^{s_m}\, (\Delta_T^{(2)})^d. \quad (20)$$

Step 3: bounds and end of proof. To finish the proof, we find upper bounds for $F(\tau_{\max})$, starting from (20), via a series of simplifications and overestimates.

First, notice that the formula can be bounded by noticing that $k - j - s_h - s_m = 2d$ and $2j + s_h + s_m + d \leq k + j - d \leq 2k$ for all terms in the RHS of (20). Pulling out a factor of Δ_T^{k-j}, we arrive at

$$\frac{F(\tau_{\max})}{\Delta_T^{k-j}} \leq \sum_{\substack{(s_h, s_m, d) \in \mathbb{Z}_{\geq 0}^3 \\ s_h + s_m + 2d = k - j}} \binom{2k}{d}\binom{j + s_h}{s_h}\binom{j - 1 + s_m}{s_m} (0.95)^{s_m} \left(\frac{\Delta_T^{(2)}}{\Delta_T^2}\right)^d.$$

We can further bound the expression, by upper bounding the terms involving d and s_m, as

$$\frac{F(\tau_{\max})}{\Delta_T^{k-j}} \leq \sum_{0 \leq s_h \leq k-j} (k+1)\binom{j + s_h}{s_h}\left[\max_{s_m \in \mathbb{Z}_{\geq 0}} \binom{j - 1 + s_m}{s_m}(0.95)^{s_m}\right]\left[\max_{0 \leq d \leq 2k}\binom{2k}{d}\left(\frac{\Delta_T^{(2)}}{\Delta_T^2}\right)^d\right],$$

where the factor $(k+1)$ is an upper bound on the maximum number of choices for (s_m, d) given s_h.

We now show that the maximum in s_m is at most e^{20j}. Notice that

$$\max_{s_m \in \mathbb{Z}_{\geq 0}} \binom{j - 1 + s}{s_m}(0.95)^{s_m}$$

is achieved at $s_m = s_* := \max\{19\,(j-1) - 1, 0\}$. For $j = 0, 1$, the maximum is achieved at $s_* = 0$ and equals 1, which is not larger than e^{20j}. For $j \geq 2$, we use the general inequality $\binom{r}{w} \leq (er/w)^w$ (valid for integers $0 \leq w \leq r$) and $1 + u \leq e^u$ (valid for all real u) to obtain:

$$\binom{j - 1 + s_*}{s_*}(0.95)^{s_*} \leq \left[\frac{19e}{20}\left(1 + \frac{j-1}{19(j-1) - 1}\right)\right]^{20\,(j-1)-1}$$

$$\leq \exp\left(\frac{19}{20}\,(19\,(j-1) - 1) + j - 1\right) \leq e^{20j}.$$

This gives our desired bound.

As for the maximum in d, we simply note that, by the binomial theorem,

$$
\binom{2k}{d} \left(\frac{\Delta_T^{(2)}}{\Delta_T^2}\right)^d \leq \left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2}\right)^{2k}.
$$

Applying these two bounds, we obtain

$$
\frac{F(\tau_{\max})}{\Delta_T^{k-j}} \leq \sum_{0 \leq s_h \leq k-j} (k+1) \binom{j+s_h}{s_h} e^{20j} \left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2}\right)^{2k}.
$$

We are left with a sum with $k - j + 1 \leq (k+1)$ terms, and the remaining binomial coefficient in our expression satisfies:

$$
\binom{j+s_h}{s_h} = \binom{j+s_h}{j} \leq \binom{k}{j}
$$

for any $0 \leq s_h \leq k - j$. We conclude that

$$
\frac{F(\tau_{\max})}{\Delta_T^{k-j}} \leq (k+1)^2 \binom{k}{j} e^{20j} \left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2}\right)^{2k}.
$$

Combining this with (16), we obtain Claim 2.8 and finish the proof.   □

3. Structural properties

In this section we establish certain structural properties of random labeled trees that hold with high probability. These properties allow us to make efficient use of the path counting techniques developed in Section 2.

We begin with a well-known fact, which shall allow us to understand the typical properties of random labeled trees by instead studying Poisson Bienaymé trees.[3] To state the fact, a little more terminology is needed. A *plane tree*, also called an *ordered rooted tree*, is a rooted tree in which the set of children of each vertex are ordered. The vertices of a plane tree are canonically labeled by strings of positive integers as follows: the root is labeled with the empty string $\varnothing$. Inductively, for a vertex labeled $a = a_1 a_2 \ldots a_k$ and with m children, its children receive labels $a1, a2, \ldots am$ in order. (Here, for a string a and a positive integer i, ai represents the string obtained by concatenating a and i.) In this way, vertices at depth d in the tree are labeled with strings from $\mathbb{N}^d$. The family trees of branching processes are (a particular sort of) random plane trees.

Given a rooted tree T with vertices labelled by $[n]$, the *plane tree corresponding to T* is constructed as follows. First, for each node v of T, order the children of v from left to right in increasing order of vertex label; this endows T with a plane structure. Then remove the original vertex labels.

Lemma 3.1. *Fix $\lambda > 0$ and $n \in \mathbb{N}$. Let $\mathbf{T}_n^\bullet$ be a uniformly random rooted tree with vertices labelled by $[n]$, and let $\mathbf{T}_n^*$ be a Poisson(λ) Bienaymé tree conditioned to have n vertices. Then the random plane tree corresponding to $\mathbf{T}_n^\bullet$ has the same distribution as $\mathbf{T}_n^*$.*

For a proof of the lemma, see [17, Section 5] or [22, Section 7].

In the remainder of the section, we bound the probability that $\mathbf{T}_n$ has certain structural properties by first proving stronger bounds on the probability that a Poisson(1) Bienaymé tree has such properties, then using Lemma 3.1 (among other tools) to transfer those bounds to $\mathbf{T}_n^\bullet$ (and thence to $\mathbf{T}_n$, when needed, by a union bound over the choices of possible root).

---

[3]Bienaymé trees are more commonly called *Galton–Watson trees*; here we use the alternative terminology proposed in [2]. Concretely, a Poisson(λ) Bienaymé tree is the family tree of a branching process with Poisson(λ) offspring distribution.

A second useful tool is the *breadth-first search* (BFS) construction of plane trees. Given a sequence of non-negative integers $d = (d_i, i \geq 0)$, the BFS construction works as follows. First, d_0 is the number of children of the root. Next, $d_1, \ldots, d_{d_0}$ are the numbers of children of the children of the root, ordered from left to right. The construction continues sequentially in this manner; so, for example, the first child of the root has children with $d_{d_0+1}, \ldots, d_{d_0+d_1}$ children, the second has children which themselves have $d_{d_0+d_1+1}, \ldots, d_{d_0+d_1+d_2}$ children, and so on. The construction may continue indefinitely or may halt; if it halts, it does so at step $\sigma = \sigma(d) = 1 + \inf\{m : d_0 + \ldots + d_m \leq m\}$, and in this case the resulting tree has σ nodes.

Given a rooted tree T and $k \in \mathbb{N}$, write $T(k)$ for the set of nodes at distance k from the root of t. Note that the BFS construction imparts a total order to the nodes of the resulting tree, which agrees with the order of the degrees in the sequence d: if $k < l$ then nodes of $T(k)$ precede those of $T(l)$, and within $T(k)$ nodes are ordered from left to right.

The following construction of Bienaymé trees is classical.

Lemma 3.2. *Let* $X = (X_i, i \geq 0)$ *be IID non-negative integer-valued random variables with law* μ. *Then the tree* $\mathbf{T}$ *built by applying the BFS construction to* X *is a Bienaymé*(μ) *tree.*

Let $S_0 = 0$ and, for $k \geq 1$ let $S_k = \sum_{0 \leq i < k}(X_i - 1)$. Then $\sigma = \inf\{k : S_k = -1\}$ is the size (number of nodes) of $\mathbf{T}$, and $(1 + S_k, 0 \leq k \leq \sigma)$ is the *BFS queue length process*, tracking the number of nodes which have been revealed in the breadth-first search process but whose degree in $\mathbf{T}$ is not yet known.

In the remainder of the section we let $X = (X_i, i \geq 0)$ be a sequence of independent Poisson(1) random variables and let $\mathbf{T}$ be the Poisson(1) Bienaymé tree built by applying the BFS construction to X.

Lemma 3.3. *For all* $k, s \in \mathbb{N}$,

$$\mathbf{P}\{\exists i \leq k : |\mathbf{T}(i)| \geq s\} \leq \mathbf{P}\left\{\max_{i \leq (k-1)s} S_i \geq s\right\}.$$

Proof. Let $\tau(0) = 0$ and, for $i \geq 0$, let $\tau(i+1) = \tau(i) + S_{\tau(i)} + 1$. It is straightforward to see, by induction, that for all $i \geq 0$, $\tau(i) = \sum_{0 \leq j < i} |\mathbf{T}(j)|$ and $|\mathbf{T}(i)| = S_{\tau(i)} + 1$.

Now let $I = \inf\{j \geq 0 : |\mathbf{T}(j)| \geq s\}$. Since $|\mathbf{T}(0)| = 1$ and $|\mathbf{T}(j)| < s$ for $0 < j < I$, it follows that $\tau(I) \leq (I-1)(s-1) + 1 \leq (I-1)s$ and $S_{\tau(I)} \geq s$. Thus if $I \leq k$ then there is $j \leq (k-1)s$ such that $S_j \geq s$. The result follows. $\qquad\square$

Proposition 3.4. *Fix positive functions* $m = m(n) = o(n)$ *and* $r = r(n) = o(n^{1/2})$. *If* τ *is any stopping time for* $S = (S_k, k \geq 0)$, *then*

$$\mathbf{P}\{\tau \leq m \mid |\mathbf{T}| = n\} = (1 + o(1))\mathbf{E}\left[S_\tau \mathbf{1}_{[\tau \leq \min(m, |\mathbf{T}|)]} \mathbf{1}_{[S_\tau \leq r]}\right] + \mathbf{P}\{\tau \leq m, S_\tau > r \mid |\mathbf{T}| = n\}$$

Proof. We first write

$$\mathbf{P}\{\tau \leq m, |\mathbf{T}| = n\} = \mathbf{P}\{\tau \leq m, S_\tau \leq r, |\mathbf{T}| = n\} + \mathbf{P}\{\tau \leq m, S_\tau > r, |\mathbf{T}| = n\}. \qquad (21)$$

To bound the first probability on the right, note that since $|\mathbf{T}| = \sigma$,

$$\mathbf{P}\{\tau \leq m, S_\tau \leq r, |\mathbf{T}| = n\} = \mathbf{P}\{\tau \leq m, S_\tau \leq r, \sigma = n\}$$

$$= \sum_{i \leq m}\sum_{s=1}^{r} \mathbf{P}\{\tau = i \leq \sigma, S_\tau = s, \sigma = n\}$$

$$= \sum_{i \leq m}\sum_{s=1}^{r} \mathbf{P}\{\tau = i \leq \sigma, S_i = s\} \mathbf{P}\{\sigma = n \mid \tau = i \leq \sigma, S_i = s\}$$

By Kemperman's formula [21, 6.3],

$$\mathbf{P}\{\sigma = n \mid \tau = i \leq \sigma, S_i = s\} = \frac{s}{n}\mathbf{P}\{S_{n-i} = -s\}$$

so

$$\mathbf{P}\left\{\tau \leq m, S_\tau \leq r, |\mathbf{T}| = n\right\} = \sum_{i \leq m} \sum_{s=1}^{r} \mathbf{P}\left\{\tau = i \leq \sigma, S_i = s\right\} \cdot \frac{s}{n}\mathbf{P}\left\{S_{n-i} = -s\right\} .$$

Uniformly over i and s with $i \leq m = o(n)$ and $s \leq r = o(n^{1/2})$, by the local central limit theorem $\mathbf{P}\left\{S_{n-i} = -s\right\} = (1 + o(1))\mathbf{P}\left\{S_n = -1\right\}$. Combined with Kemperman's formula this gives that

$$\frac{s}{n}\mathbf{P}\left\{S_{n-i} = -s\right\} = (1 + o(1))\frac{s}{n}\mathbf{P}\left\{S_n = -1\right\} = (1 + o(1))s\mathbf{P}\left\{|\mathbf{T}| = n\right\} .$$

It follows that

$$\mathbf{P}\left\{\tau \leq m, S_\tau \leq r, |\mathbf{T}| = n\right\} = (1 + o(1))\sum_{i \leq m}\sum_{s=1}^{r}\mathbf{P}\left\{\tau = i \leq \sigma, S_i = s\right\} \cdot s\mathbf{P}\left\{|\mathbf{T}| = n\right\}$$

$$= (1 + o(1))\mathbf{E}\left[S_\tau\mathbf{1}_{[\tau \leq \min(m,\sigma)]}\mathbf{1}_{[S_\tau \leq r]}\right] \cdot \mathbf{P}\left\{|\mathbf{T}| = n\right\} .$$

Using this identity, the proposition now follows from (21) and Bayes formula.   □

The next lemma controls neighbourhood sizes near a fixed vertex of $\mathbf{T}$; the following lemma uses the first to deduce additional information about the local structure of $\mathbf{T}$.

Lemma 3.5. *Let $\mathbf{T}$ be a Poisson(1) Bienaymé tree. Then for n sufficiently large, the following bounds both hold. First,*

$$\mathbf{P}\left\{\exists j \leq 4 : |\mathbf{T}(j)| \geq 37\log n\right\} \leq n^{-4}.$$

Second,

$$\mathbf{P}\left\{\exists j \leq \log n : |\mathbf{T}(j)| \geq \lfloor\log^2 n\rfloor\right\} \leq 2\exp(-(\log^2 n)/5) . \tag{22}$$

Proof. By Lemma 3.3, we have

$$\mathbf{P}\left\{\exists j \leq k : |\mathbf{T}(j)| \geq s\right\} \leq \mathbf{P}\left\{\max_{i \leq (k-1)s} S_i \geq s\right\}$$

where $(S_k, k \geq 0)$ is a random walk with steps $(X_i - 1, i \geq 1)$. Since a Poisson(λ) random variable has a median in $[\lambda - \ln 2, \lambda + 1/3]$ (see [11]) and $X_1 + \cdots + X_i$ is Poisson(i)-distributed, by Lévy's reflection principle [25, Theorem 1.4.13], for $s \geq 1$ we have

$$\mathbf{P}\left\{\max_{i \leq (k-1)s} S_i \geq s\right\} \leq 2\mathbf{P}\left\{S_{(k-1)s} \geq s - 1\right\} .$$

We now use a standard Poisson tail bound [20, Lemma 1.2]: for $\varepsilon > 0$,

$$\mathbf{P}\left\{\mathrm{Poi}(\lambda) > (1 + \varepsilon)\lambda\right\} \leq \exp(-\lambda((1 + \varepsilon)\log(1 + \varepsilon) - \varepsilon)) . \tag{23}$$

If $\varepsilon \leq 1/2$ then this upper bound is at most $\exp(-\varepsilon^2\lambda/4)$. If $k \geq 3$, so that $s - 1 \leq (k-1)s/2$, then using that $S_i + i$ is Poisson(i)-distributed, this implies that

$$\mathbf{P}\left\{S_{(k-1)s} \geq s - 1\right\} \leq 2\mathbf{P}\left\{\mathrm{Poisson}((k-1)s) \geq 3(k-1)s/2\right\} \leq \exp\left(-\frac{1}{4}\frac{(s-1)^2}{(k-1)s}\right) .$$

Taking $k = 4$ and $s = 37\log n$ now yields the first bound, and taking $k = \lfloor\log n\rfloor$ and $s = \lfloor\log^2 n\rfloor$ yields the second bound.   □

Lemma 3.6. *Let $\mathbf{T}$ be a Poisson(1) Bienaymé tree. Then for n sufficiently large, the following bounds both hold. First, writing $\deg(w) = \deg_{\mathbf{T}}(w)$ for vertices w of $\mathbf{T}$, and setting*

$$K^* := \min(k \in \mathbb{N} : \deg(w) < \log^{1/5} n \text{ for all } w \in \mathbf{T}(2k) \cup \mathbf{T}(2k+1)),$$

we have

$$\mathbf{P}\left\{K^* \geq 15(\log^{4/5} n)/\log\log n\right\} < 2n^{-4} .$$

Second, writing

$$Z = \{v \in \mathbf{T} : |v| < 2K^*, \deg(v) \geq \log^{1/5} n\},$$

we have

$$\mathbf{P}\left\{\sum_{v\in Z}\deg(v)\geq 3\log n\right\}< 4n^{-4},$$

and

$$\mathbf{P}\left\{\sum_{v\in Z}|\{w\in[n]:\mathrm{dist}(v,w)\leq 3\}|\geq 150(\log n)(\log\log n)^2\right\}< 7n^{-4}.$$

Proof. Let $\tau_0 = 0$ and for $i \geq 0$ let $\tau_{i+1} = \tau_i + S_{\tau_i} + 1$. Note that $\tau_i - 1$ is a stopping time for the filtration generated by X. Writing $f(\mathbf{T})$ for the height of $\mathbf{T}$, then for $0 \leq k < f(\mathbf{T})$, the k'th generation $\mathbf{T}(k)$ of $\mathbf{T}$ is constructed by BFS between times τ_k and τ_{k+1}, and $f(\mathbf{T}) = \inf(k > 0 : \tau_k = \tau_{k-1}) - 1$ is the height of $\mathbf{T}$. In particular, for $0 \leq k < f(\mathbf{T})$, the degrees of the vertices of $\mathbf{T}(k)$ are $X_{\tau_k}, \ldots, X_{\tau_{k+1}-1}$.

For $k \geq 0$ let

$$E_k = \{\max(X_i, \tau_{2k} \leq i \leq \min(\tau_{2k+2} - 1, \tau_{2k} + 2\lfloor \log^2 n \rfloor)) < \log^{1/5} n - 1\}.$$

Since $\tau_{2k} - 1$ is a stopping time, by the Markov property we have

$$\mathbf{P}\left\{E_k^c \mid X_i, 0 \leq i \leq \tau_{2k} - 1\right\}$$

$$= \mathbf{P}\left\{\max(X_j, 1 \leq j \leq \min(\tau_2 - 1, 1 + 2\lfloor \log^2 n \rfloor)) \geq \log^{1/5} n - 1\right\}$$

$$\leq \mathbf{P}\left\{\max(X_j, 1 \leq j \leq 1 + 2\lfloor \log^2 n \rfloor \rfloor) \geq \log^{1/5} n - 1\right\}$$

$$\leq \frac{4\log^2 n}{(\lceil \log^{1/5} n - 1\rceil)!}$$

$$\leq 4\exp\left((11/5)\log\log n - (\log^{1/5} n - 1)\log(\log^{1/5} n - 1)\right)$$

where in the third inequality we have used the bound $m! > (m/e)^m$.

Now let $K = \min(k \geq 0 : E_k \text{ occurs})$. Setting $k^* = C(\log^{4/5} n)/\log\log n$ for C sufficiently large ($C = 15$ suffices), it follows from the above bound and the Markov property that that for all large n,

$$\mathbf{P}\{K \geq k^*\} \leq \left(4\exp\left((11/5)\log\log n - (\log^{1/5} n - 1)\log(\log^{1/5} n - 1)\right)\right)^{k^*}$$

$$< n^{-4}.$$

Since the degree of a vertex in $\mathbf{T}$ is at most its number of children plus one, on the event that $|\mathbf{T}(j)| \leq \log^2 n$ for all $j \leq \lfloor \log n \rfloor$, if $K < k^*$ the maximum degree in $\mathbf{T}(2K) \cup \mathbf{T}(2K+1)$ is less than $\log^{1/5} n$, so also $K^* < k^*$. It thus follows by the above tail bound on K together with (22) that

$$\mathbf{P}\{K^* \geq k^*\} \leq 2\exp(-(\log^2 n)/5) + n^{-4},$$

which establishes the first bound.

Next, for $i \geq 1$ write $X_i^+ = (X_i + 1)\mathbf{1}_{[X_i \geq \log^{1/5} n - 1]}$. On the events that $|\mathbf{T}(j)| < \log^2 n$ for all $j \leq \lfloor \log n \rfloor$ and that $K^* \leq 15\log^{4/5} n/\log\log n$, which have combined probability at least $1 - 2n^{-4}$ for n large, we have

$$\sum_{v\in Z}\deg(v) \leq \sum_{i < \lfloor \log^2 n \rfloor \cdot 30\log^{4/5} n/\log\log n}(X_i^+ + 1) \leq \sum_{i \leq \log^{14/5} n}(X_i^+ + 1).$$

To control the latter sum, write $S^+ := \sum_{i \leq \log^{14/5} n} X_i^+$. For $t > 0$, a Chernoff bound gives

$$\mathbf{P}\left\{S^+ \geq M\right\} \leq e^{-tM}\mathbf{E}\left[e^{tS^+}\right]$$

$$= e^{-tM}(\mathbf{E}e^{tX_1^+})^{\log^{14/5} n}.$$

For any $j \geq 2$ it holds that $\mathbf{E}\left[\exp(t(X_1 + 1)\mathbf{1}_{[X_1 \geq j]})\right] \leq 1 + 2e^{tj}/j!$, so taking $t = 2$ and $j = \lceil \log^{1/5} n - 1 \rceil$ the above bound gives

$$\mathbf{P}\left\{S^+ \geq M\right\} \leq e^{-2M}\left(1 + \frac{2e^{2\lceil \log^{1/5} n - 1 \rceil}}{(\lceil \log^{1/5} n - 1 \rceil)!}\right)^{\log^{14/5} n}$$

Using that $m! \geq (m/e)^m$ and that $1 + x \leq e^x$ for $x > 0$, it follows that

$$\left(1 + \frac{2e^{\lceil \log^{1/5} n - 1 \rceil}}{(\lceil \log^{1/5} n - 1 \rceil)!}\right)^{\log^{14/5} n} < 2$$

for n large, so for n large we have

$$\mathbf{P}\left\{S^+ \geq M\right\} < 2e^{-2M}.$$

Taking $M = 2\log n$, the bound on $\mathbf{P}\left\{\sum_{v \in Z} \deg(v) \geq 3\log n\right\}$ follows.

For the third bound, we claim that for n large, for any $K \in \mathbb{N}$ and $C \geq 7\log\log n$ with $KC\log C \geq 8\log n$,

$$\mathbf{P}\left\{\exists S \subset [\lfloor \log^3 n \rfloor] : |S| \leq K, \sum_{i \in S} X_i \geq CK\right\} \leq n^{-4}. \tag{24}$$

Assuming this, we prove the third bound as follows. Let S be the set of times at which neighbours of vertices in Z (including parents of vertices of Z) are explored in the breadth-first search process. On the event G_1 that $|\mathbf{T}(j)| < \log^2 n$ for all $j \leq \lfloor \log n \rfloor$ and that $K^* < 15\log^{2/3} n/\log\log n$, we have $S \subset [\lfloor \log^3 n \rfloor]$. On the event G_2 that $\sum_{v \in Z} \deg(v) < 3\log n$, this set S also has size less than $3\log n$. It thus follows from the preceding bounds of the lemma and from (24) applied with $K = \lfloor 3\log n \rfloor$ and $C = 7\log\log n$ that

$$\mathbf{P}\left\{G_1, G_2, \sum_{i \in S} X_i \geq 3\log n \cdot 7\log\log n\right\} \leq n^{-4}.$$

Let G_3 be the event that $\sum_{i \in S} X_i < 3\log n \cdot 7\log\log n$. On $G_1 \cap G_2 \cap G_3$, the second neighbourhood of Z has size at most $3\log n \cdot (7\log\log n + 1)$. Moreover, writing S_2 for the set of times at which vertices in the second neighbourhood of Z are explored, on G_1 we have $S_2 \subset [\lfloor \log^3 n \rfloor]$. It follows from (24) applied with $K = \lfloor 3\log n(7\log\log n + 1) \rfloor$ and $C = 7\log\log n$ that

$$\mathbf{P}\left\{G_1, G_2, G_3, \sum_{i \in S_2} X_i \geq 3\log n \cdot (7\log\log n + 1)7\log\log n\right\} \leq n^{-4}.$$

For n large, if $\sum_{i \in S_2} X_i < 3\log n \cdot (7\log\log n + 1)7\log\log n$ and G_2 and G_3 both occur then

$$\sum_{v \in Z} |\{w \in [n] : \operatorname{dist}(v, w) \leq 3\}| < 150(\log n)(\log\log n)^2,$$

so the above bounds yield that

$$\mathbf{P}\left\{\sum_{v \in Z} |\{w \in [n] : \operatorname{dist}(v, w) \leq 3\}| < 150(\log n)(\log\log n)^2\right\} < 2\exp(-(\log^2 n)/5) + 6n^{-4}$$

$$< 7n^{-4},$$

the last bound holding for n large, as required.

It remains to prove (24); for this write $L = \lfloor \log^3 n \rfloor$. It suffices to consider sets $S \subset [L]$ of size exactly K, since the X_i are non-negative. By a union bound, it then follows that

$$\mathbf{P}\left\{\exists S \subset [\lfloor \log^3 n \rfloor] : |S| \le K, \sum_{i \in S} X_i \ge CK\right\}$$

$$\le \binom{L}{K} \mathbf{P}\left\{\mathrm{Poisson}(K) \ge CK\right\}$$

$$\le \exp\left(K \log L - K(C \log C - (C-1))\right),$$

where in the second inequality we have used that $\binom{L}{K} \le L^K$ and the Poisson upper tail bound (23).

Since $C > 7 \log \log n$ and $\log L \le 3 \log \log n$, for n large we have $C \log C - (C-1) - \log L > C \log C / 2$ and so the final bound is at most $e^{-(KC \log C)/2} < n^{-4}$. □

The next corollary transfers the above probability bounds from $\mathbf{T}$ to $\mathbf{T}_n$. For a vertex u of $\mathbf{T}_n$ and a positive integer j, define $\Gamma_j(u, \mathbf{T}_n)$ (or $\Gamma_j(u)$ for simplicity) as the set of vertices v with $\mathrm{dist}_{\mathbf{T}_n}(u, v) = j$ where we recall from the introduction that $\mathrm{dist}_{\mathbf{T}_n}(u, v)$ is the distance of u and v in the tree. Similarly, $\Gamma_{\le j}(u, \mathbf{T}_n)$ (or simply $\Gamma_{\le j}(u)$) denotes the set of all vertices v with $\mathrm{dist}_{\mathbf{T}_n}(u, v) \le j$.

Corollary 3.7. *For $n \ge 1$ let $\mathbf{T}_n$ be a uniformly random tree with vertices labeled by $[n]$. Then there exists a constant $C > 0$ such that the following bounds hold for all n.*

$$\mathbf{P}\left\{\exists u \in [n], \exists j \le \log n : |\Gamma_j(u)| \ge \lfloor \log^2 n \rfloor\right\} \le C n^{5/2} \exp(-(\log^2 n)/5). \qquad (25)$$

Next, for $u \in [n]$, writing

$$K^*(u) := \min(k \in \mathbb{N} : \deg(v) < \log^{1/5} n \text{ for all } v \in N_u(2k) \cup N_u(2k+1)),$$

then

$$\mathbf{P}\left\{\max(K^*(u), u \in [n]) \ge 15(\log^{4/5} n)/\log \log n\right\} \le C n^{-3/2}.$$

Finally, for $u \in [n]$ writing

$$Z(u) = \{v \in \mathbf{T}_n : \mathrm{dist}(u, v) < 2K^*(u), \deg(v) \ge \log^{1/5} n\},$$

then

$$\mathbf{P}\left\{\max_{u \in [n]} \sum_{v \in Z(u)} \deg(v) \ge 3 \log n\right\} \le C n^{-3/2},$$

and

$$\mathbf{P}\left\{\max_{u \in [n]} \sum_{v \in Z(u)} |\Gamma_{\le 3}(v)| \ge 150(\log n)(\log \log n)^2\right\} < C n^{-3/2}$$

Proof. We continue to take $\mathbf{T}$ as above. Recalling that $\sigma = \inf(k : S_k = -1)$, then

$$\mathbf{P}\left\{|\mathbf{T}| = n\right\} = \mathbf{P}\left\{\sigma = n\right\} = \frac{1}{n} \mathbf{P}\left\{X_1 + \ldots + X_n = n - 1\right\} = \frac{1 + o(1)}{\sqrt{2\pi} n^{3/2}}, \qquad (26)$$

the first identity holding by another application of Kemperman's formula and the second by the local central limit theorem (or by Stirling's formula). Combining this fact with Lemma 3.1, it follows that for any graph property $\mathscr{P}$,

$$\mathbf{P}\left\{\mathbf{T}_n^{\bullet} \text{ has property } \mathscr{P}\right\} = \mathbf{P}\left\{\mathbf{T} \text{ has property } \mathscr{P} \mid |\mathbf{T}| = n\right\}$$

$$\le (1 + o(1))\sqrt{2\pi} n^{3/2} \cdot \mathbf{P}\left\{\mathbf{T} \text{ has property } \mathscr{P}\right\}.$$

Recall that $\mathbf{T}_n^\bullet$ is a uniformly random rooted tree with vertices labeled by $[n]$. We may generate $\mathbf{T}_n^\bullet$ from $\mathbf{T}_n$ by choosing a root ρ uniformly at random, so it follows by a union bound and the above asymptotic inequality that

$$\mathbf{P}\left\{\exists u \in [n], \exists j \leq \log n : |\Gamma_j(u)| \geq \lfloor \log^2 n \rfloor\right\}$$
$$\leq n\mathbf{P}\left\{\exists j \leq \log n : |\mathbf{T}_n^\bullet(j)| \geq \lfloor \log^2 n \rfloor\right\}$$
$$\leq (1 + o(1))\sqrt{2\pi}n^{5/2}\mathbf{P}\left\{\exists j \leq \log n : |\mathbf{T}(j)| \geq \lfloor \log^2 n \rfloor\right\}.$$

The first bound of the corollary now follows from the second bound of Lemma 3.5. The other bounds of the corollary follow from Lemma 3.6 in an essentially identical fashion.   □

We also record the following tail bound on the largest degree in $\mathbf{T}_n$, which will be used, when proving the main theorem, to convert probability tail bounds to expectation bounds.

Fact 3.8. *For all n sufficiently large, $\mathbf{P}\left\{\Delta_{\mathbf{T}_n} \geq 2\log n\right\} \leq n^{-2}$.*

Proof. An argument just as in the above corollary shows that

$$\mathbf{P}\left\{\Delta_{\mathbf{T}_n} \geq 2\log n\right\} \leq (1 + o(1))\sqrt{2\pi}n^{5/2}\mathbf{P}\left\{|\mathbf{T}(1)| \geq 2\log n\right\}$$
$$= (1 + o(1))\sqrt{2\pi}n^{5/2}\mathbf{P}\left\{\text{Poisson}(1) \geq 2\log n\right\} \leq n^{-2},$$

the last bound holding for n large.   □

We also need to control the number of vertices of high degree contained in a small neighbourhood. For this, we use Proposition 3.4.

Lemma 3.9. *The probability that $\mathbf{T}_n$ contains two adjacent nodes of degree $\geq 0.9\log n/\log\log n$ is at most $n^{-0.8+o(1)}$.*

Proof. Set $h := 0.9\log n/\log\log n$. In the proof we will use the following estimates,

$$\forall i \geq 0 : \mathbf{P}\{X_i \geq h - 1\} \leq n^{-0.9+o(1)} \text{ and } \mathbf{P}\{X_i \geq \log n\} = n^{-\omega(1)}, \tag{27}$$

which both follow from the Poisson tail bound (23).

Now let E denote the event where there are two adjacent nodes in $\mathbf{T}_n$ with degrees $\geq h$. Recall that the random rooted n-vertex tree $\mathbf{T}_n^\bullet$ can be obtained from $\mathbf{T}_n$ by picking a root uniformly at random, so

$$\frac{\mathbf{P}\{E\}}{n} \leq \mathbf{P}\left\{\deg_{\mathbf{T}_n^\bullet}(\text{root}) \geq h \text{ and } \max_{v \in \mathbf{T}_n^\bullet(1)}\deg_{\mathbf{T}_n^\bullet}(v) \geq h\right\} = \mathbf{P}\{F \mid |\mathbf{T}| = n\},$$

where

$$F := \left\{\deg_{\mathbf{T}}(\text{root}) \geq h \text{ and } \max_{v \in \mathbf{T}(1)}\deg_{\mathbf{T}}(v) \geq h\right\} = \{X_0 \geq h \text{ and } \exists 1 \leq i \leq X_0 : X_i \geq h - 1\}.$$

Therefore, our goal is to show that $\mathbf{P}\{F \mid |\mathbf{T}| = n\} \leq n^{-1.8+o(1)}$. To this end, consider the events

$$G_0 := \{\exists 0 \leq i \leq \log n : X_i \geq \log n\}$$

and

$$G_1 := \{X_0 \in [h, \log n] \text{ and } \exists 1 \leq i \leq \log n : X_i \in [h - 1, \log n]\}.$$

Then $F \subset G_0 \cup G_1$. Estimate (27) and a union bound imply that G_0 has superpolynomially small probability in n if we do not condition on $\mathbf{P}\{|\mathbf{T}| = n\}$. Since $\mathbf{P}\{|\mathbf{T}| = n\} \asymp n^{-3/2}$ by (26), this also holds conditionally. It thus suffices to show that $\mathbf{P}\{G_1 \mid |\mathbf{T}| = n\} \leq n^{-1.8+o(1)}$, which we do via Proposition 3.4.

Define a stopping time τ for the $(X_i)_{i \geq 0}$ (or equivalently, for $(S_i)_{i \geq 0}$) as follows. If $X_0 \notin [h, \log n]$, or if $X_i \notin [h, \log n]$ for all $1 \leq i \leq n$, we set $\tau = n$. Otherwise, we let τ be the

smallest index $1 \leq i \leq n$ with $X_i \in [h, \log n]$. Notice that $G_1 = \{\tau \leq \log n\}$. Applying Proposition 3.4 with $d = (\log n + 1)^2$ and $m = \log n$, we obtain

$$\mathbf{P}\{G_1 \mid |\mathbf{T}| = n\} = (1 + o(1))\mathbf{E}\left[S_\tau \mathbf{1}_{[S_\tau \leq d, \tau \leq \min(\log n, |\mathbf{T}|)]}\right] + \mathbf{P}\{\tau \leq \log n, S_\tau > d \mid |\mathbf{T}| = n\}$$
$$= O(\log^2 n)\mathbf{P}\{\tau \leq \log n\} + \mathbf{P}\{\tau \leq \log n, S_\tau > (\log n + 1)^2 \mid |\mathbf{T}| = n\}.$$

The first term on the right is

$$O(\log^2 n)\,\mathbf{P}\{\tau \leq \log n\} = O(\log^2 n)\,\mathbf{P}\{X_0 \geq h \text{ and } \exists 1 \leq i \leq \log n,\, X_i \geq h - 1\} \leq n^{-1.8 + o(1)}$$

by (27) and a union bound in i.

For the second term, we note that if $\tau \leq \log n$ but $S_\tau > (\log n + 1)^2$, there must be $0 \leq i \leq \log n$ such that $X_i \geq \log n$. By (27) and a union bound, the probability of this event without the conditioning on $|\mathbf{T}| = n$ is $n^{-\omega(1)}$, and the conditioning does not change this because $\mathbf{P}\{|\mathbf{T}| = n\} \asymp n^{-3/2}$ by (26). $\qquad \square$

We close the section by proving tail bounds on the number of high-degree vertices in $\mathbf{T}_n^\bullet$. Our tool for doing so is the following distributional identity for the numbers of children of vertices in random rooted labeled trees, which is an immediate consequence of Lemmas 3.1 and 3.2.

Corollary 3.10. *Let $\mathbf{T}_n^\bullet$ be a uniformly random rooted tree with vertices labeled by $[n]$, and for $1 \leq i \leq n$ let c_i be the number of children of i in $\mathbf{T}_n^\bullet$. Then $(c_1, \ldots, c_n)$ has the same distribution as $(P_1, \ldots, P_n)$, where $P_1, \ldots, P_n$ are independent Poisson(1) random variables conditioned to satisfy $\sum_{i=1}^n P_i = n - 1$.*

Corollary 3.11. *Let $\mathbf{T}_n^\bullet$ be a uniformly random rooted tree with vertices labeled by $[n]$. Fix any $k \in \mathbb{N}$, and let $D_{\geq k}(n)$ be the number of vertices of $\mathbf{T}_n^\bullet$ with at least k children. Then for $t > 0$*

$$\mathbf{P}\{D_{\geq k}(n) \leq \mathbf{E}D_{\geq k}(n) - t\} \leq \exp(-t^2/(2\mathbf{E}D_{\geq k}(n))),$$

and uniformly over $1 \leq k \leq \lceil \log n \rceil$,

$$\mathbf{E}D_{\geq k}(n) \geq (1 - o(1))\frac{n}{k!}.$$

Proof. The vector $(c_1, \ldots, c_n)$ of numbers of children of the vertices of $\mathbf{T}_n^\bullet$ is multinomially distributed by Corollary 3.10, so its entries are negatively associated (see [13]). The lower tail bound then follows from [13, Proposition 5] and standard binomial tail estimates. For the expectation bound, let $(X_i, i \geq 1)$ be independent Poisson(1) random variables; by Corollary 3.10,

$$\mathbf{P}\{c_1 = k\} = \mathbf{P}\{X_1 = k \mid X_1 + \ldots + X_n = n - 1\}$$
$$= \mathbf{P}\{X_1 = k\}\frac{\mathbf{P}\{X_2 + \ldots + X_n = n - 1 - k\}}{\mathbf{P}\{X_1 + \ldots + X_n = n - 1\}}$$
$$= \frac{1}{k!}\frac{\mathbf{P}\{X_2 + \ldots + X_n = n - 1 - k\}}{\mathbf{P}\{X_1 + \ldots + X_n = n - 1\}}.$$

Since

$$\mathbf{E}D_{\geq k}(n) = n\mathbf{P}\{c_1 \geq k\} > n\mathbf{P}\{c_1 = k\},$$

the result follows by the local central limit theorem. $\qquad \square$

Corollary 3.12. *Let $\mathbf{T}_n^\bullet$ be a uniformly random rooted tree with vertices labeled by $[n]$, and let $a^* = a^*(n) = \max(m \in \mathbb{N} : m! \leq n)$. Then $\mathbf{E}\sqrt{\Delta_{\mathbf{T}_n^\bullet}} \leq a^*(n) + o(1)$ as $n \to \infty$.*

Proof. Note that $a^* = (1 + o(1))(\log n)/(\log \log n)$. Recycling the computation from the previous lemma, uniformly over $a^* \leq k \leq \lceil \log n \rceil$ we have

$$\mathbf{P}\{c_1 = k\} = (1 + o(1))\frac{1}{k!} \leq (1 + o(1))\frac{1}{n}\left(\frac{\log \log n}{\log n}\right)^{k - a^* - 1},$$

so writing $D_k(n)$ for the number of vertices of $\mathbf{T}_n^\bullet$ with exactly k children, we have

$$\mathbf{E}D_k(n) \leq (1 + o(1))\left(\frac{\log\log n}{\log n}\right)^{k-a^*-1} ,$$

uniformly over k in this range.

Since $\Delta_{\mathbf{T}_n^\bullet}$ is at most one greater than $\max(k : D_k(n) > 0)$, it follows that

$$\mathbf{E}\sqrt{\Delta_{\mathbf{T}_n^\bullet}} \leq \sqrt{a^* + 3} + \sqrt{1 + \log n}\,\mathbf{P}\left\{a^* + 3 \leq \Delta_{\mathbf{T}_n^\bullet} < \log n\right\} + \sqrt{n}\mathbf{P}\left\{\Delta_{\mathbf{T}_n^\bullet} \geq \log n\right\} .$$

To bound the second term, we use that

$$\mathbf{P}\left\{a^* + 3 \leq \Delta_{\mathbf{T}_n^\bullet} < \log n\right\} \leq \sum_{k=a^*+2}^{\lceil \log n\rceil - 1} \mathbf{P}\left\{D_k(n) > 0\right\}$$

$$\leq \sum_{k=a^*+2}^{\lceil \log n\rceil - 1} \mathbf{E}D_k(n) \leq (1 + o(1))\left(\frac{\log\log n}{\log n}\right) ,$$

so

$$\sqrt{1 + \log n}\,\mathbf{P}\left\{a^* + 3 \leq \Delta_{\mathbf{T}_n^\bullet} < \log n\right\} = o(1) .$$

To bound the third term, we use Corollary 3.10 to deduce that

$$\mathbf{P}\left\{c_1 \geq \lceil \log n\rceil\right\} \leq \frac{\mathbf{P}\left\{X_1 \geq \lceil \log n\rceil\right\}}{\mathbf{P}\left\{X_1 + \ldots + X_{n-1} = n - 1\right\}} ,$$

where, like above, $X_1, \ldots, X_n$ are independent and Poisson(1)-distributed. The numerator on the right-hand side is $\frac{1+o(1)}{(\lceil\log n\rceil)!}$, which decays faster than polynomially, and the denominator is $\Theta(n^{-1/2})$; so

$$\sqrt{n}\mathbf{P}\left\{\Delta_{\mathbf{T}_n^\bullet} \geq \log n\right\} \leq n^{3/2}\mathbf{P}\left\{c_1 \geq \lceil\log n\rceil\right\} = o(1) .$$

Combining the above bounds, we conclude that

$$\mathbf{E}\sqrt{\Delta_{\mathbf{T}_n^\bullet}} \leq \sqrt{a^* + 3} + o(1) = \sqrt{a^*} + o(1) ,$$

the last identity holding since $a^* = a^*(n) \to \infty$ as $n \to \infty$. The result follows since $\Delta_{\mathbf{T}_n}$ and $\Delta_{\mathbf{T}_n^\bullet}$ have the same distribution. □

4. **Tree surgery**

At this point, we already have some information about the structure of the uniformly random tree $\mathbf{T}_n$. In particular, we can show that $\mathbf{T}_n$ usually satisfies the properties $\mathbf{N1} - \mathbf{N3}$ described in the proof overview, in Section 1.2. Recall from that section that for an n-vertex tree $T = (V, E)$ and $v \in V$, the cluster $C_T(v)$ is the set of all vertices $w \in V$ such that the unique path from v to w does not contain two consecutive vertices with degree $< (\log n)^{1/5}$. We note right away the following fact.

Fact 4.1 (Proof omitted). *"$w \in C_T(v)$" is an equivalence relation over V, so that V is a disjoint union of clusters. Moreover, each cluster is a connected subset of T.*

Definition 4.2 (Typical trees). *A n-vertex tree $T = (V, E)$ is said to be* typical *if*

    N1 *The maximum degree of T is at least $0.99(\log n/\log\log n)$.*
    N2 *For all $x \in V$, $\sum_{y\in C_T(x)} \deg_T(y)\mathbf{1}_{[\deg_T(y)\geq(\log n)^{1/5}]} \leq 3\log n$.*
    N3 *There are no adjacent nodes that both have degree $\geq 0.9\log n/\log\log n$.*

We show in Section 4.1 that this terminology makes sense — a uniformly random tree is typical with high probability — and is good enough to control second neighborhood sizes. However, typicality by itself does not suffice for our path-counting arguments to work. A stronger notion is needed for this purpose.

Definition 4.3 (Nice trees). *A n-vertex tree $\tilde{T} = (\tilde{V}, \tilde{E})$ is said to be* nice *if it is typical and satisfies the following additional property:*

N4 *For all $x \in \tilde{V}$, $C_{\tilde{T}}(x)$ contains less than $9 \log \log n$ vertices of degree $\geq (\log n)^{1/5}$.*

Thus a nice tree is a typical tree where there are very few nodes of (relatively) high degree in any cluster. The main result of this section is that any upper bound on the top eigenvalue of all nice trees extends to all typical trees. The proof of the next lemma is given in Section 4.3.

Lemma 4.4. *For $n > 10$, for any typical n-vertex tree $T = (V, E)$, there exists a nice n-vertex tree $\tilde{T}$ with $\lambda_1(T) \leq \lambda_1(\tilde{T})$ and with $\Delta_{\tilde{T}} = \Delta_T$.*

To prove Lemma 4.4 we use the rewiring result, Lemma 1.5, from Section 1.2, to perform a sequence of rewirings which can only increase the top eigenvalue. Moreover, Section 4.2 establishes that "good rewirings" preserve the typicality and and maximum degree of the tree, while decreasing the number of high-degree vertices in a cluster. Lemma 4.4 will follow from iterating "good rewirings" until they are no longer possible, which will be shown to imply that the resulting tree is nice.

4.1. **Basic facts on typical trees.** We prove here two facts pertaining to typical trees. The first one is that uniformly random trees are usually typical.

Lemma 4.5. *The uniformly random tree $\mathbf{T}_n$ is typical (i.e., satisfies Definition 4.2) with probability $1 - n^{-0.8 + o(1)}$.*

Proof. It suffices to show that $\mathbf{T}_n$ satisfies each property **N1**-**N3** with probability $\geq 1 - n^{-0.8+o(1)}$.

For property **N1**, it follows from results of Moon [19, Theorem 2 and Lemma 4] that $\Delta_{\mathbf{T}_n} \geq 0.99 \, (\log n / \log \log n)$ with probability $1 - e^{-(1+o(1))n^{0.01}}$.

For property **N3**, the desired probability bound is given by Lemma 3.9.

Property **N2** can be dealt with via Corollary 3.7. For each $u \in [n]$, let

$$K^*(u) := \min(k \in \mathbb{N} : \deg(v) < \log^{1/5} n \text{ for all } v \in N_u(2k) \cup N_u(2k+1))$$

and

$$Z(u) = \{v \in \mathbf{T}_n : \text{dist}(u,v) < 2K^*(u), \deg(v) \geq \log^{1/5} n\}$$

be as in the corollary. By definition of $K^*(u)$, all vertices in $C_{\mathbf{T}_n}(u)$ lie within distance less than $2K^*(u)$ from u. As a consequence, all nodes in $v \in C_{\mathbf{T}_n}(u)$ with degree $\det_{\mathbf{T}_n}(v) \geq (\log n)^{1/5}$ lie in the set $Z(u)$. It follows that

$$\max_{u \in [n]} \sum_{v \in C_{\mathbf{T}_n}(u)} \deg_{\mathbf{T}_n}(v) \mathbf{1}_{[\deg_{\mathbf{T}_n}(v) \geq (\log n)^{1/5}]} \leq \max_{u \in [n]} \sum_{v \in Z(u)} \deg_{\mathbf{T}_n}(u).$$

Corollary 3.7 implies that the right-hand side is bounded by $3 \log n$, with probability at least $1 - C n^{-3/2}$. Therefore, **N2** holds with at least this probability. $\square$

Our second result on typical trees is that their second neighborhoods can never be two large.

Proposition 4.6. *In a typical tree with maximum degree $\Delta_T \geq 0.99 \, (\log n / \log \log n)$, the second neighborhood of any vertex has size at most*

$$L \Delta_T^{6/5} (\log \Delta_T)^{1/5},$$

where $L > 0$ is a universal constant.

Proof. Fix a vertex $x \in V(T)$. All neighbors of x with degree at least $(\log n)^{1/5}$ lie in $C_T(x)$. Therefore, property **N2** implies that the total contribution of these neighbors to the second neighborhood of x is at most $3 \log n = O(\Delta_T \log \Delta_T)$. On the other hand, x has at most

Δ_T neighbors of degree less than $(\log n)^{1/5}$ (by **N1**), and each of these contribute at most $(\log n)^{1/5} = O((\Delta_T \log \Delta_T)^{1/5})$ to the second neighborhood. Therefore, there are at most

$$O\left(\Delta_T \log \Delta_T + \Delta_T^{6/5} \log^{1/5} \Delta_T\right) = O\left(\Delta_T^{6/5} \log^{1/5} \Delta_T\right)$$

second neighbors to x.   □

4.2. **Safe and good rewirings**. The idea of the proof of Lemma 4.4 is to perform a sequence of rewirings from a typical tree until it becomes nice. However, we have to ensure that the tree remains typical and preserves its maximum degree along the way. This motivates the next definition.

We fix a n-vertex tree $T = (V, E)$ for the remainder of the subsection.

Definition 4.7 (Safe/good to rewire). *A pair* $(v, w) \in V^2$ *is* safe to rewire *for the tree T if $v \neq w$ and the following three conditions are satisfied*

   S1 $\min\{\deg_T(v), \deg_T(w)\} \geq (\log n)^{1/5}$;
   S2 $\deg_T(v) + \deg_T(w) \leq 0.9 (\log n / \log \log n)$;
   S3 v *and w belong to the same cluster in T*.

If moreover $\lambda_1(T_{v,w}) \geq \lambda_1(T)$, (v, w) is said to be good to rewire *for T*.

The following is an immediate consequence of Lemma 1.5 and the fact that the definition of "safe to rewire" is symmetric in v and w.

Fact 4.8. *If $(v, w) \in V^2$ is safe to rewire for T, then either (v, w) or (w, v) is good to rewire.*

It turns out that rewiring a pair that it is safe to rewire has a constrained effect on the cluster structure of T. This is the content of the following Lemma.

Lemma 4.9. *Let T be a tree and (v, w) be safe to rewire. Then any cluster of $T_{v,w}$ is fully contained in a cluster of T.*

Proof. By Fact 4.1, we must show that any two vertices $a, b \in V$ that belong to the same cluster in $T_{v,w}$ also belong to the same cluster in T. To do this, fix $a, b \in V$ in the same cluster in $T_{v,w}$, and consider the unique path $Y = y_0 y_1 \ldots y_m$ connecting $a = y_0$ to $b = y_m$ in $T_{v,w}$. The fact that a, b lie in the same cluster of $T_{v,w}$ is equivalent to the statement that

$$\max\{\deg_{T_{v,w}}(y_{i-1}), \deg_{T_{v,w}}(y_i)\} \geq (\log n)^{1/5} \text{ for each } i \in [m]. \tag{28}$$

Recall from Section 1.2 that $T_{v,w}$ is defined by letting u be the unique neighbor of v in T along the vw path, and replacing each edge sv with $s \in N_T(v) - \{u\}$ with an edge sw. Fixing this vertex u, we see that there are two possibilities.

Case 1: Y does not contain w. In this case, Y is also the path in T connecting a, b. Therefore, a and b lie in the same cluster in T if there are no two consecutive vertices in Y have degree $< (\log n)^{1/5}$ in T. To see that this is indeed true, notice that all vertices in $V - \{v, w\}$ have the same degree in T as in $T_{v,w}$. Moreover, **S1** ensures that $\deg_T(v) \geq (\log n)^{1/5}$. Therefore, condition (28) for Y as a path $T_{v,w}$ also applies to Y as a path in T, and we are done.

Case 2: Y contains w. In this case, it suffices to argue that $a, b \in C_T(w)$. In fact, we only prove $a \in C_T(w)$, as $b \in C_T(w)$ follows from "reversing the path."

Since Y is a path in $T_{v,w}$, there exists exactly one index $j \in [m] \cup \{0\}$ with $y_j = w$. We further split into cases.

   • **Case 2.1:** $j = 0$. Then $a = w$ and we are done.
   • **Case 2.2:** $j > 0$. Then $y_0 \ldots y_{j-1}$ is a path in $T_{v,w}$ not containing w, so $a = y_0$ and y_{j-1} lie in the same cluster of T by Case 1. It now suffices to show that $y_{j-1} \in C_T(w)$, which again requires a considering two cases.

- **Case 2.2.1**: y_{j-1} **is a neighbor of** w **in** T. Then $y_{j-1} \in C_T(w)$ because **S1** guarantees $\deg_T(w) \geq (\log n)^{1/5}$;
- **Case 2.2.2**: y_{j-1} **is not a neighbor of** w **in** T. It is still true that $y_{j-1}w$ is an edge in $T_{v,w}$, so it must be that $y_{j-1} \in N_T(v) - \{u\}$, with u as above. In this case, **S3** guarantees that $v \in C_T(w)$, and **S1** implies that v has degree $\geq (\log n)^{1/5}$, so $y_{j-1} \in C_T(v) = C_T(w)$, as desired.

$\square$

We now argue that the performing rewirings on safe-to-rewire vertices preserves the typicality and maximum degrees of a tree.

Lemma 4.10. *Assume T is typical and $(v,w) \in V^2$ is safe to rewire for T. Then $T_{v,w}$ and $T_{w,v}$ both have the same vertex set and the same maximum degree as T. Moreover, $T_{v,w}$ and $T_{w,v}$ are both typical.*

Proof. We only consider $T_{v,w}$ without loss of generality. Notice that the vertex set is preserved by construction. As for the maximum degree, we have $\Delta_T \geq 0.99 (\log n/\log\log n)$ since T is typical, and **S2** means that neither v or w achieve the maximum degree. For $T_{v,w}$, we have

$$\deg_{T_{v,w}}(v) = 1, \deg_{T_{v,w}}(w) = \deg_T(v) + \deg_T(w) - 1 < 0.9 (\log n/\log\log n), \qquad (29)$$

and all vertices $x \in V \setminus \{v,w\}$ have the same degree as in T. It follows that $\Delta_{T_{v,w}} = \Delta_T$ is still achieved by some vertex that is neither v nor w.

We now argue that $T_{v,w}$ is typical, as per Definition 4.2. Property **N1** is automatic because T and $T_{v,w}$ have the same maximal degree.

To prove property **N2** for $T_{v,w}$, first note that Lemma 4.9 implies $C_{T_{v,w}}(x) \subset C_T(x)$ for all $x \in V$. Since T is typical (and in particular satisfies **N2**), we obtain

$$\sum_{y \in C_T(x)} \deg_T(y) \mathbf{1}_{[\deg_T(y) \geq (\log n)^{1/5}]} \leq 3\log n.$$

Therefore, it suffices to show that

$$\sum_{y \in C_T(x)} \deg_{T_{v,w}}(y) \mathbf{1}_{[\deg_{T_{v,w}}(y) \geq (\log n)^{1/5}]} \leq \sum_{y \in C_T(x)} \deg_T(y) \mathbf{1}_{[\deg_T(y) \geq (\log n)^{1/5}]}, \qquad (30)$$

where $x \in V$ is arbitrary. This property is automatic if $v, w \notin C_T(x)$, because in this case the degrees in the left-hand side of (30) are the same as in the right-hand side. If on the other hand $v \in C_T(x)$ or $w \in C_T(x)$, it must be that $v, w \in C_T(x)$, because **S3** guarantees that v, w lie in the same cluster of T. Property **S1** and equation (29) then imply that the left-hand side of (30) is strictly smaller than the right-hand side: the combined contribution of the terms $y = u$ and $y = v$ decreases by 1, and all other terms remain unchanged.

Finally, equation (29) implies that no nodes of degree $\geq 0.9 (\log n/\log\log n)$ are ever created by the rewiring operation. Since T satisfies **N3**, it follows that property **N3** also holds for $T_{v,w}$. $\square$

4.3. **From typical to nice trees.** We now have all the tools we need to prove Lemma 4.4.

Proof of Lemma 4.4. Consider the following procedure applied to T.

(1) Set $\tilde{T} \leftarrow T$;
(2) While there exists a good-to-rewire pair (v,w) for $\tilde{T}$, set $\tilde{T} \leftarrow T_{v,w}$ $\tilde{T} \leftarrow \tilde{T}_{v,w}$.
(3) Output $\tilde{T}$.

Note that the vertex set of $\tilde{T}$ is always V; only the edge set changes. The above procedure must terminate, since each rewiring operation increases the number of leaves in the tree (cf. (4) and the fact that v is not itself a leaf by condition **S1** in Definition 4.7). By Lemma 4.10, the final tree $\tilde{T}$ is typical, and has the same vertex set and the same maximum degree as T. We also know that $\lambda_1(\tilde{T}) \geq \lambda_1(T)$ and that $\tilde{T}$ cannot contain any pair $(v,w) \in V^2$ that is

good to rewire. Therefore, by Fact 4.8, it also cannot contain any pair that is safe to rewire (otherwise the procedure could continue). It follows that $\tilde{T}$ has the following property:

Ñ4 For any two vertices $v, w \in V$ that lie in the same cluster in $\tilde{T}$,

$$\min\{\deg_{\tilde{T}}(v), \deg_{\tilde{T}}(w)\} \geq (\log n)^{1/5} \Rightarrow \deg_{\tilde{T}}(v) + \deg_{\tilde{T}}(w) > 0.9 \, (\log n / \log \log n).$$

To finish, we show the following claim.

Claim 4.11. *If $\tilde{T}$ is typical and satisfies $\tilde{\mathbf{N}}\mathbf{4}$, then for all $x \in V$, $C_{\tilde{T}}(x)$ contains $< 9 \log \log n$ vertices of degree $\geq (\log n)^{1/5}$; i.e. $\tilde{T}$ is nice.*

To see this, fix any $x \in V$, and consider the sum

$$\sum_{y \in C_{\tilde{T}}(x)} \deg_{\tilde{T}}(y) \mathbf{1}_{[\deg_{\tilde{T}}(y) \geq (\log n)^{1/5}]}.$$

On the one hand, this sum is bounded from above by $3 \log n$ because $\tilde{T}$ is typical (cf. property **N2** in Definition 4.2). On the other hand, $\tilde{\mathbf{N}}\mathbf{4}$ implies that, for any two vertices $v, w \in C_{\tilde{T}}(x)$ with $\min\{\deg_{\tilde{T}}(v), \deg_{\tilde{T}}(w)\} \geq (\log n)^{1/5}$, their joint contribution to the sum is at least $0.9 \, (\log n / \log \log n)$. Therefore

$$0.9 \, (\log n / \log \log n) \left\lfloor \frac{\#\{v \in C_{\tilde{T}}(x) \, : \, \deg_{\tilde{T}}(v) \geq (\log n)^{1/5}\}}{2} \right\rfloor \leq 3 \log n,$$

from which the claim follows after some simple estimates (this is where the condition that $n > 10$ is required). $\square$

5. Putting it all together

In this section we finish the proof of the main result. We will use eigenvalue bound from Corollary 2.7, and we encourage the reader to refresh themselves on the statement of that corollary (and that of Proposition 2.6, on which the corollary relies).

The main missing technical ingredient is the construction of a good enough (M, H)-code for nice trees. The code we shall use appears in Figure 4. (Like in Section 2.2, in Figure 4 we use the $*$ symbol as a wildcard, meaning that any of ℓ, m, h could be there.) The following claim asserts that it is indeed good enough for our purposes.

Claim 5.1. *For the (M, H)-code C in Figure 4, the parameter ε in Proposition 2.6 satisfies:*

$$\varepsilon \leq L \left(\frac{\log^{2/15} \Delta_T}{\Delta_T^{8/15}} \right).$$

for any nice tree T on $n \geq 3$ vertices (cf. Definition 4.3) with maximum degree Δ_T. Here, $L > 0$ is a universal constant.

Proof. We assume that the tree T has maximum degree Δ_T and is nice, as per Definition 4.3. To recapitulate, this means that

N1 The maximum degree of T is $\Delta_T \geq 0.99 \, (\log n / \log \log n)$.
N2 For all $x \in V$, $\sum_{y \in} \deg_T(y) \mathbf{1}_{[\deg_T(y) \geq (\log n)^{1/5}]} \leq 3 \log n$.
N3 There are no adjacent nodes that both have degree $\geq 0.9 \log n / \log \log n$.
N4 For all $x \in V$, $C_T(x)$ contains $< 9 \log \log n$ vertices of degree $\geq (\log n)^{1/5}$.

To apply our combinatorial arguments, we let H denote the set of vertices $z \in V$ with degrees $\deg_T(z) \geq (\log n)^{1/5}$.

For the code C shown in Figure 4, our task is to show that

$$\varepsilon^{(b)} := \frac{\Delta_T (c^{(b)})^{\frac{2}{(\mathrm{len}(c^{(b)}) + f(c^{(b)}))}}}{\Delta_T} \leq L \frac{\log^{2/15} \Delta_T}{\Delta_T^{8/15}}$$
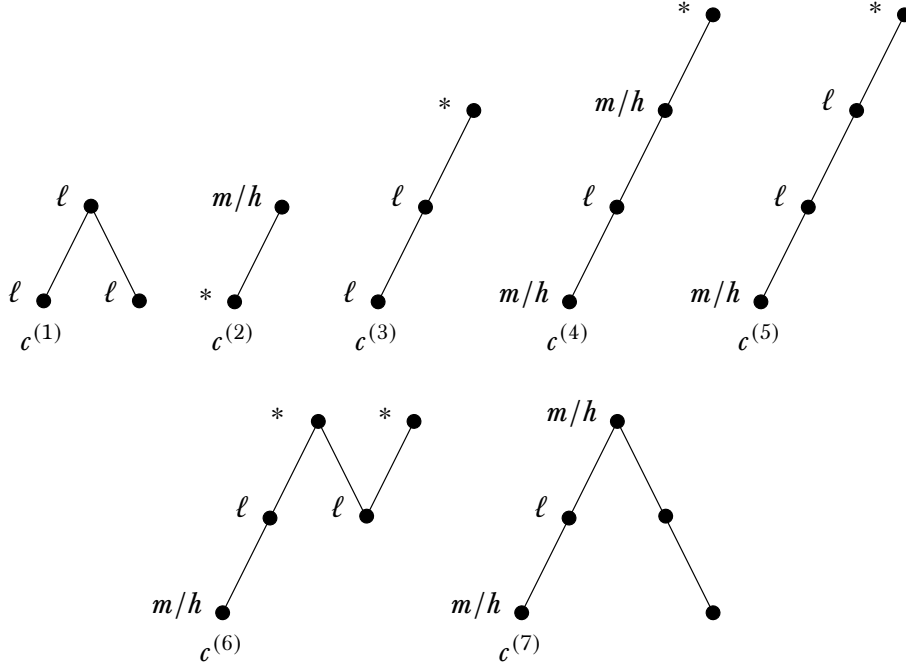
FIGURE 4. The code we use in our analysis. Each of $c^{(2)}, \ldots, c^{(7)}$ in the above figure corresponds to multiple code words, as there are two choices at each location marked m/h and three choices at each location marked $*$. It follows that the total number of words in the above code is $1+6+3+12+6+18+4 = 50$.

for each index $b \leq 7$. Here, we recall that $\Delta_T(c^{(b)})$ is the maximum number of ways one can insert $c^{(b)}$ in T (starting from a vertex of the appropriate kind); $f(c^{(b)})$ is the final value of the codeword; and $\mathrm{len}(c^{(b)})$ is the length of the codeword. The parameter ε in Claim 5.1 is (up to a universal constant $(2a)^2 = 100^2$) the maximum of the $\varepsilon^{(b)}$. (Above we are slightly abusing notation since there are actually 50 codewords, as noted in the caption of Figure 4, but this should not cause confusion in the sequel; for each $b \leq 7$, the bound we prove holds for any of the codewords corresponding to $c^{(b)}$.)

Note that, by condition **N1**, $\log n \leq L\Delta \log \Delta$ and $\log \log n \leq L \log \Delta$. We will use these bounds repeatedly in what follows.

Codeword $c^{(1)}$. Here $f(c^{(1)}) = 0$ and $\mathrm{len}(c^{(1)}) = 1$. Also, $\Delta_T(c^{(1)}) \leq (\log n)^{1/5} \leq L\Delta_T^{1/5} \log^{1/5} \Delta_T$. We conclude that:

$$\varepsilon^{(1)} \leq L \frac{\log^{2/5} \Delta_T}{\Delta_T^{3/5}}.$$

Codeword $c^{(2)}$. In this case $f(c^{(2)}) = \mathrm{len}(c^{(2)}) = 1$. To compute $\Delta_T(c^{(2)})$, let x be the first vertex in the path. Then the second vertex y belongs to $C_T(x)$ because y has degree $\geq (\log n)^{1/5}$. Condition **N4** means that, given x, there are at most $9 \log \log n \leq L \log \Delta_T$ choices for y. We conclude that

$$\varepsilon^{(2)} \leq \frac{L \log \Delta_T}{\Delta_T}.$$

Codeword $c^{(3)}$. Final value and length are given by $f(c^{(3)}) = 2$ and $\mathrm{len}(c^{(3)}) = 2$. Moreover, given the initial vertex x, there are $((\log n)^{1/5})^2 \leq L\Delta_T^{2/5} \log^{2/5} \Delta_T$ choices for vertices $y \sim x$ and $z \sim y$. We conclude that:

$$\varepsilon^{(3)} \leq L\frac{\log^{1/5}\Delta_T}{\Delta_T^{4/5}}.$$

Codeword $c^{(4)}$. In this case $f(c^{(4)}) = \text{len}(c^{(4)}) = 3$. To compute $\Delta_T(c^{(4)})$, let x, y, z, w be the vertices along the path; our goal is to bound the number of choices for w. To do this, notice that $z \in C_T(x)$ because both x and z have degree $\geq (\log n)^{1/5}$. Therefore, w is a neighbor of a node $z \in C_T(x)$ with degree $\deg_T(z) \geq (\log n)^{1/5}$. Condition **N2** guarantees that there are at most $3 \log n \leq L\Delta_T \log \Delta_T$ choices for w. We conclude

$$\varepsilon^{(4)} \leq L\frac{\log^{1/3}\Delta_T}{\Delta_T^{2/3}}.$$

Codeword $c^{(5)}$. In this case again $f(c^{(5)}) = \text{len}(c^{(5)}) = 3$. We again let x, y, z, w be the vertices where the codeword is placed, in order they appear. The number of choices for y given x is at most Δ_T, and one can then choose z and w in $((\log n)^{1/5})^2 \leq L(\Delta_T \log \Delta_T)^{2/5}$ ways. We conclude that $\Delta_T(c^{(5)}) \leq L\Delta_T^{7/5}\log^{2/5}\Delta_T$ and

$$\varepsilon^{(5)} \leq L\frac{\log^{2/15}\Delta_T}{\Delta_T^{8/15}}.$$

Codeword $c^{(6)}$. In this case, $f(c^{(6)}) = 2$ and $\text{len}(c^{(6)}) = 4$. To count $\Delta_T(c^{(6)})$, notice that this codeword corresponds to a walk visiting vertices x, y, z, returning to y and then visiting some vertex w. Given x, one can choose y in at most Δ_T ways; having chosen y, one can choose the pair z, w of neighbours of y in at most $((\log n)^{1/5})^2 \leq L(\Delta_T \log \Delta_T)^{2/5}$ ways since y must have low degree. Therefore, $\Delta_T(c^{(6)}) \leq L\Delta_T^{7/5}\log^{2/5}\Delta_T$ and

$$\varepsilon^{(6)} \leq L\frac{\log^{2/15}\Delta_T}{\Delta_T^{8/15}}.$$

Codeword $c^{(7)}$. In this final case we have $f(c^{(7)}) = 2$ and $\text{len}(c^{(7)}) = 4$. This codeword corresponds to a walk of the form $xyzyx$ where both x and z have degree at least $(\log n)^{1/5}$. But then x and z belong to same cluster and as in the case of codeword $c^{(2)}$, we can argue that by property **N4**, that there are at most $9 \log \log n \leq L \log \Delta_T$ choices for z. Thus $\Delta_T(c^{(7)}) \leq L \log \Delta_T$ and so

$$\varepsilon^{(7)} \leq L\frac{\log^{1/3}\Delta_T}{\Delta_T}. \qquad \square$$

We now turn to the proofs of the main theorems.

Proof of Theorem 1.1. We first use the probability bound of the theorem to prove the expectation bound, then prove the probability bound.

Since $0 \leq \lambda_1(\mathbf{T}_n) \leq 2\sqrt{\Delta_{\mathbf{T}_n}} \leq 2\sqrt{n-1}$, if $K \in \mathbb{N}$ is large enough that $(\log^2 x)/\sqrt{x}$ is decreasing for $x \geq K$ then we have

$$\mathbf{E}|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| \leq \sqrt{K}\mathbf{P}\left\{\Delta_{\mathbf{T}_n} \leq K\right\}$$

$$+ L\left(\frac{\log^2 K}{\sqrt{K}}\right)^{1/15}\mathbf{P}\left\{\Delta_{\mathbf{T}_n} \geq K, |\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| \leq L\left(\frac{\log^2 \Delta_{\mathbf{T}_n}}{\sqrt{\Delta_{\mathbf{T}_n}}}\right)^{1/15}\right\}$$

$$+ \sqrt{2\log n}\mathbf{P}\left\{\Delta_{\mathbf{T}_n} < 2\log n, |\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| > L\left(\frac{\log^2 \Delta_{\mathbf{T}_n}}{\sqrt{\Delta_{\mathbf{T}_n}}}\right)^{1/15}\right\}$$

$$+ 2\sqrt{n-1}\mathbf{P}\left\{\Delta_{\mathbf{T}_n} \geq 2\log n\right\}.$$

The first term of this upper bound tends to 0 since $\Delta_{\mathbf{T}_n} \to \infty$ in probability as $n \to \infty$; this follows from the fact, noted in the introduction, that $\Delta_{\mathbf{T}_n}$ is approximately distributed as the maximum of n independent Poisson(1) random variables. The probability in the second term on the right is at most 1. The third term on the right tends to 0 by the probability bound of the theorem. The fourth term on the right tends to 0 by Fact 3.8. We deduce that

$$\mathbf{E}|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| \le L \left(\frac{\log^2 K}{\sqrt{K}} \right)^{1/15} + o(1) \,.$$

Since we can take $K \in \mathbb{N}$ arbitrarily large, it follows that $\mathbf{E}|\lambda_1(\mathbf{T}_n) - \sqrt{\Delta_{\mathbf{T}_n}}| \to 0$.

We now turn to proving the probability bound of the theorem. From Lemma 4.5 we know that the random labelled tree $\mathbf{T}_n$ is typical (Definition 4.2) with probability at least $1 - n^{-0.8+o(1)}$. Therefore, it suffices to show the following deterministic statement: for any typical tree T on n vertices and maximum degree Δ_T,

$$\sqrt{\Delta_T} \le \lambda_1(T) \le \sqrt{\Delta_T} + L \frac{\log^{2/15} \Delta_T}{\Delta_T^{1/30}}, \tag{31}$$

with a universal constant $L > 0$. The lower bound is true for all graphs with maximal degree Δ_T, so it suffices to prove the upper bound. By adjusting the value of L, we may also assume that n is large. In the remainder of this section, L always denotes a universal constant (not depending on n) whose value may change line to line.

Applying Lemma 4.4, we obtain a nice tree $\tilde{T}$ on n vertices with the same maximum degree

$$\Delta_T \ge \frac{0.99 \log n}{\log \log n}$$

as T and with $\lambda_1(T) \le \lambda_1(\tilde{T})$.

Properties **N1** and **N3**, together with the preceding bound on Δ_T, imply that $\tilde{T}$ does not have adjacent nodes of degree $\ge 0.95\Delta_T$. Corollary 2.7 applies to the tree $\tilde{T}$ with the (M, H)-code C in Figure 4, which has $a = 50$ codewords. It follows that

$$\lambda_1(\tilde{T}) = \sqrt{\Delta_T} \left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2} \right) \sqrt{1 + e^{20} a^2 \varepsilon}.$$

Using the value of ε given by Claim 5.1, we obtain that

$$\lambda_1(\tilde{T}) \le \sqrt{\Delta_T} \left(1 + \frac{\Delta_T^{(2)}}{\Delta_T^2} \right) \left(1 + L \frac{\log^{2/15} \Delta_T}{\Delta_T^{8/15}} \right).$$

Proposition 4.6 gives us

$$\Delta_T^{(2)} \le L \Delta_T^{6/5} (\log \Delta_T)^{1/5},$$

and the upper bound in (31) now follows after some simple estimates. □

For the second theorem, we require a final lemma, relating eigenvalues and the degrees of nodes in trees.

Lemma 5.2. *Let T be a tree with n vertices. List the vertices of T in decreasing order of degree as $v_1, \dots, v_n$, and write d_i for the degree of v_i. Then for any integer $k \le n/2$ we have $\lambda_k(T) \ge \sqrt{d_{2k} - 1}$.*

Proof. Fix $k \le n/2$ and choose a set $S \subset [2k]$ of size k such that $\{v_i, i \in S\}$ is an independent set in T; such S exists since T is bipartite. Then fix a root for T, and for each $i \in S$ write C_i for the resulting set of children of v_i in T. Note that $|C_i| = d_i$ if v_i is the root of T and otherwise $|C_i| = d_i - 1$.

Then for $i \in S$ define a vector $x^{(i)}$ by

$$x_v^{(i)} = \begin{cases} \sqrt{|C_i|} & \text{if } v = v_i \\ 1 & \text{if } v \in C_i \\ 0 & \text{otherwise.} \end{cases}$$

Then $\langle x^{(i)}, Ax^{(i)} \rangle \geq \sqrt{|C_i|}|x^{(i)}|^2 \geq \sqrt{d_i - 1}|x^{(i)}|^2$. Moreover, $\langle x^{(i)}, x^{(j)} \rangle = 0$ for $i \neq j$ with $i, j \in S$, since $\{v_i, i \in S\}$ is an independent set. It follows by Rayleigh's principle that $\lambda_k(T) \geq \min(\sqrt{d_i - 1}, i \in S) \geq \sqrt{d_{2k} - 1}$. $\qquad \square$

Proof of Theorem 1.2. Recall from Corollary 3.12 that we define $a^*(n) = \max(m \in \mathbb{N} : m! \leq n)$. By [10, Theorem 1], the median a_n of $\Delta_{\mathbf{T}_n}$ satisfies that

$$|a_n - a^*| = O(1), \tag{32}$$

so in particular,

$$a_n = (1 + o(1))(\log n)/\log \log n.$$

It then follows from Corollary 3.12 that

$$\mathbf{E}\Delta_{\mathbf{T}_n} \leq \sqrt{a^*(n)} + o(1) = \sqrt{a_n} + o(1),$$

so, by Theorem 1.1 and the triangle inequality, $\mathbf{E}\lambda_1(\mathbf{T}_n) \leq \sqrt{a_n} + o(1)$. Since $\lambda_1(\mathbf{T}_n) \geq \lambda_{k(n)}(\mathbf{T}_n)$, to prove the theorem it remains to show that $\mathbf{E}\lambda_{k(n)}(\mathbf{T}_n) \geq \sqrt{a_n} - o(1)$, where $k(n) = \lceil e^{(\log n)^\beta} \rceil$ for some $\beta \in (0, 1/2)$ fixed.

Let $\alpha \in (\beta, 1/2)$ and take $a = \lfloor a_n - \log^\alpha n \rfloor$; then it follows from Corollary 3.11 that

$$\mathbf{E}D_{\geq a}(n) \geq (1 - o(1))\frac{n}{a!} > 4e^{(\log n)^\beta},$$

the second bound holding for n sufficiently large. Using the lower tail bound from Corollary 3.11 it then follows that

$$\mathbf{P}\left\{D_{\geq a}(n) \leq 2e^{(\log n)^\beta}\right\} \leq \exp(-2e^{(\log n)^\beta}).$$

Since the degree of a node in $\mathbf{T}_n$ is at least its number of chidren in $\mathbf{T}_n^\bullet$, it then follows from Lemma 5.2 that

$$\mathbf{P}\left\{\lambda_{\lceil e^{(\log n)^\beta} \rceil}(\mathbf{T}_n) < \sqrt{a - 1}\right\} \leq \exp(-2e^{(\log n)^\beta}).$$

Since $\sqrt{a - 1} \geq \sqrt{a_n} - o(1)$ and $k(n) = \lceil e^{(\log n)^\beta} \rceil$, it follows that

$$\mathbf{E}\lambda_{k(n)}(\mathbf{T}_n) \geq \sqrt{a_n} - o(1),$$

as required. $\qquad \square$

6. Acknowledgements

REFERENCES

[1] Miklós Abért, Andreas Thom, and Bálint Virág. Benjamini–Schramm convergence and pointwise convergence of the spectral measure. Preprint, 2016. URL https://tu-dresden.de/mn/math/geometrie/thom/ressourcen/dateien/forschung/preprint_app. 2

[2] Louigi Addario-Berry, Anna Brandenberger, Jad Hamdan, and Céline Kerriou. Universal height and width bounds for random trees. *Electron. J. Probab.*, 27:Paper No. 118, 24, 2022. doi: 10.1214/22-ejp842. URL https://doi.org/10.1214/22-ejp842. 22

[3] Johannes Alt, Raphael Ducatez, and Antti Knowles. Delocalization transition for critical Erdős-Rényi graphs. *Comm. Math. Phys.*, 388(1):507–579, 2021. ISSN 0010-3616. doi: 10.1007/s00220-021-04167-y. URL https://mathscinet.ams.org/mathscinet-getitem?mr=4328063. 2

[4] Johannes Alt, Raphaël Ducatez, and Antti Knowles. Extremal eigenvalues of critical Erdős-Rényi graphs. *Ann. Probab.*, 49(3):1347–1401, 2021. ISSN 0091-1798,2168-894X. doi: 10.1214/20-aop1483. URL https://doi.org/10.1214/20-aop1483.

[5] Johannes Alt, Raphael Ducatez, and Antti Knowles. Poisson statistics and localization at the spectral edge of sparse Erdős-Rényi graphs. *Ann. Probab.*, 51(1):277–358, 2023. ISSN 0091-1798,2168-894X. doi: 10.1214/22-aop1596. URL https://doi.org/10.1214/22-aop1596. 2, 3

[6] Adam Arras and Charles Bordenave. Existence of absolutely continuous spectrum for Galton-Watson random trees. *Comm. Math. Phys.*, 403(1):495–527, 2023. ISSN 0010-3616,1432-0916. doi: 10.1007/s00220-023-04798-3. URL https://doi.org/10.1007/s00220-023-04798-3. 2

[7] Shankar Bhamidi, Steven N. Evans, and Arnab Sen. Spectra of large random trees. *J. Theoret. Probab.*, 25(3):613–654, 2012. ISSN 0894-9840,1572-9230. doi: 10.1007/s10959-011-0360-9. URL https://doi.org/10.1007/s10959-011-0360-9. 2

[8] Charles Bordenave. Spetrum of large random graphs. Preprint, 2016. URL https://www.i2m.univ-amu.fr/perso/charles.bordenave/_media/courssrg.pdf. 2

[9] Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures Algorithms*, 37(3):332–352, 2010. ISSN 1042-9832,1098-2418. doi: 10.1002/rsa.20313. URL https://doi.org/10.1002/rsa.20313. 2

[10] Robin Carr, William MY Goh, and Eric Schmutz. The maximum degree in a random tree and related problems. *Random Structures & Algorithms*, 5(1):13–24, 1994. 38

[11] K. P. Choi. On the medians of gamma distributions and an equation of Ramanujan. *Proc. Amer. Math. Soc.*, 121(1):245–251, 1994. ISSN 0002-9939. doi: 10.2307/2160389. URL https://doi.org/10.2307/2160389. 24

[12] Simon Coste and Justin Salez. Emergence of extended states at zero in the spectrum of sparse random graphs. *Ann. Probab.*, 49(4):2012–2030, 2021. ISSN 0091-1798,2168-894X. doi: 10.1214/20-aop1499. URL https://doi.org/10.1214/20-aop1499. 2

[13] D. Dubhashi and D. Ranjan. Balls and bins: A study in negative dependence. *Random Structures and Algorithms*, 13(2):99–124, 1998. 29

[14] CD Godsil. Spectra of trees. In *North-Holland Mathematics Studies*, volume 87, pages 151–159. North-Holland, 1984. 4

[15] Alice Guionnet. Bernoulli random matrices. arXiv:2112.05506v1 [math.PR], 2021. 3

[16] Ella Hiesmayr and Theo McKenzie. The spectral edge of constant degree Erdős-Rényi graphs. arXiv:2309.11007 [math.PR], September 2023. URL https://arxiv.org/abs/2309.11007. 3

[17] V. F. Kolčin. Branching processes, random trees and a generalized particle distribution scheme. *Mat. Zametki*, 21(5):691–705, 1977. ISSN 0025-567X. 22

[18] Christian Krattenthaler. Lattice path enumeration. In *Handbook of enumerative combinatorics*, Discrete Math. Appl. (Boca Raton), pages 589–678. CRC Press, Boca Raton, FL, 2015. 13

[19] J. W. Moon. On the maximum degree in a random tree. *Michigan Mathematical Journal*, 15(4):429–432, 1968. doi: 10.1307/mmj/1029000098. URL https://doi.org/10.1307/mmj/1029000098. 1, 31

[20] Mathew Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003. ISBN 0-19-850626-0. doi: 10.1093/acprof:oso/9780198506263.001.0001. URL https://doi.org/10.1093/acprof:oso/9780198506263.001.0001. 24

[21] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. ISBN 978-3-540-30990-1; 3-540-30990-X. URL https://mathscinet.ams.org/mathscinet-getitem?mr=2245368. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. 23

[22] Jim Pitman. Enumerations of trees and forests related to branching processes and random walks. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 163–180. Amer. Math. Soc., Providence, RI, 1998. ISBN 0-8218-0827-3. doi: 10.1090/dimacs/041/08. URL https://doi.org/10.1090/dimacs/041/08. 22

[23] Justin Salez. Spectral atoms of unimodular random trees. *J. Eur. Math. Soc. (JEMS)*, 22(2):345–363, 2020. ISSN 1435-9855,1435-9863. doi: 10.4171/jems/923. URL https://doi.org/10.4171/jems/923. 2

[24] Dragan Stevanović. Bounding the largest eigenvalue of trees in terms of the largest vertex degree. *Linear algebra and its applications*, 360:35–42, 2003. 4

[25] Daniel W. Stroock. *Probability theory*. Cambridge University Press, Cambridge, second edition, 2011. ISBN 978-0-521-13250-3. An analytic view. 24

DEPARTMENT OF MATHEMATICS AND STATISTICS, MCGILL UNIVERSITY, 805 SHERBROOKE STREET WEST, MONTRÉAL, QUÉBEC, H3A 2K6, CANADA
Email address: louigi.addario@mcgill.ca
URL: http://problab.ca/louigi/

DEPARTMENT OF ECONOMICS AND BUSINESS, POMPEU FABRA UNIVERSITY, BARCELONA, SPAIN; ICREA, PG. LLUÍS COMPANYS 23, 08010 BARCELONA, SPAIN; BARCELONA GRADUATE SCHOOL OF ECONOMICS
Email address: gabor.lugosi@gmail.com
URL: http://www.econ.upf.edu/~lugosi/

INSTITUTO DE MATEMÁTICA PURA E APLICADA (IMPA), ESTRADA DONA CASTORINA 110, RIO DE JANEIRO, RJ, 22460-320, BRAZIL
Email address: rimfo@impa.br
URL: https://sites.google.com/view/roboliv