

# On the distribution of eigenvalues of increasing trees

Kenneth Dadedzi<sup>a,b,1</sup>, Stephan Wagner<sup>c,b,\*,2</sup>

<sup>a</sup> Department of Mathematical Sciences, University of Ghana, Post Office Box LG 62, Legon, Accra, Ghana

<sup>b</sup> Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

<sup>c</sup> Department of Mathematics, Uppsala University, Box 480, 751 06 Uppsala, Sweden



## ARTICLE INFO

### Article history:

Received 24 August 2022

Received in revised form 20 June 2023

Accepted 14 October 2023

Available online 7 November 2023

### Keywords:

Recursive tree

Binary increasing tree

Eigenvalues

Additive parameter

Central limit theorem

## ABSTRACT

We prove that the multiplicity of a fixed eigenvalue  $\alpha$  in a random recursive tree on  $n$  vertices satisfies a central limit theorem with mean and variance asymptotically equal to  $\mu_\alpha n$  and  $\sigma_\alpha^2 n$  respectively. It is also shown that  $\mu_\alpha$  and  $\sigma_\alpha^2$  are positive for every totally real algebraic integer. The proofs are based on a general result on additive tree functionals due to Holmgren and Janson. In the case of the eigenvalue 0, the constants  $\mu_0$  and  $\sigma_0^2$  can be determined explicitly by means of generating functions. Analogous results are also obtained for Laplacian eigenvalues and binary increasing trees.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The spectra of matrices associated with graphs are known to carry substantial information about the graphs. The study of graph eigenvalues has found many applications in mathematics, physics, chemistry, biology, economics and computer science (see for example [12, Chapter 9]).

Eigenvalues of random graphs under different models of randomness have been studied extensively due to their connections to random matrix theory, see for example [4,5,29]. Wigner's semicircle law [31] famously appears in the limit of the eigenvalue distribution of dense random graphs.

The eigenvalues of trees are quite different from those of arbitrary graphs. A famous early result due to Schwenk [28] states that the probability that a (uniformly) random tree has a cospectral mate (a non-isomorphic tree with the same eigenvalues) tends to 1 as the number of vertices goes to infinity. The actual distribution of eigenvalues in random trees was more recently studied by Bhamidi, Evans and Sen in [9]. In particular, they prove (see [9, Theorem 4.1]) that the adjacency spectrum converges to a well-defined limit under different random tree models, including e.g. conditioned Galton–Watson trees and recursive trees. However, they do not fully characterise this limit. Recently, Salez [27] studied the spectral distribution of *unimodular Galton–Watson trees*, which occur prominently as weak limits of large random graphs with given asymptotic degree distribution.

Spectral convergence can be derived from *Benjamini–Schramm convergence* [7]: a sequence of (possibly random) graphs  $G_n$  is said to converge to a (possibly random) rooted graph  $G$  in the Benjamini–Schramm sense if for every positive in-

\* Corresponding author.

E-mail addresses: [kdadedzi@ug.edu.gh](mailto:kdadedzi@ug.edu.gh) (K. Dadedzi), [stephan.wagner@math.uu.se](mailto:stephan.wagner@math.uu.se) (S. Wagner).

<sup>1</sup> The first author was supported by the German Academic Exchange Service DAAD (91628205).

<sup>2</sup> The second author was supported by the National Research Foundation of South Africa, grant 96236, and the Knut and Alice Wallenberg Foundation (KAW 2017.0112).

teger  $r$ , the distribution of the  $r$ -neighbourhood of a uniformly random vertex of  $G_n$  converges to the distribution of the  $r$ -neighbourhood of the root of  $G$ . Indeed, it was shown by Ábert, Thom and Virág [2] that Benjamini–Schramm convergence implies pointwise convergence of the spectral measure for graphs with bounded degrees or more generally bounded expected degrees. It is worth noting that an analogous result also holds for the matching measure [1] (where the roots of the matching polynomial are considered instead of the eigenvalues of the adjacency matrix), which is equivalent for trees. The tree models we consider converge in this sense; this is already implicit in the work of Aldous on fringe distributions [3]. In the appendix, we provide a description of the Benjamini–Schramm limits with a simple self-contained combinatorial proof in one of the cases.

The focus of this paper is the specific model of random recursive trees. A random recursive tree with  $n$  vertices is constructed by a simple growth process: start with a root labelled 1. In the  $k$ -th step, a vertex labelled  $k$  is attached to one of the previous  $k - 1$  vertices, chosen uniformly at random. There are  $(n - 1)!$  different labelled trees that can be obtained in this way; each of them has the same probability  $\frac{1}{(n-1)!}$  in our model.

We provide a refinement of the result of Bhamidi, Evans and Sen for individual eigenvalues by studying the multiplicity of a fixed eigenvalue  $\alpha$  in the adjacency spectrum of a large random recursive tree. In particular, we prove that this multiplicity has linear mean and variance in the size of the tree, and also show that a central limit theorem holds. Analogous results are also obtained for the Laplacian spectrum as well as binary increasing trees, which are similar to recursive trees and also belong to the general family of increasing trees that was introduced by Bergeron, Flajolet and Salvy [8]. The proof of these results is based on tools from linear algebra combined with a general result due to Holmgren and Janson [19] on additive tree functionals (for a precise definition, see Section 2.3). In a forthcoming paper [13], we will use rather different methods to establish a similar central limit theorem for simply generated trees.

In the special case of the eigenvalue 0, which is also related to other parameters of a tree (namely the independence number and the matching number), we can provide a more detailed analysis by means of generating functions and determine the mean and variance more precisely: specifically, the mean is  $0.192694n + \mathcal{O}(1)$ , and the variance is  $0.138629n + \mathcal{O}(n/\log n)$  for recursive trees with  $n$  vertices. For binary increasing trees, mean and variance are  $0.085753n + \mathcal{O}(1)$  and  $0.057162n + \mathcal{O}(1)$ , respectively. See Section 4 for details. Similar results specifically for the eigenvalue 0 were obtained by Bauer and Golinelli [6] for random labelled trees, and by Bordenave, Lelarge and Salez [10] for Galton–Watson trees occurring as local limits of sparse Erdős–Rényi graphs.

## 2. Preliminaries

Before we get to the main theorems, we first provide some important definitions and gather auxiliary results on tree spectra, increasing trees, and additive tree functionals in this section.

### 2.1. Eigenvalues of graphs and trees

Let us briefly review some simple but relevant facts on the spectra of graphs and trees. The eigenvalues of a graph are the eigenvalues of its adjacency matrix, the Laplacian eigenvalues the eigenvalues of its Laplacian matrix (which has the vertex degrees as diagonal entries, and otherwise  $-1$  or  $0$  depending on whether or not vertices are adjacent). Since both the adjacency matrix and the Laplacian matrix are symmetric, all eigenvalues are real. Moreover, the Laplacian eigenvalues are always non-negative.

The eigenvalues of bipartite graphs, and thus in particular trees, are known to be symmetric (see for instance [11, Theorem 3.11]). In other words,  $\alpha$  is an eigenvalue if and only if  $-\alpha$  is, and the multiplicities agree.

An important result that we will make use of in our arguments is Cauchy’s interlacing theorem [12, Corollary 1.3.12]: if  $M$  is a symmetric matrix with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , and  $\beta_1 \geq \beta_2 \geq \dots \geq \beta_{n-1}$  are the eigenvalues of the matrix obtained by removing one of its rows and the corresponding column, then we have

$$\lambda_1 \geq \beta_1 \geq \lambda_2 \geq \beta_2 \geq \dots \geq \lambda_{n-1} \geq \beta_{n-1} \geq \lambda_n.$$

This is specifically relevant for us in the case where  $M$  is the adjacency matrix of a graph. Removing a row and the corresponding column is equivalent to removing one of the vertices.

### 2.2. Increasing trees

Increasing trees are labelled rooted trees with the characteristic property that the labels of nodes increase as one moves along any path from the root to a leaf. Bergeron, Flajolet and Salvy [8] were the first to study general varieties of increasing trees. Such varieties are characterised by weights associated with increasing trees.

Let a sequence  $(w_k)_{k \geq 0}$  of non-negative real numbers with  $w_0 > 0$  (called the *weight sequence*) be given. We define the weight  $W(T)$  of an increasing tree  $T$  by

$$W(T) = \prod_{v \in V(T)} w_{d^+(v)} = \prod_{k \geq 0} w_k^{D_k(T)},$$

where  $d^*(v)$  denotes the outdegree of  $v$  and  $D_k(T)$  is accordingly the number of vertices in  $T$  with outdegree  $k$ . We associate a generating series with the weight sequence by  $\Phi(t) = \sum_{k \geq 0} w_k t^k$ .

Let  $\mathcal{T}_n$  be the set of all increasing rooted ordered trees of order  $n$ . We define the exponential generating function for increasing trees associated with the weight sequence  $(w_k)_{k \geq 0}$  by

$$G(x) = \sum_{T \in \mathcal{T}_n} \frac{W(T)x^{|T|}}{|T|!}.$$

It is known [8, Theorem 1] that this generating function satisfies the differential equation

$$G'(x) = \Phi(G(x)), \quad G(0) = 0.$$

Once the weight has been defined, *random trees* from a specific variety are obtained by selecting a tree  $T$  with a probability that is proportional to its weight  $W(T)$ .

For certain special varieties of increasing trees, random trees can equivalently be constructed by means of a tree evolution (growth) process. These have been characterised by Panholzer and Prodinger in [24]: they belong to three particular groups characterised by the weight generating series  $\Phi(t)$ .

- (1)  $\Phi(t) = w_0 e^{\frac{w_1}{w_0} t}$ , where  $w_0 > 0$  and  $w_1 > 0$ .
- (2)  $\Phi(t) = w_0 \left(1 - \frac{w_1}{r w_0} t\right)^{-r}$ , where  $r > 0$ ,  $w_0 > 0$  and  $w_1 > 0$ .
- (3)  $\Phi(t) = w_0 \left(1 + \frac{w_1}{d w_0} t\right)^d$ , where  $d > 1$ ,  $w_0 > 0$  and  $w_1 > 0$ .

These correspond to the varieties of *recursive trees*, *generalised plane-oriented recursive trees* and *d-ary increasing trees*, respectively. In this paper, we will focus on two particularly important varieties, namely recursive trees and binary (2-ary) increasing trees. It is well known that the probabilistic model of random binary increasing trees is also equivalent to that of random *binary search trees*, see for instance [14, Chapter 6]. One usually uses the following normalised versions of the weights above (for the probabilistic model, the choice of  $w_0$  and  $w_1$  is actually irrelevant):

- (1) Recursive trees are associated with  $\Phi(t) = e^t$  ( $w_0 = w_1 = 1$ ).
- (2) Binary increasing trees are associated with  $\Phi(t) = (1 + t)^2$  ( $d = 2$ ,  $w_0 = 1$  and  $w_1 = 2$ ).

The growth process for recursive trees has already been described in the introduction. For binary increasing trees, it is very similar, except that every vertex can have at most two children: a left child, a right child, or both.

### 2.3. Additive tree functionals

Many invariants associated with rooted trees satisfy a type of recursion in which the invariant is summed over all root branches. The notion of *additive functionals* [19,20,25,30] provides a unifying framework for invariants of this kind. Let  $F$  be an invariant that assigns a value  $F(T)$  to every rooted tree  $T$ . This invariant is said to be *additive* with *toll function*  $f$  if it satisfies the recursion

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

for all rooted trees  $T$ , where  $T_1, T_2, \dots, T_k$  are the root branches of  $T$ , i.e., the components obtained by removing the root from  $T$ , endowed with their natural roots (the children of  $T$ 's root).

Many important examples of tree invariants satisfy such a recursion. A typical example is the number of leaves, whose toll function is

$$f(T) = \begin{cases} 1 & T \text{ is a single vertex,} \\ 0 & \text{otherwise.} \end{cases}$$

A generalisation of this example concerns *fringe subtrees*: a fringe subtree of a rooted tree is a subtree consisting of a vertex and all its descendants. Thus a single leaf can also be regarded as a fringe subtree. The number of fringe subtrees of a given shape is also an additive functional in this sense, and this example will be important for us later. As we will see, the multiplicity of a fixed eigenvalue also fits this framework.

There are limit theorems for additive tree functionals for several models of random trees and under different technical assumptions, see [19,20,25,30]. For our purposes, we particularly rely on a general central limit theorem due to Holmgren

and Janson [19]: as usual, let  $\mathcal{N}(\mu, \sigma^2)$  denote a normal distribution with mean  $\mu$  and variance  $\sigma^2$ , and let  $\xrightarrow{d}$  denote convergence in distribution. Moreover, let  $\mathcal{R}_n$  and  $\mathcal{B}_n$  denote a random recursive tree of order  $n$  and a random binary increasing tree of order  $n$  respectively.

**Theorem 2.1** (see [19, Theorem 1.14]).

(1) For random recursive trees, assume that

$$\sum_{k=1}^{\infty} \frac{\sqrt{\mathbb{V} f(\mathcal{R}_k)}}{k^{3/2}} < \infty,$$

$$\lim_{k \rightarrow \infty} \frac{\mathbb{V} f(\mathcal{R}_k)}{k} = 0,$$

$$\sum_{k=1}^{\infty} \frac{(\mathbb{E}(f(\mathcal{R}_k)))^2}{k^2} < \infty.$$

Then, as  $n \rightarrow \infty$ ,

$$\frac{\mathbb{E}(F(\mathcal{R}_n))}{n} \rightarrow \hat{\mu}_F := \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \mathbb{E}(f(\mathcal{R}_k)),$$

$$\frac{\mathbb{V}(F(\mathcal{R}_n))}{n} \rightarrow \hat{\sigma}_F^2 < \infty,$$

and

$$\frac{F(\mathcal{R}_n) - \mathbb{E}(F(\mathcal{R}_n))}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}_F^2).$$

(2) For binary increasing trees, assume that

$$\sum_{k=1}^{\infty} \frac{\sqrt{\mathbb{V}(f(\mathcal{B}_k))}}{k^{3/2}} < \infty,$$

$$\lim_{k \rightarrow \infty} \frac{\mathbb{V}(f(\mathcal{B}_k))}{k} = 0,$$

$$\sum_{k=1}^{\infty} \frac{(\mathbb{E}(f(\mathcal{B}_k)))^2}{k^2} < \infty.$$

Then, as  $n \rightarrow \infty$ ,

$$\frac{\mathbb{E}(F(\mathcal{B}_n))}{n} \rightarrow \mu_F := \sum_{k=1}^{\infty} \frac{2}{(k+1)(k+2)} \mathbb{E}(f(\mathcal{B}_k)),$$

$$\frac{\mathbb{V}(F(\mathcal{B}_n))}{n} \rightarrow \sigma_F^2 < \infty,$$

and

$$\frac{F(\mathcal{B}_n) - \mathbb{E}(F(\mathcal{B}_n))}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_F^2).$$

### 2.4. The toll function

In this section, we show that the multiplicity of an eigenvalue in the spectrum of a rooted tree can be viewed as an additive parameter with bounded toll function. A similar statement is also obtained for the Laplacian spectrum. We let  $N_\alpha(T)$  denote the multiplicity of the eigenvalue  $\alpha$  in the spectrum of a rooted tree  $T$ . Let  $T - r$  be the forest obtained from a rooted tree  $T$  by deleting the root  $r$ .

Now we let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  and  $\beta_1 \geq \beta_2 \geq \dots \geq \beta_{n-1}$  be the eigenvalues of the tree  $T$  and the forest  $T - r$  respectively. Suppose that  $\alpha = \lambda_{k+1}$  has multiplicity  $l$ , that is

$$\lambda_1 \geq \dots \geq \lambda_k > \lambda_{k+1} = \dots = \lambda_{k+l} > \lambda_{k+l+1} \geq \dots \geq \lambda_n.$$

By the Cauchy interlacing theorem we have the following three possibilities.

**Case 1:**

$$\lambda_k \geq \beta_k > \lambda_{k+1} = \beta_{k+1} = \dots = \beta_{k+l-1} = \lambda_{k+l} > \beta_{k+l} \geq \lambda_{k+l+1}$$

so that  $N_\alpha(T) - N_\alpha(T - r) = l - (l - 1) = 1$ ,

**Case 2:**

$$\lambda_k > \beta_k = \lambda_{k+1} = \beta_{k+1} = \dots = \beta_{k+l-1} = \lambda_{k+l} = \beta_{k+l} > \lambda_{k+l+1}$$

so that  $N_\alpha(T) - N_\alpha(T - r) = l - (l + 1) = -1$ ,

**Case 3:**

$$\lambda_k > \beta_k = \lambda_{k+1} = \beta_{k+1} = \dots = \beta_{k+l-1} = \lambda_{k+l} > \beta_{k+l} \geq \lambda_{k+l+1}$$

or

$$\lambda_k \geq \beta_k > \lambda_{k+1} = \beta_{k+1} = \dots = \beta_{k+l-1} = \lambda_{k+l} = \beta_{k+l} > \lambda_{k+l+1}$$

so that  $N_\alpha(T) - N_\alpha(T - r) = l - l = 0$ .

So we see that  $N_\alpha(T) - N_\alpha(T - r) \in \{-1, 0, 1\}$ . The forest  $T - r$  has the branches of  $T$  as its components; let them be denoted by  $T_1, T_2, \dots, T_k$ . Since the multiplicity of an eigenvalue in the adjacency matrix of a graph can be obtained by summing the multiplicity over the connected components (if we choose an appropriate vertex order, the adjacency matrix becomes a block diagonal matrix with the blocks corresponding to the components), we have

$$N_\alpha(T - r) = \sum_{i=1}^k N_\alpha(T_i).$$

Altogether, we thus obtain the recursion

$$N_\alpha(T) = \sum_{i=1}^k N_\alpha(T_i) + n_\alpha(T),$$

where the toll function  $n_\alpha(T)$  is given by

$$n_\alpha(T) = N_\alpha(T) - N_\alpha(T - r) \in \{-1, 0, 1\}.$$

For the Laplacian spectrum, we can argue in a similar manner, but we have to modify the definition slightly. Instead of considering the multiplicity of  $\alpha$  as a Laplacian eigenvalue of  $T$ , let  $M_\alpha(T)$  be the multiplicity of  $\alpha$  as an eigenvalue of the matrix obtained from the Laplacian matrix by adding 1 to the diagonal entry in the row and column that correspond to the root (i.e., we artificially increase the root degree by 1). Let this be called the *modified Laplacian matrix*. As changing a single entry of a matrix can change its rank by at most 1,  $M_\alpha(T)$  differs from the multiplicity of  $\alpha$  as an eigenvalue of the Laplacian matrix by at most 1.

If we remove the row and column that correspond to the root from the modified Laplacian matrix, then we obtain (up to reordering) a block diagonal matrix, where each block is the modified Laplacian matrix of one of the root branches. Here, it is important that we are using the modified Laplacian as opposed to the normal Laplacian matrix, since the degree of the root in each branch  $T_i$  is exactly 1 less than its degree in the whole tree  $T$ . Using the Cauchy interlacing theorem once again, we find that

$$M_\alpha(T) = \sum_{i=1}^k M_\alpha(T_i) + m_\alpha(T),$$

where  $m_\alpha(T) \in \{-1, 0, 1\}$ .

**3. The central limit theorem**

From the discussion in the previous section, we know that  $N_\alpha(T)$  can be regarded as an additive parameter whose toll function  $n_\alpha(T) = N_\alpha(T) - N_\alpha(T - r)$  satisfies  $n_\alpha(T) \in \{-1, 0, 1\}$ . So in particular, if  $\mathcal{R}_n$  and  $\mathcal{B}_n$  denote a random recursive tree of order  $n$  and a random binary increasing tree of order  $n$  respectively,

$$\mathbb{E}(n_\alpha(\mathcal{R}_k)), \mathbb{V}(n_\alpha(\mathcal{R}_k))$$

as well as

$$\mathbb{E}(n_\alpha(\mathcal{B}_k)), \mathbb{V}(n_\alpha(\mathcal{B}_k))$$

are all trivially bounded by constants, showing that the technical conditions of Theorem 2.1 are satisfied. Hence we have the following theorem:

**Theorem 3.1.** Fix a real number  $\alpha$  that can occur as an eigenvalue of a tree. There exist constants  $\mu_{\text{rec},\alpha}$  and  $\sigma_{\text{rec},\alpha}^2$  such that the multiplicity  $N_\alpha(\mathcal{R}_n)$  of  $\alpha$  as an eigenvalue of the random recursive tree  $\mathcal{R}_n$  with  $n$  vertices satisfies

$$\frac{N_\alpha(\mathcal{R}_n) - \mu_{\text{rec},\alpha}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{rec},\alpha}^2).$$

Likewise, suppose that  $\alpha$  can occur as an eigenvalue of a binary tree. There exist constants  $\mu_{\text{bin},\alpha}$  and  $\sigma_{\text{bin},\alpha}^2$  such that the multiplicity  $N_\alpha(\mathcal{B}_n)$  of  $\alpha$  as an eigenvalue of the random binary increasing tree  $\mathcal{B}_n$  with  $n$  vertices satisfies

$$\frac{N_\alpha(\mathcal{B}_n) - \mu_{\text{bin},\alpha}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{bin},\alpha}^2).$$

**Remark 3.2.** It is known that a real number  $\alpha$  can be an eigenvalue of a tree if and only if it is a totally real algebraic integer, i.e., it is a zero of a monic polynomial whose coefficients are integers and whose zeros are all real (in other words, the conjugates of  $\alpha$  must also all be real). It is clear that this is a necessary condition, and it was also shown to be sufficient by Salez [26]. Since eigenvalues of a graph are bounded by the maximum degree, not all such real numbers are also eigenvalues of a binary tree. We are not aware of any explicit characterisation of the possible eigenvalues of binary trees.

A completely analogous statement holds for the Laplacian spectrum. Note here that we are applying Theorem 2.1 to the invariant  $M_\alpha(T)$ , which is not exactly the multiplicity of  $\alpha$  as an eigenvalue of the Laplacian of  $T$ . However, since the difference is at most 1, it vanishes in the limit due to the normalising factor  $\sqrt{n}$ .

**Theorem 3.3.** Fix a real number  $\alpha$  that can occur as an eigenvalue of the modified Laplacian matrix of a rooted tree. There exist constants  $\nu_{\text{rec},\alpha}$  and  $\tau_{\text{rec},\alpha}^2$  such that the multiplicity  $L_\alpha(\mathcal{R}_n)$  of  $\alpha$  as a Laplacian eigenvalue of the random recursive tree  $\mathcal{R}_n$  with  $n$  vertices satisfies

$$\frac{L_\alpha(\mathcal{R}_n) - \nu_{\text{rec},\alpha}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \tau_{\text{rec},\alpha}^2).$$

Likewise, there exist constants  $\nu_{\text{bin},\alpha}$  and  $\tau_{\text{bin},\alpha}^2$  such that the multiplicity  $L_\alpha(\mathcal{B}_n)$  of  $\alpha$  as a Laplacian eigenvalue of the random binary increasing tree  $\mathcal{B}_n$  with  $n$  vertices satisfies

$$\frac{L_\alpha(\mathcal{B}_n) - \nu_{\text{bin},\alpha}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \tau_{\text{bin},\alpha}^2).$$

**Remark 3.4.** Any finite linear combination of multiplicities  $N_\alpha$  can also be regarded as an additive functional whose toll function is still bounded. Therefore Theorem 2.1 still applies, which means that any finite linear combination  $\sum_{i=1}^k c_i N_{\alpha_i}(T)$  also satisfies a central limit theorem. Applying the Cramér–Wold device [18, Chapter 5, Theorem 10.5], we see that the multiplicities of any finite set of eigenvalues satisfy a multidimensional central limit theorem. An analogous statement also holds for Laplacian eigenvalues.

**Remark 3.5.** Note that we can express the mean constants as

$$\mu_{\text{rec},\alpha} = \sum_{k=1}^{\infty} \frac{\mathbb{E}(n_\alpha(\mathcal{R}_k))}{k(k+1)} \text{ and } \mu_{\text{bin},\alpha} = \sum_{k=1}^{\infty} \frac{2\mathbb{E}(n_\alpha(\mathcal{B}_k))}{(k+1)(k+2)},$$

respectively. Analogous formulas hold for  $\nu_{\text{rec},\alpha}$  and  $\nu_{\text{bin},\alpha}$ .

It is clear that the sums

$$\sum_{\alpha} \mu_{\text{rec},\alpha} \text{ and } \sum_{\alpha} \mu_{\text{bin},\alpha}$$

and their counterparts for Laplacian eigenvalues are less than or equal to 1. However, it is not possible to interchange the order of summation in

$$\sum_{\alpha} \mu_{\text{rec},\alpha} = \sum_{\alpha} \sum_{k=1}^{\infty} \frac{\mathbb{E}(n_\alpha(\mathcal{R}_k))}{k(k+1)}$$

since the double sum may not be absolutely convergent. So it is not clear whether or not the sum is indeed equal to 1, which would mean that the limiting spectral measure is purely discrete. The same comment applies to binary increasing trees and Laplacian eigenvalues.

### 3.1. Forcing subtrees and positivity of the mean constants

Even if a real number  $\alpha$  can occur as an eigenvalue of a tree, it is not a priori clear that the corresponding mean constant  $\mu_{\text{rec}}(\alpha)$  is not equal to 0. In this section, we prove that this is always the case. In fact, we are able to provide lower bounds on  $\mu_{\text{rec}}(\alpha)$  and  $\mu_{\text{bin}}(\alpha)$ . The main idea is to establish a relationship between the number of occurrences of a subtree  $H$  in a large tree  $T$  and specific eigenvalues in the spectrum of  $T$ .

Recall that a fringe subtree of a rooted tree  $T$  is a subtree that consists of a vertex  $v$  and all its descendants. Let us denote this subtree by  $T_v$ . It is easy to prove by induction that any additive functional  $F$  with toll function  $f$  satisfies

$$F(T) = \sum_v f(T_v)$$

if we define  $f(\bullet) = F(\bullet)$  for the single-vertex tree  $\bullet$ . The number of occurrences of a fixed rooted tree  $H$  as a fringe subtree (i.e., the number of fringe subtrees isomorphic to  $H$ ) is an additive functional with a toll function given by  $f(T) = 1$  whenever  $T$  is isomorphic to  $H$  and  $f(T) = 0$  otherwise. Therefore, Theorem 2.1 applies to this functional (see also [30]), and we have the following:

**Proposition 3.6.** *For any fixed rooted tree  $H$ , the number of occurrences of  $H$  as a fringe subtree of a random recursive tree with  $n$  vertices satisfies a central limit theorem with mean  $\mu_{\text{rec}, H}n + \mathcal{O}(1)$ . The constant is given by  $\mu_{\text{rec}, H} = \frac{\beta_{\text{rec}}(H)}{|H|(|H|+1)}$ , where  $\beta_{\text{rec}}(H)$  is the probability that a random recursive tree with  $|H|$  vertices is isomorphic to  $H$ .*

*Likewise, for any fixed rooted binary tree  $H$ , the number of occurrences of  $H$  as a fringe subtree of a random binary increasing tree with  $n$  vertices satisfies a central limit theorem with mean  $\mu_{\text{bin}, H}n + \mathcal{O}(1)$ . The constant is given by  $\mu_{\text{bin}, H} = \frac{2\beta_{\text{bin}}(H)}{(|H|+1)(|H|+2)}$ , where  $\beta_{\text{bin}}(H)$  is the probability that a random binary increasing tree with  $|H|$  vertices is isomorphic to  $H$ .*

Now, let us consider the relationship between the spectrum of a rooted tree  $T$  and fringe trees found in  $T$ . When we join  $k_i$  copies of a tree  $H$  to different vertices  $u_i$  ( $i \in \{1, 2, \dots, l\}$ ) in a tree  $T$ , we can bound the multiplicities of certain eigenvalues in the resulting tree from below. This is captured in the following results.

**Theorem 3.7.** *Let  $T$  be a tree obtained from  $G$  by joining  $k_i$  copies of the rooted tree  $H$  to the vertices  $u_i \in V(G)$ ,  $i \in \{1, 2, \dots, l\}$ . That is, for each  $i$ , we take  $k_i$  copies of  $H$  and connect each of their roots to  $u_i$  by an edge. Then if  $\sum_{i=1}^l (k_i - 1) > 0$ , each eigenvalue of  $H$  is an eigenvalue of the resulting tree, and the multiplicity of each of these eigenvalues is at least  $\sum_{i=1}^l (k_i - 1)$ .*

**Proof.** Consider the forest  $T \setminus \{u_1, u_2, \dots, u_l\}$ . Obviously, it has  $k_1 + k_2 + \dots + k_l$  components isomorphic to  $H$ . So if  $\alpha$  is an eigenvalue of  $H$ , then it is an eigenvalue of  $T \setminus \{u_1, u_2, \dots, u_l\}$  whose multiplicity is at least  $k_1 + k_2 + \dots + k_l$ . The interlacing theorem shows that the multiplicity of  $\alpha$  as an eigenvalue of  $T$  differs from the multiplicity as an eigenvalue of  $T \setminus \{u_1, u_2, \dots, u_l\}$  by at most  $l$ . Thus,  $\alpha$  is an eigenvalue of  $T$  with multiplicity at least

$$k_1 + k_2 + \dots + k_l - l = (k_1 - 1) + (k_2 - 1) + \dots + (k_l - 1). \quad \square$$

As a simple corollary, we have the following known bound for the eigenvalue 0. Let  $l(T)$  be the number of leaves in the tree  $T$ , and let  $q(T)$  be the number of quasipendant vertices (a vertex is called quasipendant if it is adjacent to a leaf) in the tree  $T$ .

**Corollary 3.8** (see e.g. [11, p. 258]). *The multiplicity of the eigenvalue 0 in the adjacency spectrum of a tree  $T$  is at least  $l(T) - q(T)$ .*

**Proof.** This is the special case of Theorem 3.7 where  $H$  is a single vertex and  $u_1, u_2, \dots$  are the quasipendant vertices of  $T$ . If there are  $l_i$  leaves attached to  $u_i$ , then by Theorem 3.7, the multiplicity of the eigenvalue 0 is at least

$$\sum_{i=1}^m (l_i - 1) = \sum_{i=1}^m l_i - m = l(T) - q(T). \quad \square$$

In particular, let  $H$  be a rooted tree that has  $\alpha$  as an eigenvalue, and let  $K$  be obtained by joining two copies of  $H$  to a common root. The presence of  $K$  as a fringe subtree in  $T$  “forces”  $\alpha$  to be an eigenvalue of  $T$ . More generally, if  $K$  occurs  $r$  times as a fringe subtree of  $T$ , then Theorem 3.7 shows that  $\alpha$  is an eigenvalue of multiplicity at least  $r$ :  $N_\alpha(T) \geq r$ . The following is now immediate from Proposition 3.6:

**Corollary 3.9.** *Let  $H$  be a rooted tree that has  $\alpha$  as an eigenvalue, and let  $K$  be obtained by joining two copies of  $H$  to a common root. Then we have*

$$\mu_{\text{rec},K} \leq \mu_{\text{rec},\alpha}.$$

In particular,  $\mu_{\text{rec},\alpha} > 0$  as soon as  $\alpha$  is an eigenvalue of some tree. Analogous corollaries hold for binary increasing trees and for Laplacian eigenvalues. However, it is worth noting that the bound above is generally quite weak: while the presence of  $K$  as a fringe subtree implies that  $\alpha$  is an eigenvalue, the converse is not true. Moreover, there generally exist several non-isomorphic “forcing subtrees” like  $K$  corresponding to any particular eigenvalue  $\alpha$ .

### 3.2. Positivity of the variance constants

We can use a similar argument to show that the variance constants in Theorem 3.1 are always strictly positive (in the general theorem of Holmgren and Janson, it is possible that the variance constants  $\sigma_F$  and  $\hat{\sigma}_F$  are 0). We exhibit this for eigenvalues of recursive trees. The case of Laplacian eigenvalues is similar, as is the argument for binary increasing trees.

Let  $\alpha$  be a fixed real number that is a possible eigenvalue of a tree. Using the “forcing subtrees” argument from the previous section, we can construct trees for which the multiplicity of  $\alpha$  is arbitrarily large. In particular, we can find a rooted tree  $S$  with  $N_\alpha(S) \geq 4$ . On the other hand, we let  $S'$  be a path with the same number of vertices as  $S$ . Since paths do not have repeated eigenvalues (the eigenvalues of the  $n$ -vertex path are  $2 \cos \frac{\pi k}{n+1}$ ,  $k \in \{1, 2, \dots, n\}$ , see e.g. [12, p. 47]), we have  $N_\alpha(S') \leq 1$ .

Now let  $T$  be an arbitrary tree, and suppose that  $S$  occurs as a fringe subtree of  $T$ ; in other words,  $T$  can be obtained by attaching  $S$  to a vertex  $v$  of a rooted tree  $R$ . From arguments that we used before, it follows that

$$N_\alpha(T) \geq N_\alpha(R - v) + N_\alpha(S) - 1.$$

Let  $T'$  be obtained by replacing  $S$  with  $S'$ . By the same reasoning, we have

$$N_\alpha(T') \leq N_\alpha(R - v) + N_\alpha(S') + 1.$$

Combined with our assumptions on  $N_\alpha(S)$  and  $N_\alpha(S')$ , this gives us

$$N_\alpha(T') \leq N_\alpha(T) - 1.$$

Now let us consider a random recursive tree  $\mathcal{R}_n$  with  $n$  vertices, and let us replace all occurrences of either  $S$  or  $S'$  as a fringe subtree by a marked leaf to obtain a reduced tree  $\mathcal{R}'_n$ . As a consequence of Theorem 2.1 and Proposition 3.6, the number of marked leaves in this reduced tree satisfies a central limit theorem with mean  $(\mu_{\text{rec},S} + \mu_{\text{rec},S'})n + \mathcal{O}(1)$ . Conditioned on its size, each fringe subtree of a random recursive tree is again a random recursive tree. Therefore, if we condition on the shape of the reduced tree  $\mathcal{R}'_n$  obtained from a random recursive tree  $\mathcal{R}_n$  and let the number of marked leaves in the reduced tree be  $M_n$ , then the number of fringe subtrees represented by marked leaves that are isomorphic to  $S$  follows a binomial distribution  $\text{Bin}(M_n, p)$ , where  $p = \frac{\beta_{\text{rec},S}}{\beta_{\text{rec},S} + \beta_{\text{rec},S'}}$ . Since replacing a fringe subtree isomorphic to  $S$  by a fringe subtree isomorphic to  $S'$  decreases the multiplicity of  $\alpha$  as an eigenvalue by at least 1, the variance of  $N_\alpha(\mathcal{R}_n)$ , conditioned on the value of  $M_n$ , is at least the variance of this binomial distribution, which is  $p(1 - p)M_n$ . It now follows from the law of total variance that

$$\begin{aligned} \mathbb{V}(N_\alpha(\mathcal{R}_n)) &= \mathbb{E}(\mathbb{V}(N_\alpha(\mathcal{R}_n)|M_n)) + \mathbb{V}(\mathbb{E}(N_\alpha(\mathcal{R}_n)|M_n)) \\ &\geq \mathbb{E}(\mathbb{V}(N_\alpha(\mathcal{R}_n)|M_n)) \\ &\geq \mathbb{E}(p(1 - p)M_n) \\ &= p(1 - p)(\mu_{\text{rec},S} + \mu_{\text{rec},S'})n + \mathcal{O}(1), \end{aligned}$$

which implies that  $\sigma_{\text{rec},\alpha}^2 \geq p(1 - p)(\mu_{\text{rec},S} + \mu_{\text{rec},S'}) > 0$ . This shows that the variance constant is always positive, so that the multiplicity of  $\alpha$  as an eigenvalue, suitably normalised, converges weakly to a non-degenerate normal distribution.

## 4. The special case of the eigenvalue 0

The eigenvalue 0 of the adjacency matrix is special for several reasons. Since the spectrum of every tree is symmetric, the multiplicity of 0 as an eigenvalue is even if and only if the number of vertices is. This implies that the toll function  $n_0$  defined in Section 2.4 can only take the values  $-1$  and  $1$ , as opposed to the generic case where  $n_\alpha$  can also attain the value 0.

One can also recursively characterise the toll function  $n_0$ , which will allow us to explicitly determine the mean constants  $\mu_{\text{rec},0}$  and  $\mu_{\text{bin},0}$  as well as the variance constants  $\sigma_{\text{rec},0}^2$  and  $\sigma_{\text{bin},0}^2$  using methods from analytic combinatorics.



To this end, let  $T$  be a rooted tree with root  $r$ , and let  $\Psi(T, z)$  and  $\Psi_r(T, z)$  be the characteristic polynomials of  $T$  and  $T - r$  respectively. Note that  $n_0(T) = 1$  if and only if  $\lim_{z \rightarrow 0^+} \frac{\Psi_r(T, z)}{\Psi(T, z)} = \pm\infty$ , and  $n_0(T) = -1$  if and only if  $\lim_{z \rightarrow 0^+} \frac{\Psi_r(T, z)}{\Psi(T, z)} = 0$ .

The ratio  $\frac{\Psi_r(T, z)}{\Psi(T, z)}$  can be related to the angles of  $T$ . Suppose that  $\{e_1, e_2, \dots, e_n\}$  are the natural basis vectors, that  $\lambda_1, \lambda_2, \dots, \lambda_m$  are the distinct eigenvalues of  $T$ , and that the adjacency matrix  $A(T)$  has the spectral decomposition

$$A(T) = \lambda_1 P_1 + \lambda_2 P_2 + \dots + \lambda_m P_m,$$

where  $P_i$  represents the orthogonal projection of  $\mathbb{R}^n$  onto the eigenspace of  $\lambda_i$  with respect to the basis  $\{e_1, e_2, \dots, e_n\}$ . The values of  $\theta_{ij} = \|P_i e_j\|$ , where  $1 \leq i \leq m$  and  $1 \leq j \leq n$ , are called the angles of  $T$ . If the  $j$ -th column corresponds to a specific vertex  $v$ , then we say that  $\theta_{ij}$  are the angles corresponding to  $v$ .

**Theorem 4.1** (see [12, p. 33]). *Let  $\Psi(T, z)$  and  $\Psi_r(T, z)$  be the characteristic polynomials of a rooted tree  $T$  with root  $r$  and the forest  $T - r$  respectively. Suppose that  $\lambda_1, \lambda_2, \dots, \lambda_m$  are the distinct eigenvalues of  $T$  and  $\theta_{ir}$  are the angles corresponding to  $r$ . Then*

$$\frac{\Psi_r(T, z)}{\Psi(T, z)} = \sum_{i=1}^m \frac{\theta_{ir}^2}{z - \lambda_i}.$$

This implies in particular that  $\lim_{z \rightarrow 0^+} \frac{\Psi_r(T, z)}{\Psi(T, z)}$  cannot be  $-\infty$ , since all numerators are non-negative. The limit is therefore either 0 or  $+\infty$ .

Now we apply the following recursion for the characteristic polynomial due to Mohar [23]:

**Theorem 4.2** (see [23]). *Let  $\Psi(T, z)$  and  $\Psi_r(T, z)$  be the characteristic polynomials of a tree  $T$  and the forest  $T - r$  respectively, where  $r$  is the root of  $T$ . Suppose  $T_1, T_2, \dots, T_k$  are the branches of  $T$  with  $v_1, v_2, \dots, v_k$  their respective roots. Then*

$$\frac{\Psi_r(T, z)}{\Psi(T, z)} = \frac{1}{z - \sum_{j=1}^k \frac{\Psi_{v_j}(T_j, z)}{\Psi(T_j, z)}}. \tag{1}$$

This readily shows that if  $\lim_{z \rightarrow 0^+} \frac{\Psi_{v_j}(T_j, z)}{\Psi(T_j, z)} = 0$  for all  $j$ , we have  $\lim_{z \rightarrow 0^+} \frac{\Psi_r(T, z)}{\Psi(T, z)} = +\infty$ . In other words, if  $n_0(T_j) = -1$  for all  $j$ , then  $n_0(T) = 1$ . On the other hand, if at least one of the fractions  $\frac{\Psi_{v_j}(T_j, z)}{\Psi(T_j, z)}$  tends to  $+\infty$  as  $z \rightarrow 0^+$  (equivalently, if  $n_0(T_j) = 1$  for at least one index  $j$ ), then the denominator in (1) tends to  $-\infty$ , and we have  $\lim_{z \rightarrow 0^+} \frac{\Psi_r(T, z)}{\Psi(T, z)} = 0$ , thus  $n_0(T) = -1$ .

In summary,

$$n_0(T) = \begin{cases} 1 & \text{if } n_0(T_j) = -1 \text{ for all } j, \\ -1 & \text{if } n_0(T_j) = 1 \text{ for at least one } j. \end{cases} \tag{2}$$

This recursion will be translated to functional equations for generating functions in the following sections.

#### 4.1. Recursive trees

In this section, we focus on the multiplicity of 0 in the spectrum of recursive trees. As indicated earlier, recursive trees are increasing trees that can be constructed by a growth or evolution process, where each new vertex is attached to one of the previous vertices uniformly at random. As there are  $n - 1$  possibilities when the  $n$ -th vertex is attached, there are  $(n - 1)!$  different recursive trees.

By marking the root of a recursive tree with the lowest label, a recursive tree can symbolically be defined by a boxed product. That is, a marked node attached to a set of recursive trees. If we let  $Y(x)$  be the exponential generating function, this symbolic definition translates to the differential equation

$$Y'(x) = \exp(Y(x)), \quad \text{where } Y(0) = 0,$$

see Section 2.2. This differential equation has the explicit solution  $Y(x) = -\log(1 - x)$ , which is consistent with the earlier observation that there are  $(n - 1)!$  recursive trees with  $n$  vertices.

We now define a bivariate exponential generating function  $Y(x, t)$  that takes into account the multiplicity of the eigenvalue 0:

$$Y(x, t) = \sum_{T \in \mathfrak{R}} e^{tN_0(T)} \frac{x^{|T|}}{|T|!},$$

where the sum is over the set  $\mathfrak{R}$  of all recursive trees. Of course,  $Y(x, 0) = Y(x) = -\log(1 - x)$ .

Based on the recursive characterisation of  $n_0$  in (2), we divide the set of all recursive trees into the sets  $\mathfrak{R}_+$  and  $\mathfrak{R}_-$  of trees satisfying  $n_0(T) = 1$  and  $n_0(T) = -1$ , respectively. We also let  $Y_+(x, t)$  and  $Y_-(x, t)$  be the respective associated exponential generating functions, defined in the same way as  $Y(x, t)$  above. The recursion (2) translates to the system of equations

$$\begin{aligned} \frac{\partial}{\partial x} Y_+(x, t) &= \exp(t + Y_-(x, t)), \\ \frac{\partial}{\partial x} Y_-(x, t) &= \exp(-t + Y(x, t)) - \exp(-t + Y_-(x, t)) \\ &= \exp(-t + Y_-(x, t))(\exp(Y_+(x, t)) - 1). \end{aligned}$$

This system of partial differential equations does not seem to have an explicit solution. However, in order to determine mean and variance of  $N_0(\mathcal{R}_n)$ , we only need the coefficients of  $x^n$  in the partial derivatives  $\left. \frac{\partial}{\partial t} Y(x, t) \right|_{t=0}$  and  $\left. \left( \frac{\partial}{\partial t} \right)^2 Y(x, t) \right|_{t=0}$ , and these satisfy ordinary differential equations that we will actually be able to solve. This eventually leads to the following theorem.

**Theorem 4.3.** *We have the following generating function for the mean multiplicity  $\mathbb{E}(N_0(\mathcal{R}_n))$  of the eigenvalue 0 of a random recursive tree with  $n$  vertices:*

$$\sum_{n \geq 1} \frac{x^n}{n} \mathbb{E}(N_0(\mathcal{R}_n)) = \frac{1}{1-x} \int_0^x \frac{1 + \log(1-u)}{1 - \log(1-u)} du.$$

Asymptotically, the mean multiplicity of the eigenvalue 0 of a random recursive tree with  $n$  vertices is

$$\mathbb{E}(N_0(\mathcal{R}_n)) = (2G - 1)n + \mathcal{O}\left(\frac{1}{\log n}\right),$$

where  $G$  denotes the Euler–Gompertz constant [15, pp. 425–426]. Moreover, the variance of the multiplicity of the eigenvalue 0 of a random recursive tree with  $n$  vertices is

$$\mathbb{V}(N_0(\mathcal{R}_n)) = K_1 n + \mathcal{O}\left(\frac{n}{\log n}\right)$$

for a constant  $K_1 \approx 0.138629$  (see Remark 4.4 for precise expressions).

**Proof.** In order to simplify some of the equations that follow, we introduce the abbreviations  $F(x) = Y(x, 0) = -\log(1 - x)$ ,  $F_{\pm}(x) = Y_{\pm}(x, 0)$  as well as

$$G(x) = \left. \frac{\partial}{\partial t} Y(x, t) \right|_{t=0}, \quad G_{\pm}(x) = \left. \frac{\partial}{\partial t} Y_{\pm}(x, t) \right|_{t=0}$$

and

$$H(x) = \left. \left( \frac{\partial}{\partial t} \right)^2 Y(x, t) \right|_{t=0}, \quad H_{\pm}(x) = \left. \left( \frac{\partial}{\partial t} \right)^2 Y_{\pm}(x, t) \right|_{t=0}.$$

Let us start by determining  $F_+(x)$  and  $F_-(x)$ . Setting  $t = 0$  in the system of differential equations gives us

$$\begin{aligned} F'_+(x) &= \exp(F_-(x)), \\ F'_-(x) &= \exp(F_-(x))(\exp(F_-(x)) - 1). \end{aligned}$$

Since we know that  $F_-(x) = F(x) - F_+(x) = -\log(1 - x) - F_+(x)$ , the first of these equations yields

$$F'_+(x) = \frac{\exp(-F_+(x))}{1-x}.$$

This separable differential equation with initial condition  $F_+(0) = 0$  has the explicit solution

$$F_+(x) = \log(1 - \log(1 - x)),$$

and we also get

$$F_-(x) = F(x) - F_+(x) = -\log((1-x)(1 - \log(1-x))).$$

Next we consider the first derivatives. Differentiating the system of differential equations for  $Y_+(x, t)$  and  $Y_-(x, t)$  with respect to  $t$  and plugging in  $t = 0$ , we obtain

$$G'_+(x) = \exp(F_-(x))(1 + G_-(x)),$$

$$G'_-(x) = \exp(F_-(x) + F_+(x))(-1 + G_-(x) + G_+(x)) - \exp(F_-(x))(-1 + G_-(x)).$$

Adding these two and plugging in the formulas for  $F(x)$  and  $F_-(x)$ , we obtain

$$G'(x) = \exp(F(x))(G(x) - 1) + 2 \exp(F_-(x)) = \frac{G(x)}{1-x} - \frac{1}{1-x} + \frac{2}{(1-x)(1-\log(1-x))}.$$

Multiplying by  $1 - x$  and regrouping gives

$$\frac{d}{dx}((1-x)G(x)) = (1-x)G'(x) - G(x) = \frac{1 + \log(1-x)}{1 - \log(1-x)},$$

hence we have (taking the initial condition  $G(0) = 0$  into account)

$$G(x) = \frac{1}{1-x} \int_0^x \frac{1 + \log(1-u)}{1 - \log(1-u)} du.$$

This already proves the first statement of the theorem. At this point, let us introduce the function

$$A(x) = \int_0^x \frac{1}{1 - \log(1-u)} du$$

and note some of its properties: we have  $A(0) = 0$ ,  $A(1) = G$ , as well as  $A'(x) = \frac{1}{1-\log(1-x)}$ . For later use, we also need its asymptotic behaviour as  $x \rightarrow 1$ . Integration by parts yields

$$\begin{aligned} A(x) &= G - \int_x^1 \frac{1}{1 - \log(1-u)} du = G - \int_0^{1-x} \frac{1}{1 - \log u} du \\ &= G - \frac{u}{1 - \log u} \Big|_0^{1-x} + \int_0^{1-x} \frac{1}{(1 - \log u)^2} du \\ &= G - \frac{1-x}{1 - \log(1-x)} + \int_0^{1-x} \frac{1}{(1 - \log u)^2} du. \end{aligned}$$

So it follows that

$$A(x) = G + \frac{1-x}{\log(1-x)} + \mathcal{O}\left(\left|\frac{1-x}{\log^2(1-x)}\right|\right)$$

as  $x \rightarrow 1$ . In terms of this function, we can express  $G(x)$  as

$$G(x) = \frac{1}{1-x} \int_0^x \frac{1 + \log(1-u)}{1 - \log(1-u)} du = \frac{1}{1-x} \int_0^x \left(\frac{2}{1 - \log(1-u)} - 1\right) du = \frac{2A(x) - x}{1-x}.$$

We will also need  $G_-(x)$  later, so let us return to the differential equation

$$G'_+(x) = \exp(F_-(x))(1 + G_-(x)),$$

which is equivalent to

$$\begin{aligned} G'_-(x) = G'(x) - G'_+(x) &= \frac{2A'(x) - 1}{1-x} + \frac{2A(x) - x}{(1-x)^2} - \frac{1}{(1-x)(1-\log(1-x))}(1 + G_-(x)) \\ &= \frac{2A(x) - 1}{(1-x)^2} + \frac{1}{(1-x)(1-\log(1-x))}(1 - G_-(x)). \end{aligned}$$

Multiplying by the integrating factor  $1 - \log(1-x)$  and rearranging yields

$$\frac{d}{dx}((1 - \log(1 - x))G_-(x)) = \frac{(2A(x) - 1)(1 - \log(1 - x))}{(1 - x)^2} + \frac{1}{1 - x}.$$

We integrate both sides and apply integration by parts to obtain

$$\begin{aligned} (1 - \log(1 - x))G_-(x) &= -\frac{(2A(t) - 1)\log(1 - t)}{1 - t} \Big|_0^x + \int_0^x \frac{2A'(u)\log(1 - u)}{1 - u} du - \log(1 - x) \\ &= -\frac{(2A(x) - 1)\log(1 - x)}{1 - x} + \log(1 - x) + 2\log(1 - \log(1 - x)) \end{aligned}$$

and finally

$$G_-(x) = -\frac{(2A(x) - 1)\log(1 - x)}{(1 - x)(1 - \log(1 - x))} + \frac{\log(1 - x) + 2\log(1 - \log(1 - x))}{(1 - \log(1 - x))}. \tag{3}$$

Now we have all ingredients to determine  $H(x) = \left(\frac{\partial}{\partial t}\right)^2 Y(x, t) \Big|_{t=0}$  as well. We differentiate the system of differential equations for  $Y_{\pm}(x, t)$  twice with respect to  $t$ , then plug in  $t = 0$  and add the two equations. After some simplifications, this gives us

$$\begin{aligned} H'(x) &= e^{F(x)}(H(x) + (G(x) - 1)^2) + 4e^{F(x)}G_-(x) \\ &= \frac{H(x)}{1 - x} + \frac{(2A(x) - 1)^2}{(1 - x)^3} + \frac{4G_-(x)}{(1 - x)(1 - \log(1 - x))}. \end{aligned}$$

Again, we can solve this differential equation by multiplying by the integrating factor  $1 - x$  and integrating, which finally gives us the expression

$$H(x) = \frac{1}{1 - x} \int_0^x \left( \frac{(2A(u) - 1)^2}{(1 - u)^2} + \frac{4G_-(u)}{1 - \log(1 - u)} \right) du.$$

Using (3) and the properties of the function  $A(x)$  mentioned previously, we find that

$$\frac{(2A(u) - 1)^2}{(1 - u)^2} + \frac{4G_-(u)}{1 - \log(1 - u)} = \frac{(2G - 1)^2}{(1 - u)^2} + \mathcal{O}\left(\left|\frac{1}{(1 - u)\log^2(1 - u)}\right|\right)$$

as  $u \rightarrow 1$ . This means in particular that the integral

$$K_1 = \int_0^1 \left( \frac{(2A(u) - 1)^2}{(1 - u)^2} + \frac{4G_-(u)}{1 - \log(1 - u)} - \frac{(2G - 1)^2}{(1 - u)^2} \right) du$$

is convergent, and we can write

$$\begin{aligned} H(x) &= \frac{(2G - 1)^2x}{(1 - x)^2} + \frac{1}{1 - x} \int_0^x \left( \frac{(2A(u) - 1)^2}{(1 - u)^2} + \frac{4G_-(u)}{1 - \log(1 - u)} - \frac{(2G - 1)^2}{(1 - u)^2} \right) du \\ &= \frac{(2G - 1)^2x}{(1 - x)^2} + \frac{K_1}{1 - x} + \frac{1}{1 - x} \int_x^1 \left( \frac{(2A(u) - 1)^2}{(1 - u)^2} + \frac{4G_-(u)}{1 - \log(1 - u)} - \frac{(2G - 1)^2}{(1 - u)^2} \right) du. \end{aligned}$$

We are now in a position to determine the asymptotic behaviour of the coefficients by means of singularity analysis [16, Chapter VI]. All functions we have been dealing with have a dominant singularity at 1 (i.e., there is no other singularity whose modulus is less than or equal to 1). Due to the closure properties of functions amenable to singularity analysis (see [16, Section VI.10]), we can apply singularity analysis to all of them. Specifically, the asymptotic expansion of  $A(x)$  translates to an expansion for  $G(x)$ :

$$G(x) = \frac{2A(x) - x}{1 - x} = \frac{2G - 1}{1 - x} + 1 + \frac{1}{\log(1 - x)} + \mathcal{O}\left(\left|\frac{1}{\log^2(1 - x)}\right|\right).$$

This translates to an asymptotic formula for the coefficients of  $G(x)$  and thus the mean of  $N_0$ . We have

$$\mathbb{E}(N_0(\mathcal{R}_n)) = \frac{[x^n]G(x)}{[x^n]F(x)} = n[x^n]G(x) = (2G - 1)n + \mathcal{O}\left(\frac{1}{\log n}\right).$$

In the same way, we can also deal with the variance. Since

$$H(x) = \frac{(2G - 1)^2 x}{(1 - x)^2} + \frac{K_1}{1 - x} + \mathcal{O}\left(\left|\frac{1}{(1 - x) \log^2(1 - x)}\right|\right),$$

we have

$$\mathbb{E}(N_0^2(\mathcal{R}_n)) = \frac{[x^n]H(x)}{[x^n]F(x)} = n[x^n]H(x) = (2G - 1)^2 n^2 + K_1 n + \mathcal{O}\left(\frac{n}{\log^2 n}\right).$$

So the result on the variance now follows as well.  $\square$

**Remark 4.4.** Since  $2G - 1 \approx 0.192694$ , we infer that approximately 19.3% of the spectrum of a large random recursive tree consist of the eigenvalue 0.

The integral representation for  $K_1$  can be simplified using integration by parts. There are many different ways to represent  $K_1$ , for example

$$\begin{aligned} K_1 &= 4G(G - 1) + 8 \int_0^1 \frac{\log(1 - \log w)}{(1 - \log w)^2} dw \\ &= 4(G - 1)(G - 2) - 4 \int_0^\infty \log^2(1 + u)e^{-u} du \\ &\approx 0.138629. \end{aligned}$$

#### 4.2. Binary increasing trees

Now we consider the multiplicity of 0 in the spectrum of binary increasing trees. The approach is very similar to recursive trees.

A binary increasing tree has the property that each vertex has two possible places to which a child can be attached, that is, a left or right child. In view of that, to attach a node labelled  $n$  to an existing binary increasing tree of order  $n - 1$ , there are  $n$  possible places to do so. It follows that there are  $n!$  different binary increasing trees with  $n$  vertices.

Let  $\mathfrak{B}$  be the set of all binary increasing trees. As in the previous subsection, we denote the exponential generating function by  $Y(x)$ . In view of the recursive decomposition into root, left subtree and right subtree, we have

$$Y'(x) = (1 + Y(x))^2, \quad \text{where } Y(0) = 0,$$

see again Section 2.2. This differential equation has the explicit solution  $Y(x) = \frac{x}{1-x}$ , in agreement with our earlier observation that there are  $n!$  binary increasing trees with  $n$  vertices.

As in the case of recursive trees, we can now define a bivariate exponential generating function  $Y(x, t)$  in the following way:

$$Y(x, t) = \sum_{T \in \mathfrak{B}} e^{tN_0(T)} \frac{x^{|T|}}{|T|!},$$

where the sum is over the set  $\mathfrak{B}$  of all binary increasing trees. Of course, we have  $Y(x, 0) = Y(x) = \frac{x}{1-x}$ .

Now we exploit the recursive characterisation of  $n_0$  in (2) again, by dividing the set of all binary increasing trees into the sets  $\mathfrak{B}_+$  and  $\mathfrak{B}_-$  of trees satisfying  $n_0(T) = 1$  and  $n_0(T) = -1$ , respectively. We also let  $Y_+(x, t)$  and  $Y_-(x, t)$  be the respective associated exponential generating functions again. The recursion (2) now gives us the system of equations

$$\begin{aligned} \frac{\partial}{\partial x} Y_+(x, t) &= e^t (1 + Y_-(x, t))^2, \\ \frac{\partial}{\partial x} Y_-(x, t) &= e^{-t} (1 + Y_+(x, t))^2 - e^{-t} (1 + Y_-(x, t))^2. \end{aligned}$$

Again, we cannot find an explicit solution for the bivariate generating function, but we can determine the partial derivatives at 0. This leads to the following analogue of Theorem 4.3:

**Theorem 4.5.** *We have the following generating function for the mean multiplicity  $\mathbb{E}(N_0(\mathcal{B}_n))$  of the eigenvalue 0 of a random binary increasing tree with  $n$  vertices:*

$$\sum_{n \geq 1} x^n \mathbb{E}(N_0(\mathcal{B}_n)) = \frac{1}{(1-x)^2} \int_0^x Q((1-u)\sqrt{5}) du,$$

where  $Q(t) = \frac{(8+4\sqrt{5})t^2+(84+36\sqrt{5})t-(22+10\sqrt{5})}{(7+3\sqrt{5}-2t)^2}$ . Asymptotically, the mean multiplicity of the eigenvalue 0 of a random binary increasing tree with  $n$  vertices is

$$\mathbb{E}(N_0(\mathcal{B}_n)) = C_1(n+1) + \sqrt{5} - 2 + \mathcal{O}(n^{-\sqrt{5}}),$$

where  $C_1 \approx 0.085753$ . Moreover, the variance of the multiplicity of the eigenvalue 0 of a random binary increasing tree with  $n$  vertices is

$$\mathbb{V}(N_0(\mathcal{B}_n)) = K_2 n + \mathcal{O}(1),$$

with  $K_2 \approx 0.057162$ .

**Proof.** We use the same abbreviations as in the previous section, namely  $F(x) = Y(x, 0) = \frac{x}{1-x}$ ,  $F_{\pm}(x) = Y_{\pm}(x, 0)$  as well as

$$G(x) = \frac{\partial}{\partial t} Y(x, t) \Big|_{t=0}, \quad G_{\pm}(x) = \frac{\partial}{\partial t} Y_{\pm}(x, t) \Big|_{t=0}$$

and

$$H(x) = \left(\frac{\partial}{\partial t}\right)^2 Y(x, t) \Big|_{t=0}, \quad H_{\pm}(x) = \left(\frac{\partial}{\partial t}\right)^2 Y_{\pm}(x, t) \Big|_{t=0}.$$

We start again by determining  $F_+(x)$  and  $F_-(x)$ . Setting  $t = 0$  in the system of differential equations, we obtain

$$\begin{aligned} F'_+(x) &= (1 + F_-(x))^2, \\ F'_-(x) &= (1 + F(x))^2 - (1 + F_-(x))^2. \end{aligned}$$

Now  $F_-(x) = F(x) - F_+(x) = \frac{x}{1-x} - F_+(x)$ , thus

$$F'_+(x) = \left(\frac{1}{1-x} - F_+(x)\right)^2.$$

This Riccati-type differential equation (with the initial value  $F_+(0) = 0$ ) has the explicit solution

$$F_+(x) = \frac{3 - \sqrt{5}}{2(1-x)} + \frac{2\sqrt{5}}{(1-x)(2 - (7 + 3\sqrt{5})(1-x)^{-\sqrt{5}})}.$$

Consequently, since  $F_+(x) + F_-(x) = F(x) = \frac{x}{1-x}$ ,

$$F_-(x) = \frac{\sqrt{5} - 1}{2(1-x)} - \frac{2\sqrt{5}}{(1-x)(2 - (7 + 3\sqrt{5})(1-x)^{-\sqrt{5}})} - 1.$$

Now we take the first derivative with respect to  $t$  again. Differentiating the system of differential equations and plugging in  $t = 0$  gives

$$\begin{aligned} G'_+(x) &= F'_+(x) + 2(1 + F_-(x))G_-(x), \\ G'_-(x) &= -F'_-(x) + 2(1 + F(x))G(x) - 2(1 + F_-(x))G_-(x). \end{aligned}$$

Adding the two yields

$$G'(x) = F'_+(x) - F'_-(x) + 2(1 + F(x))G(x).$$

Since  $1 + F(x) = \frac{1}{1-x}$ , we can solve this differential equation by multiplying by the integrating factor  $(1-x)^2$  and integrating, which results in

$$\begin{aligned} G(x) &= \frac{1}{(1-x)^2} \int_0^x (1-u)^2 (F'_+(u) - F'_-(u)) du \\ &= \frac{1}{(1-x)^2} \int_0^x \frac{(8 + 4\sqrt{5})(1-u)^{2\sqrt{5}} + (84 + 36\sqrt{5})(1-u)^{\sqrt{5}} - (22 + 10\sqrt{5})}{(7 + 3\sqrt{5} - 2(1-u)^{\sqrt{5}})^2} du \end{aligned}$$

$$= \frac{1}{(1-x)^2} \int_0^1 \frac{(8+4\sqrt{5})v^{2\sqrt{5}} + (84+36\sqrt{5})v^{\sqrt{5}} - (22+10\sqrt{5})}{(7+3\sqrt{5}-2v^{\sqrt{5}})^2} dv - \frac{1}{(1-x)^2} \int_0^{1-x} \frac{(8+4\sqrt{5})v^{2\sqrt{5}} + (84+36\sqrt{5})v^{\sqrt{5}} - (22+10\sqrt{5})}{(7+3\sqrt{5}-2v^{\sqrt{5}})^2} dv.$$

In particular, we have established the first statement of the theorem. The integrand has the expansion

$$(2 - \sqrt{5}) + (50 - 22\sqrt{5})v^{\sqrt{5}} + \mathcal{O}(|v|^{2\sqrt{5}}),$$

so it follows immediately that

$$G(x) = \frac{C_1}{(1-x)^2} - \frac{2 - \sqrt{5}}{1-x} + \mathcal{O}(|1-x|^{\sqrt{5}-1}),$$

where

$$C_1 = \int_0^1 \frac{(8+4\sqrt{5})v^{2\sqrt{5}} + (84+36\sqrt{5})v^{\sqrt{5}} - (22+10\sqrt{5})}{(7+3\sqrt{5}-2v^{\sqrt{5}})^2} dv \approx 0.085753.$$

Singularity analysis now shows that

$$\mathbb{E}(N_0(\mathcal{B}_n)) = \frac{[x^n]G(x)}{[x^n]F(x)} = C_1(n+1) + \sqrt{5} - 2 + \mathcal{O}(n^{-\sqrt{5}}).$$

It remains to consider the variance. For this purpose, we will need  $G_-(x)$ . Recall that  $G_-(x)$  satisfies the linear differential equation

$$G'_-(x) = -F'_-(x) + 2(1 + F(x))G(x) - 2(1 + F_-(x))G_-(x) = -F'_-(x) + \frac{2G(x)}{1-x} - 2(1 + F_-(x))G_-(x).$$

We can solve it by multiplying by the integrating factor

$$\exp\left(\int 2(1 + F_-(x)) dx\right) = \frac{(7 + 3\sqrt{5} - 2(1-x)^{\sqrt{5}})^2}{(1-x)^{\sqrt{5}-1}}$$

and integrating both sides. This gives us

$$G_-(x) = \frac{(1-x)^{\sqrt{5}-1}}{(7+3\sqrt{5}-2(1-x)^{\sqrt{5}})^2} \int_0^x \frac{(7+3\sqrt{5}-2(1-u)^{\sqrt{5}})^2}{(1-u)^{\sqrt{5}-1}} \left(\frac{2G(u)}{1-u} - F'_-(u)\right) du. \tag{4}$$

Lastly,  $H(x)$  is determined by first differentiating the system of equations for  $Y_+(x, t)$  and  $Y_-(x, t)$  twice with respect to  $t$ , plugging in  $t = 0$  and adding the equations, which yields

$$H'(x) = (1 + F(x))^2 - 4(1 + F(x))G(x) + 2G(x)^2 + 8(1 + F_-(x))G_-(x) + 2(1 + F(x))H(x).$$

Here, the integrating factor is simply  $(1-x)^2$  as in the differential equation for  $G(x)$ . This eventually gives us

$$H(x) = \frac{1}{(1-x)^2} \int_0^x (1 - 4(1-u)G(u) + 2(1-u)^2G(u)^2 + 8(1-u)^2(1+F_-(u))G_-(u)) du.$$

In order to apply singularity analysis, we need the expansion around the dominant singularity  $x = 1$ . Recall that

$$G(x) = \frac{C_1}{(1-x)^2} - \frac{2 - \sqrt{5}}{1-x} + \mathcal{O}(|1-x|^{\sqrt{5}-1}).$$

Moreover,

$$F_-(x) = \frac{\sqrt{5} - 1}{2(1-x)} - 1 + \mathcal{O}(|1-x|^{\sqrt{5}-1}).$$

Plugging these into the representation (4) for  $G_-(x)$ , we also find that

$$G_-(x) = \frac{C_1(\sqrt{5}-1)}{2(1-x)^2} - \frac{7-3\sqrt{5}}{2\sqrt{5}(1-x)} + \mathcal{O}(1).$$

We can now plug the asymptotic expansions of  $G, F_-$  and  $G_-$  into the integral representation for  $H$ , which shows that the integrand is

$$1 - 4(1-u)G(u) + 2(1-u)^2G(u)^2 + 8(1-u)^2(1+F_-(u))G_-(u) = \frac{2C_1^2}{(1-u)^2} + \mathcal{O}(1).$$

In particular, it follows that the integral

$$C_2 = \int_0^1 \left( 1 - 4(1-u)G(u) + 2(1-u)^2G(u)^2 + 8(1-u)^2(1+F_-(u))G_-(u) - \frac{2C_1^2}{(1-u)^2} \right) du$$

is convergent, and we obtain

$$H(x) = \frac{2C_1^2x}{(1-x)^3} + \frac{C_2}{(1-x)^2} + \mathcal{O}\left(\frac{1}{|1-x|}\right).$$

Singularity analysis gives us the asymptotic behaviour of the second moment:

$$\mathbb{E}(N_0^2(\mathcal{B}_n)) = \frac{[x^n]H(x)}{[x^n]F(x)} = C_1^2n(n+1) + C_2n + \mathcal{O}(1),$$

so finally

$$\mathbb{V}(N_0(\mathcal{B}_n)) = (4C_1 - 2\sqrt{5}C_1 - C_1^2 + C_2)n + \mathcal{O}(1),$$

proving the last statement of the theorem with  $K_2 = (4C_1 - 2\sqrt{5}C_1 - C_1^2 + C_2) \approx 0.057162$ .  $\square$

### 4.3. Connections to other graph parameters

It is well known that the characteristic polynomial of the adjacency matrix of a forest coincides with its matching polynomial (a special case of a general result due to Sachs, see [11, Theorem 1.3]), which in turn implies that the multiplicity of 0 as an eigenvalue is given by  $N_0(T) = |T| - 2m(T)$ , where  $m(T)$  is the matching number (the greatest cardinality of a matching) of  $T$ . Moreover, it is a known consequence of König's theorem that independence number (i.e., the greatest cardinality of an independent set) and matching number of any bipartite graph (thus in particular any tree) add up to the order of the graph. Consequently, we can express the independence number  $i(T)$  of a tree  $T$  in terms of the multiplicity of 0 as well:  $i(T) = |T| - m(T) = \frac{1}{2}(N_0(T) + |T|)$ .

Therefore, we obtain central limit theorems for independence number and matching number as simple corollaries: with the exception of the variance constants, these have been proven before [17,21], and the mean for recursive trees has been determined even earlier [22].

**Corollary 4.6.** *The independence number and matching number of a random recursive tree with  $n$  vertices satisfy the following central limit theorem:*

$$\frac{i(\mathcal{R}_n) - c_{\text{rec}}n}{s_{\text{rec}}\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ and } \frac{m(\mathcal{R}_n) - (1 - c_{\text{rec}})n}{s_{\text{rec}}\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where  $c_{\text{rec}} = G \approx 0.596347$  and  $s_{\text{rec}} = \frac{\sqrt{K_1}}{2} \approx 0.186165$ .

Likewise, the independence number and matching number of a random binary increasing tree with  $n$  vertices satisfy the following central limit theorem:

$$\frac{i(\mathcal{B}_n) - c_{\text{bin}}n}{s_{\text{bin}}\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ and } \frac{m(\mathcal{B}_n) - (1 - c_{\text{bin}})n}{s_{\text{bin}}\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where  $c_{\text{bin}} = \frac{1+C_1}{2} \approx 0.542876$  and  $s_{\text{bin}} = \frac{\sqrt{K_2}}{2} \approx 0.119543$ .



### 5. Other eigenvalues

For other eigenvalues, determining the mean and variance constants numerically is substantially more difficult as the recursive characterisation of the toll function is no longer available. It was remarked earlier that the mean constants can be determined from the convergent series

$$\mu_{\text{rec},\alpha} = \sum_{k=1}^{\infty} \frac{\mathbb{E}(n_{\alpha}(\mathcal{R}_k))}{k(k+1)} \text{ and } \mu_{\text{bin},\alpha} = \sum_{k=1}^{\infty} \frac{2\mathbb{E}(n_{\alpha}(\mathcal{B}_k))}{(k+1)(k+2)}.$$

However, the convergence of these series is rather slow, and the numerators are not easy to compute either. To illustrate this, let us consider the example of the eigenvalue 1 (or, by symmetry, the eigenvalue  $-1$ ).

By means of a computer program, we were able to compute the values  $\mathbb{E}(n_1(\mathcal{R}_k))$  explicitly for  $k \leq 30$ . We have

$$\sum_{k=1}^{30} \frac{\mathbb{E}(n_1(\mathcal{R}_k))}{k(k+1)} \approx 0.048771.$$

However, the only a-priori estimate that we have to deal with the tail of the series is  $|\mathbb{E}(n_1(\mathcal{R}_k))| \leq 1$ , so

$$\left| \sum_{k=31}^{\infty} \frac{\mathbb{E}(n_1(\mathcal{R}_k))}{k(k+1)} \right| \leq \sum_{k=31}^{\infty} \frac{1}{k(k+1)} = \frac{1}{31},$$

which means that we only get the very weak bound  $\mu_{\text{rec},1} \in [0.016512, 0.081029]$ . By extrapolation from the available numerical data, we obtain heuristically that  $\mu_{\text{rec},1}$  is about 0.03. It would certainly be desirable to have better methods to calculate the constants in our central limit theorems numerically.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

No data was used for the research described in the article.

#### Appendix A. Benjamini–Schramm convergence

Recall that a sequence of graphs  $G_n$  is said to converge to  $G$  in the Benjamini–Schramm sense if for every positive integer  $r$ , the distribution of the  $r$ -neighbourhood of a uniformly random vertex of  $G_n$  converges to the distribution of the  $r$ -neighbourhood of the root of  $G$ . In this appendix, we explicitly describe the Benjamini–Schramm limit of recursive trees and binary increasing trees, with a proof in the former case. As mentioned in the introduction, existence of these limits implies pointwise convergence of the spectral measure for both models.

**Theorem A.1.** *Let  $\mathcal{R}_n$  denote a random recursive tree of order  $n$ . The sequence  $(\mathcal{R}_n)_{n \geq 1}$  converges in the Benjamini–Schramm sense to a semi-infinite tree  $\mathcal{R}$  that can be constructed as follows:*

- Start with a one-ended path of vertices  $v_1, v_2, \dots$ .
- Determine positive integer random variables  $X_1, X_2, \dots$  recursively in the following way:  $P(X_1 = k) = \frac{1}{k(k+1)}$  for all  $k \geq 1$ , and given  $X_1, X_2, \dots, X_{r-1}$ ,

$$P(X_r = k) = \frac{Y_{r-1} + 1}{(k + Y_{r-1})(k + Y_{r-1} + 1)}, \text{ with } Y_{r-1} = X_1 + \dots + X_{r-1}.$$

- For every  $i \geq 1$ , generate a random recursive tree with  $X_i$  vertices and identify its root with  $v_i$  (labels can be ignored).

**Proof.** We use an elementary counting argument: let  $x_1, x_2, \dots, x_r$  be given positive integers. We determine the total number of times a vertex  $v_1$  occurs in all  $(n - 1)!$  recursive trees of order  $n$  such that the following holds: if  $v_2, v_3, \dots, v_r$  are the direct ancestors of  $v_1$  (i.e.,  $v_i$  is a child of  $v_{i+1}$  for every  $i < r$ ), then the tree  $S_i$  consisting of  $v_i$  and all its descendants, except for  $v_{i-1}$  and its descendants if  $i > 1$ , has  $x_i$  vertices. See Fig. 1.

Set  $y_i = x_1 + \dots + x_i$  for all  $i$ . We pick the label of  $v_r$  first and denote it by  $\ell$ . All descendants of  $v_r$  then need to have a label greater than  $\ell$ . So we have  $\binom{n-\ell}{y_{r-1}}$  choices for these labels. Once we have selected them, there are  $\binom{y_{r-1}}{x_{r-1}}$  choices for

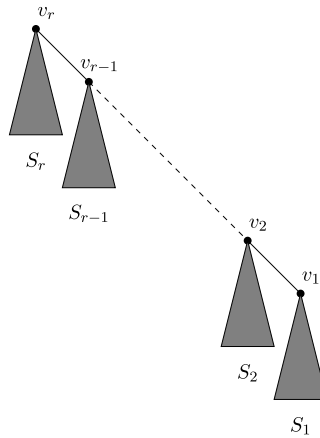


Fig. 1. The vertices  $v_1, \dots, v_r$  and the trees attached to them.

the labels of vertices in  $S_r$  (except for  $v_r$ ). The smallest of the remaining labels needs to be the label of  $v_{r-1}$ , and we have  $\binom{y_{r-1}-1}{x_{r-1}-1}$  choices for the labels of vertices in  $S_{r-1}$  (except for  $v_{r-1}$ ), and so on. This gives us

$$\binom{n-\ell}{y_r-1} \binom{y_r-1}{x_r-1} \binom{y_{r-1}-1}{x_{r-1}-1} \cdots \binom{y_1-1}{x_1-1}$$

choices. Since there are  $(x_i - 1)!$  recursive trees of order  $x_i$  (given the labels), there are  $\prod_{i=1}^r (x_i - 1)!$  possibilities for the trees  $S_1, S_2, \dots, S_r$ , which yields

$$\binom{n-\ell}{y_r-1} \frac{(y_r-1)!}{y_{r-1}y_{r-2} \cdots y_1}$$

combinations. Lastly, there are  $(n - y_r - 1)!$  different recursive trees that can be formed by the remaining vertices, and  $\ell - 1$  choices for the parent vertex of  $v_r$ . We sum the resulting product over all  $\ell$  and make use of the identity

$$\sum_{\ell=1}^n \binom{n-\ell}{y_r-1} (\ell-1) = [u^{y_r-1}] \sum_{\ell=1}^n (1+u)^{n-\ell} (\ell-1) = [u^{y_r-1}] \frac{(1+u)^n - nu - 1}{u^2} = \binom{n}{y_r+1}$$

for  $n > y_r$ . Putting everything together, we obtain

$$\binom{n}{y_r+1} (n - y_r - 1)! \frac{(y_r-1)!}{y_{r-1}y_{r-2} \cdots y_1} = \frac{n!}{(y_r+1)y_r y_{r-1} \cdots y_1}$$

occurrences of such a vertex  $v_1$ .

This means that if a random vertex  $v_1$  is chosen in a random recursive tree of order  $n$ , the probability that it and its immediate ancestors  $v_2, \dots, v_r$  satisfy the conditions is

$$\frac{1}{(y_r+1)y_r y_{r-1} \cdots y_1}.$$

In particular, the probability that  $v_1$  has  $x_1$  descendants (including itself) is  $\frac{1}{y_1(y_1+1)} = \frac{1}{x_1(x_1+1)}$  (this is a well-known fact, see [3, Section 3.2]), and given  $x_1, \dots, x_{r-1}$ , the probability that  $v_r$  has  $x_r$  descendants that are not descendants of  $v_{r-1}$  is

$$\frac{(y_{r-1}+1)y_{r-1}y_{r-2} \cdots y_1}{(y_r+1)y_r y_{r-1} \cdots y_1} = \frac{y_{r-1}+1}{y_r(y_r+1)}$$

for every  $n > y_r$ , in agreement with the construction of the random variables  $X_1, X_2, \dots$  in the statement of the theorem. Finally, the combinatorial argument also shows that all  $\prod_{i=1}^r (x_i - 1)!$  combinations of recursive trees attached to  $v_1, \dots, v_r$  are equally likely, so conditioned on  $x_1, \dots, x_r$  they can be regarded as independent random recursive trees. So we see that the distribution of every fixed  $r$ -neighbourhood of a random vertex of  $\mathcal{R}_n$  converges to the distribution of the  $r$ -neighbourhood of  $v_1$  in  $\mathcal{R}$ . This completes the proof.  $\square$

The following theorem is proven in a completely analogous fashion:

**Theorem A.2.** Let  $\mathcal{B}_n$  denote a random binary increasing tree of order  $n$ . The sequence  $(\mathcal{B}_n)_{n \geq 1}$  converges in the Benjamini–Schramm sense to a semi-infinite tree  $\mathcal{B}$  that can be constructed as follows:

- Start with a one-ended path of vertices  $v_1, v_2, \dots$ .
- Determine positive integer random variables  $X_1, X_2, \dots$  recursively in the following way:  $P(X_1 = k) = \frac{2}{(k+1)(k+2)}$  for all  $k \geq 1$ , and given  $X_1, X_2, \dots, X_{r-1}$ ,

$$P(X_r = k) = \frac{2(Y_{r-1} + 1)(Y_{r-1} + 2)}{(k + Y_{r-1})(k + Y_{r-1} + 1)(k + Y_{r-1} + 2)}, \quad \text{with } Y_{r-1} = X_1 + \dots + X_{r-1}.$$

- Generate a random binary increasing tree with  $X_1$  vertices and identify its root with  $v_1$ . For every  $i \geq 2$ , generate a random binary increasing tree with  $X_i - 1$  vertices and attach its root to  $v_i$  by an edge (labels can be ignored).

## References

- [1] M. Abért, P. Csikvári, T. Hubai, Matching measure, Benjamini-Schramm convergence and the monomer-dimer free energy, *J. Stat. Phys.* 161 (1) (2015) 16–34.
- [2] M. Abért, A. Thom, B. Virág, Benjamini-Schramm convergence and pointwise convergence of the spectral measure, Preprint, <https://users.renyi.hu/~abert/luckapprox.pdf>, 2015.
- [3] D. Aldous, Asymptotic fringe distributions for general families of random trees, *Ann. Appl. Probab.* (1991) 228–266.
- [4] G.W. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.
- [5] Z. Bai, J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, second edition, Springer Series in Statistics, Springer, New York, 2010.
- [6] M. Bauer, O. Golinelli, On the kernel of tree incidence matrices, *J. Integer Seq.* 3 (1) (2000) 00.14.
- [7] I. Benjamini, O. Schramm, Recurrence of distributional limits of finite planar graphs, *Electron. J. Probab.* 6 (2001) 23.
- [8] F. Bergeron, P. Flajolet, B. Salvy, Varieties of increasing trees, in: *Colloquium on Trees in Algebra and Programming*, Springer, 1992, pp. 24–48.
- [9] S. Bhamidi, S.N. Evans, A. Sen, Spectra of large random trees, *J. Theor. Probab.* 25 (3) (2012) 613–654.
- [10] C. Bordenave, M. Lelarge, J. Salez, The rank of diluted random graphs, *Ann. Probab.* 39 (3) (2011) 1097–1121.
- [11] D. Cvetković, M. Doob, H. Sachs, *Spectra of Graphs: Theory and Application*, Pure Appl. Math., Academic Press, 1980.
- [12] D. Cvetković, P. Rowlinson, S. Simić, *An Introduction to the Theory of Graph Spectra*, London Mathematical Society Student Texts, vol. 75, Cambridge University Press, 2010.
- [13] K. Dadedzi, S. Wagner, On the spectrum of simply generated trees, 2023, in preparation.
- [14] M. Drmota, *Random Trees: An Interplay Between Combinatorics and Probability*, Mathematics and Statistics, Springer, 2009.
- [15] S.R. Finch, *Mathematical Constants*, Encyclopedia of Mathematics and Its Applications., vol. 94, Cambridge University Press, Cambridge, 2003.
- [16] P. Flajolet, R. Sedgewick, *Analytic Combinatorics*, 1st edition, Cambridge University Press, New York, USA, 2009.
- [17] M. Fuchs, C. Holmgren, D. Mitsche, R. Neininger, A note on the independence number, domination number and related parameters of random binary search trees and random recursive trees, *Discrete Appl. Math.* 292 (2021) 64–71.
- [18] A. Gut, *Probability: a Graduate Course*, second edition, Springer Texts in Statistics, Springer, New York, 2013.
- [19] C. Holmgren, S. Janson, Limit laws for functions of fringe trees for binary search trees and random recursive trees, *Electron. J. Probab.* (2015).
- [20] S. Janson, Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees, *Random Struct. Algorithms* 48 (1) (2016) 57–101.
- [21] S. Janson, On the independence number of some random trees, *Electron. Commun. Probab.* 25 (2020) 63.
- [22] A. Meir, J.W. Moon, The expected node-independence number of various types of trees, in: *Recent Advances in Graph Theory*, Proc. Second Czechoslovak Sympos., Prague, 1974, Academia, Prague, 1975, pp. 351–363.
- [23] B. Mohar, Computing the characteristic polynomial of a tree, *J. Math. Chem.* 3 (1989) 403–406.
- [24] A. Panholzer, H. Prodinger, Level of nodes in increasing trees revisited, *Random Struct. Algorithms* 31 (2) (2007) 203–226.
- [25] D. Ralaivaosaona, S. Wagner, A central limit theorem for additive functionals of increasing trees, *Comb. Probab. Comput.* 28 (4) (2019) 618–637.
- [26] J. Salez, Every totally real algebraic integer is a tree eigenvalue, *J. Comb. Theory, Ser. B* 111 (2015) 249–256.
- [27] J. Salez, Spectral atoms of unimodular random trees, *J. Eur. Math. Soc.* 22 (2) (2020) 345–363.
- [28] A.J. Schwenk, Almost all trees are cospectral, in: *New Directions in the Theory of Graphs*, Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich., 1971, Academic Press, New York, 1973, pp. 275–307.
- [29] T. Tao, *Topics in Random Matrix Theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012.
- [30] S. Wagner, Central limit theorems for additive tree parameters with small toll functions, *Comb. Probab. Comput.* 24 (1) (2015) 329–353.
- [31] E.P. Wigner, On the distribution of the roots of certain symmetric matrices, *Ann. Math.* (2) 67 (1958) 325–327.