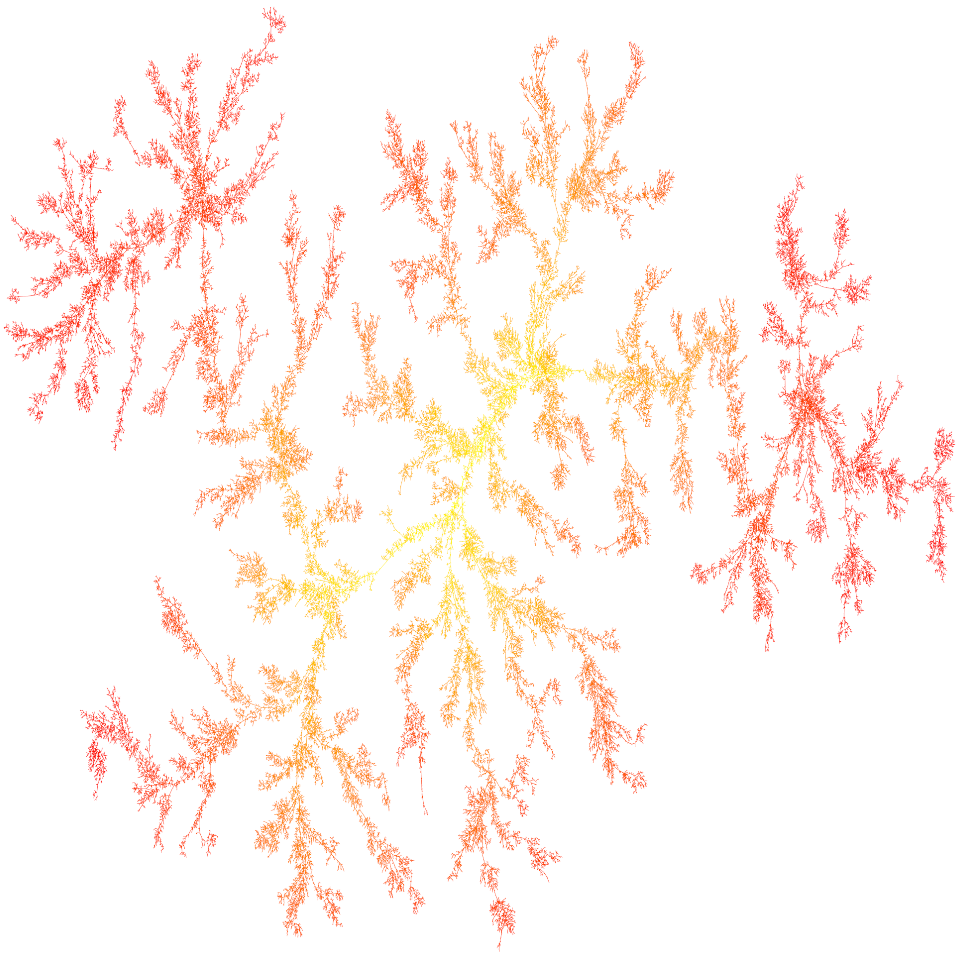
AN INTRODUCTION TO RANDOM TREES AND THEIR LOCAL LIMITS

by

Benedikt Stufler



Introduction

These are lecture notes for the ISM Discovery School 2023 on random trees, graphs and maps. We cover a selection of standard topics presented in the books [6, 5] and papers [2, 3, 4, 1, 9, 7].

Shouldn't trees be growing outdoors? — In combinatorics, the term *tree* refers to a specific type of graph. A tree is a graph that is acyclic (has no cycles) and connected (there is a path between any two vertices). Equivalently, a tree is a collection of vertices (also called nodes) joined by edges, so that there is a unique path between any pair of nodes. When talking about combinatorial trees it is custom to borrow vocabulary from botany, and refer to leaves, branches, and roots. Family trees serve as an additional source of vocabulary. In rooted trees we can refer to specific nodes as parents, children, siblings, and so on.

Are there any applications? — Trees serve as the foundation for various important data structures and algorithms in computer science. Examples include binary trees, heaps, trie structures (used in efficient string search), segment trees (used for range queries), and B-trees (used in databases). These structures provide efficient storage, retrieval, and manipulation of data, enabling faster and more optimized algorithms.

Why should we generate trees at random? — Any mathematician will tell you that buying lottery tickets is in general not a smart investment. This is a good example of how studying the average-case is just as important as performing a worst-case and best-case analysis. Similarly, studying the typical behaviour of combinatorial objects related to data structures helps to make prediction on the typical performance of algorithms. Equally significant, mathematics is driven by the pursuit of solving complex and challenging problems, and the field of random trees offers several such challenges.

1. Plane trees

Trees can be rooted or unrooted, labelled or unlabelled, and ordered or unordered. Any of the eight combinations of these adjectives refers to a class of trees that has been studied, and there are many more. In order to showcase the methods used to study random trees we focus on the class of plane trees, that is, trees that are rooted, ordered, and unlabelled.

That is, a plane tree has a unique root vertex, the set of children of each vertex is endowed with a linear order, and we view any two such trees as equivalent if one can be obtained from the other by relabelling vertices. This informal definition is best understood by looking at Figure 1.

1.1. Enumeration. — One of the first questions when studying a class of combinatorial structures is how many such structures with a given size are there? Hence, we would like to calculate the number t_n of plane trees with n vertices.

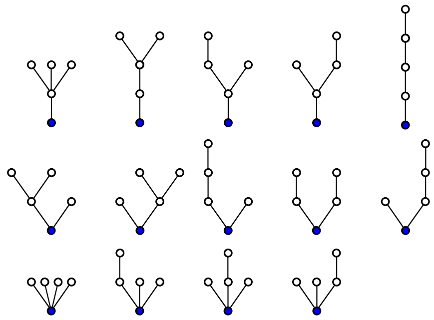


FIGURE 1. All plane trees with five vertices. The root vertices are coloured blue.

In enumerative combinatorics, such a problem is often solved trough the use of power series. We define the ordinary generating series

$$T(z) = \sum_{n \geq 1} t_n z^n = z + z^2 + 2z^3 + 5z^4 + 14z^5 + 42z^6 + 132z^7 + 429 + \ldots$$

whose coefficients enumerate plane trees. Observe that the nth coefficient

$$[z^n]T(z)^2 = \sum_{n \geq 1} \left(\sum_{i+j=n} t_i t_j \right) z^n$$

equals the number of pairs of plane trees whose sizes sum up to n. Likewise, for any integer $k \geq 1$ the kth power $T(z)^k$ enumerates sequences of k plane trees according to their total size. Since an arbitrary plane tree with n vertices consists of a root vertex that is joined to a sequence of smaller plane trees with total size $n-1$ we obtain

$$T(z) = z(1 + T(z) + T(z)^2 + \ldots) = z/(1 - T(z)).$$

This quadratic equation in the ring $\mathbb{R}[[z]]$ of formal power series has two solutions:

$$T(z) \in \left\{ \frac{1}{2}\left(1 - \sqrt{1-4z}\right), \frac{1}{2}\left(1 + \sqrt{1-4z}\right) \right\},$$

with the convention that the square root represents the binomial series

$$\sqrt{1-4z} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-4z)^n.$$

The second solution can't be equal to $T(z)$ since it has negative coefficients. Hence we obtain

$$T(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right)$$

and, after a quick calculation,

$$t_n = \frac{1}{n}\binom{2n-2}{n-1}.$$

Stirling's formula $n! \sim \sqrt{2\pi n}(n/e)^n$ allows us to deduce

$$t_n \sim \frac{1}{4\sqrt{\pi}} n^{-3/2} 4^n$$

as $n \to \infty$.

1.2. Lagrange inversion and forests of plane trees. — We use the notation $[z^n]a(z)$ to denote the nth coefficient of some power series $a(z)$. The Lagrange inversion theorem provides information on the coefficients on the functional composition of power series:

Lemma 1.1 (Lagrange inversion, analytic version)

Let $a(z)$ and $g(z)$ denote power series with positive radii of convergence such that $a(0) = 0$ and $a'(0) \neq 0$. Let $b(z)$ denote the compositional inverse of $a(z)$, whose existence and uniqueness on some neighbourhood of the origin is ensured by the inverse function theorem. Then

$$[z^n]g(a(z)) = \frac{1}{n}[z^{n-1}]g'(z)\left(\frac{z}{b(z)}\right)^n.$$

Proof. — On suitable open neighbourhoods U and V of the origin the functions $a : U \to V$ and $b : V \to U$ are analytic, inverse to each other, and have non-vanishing first derivatives. Let $\gamma : [0,1] \to U, t \mapsto re^{2\pi t}$ denote a positively oriented circle with sufficiently small radius $r > 0$ so that γ takes values in U. Then the image $\tilde{\gamma} = a(\gamma)$ is a closed path in V. Cauchy's formula and our assumptions on $a(z)$ ensure that

$$\begin{aligned}[z^n]g(a(z)) &= \frac{1}{2\pi i}\int_\gamma \frac{g(a(z))}{z^{n+1}}\,\mathrm{d}z \\ &= \frac{1}{2\pi i}\int_{\tilde{\gamma}} \frac{g(u)b'(u)}{b(u)^{n+1}}\,\mathrm{d}u.\end{aligned}$$

Using

$$\frac{\mathrm{d}}{\mathrm{d}u}\frac{g(u)}{b(u)^n} = \frac{g'(u)}{b(u)^n} - n\frac{g(u)b'(u)}{b(u)^{n+1}},$$

it follows that

$$\frac{1}{2\pi i}\int_{\tilde{\gamma}} \frac{g(u)b'(u)}{b(u)^{n+1}}\,\mathrm{d}u = \frac{1}{2n\pi i}\int_{\tilde{\gamma}} \frac{g'(u)}{b(u)^n}\,\mathrm{d}u.$$

Our assumptions on $a(z)$ ensure that $\tilde{\gamma}$ has winding number

$$n(\tilde{\gamma}, 0) = \frac{1}{2\pi i}\int_{\tilde{\gamma}} \frac{1}{z}\,\mathrm{d}z = \frac{1}{2\pi i}\int_\gamma \frac{a'(z)}{a(z)}\,\mathrm{d}z = 1.$$

By Cauchy's formula, it follows that

$$[z^n]g(a(z)) = \frac{1}{n}[u^{n-1}]g'(u)\left(\frac{u}{b(u)}\right)^n.$$

□

Forests of $k \geq 1$ plane trees are enumerated by $T(z)^k$, and since $T(z)(1-T(z)) = z$ we may apply Lagrange inversion with $a(z) = T(z)$, $b(z) = z(1-z)$, and $g(z) = z^k$, yielding

$$[z^n]T(z)^k = \frac{k}{n}[u^{n-k}]\left(\frac{1}{1-u}\right)^n.$$

Enumerating in how many ways we can write a number $\ell \geq 0$ as sum of an ordered list of n non-negative integers we obtain

$$\left(\frac{1}{1-u}\right)^n = \sum_{\ell=0}^{\infty} \binom{\ell+n-1}{\ell} u^\ell.$$

Hence

$$[z^n]T(z)^k = \frac{k}{n}\binom{2n-k-1}{n-1}.$$

Remark 1.2. — *The same formula actually holds also for power series with radius of convergence zero. Care has to be taken when performing the composition operation* $(a \circ b)(z) = a(b(z))$ *in the ring of formal power series. For example, the identity*

$$T(z(1-z)) = z$$

holds in the ring of formal power series as well. However, for $F = T(z)$, $G = z(1-z)$ *and* $H = 1$ *we have*

$$(F \circ G) \circ H = 1 \neq 0 = F \circ (G \circ H).$$

1.3. Shape analysis of random plane trees. — Having enumerated plane trees according to their number of vertices, the next question is what can we say about their shape, in particular when the trees are large.

1.3.1. *The root degree.* — Let us start with an easy example. What is the probability for a uniform random n-vertex plane tree T_n to have the property, that the number $d^+_{\mathsf{T}_n}(o)$ of children of the root vertex o is equal to some given integer $k \geq 1$?

A plane with root degree k consists of a root vertex and a forest of k plane trees. Thus, dividing the number of such trees with n vertices by the total number of plane trees t_n with n vertices, we obtain

$$\mathbb{P}(d^+_{\mathsf{T}_n}(o) = k) = \frac{[z^{n-1}]T(z)^k}{t_n}.$$

Plugging in the exact formulas for obtained in the previous section we obtain after a short calculation

$$\lim_{n\to\infty} \mathbb{P}(d^+_{\mathsf{T}_n}(o) = k) = k2^{-k-1}.$$

The factor k suggests that the limiting probability corresponds to k equally likely events. We are going to see that this is because among the branches attached to the root there typically is a unique giant component. Hence, if the root has k children, there are k choices for which branch is macroscopic.

1.3.2. *Height of a random vertex.* — Say, we take a uniformly at random select vertex v_n of our random n-vertex plane tree T_n. What can we say of about the height $\mathrm{h}_{\mathsf{T}_n}(v_n)$?

It will be convenient to introduce some terminology. We call a plane tree that apart from its root vertex has an additional marked vertex (possibly equal to the root vertex) a marked plane tree. The path between the root and the marked vertex is called the spine. It's length is hence the height of the marked vertex. The marked vertex and all its descendants form the fringe subtree at the marked vertex.

Marked plane trees whose specified node has height $h \geq 0$ are enumerated by

$$T^{[h]}(z) := (T(z)^2/z)^h T(z) = T(z)^{2h+1}/z^h.$$

This is because each of the h non-marked spine vertices can be seen as the root of a tree to the left of the spine and as the root of a tree to the right of the spine. Since this counts the spine vertex twice, we need to divide by z, so that each non-marked vertex contributes a factor $T(z)^2/z$. The final factor $T(z)$ is due to the fringe subtree at the marked vertex.

Since each plane tree with n vertices has n possible marked vertices, the pair (T_n, v_n) is uniformly distributed among all n-vertex marked plane trees, and

$$\mathbb{P}(\mathrm{h}_{\mathsf{T}_n}(v_n) = h) = \frac{[z^n]T^{[h]}(z)}{nt_n} = \frac{[z^{n+h}]T(z)^{2h+1}}{nt_n}.$$

Here is a fact that we are going to use repeatedly:

Proposition 1.3. — *Suppose that $k = k(m)$ and $n = n(m)$ are sequences of positive integers that depend on some integer $m \geq 1$ and satisfy $k \to \infty$ and $n \to \infty$. If $x := k/\sqrt{n}$ remains bounded away from zero and infinity, then*

$$[z^n]T(z)^k \sim \frac{x}{n\sqrt{\pi}} 2^{2n-k-1} \exp(-x^2/4)$$

as $m \to \infty$.

Proof. — Using Stirling's formula we obtain

$$[z^n]T(z)^k = \frac{k}{n}\binom{2n-k-1}{n-1} \sim \frac{x}{n\sqrt{\pi}} \frac{(2n-k-1)^{2n-k-1}}{(n-1)^{n-1}(n-k)^{n-k}}.$$

We may rewrite this by

$$\frac{x}{n\sqrt{\pi}} 2^{2n-k-1} \exp\Bigg((2n-k-1)\log\left(1 - \frac{k+1}{2n}\right) - (n-1)\log\left(1 - \frac{1}{n}\right) - (n-k)\log\left(1 - \frac{k}{n}\right)\Bigg).$$

Using the Taylor approximation $\log(1+z) = z - \frac{z}{2} + O(z^3)$ for small z this simplifies asymptotically to the claimed expression. □

Given $0 < a < b$ we have hence uniformly for all $a \leq x \leq b$ such that $h := x\sqrt{n}$ is an integer that $\frac{2h+1}{\sqrt{n+h}} \sim 2x$ and hence

$$\mathbb{P}(\mathrm{h}_{\mathsf{T}_n}(v_n) = h) \sim \frac{2x}{n^2 t_n \sqrt{\pi}} 2^{2n-2} \exp(-x^2) \sim \frac{2x}{\sqrt{n}} \exp(-x^2).$$

There are approximately $\sqrt{n}(b-a)$ values $x \in [a,b]$ so that $x\sqrt{n}$ is an integer, and the distance between two such consecutive numbers is precisely $\frac{1}{\sqrt{n}}$. By Riemannian sum approximation it follows that

$$\mathbb{P}(a \le \mathrm{h}_{\mathsf{T}_n}(v_n)/\sqrt{n} \le b) = \sum_{\substack{a \le x \le b \\ x\sqrt{n} \in \mathbb{Z}}} \frac{2x}{\sqrt{n}} \exp(-x^2) \to \int_a^b 2x \exp(-x^2) \,\mathrm{d}x.$$

This standard argument shows that the local limit theorem we have verified implies distributional convergence

$$\mathrm{h}_{\mathsf{T}_n}(v_n)/\sqrt{n} \xrightarrow{d} Z,$$

for a random variable Z with density $2x \exp(-x^2)$.(1) This is the Rayleigh distribution with scale parameter $2^{-1/2}$.

1.3.3. *The Boltzmann model.* — The result on the height of specified vertices may be actually be strengthened. In order to state the strengthened version we define the Boltzmann model of plane trees as the random plane tree T that assume any finite plane tree T with probability

$$\mathbb{P}(\mathsf{T} = T) = 4^{-|T|}2.$$

Here $|T|$ denotes the number of vertices of a plane tree T. This way,

$$\mathbb{P}(|\mathsf{T}| = n) = 2t_n 4^{-n} \sim \frac{1}{2\sqrt{\pi}} n^{-3/2}$$

and

$$\mathbb{E}[z^{|\mathsf{T}|}] = 2T(z/4).$$

We let $(\mathsf{T}(i))_{i \ge 1}$ denote a family of independent copies of T. Proposition 1.3 ensures that given a compact set $K \subset]0, \infty[$ for all $x \in K$ such that $k = xn^2$ is an integer we have

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n |\mathsf{T}(i)| = k \right) &= [z^k] \left(2T(z/4)\right)^n \\ &= 2^n 4^{-k} [z^k] T(z)^n \\ &\sim n^{-2} \frac{1}{2\sqrt{\pi} x^{3/2}} \exp(-1/(4x)). \end{aligned}$$

As before, this local limit entails distributional convergence

$$n^{-2} \sum_{i=1}^n |\mathsf{T}(i)| \xrightarrow{d} X_{1/2}$$

for a random variable $X_{1/2} > 0$ having density $\frac{1}{2\sqrt{\pi} x^{3/2}} \exp(-1/(4x))$. The index $1/2$ indicates that this is a (strictly) $1/2$-stable law.

(1) The altitude of nodes in socalled simply generated families of trees was investigated by Meir and Moon [8]. The standard source on analytic combinatorics [6, Prop. IX.23] cites this result, but there appears to be a typographical error in the expression for A: It should read $A = \rho\tau\phi''(\tau)$, not $A = \tau\phi''(\tau)$.

1.3.4. *The skeleton of random plane trees.* — We aim to show a result that describes the asymptotic global and local structure of the random n-vertex plane tree T_n. The idea is as follows: In the section on the height of a marked vertex we considered the event that $2h+1 \approx \sqrt{n}$ trees have size $n-h \approx n$. The calculations leave some "wiggle room": if we perturb h by $o(\sqrt{n})$ and n by $o(n)$ then the asymptotic probability stays almost the same, just with an additional power of 2 as a factor. We are going to interpret this factor probabilistically in order to simultaneously describe the asymptotic behaviour of the $o(\sqrt{n})$-neighbourhoods of the root vertex and the uniformly selected vertex v_n in T_n.

Recall that to each spine vertex in a marked tree (T, v) correspond a left and right branch, except for the tip of the spine, where only the fringe subtree of the marked vertex is attached. For any integer $0 \le k < \mathrm{h}_T(v)$ we let $S_{\mathrm{root}}((T,v),k)$ denote the $2k$ branches corresponding to the first k spine vertices (in some canonical order), and $S_{\mathrm{tip}}((T,v),k)$ the $2k-1$ branches corresponding to the last k spine vertices. For $k \ge \mathrm{h}_T(v)$ we set these sequences to some placeholder value.

Proposition 1.4. — *Given a compact set $K \subset]0,\infty[$, consider any element $x \in K$ such that $h := x\sqrt{n}$ is an integer. Additionally, let $k = k_n$ denote a sequence of positive integers with $k = o(\sqrt{n})$, and choose a sequence $s' = s'_n$ so that $s' = o(n)$ but $s'/k^2 \to \infty$. Then uniformly for all sequences $S = (T_1, \ldots, T_{2k})$ and $S' = (T'_1, \ldots, T'_{2k+1})$ of plane trees with total size at most $s \le s'$ and all admissible $x \in K$ we have*

$$\begin{aligned}\mathbb{P}\left(S_{\mathrm{root}}(\mathsf{T}_n, v_n) = S, S_{\mathrm{tip}}(\mathsf{T}_n, v_n) = S', \mathrm{h}_{\mathsf{T}_n}(v_n) = h\right) \\ \sim \left(\prod_{i=1}^{2k} \mathbb{P}(\mathsf{T} = T_i)\right)\left(\prod_{i=1}^{2k+1} \mathbb{P}(\mathsf{T} = T'_i)\right) \frac{2x}{\sqrt{n}} \exp(-x^2)\end{aligned}$$

with, as we have shown before,

$$\mathbb{P}\left(\mathrm{h}_{\mathsf{T}_n}(v_n) = h\right) \sim \frac{2x}{\sqrt{n}} \exp(-x^2).$$

Furthermore,

$$d_{\mathrm{TV}}\left(\left(S_{\mathrm{root}}(\mathsf{T}_n, v_n), S_{\mathrm{tip}}(\mathsf{T}_n, v_n)\right), (\mathsf{T}(i))_{1 \le i \le 4k+1} \right) \to 0.$$

Proof. — In the event under consideration, the trees have a midsection of length $h-2k$. The remainder of the tree with the specified branches has total size $s-(2k-1)$, since we need to subtract the root vertices of the branches that we counted twice. Hence, similarly as for determining the height of the random vertex v_n,

$$\mathbb{P}\left(S_{\mathrm{root}}(\mathsf{T}_n, v_n) = S, S_{\mathrm{tip}}(\mathsf{T}_n, v_n) = S', \mathrm{h}_{\mathsf{T}_n}(v_n) = h\right) = \frac{[z^{\tilde{n}}](T(z)^2/z)^{\tilde{h}}}{n t_n}$$

with $\tilde{n} = n - s + 2k = n + o(n)$ and $\tilde{h} = h - 2k = h + o(\sqrt{n})$, so that

$$\frac{2\tilde{h}}{\sqrt{\tilde{n} + \tilde{h}}} \to 2x$$

and

$$2(\tilde{n} + \tilde{h}) - 2\tilde{h} - 1 = 2(n - s + 2k) - 1$$

By Proposition 1.3 it follows that

$$\begin{aligned}\frac{[z^{\tilde{n}}](T(z)^2/z)^{\tilde{h}}}{nt_n} &= \frac{[z^{\tilde{n}+\tilde{h}}]T(z)^{2\tilde{h}}}{nt_n} \\ &\sim \frac{2x}{n^2 t_n \sqrt{\pi}} 2^{2(n-s+2k)-1} \exp(-x^2) \\ &\sim \frac{2x}{\sqrt{n}} 4^{-s} 2^{4k+1} \exp(-x^2) \\ &\sim \left(\prod_{i=1}^{2k} \mathbb{P}(\mathsf{T} = T_i)\right) \left(\prod_{i=1}^{2k+1} \mathbb{P}(\mathsf{T} = T_i')\right) \frac{2x}{\sqrt{n}} \exp(-x^2).\end{aligned}$$

We already know that the total size of $4k+1$ independent copies of the Boltzmann model behaves asymptotically like $(4k+1)^2 X_{1/2}$. By our assumption $s/k^2 \to \infty$ it follows that the set $\mathcal{E}_n$ of forests of $4k+1$ trees with total size at most s satisfies

$$\mathbb{P}\left((\mathsf{T}(i))_{1 \le i \le 4k+1} \in \mathcal{E}_n\right) = 1 + o(1)$$

and

$$\sup_{A \subset \mathcal{E}_n} \left|\mathbb{P}\left((S_{\mathrm{root}}(\mathsf{T}_n, v_n), S_{\mathrm{tip}}(\mathsf{T}_n, v_n)) \in A\right) - \mathbb{P}\left((\mathsf{T}(i))_{1 \le i \le 4k+1} \in A\right)\right| \to 0.$$

This implies

$$\mathbb{P}\left((S_{\mathrm{root}}(\mathsf{T}_n, v_n), S_{\mathrm{tip}}(\mathsf{T}_n, v_n)) \in \mathcal{E}_n\right) = 1 + o(1)$$

and hence the restriction $A \subset \mathcal{E}_n$ may be dropped, yielding the claimed approximation in total variation. □

In other words, the tree T_n with a uniform marked vertex v_n looks like a spine of length about $Z\sqrt{n}$ with trees glued to it. The total size of these trees needs to add up so that the whole thing consists of n vertices, and any specified $o(\sqrt{n})$ number of those behave asymptotically like independent copies of the Boltzmann model. In particular, this describes the asymptotic local shape of T_n near its root and near a typical vertex.

The proof of Proposition 1.4 extends easily to any fixed number $\ell \ge 1$ of independent random vertices of T_n. Let us call an ℓ-tree a plane tree with ℓ leaves, labelled from 1 to ℓ, such that the root has one child and any other internal vertex (that is, any other non-leaf) has exactly two children. Hence, an ℓ-tree has $2\ell - 1$ edges. An $(\ell+1)$-tree can be obtained from an ℓ-tree in a unique way by selecting one of its edges e, adding a vertex v in its middle, and adding a child with label $\ell+1$ to v that lies either to the left or to the right of the former edge e. Thus, the number $t(\ell)$ of ℓ-trees satisfies $t(1) = 1$, $t(\ell+1) = (2\ell-1)2t(\ell)$, and hence

$$t(\ell) = 2^{\ell-1} \prod_{i=1}^{\ell-1} (2i-1).$$

Let $T^{(\ell)}$ be plane tree with ℓ marked vertices such that the subtree spanned the root vertex and the marked vertices may be obtained from an ℓ-tree T_{reduced} by blowing up each edge into a path of length larger than k. We let $\mathrm{h}(T^{(\ell)})$ denote the lengths of this $2\ell - 1$ paths. We call T_{reduced} the reduced tree of $T^{(\ell)}$ and denote it by

$R(T^{(\ell)})$. Similarly to the case $\ell = 1$, we may define a sequence $S(T^{(\ell)})$ of plane trees that glued to this subtree in $T^{(\ell)}$. Here the root contributes $2k$ such trees, each leaf contributes $2k+1$ such trees, and each branchpoint (or lowest common ancestor of distinct leaves) contributes $3+6k$ trees, exactly three of which are glued exactly to this lowest common ancestor. There are ℓ leaves, one root, and $\ell - 1$ branchpoints to consider, thus $S(T^{(\ell)})$ consists of

$$f(k,\ell) := (2k+1)\ell + 2k + (3+6k)(\ell-1) = 4k(2\ell-1) + 4\ell - 3$$

trees. Whenever the tree $T^{(\ell)}$ is not of the form considered here we set $S(T^{(\ell)})$, $\mathrm{h}(T^{(\ell)})$, and $R(T^{(\ell)})$ to some placeholder values.

Proposition 1.5. — *Let $\ell \geq 1$ be an integer. Given a compact set $K \subset]0,\infty[$, consider a sequence $x = (x_1, \ldots, x_{2\ell-1}) \in K^{2\ell-1}$ such that $h_i := x_i\sqrt{n}$ is an integer for all $1 \leq i \leq 2\ell - 1$. Additionally, let $k = k_n$ denote a sequence of positive integers with $k = o(\sqrt{n})$, and choose a sequence $s = s_n$ so that $s = o(n)$ but $s/k^2 \to \infty$. Let $\mathsf{T}_n^{(\ell)}$ denote the tree T_n equipped with ℓ uniformly and independently selected marked vertices. Then uniformly for all such sequences $x \in K^{2\ell-1}$, all sequences $S = (T_i)_{1 \leq i \leq f(k,\ell)}$ of $f(k,\ell)$ plane trees with total size at most s, and all ℓ-trees T_{reduced} we have*

$$\mathbb{P}\left(R(\mathsf{T}_n^{(\ell)}) = T_{\text{reduced}}, S(\mathsf{T}_n^{(\ell)}) = S, \mathrm{h}(\mathsf{T}_n^{(\ell)}) = x\right) \sim \left(\prod_{i=1}^{f(k,\ell)} \mathbb{P}(\mathsf{T} = T_i)\right) \frac{2\|x\|_1}{\sqrt{n}} \exp(-\|x\|_1^2).$$

In other words, each reduced tree is assumed equally likely with probability $1/t(\ell)$, the specified trees behave asymptotically like independent copies of the Boltzmann model, and the distances between leaves, branchpoints, and the root admit a joint limit law with density $t(\ell)\frac{2\|x\|_1}{\sqrt{n}} \exp(-\|x\|_1^2)$.

1.3.5. *Local convergence.* — Given a (finite) plane tree T and an integer $k \geq 0$ let the neighbourhood $U_k(T)$ denote the plane tree obtained by cutting away all vertices with height larger than k. Let $\hat{\mathsf{T}}$ denotes the infinite plane tree obtained by taking a half-infinite path and gluing to each of its vertices two independent copies of the Boltzmann model T. The first vertex of the path is considered the root of $\hat{\mathsf{T}}$. Proposition 1.4 shows that if $k_n = o(\sqrt{n})$ is a sequence of non-negative integers, then

$$d_{\mathrm{TV}}(U_{k_n}(\mathsf{T}_n), U_{k_n}(\hat{\mathsf{T}})) \to 0.$$

This entails (for k_n constant) local convergence

$$\mathsf{T}_n \xrightarrow{d} \hat{\mathsf{T}}$$

when viewed as random elements in a suitable space.

For any vertex v of T we let $f^{[k]}(T,v)$ denote the plane tree consisting of the kth ancestor of v and all of its descendants. We view $f^{[k]}(T,v)$ as marked at v and call it an extended fringe subtree. If the height of v is less than k than we set $f^{[k]}(T,v)$ to some place-holder value.

We let T^* denote the infinite plane tree that we construct starting with an half-infinite path $u_0, u_1, \ldots$ that we call its spine, such that u_i is the parent of u_{i-1} for all $i \geq 0$. We glue an independent copy of T to u_0, and for each $i \geq 1$ we glue two independent copies of T to u_i, one to the left of the spine and one to its right. Proposition 1.4 entails for $k_n = o(\sqrt{n})$

$$d_{\mathrm{TV}}\left(f^{[k_n]}(\mathsf{T}_n, v_n), f^{[k_n]}(\mathsf{T}^*)\right) \to 0.$$

For k_n constant this is called annealed convergence, and may also be rephrased as

$$(\mathsf{T}_n, v_n) \xrightarrow{d} \mathsf{T}^*$$

when viewed as random elements in a suitable space.

Proposition 1.5 also entails that when $v_n(1)$ and $v_n(2)$ are two independent random vertices of T_n, then for each $k \geq 0$

$$\mathbb{P}(f^{[k]}(\mathsf{T}_n, v_n(1)) = (T, v), f^{[k]}(\mathsf{T}_n, v_n(2)) = (T, v)) \to \mathbb{P}(f^{[k]}(\mathsf{T}^*) = (T, v))^2.$$

Setting

$$N_{(T,v)}(\mathsf{T}_n) = |\{u \in \mathsf{T}_n \mid f^{[k]}(\mathsf{T}_n, u) = (T, v)\}|,$$

we have

$$\mathbb{P}(f^{[k]}(\mathsf{T}_n, v_n(1)) = (T, v), f^{[k]}(\mathsf{T}_n, v_n(2)) = (T, v)) = n^{-2}\mathbb{E}[(N_{(T,v)}(\mathsf{T}_n))^2],$$

and hence $\mathbb{V}[N_{(T,v)}(\mathsf{T}_n)] = o(n^2)$. By Chebyshev's inequality it follows that

$$\frac{N_{(T,v)}(\mathsf{T}_n)}{n} \xrightarrow{p} \mathbb{P}(f^{[k]}(\mathsf{T}^*) = (T, v)).$$

This called quenched convergence and may actually be rephrased as convergence

$$\mathfrak{L}((\mathsf{T}_n, v_n) \mid \mathsf{T}_n) \xrightarrow{p} \mathfrak{L}(\mathsf{T}^*)$$

of random elements of a space of probability measures, with the limit being in fact deterministic.

1.3.6. *A connection to branching processes.* — The Boltzmann model admits a construction as genealogy tree of a branching process. To this end, let ξ denote a geometric random variable with distribution

$$\mathbb{P}(\xi = k) = 2^{-k-1}, \qquad k \geq 0.$$

Let $\tilde{\mathsf{T}}$ denote the tree obtained as follows. We start with a single root vertex that receives children according to an independent copy of ξ. If this value is zero we stop. Otherwise, each child of the root receives offspring according to an independent copy of ξ. That is, one copy for each child. We then proceed in the same way for the grandchildren of the root vertex, and so on. Thus, $\tilde{\mathsf{T}}$ is the genealogical tree of a branching process with reproduction mechanism governed by ξ.

For any finite plane tree T we have

$$\begin{aligned}\mathbb{P}(\tilde{\mathsf{T}} = T) &= \prod_{v \in T} \mathbb{P}(\xi = d_T^+(v)) \\ &= 2^{-\sum_{v \in T}(d_T^+(v)+1)} \\ &= 4^{-|T|}2 \\ &= \mathbb{P}(\mathsf{T} = T),\end{aligned}$$

with v ranging over the vertices of T, and $d_T^+(v)$ denoting the number of children of v. Hence, the tree $\tilde{\mathsf{T}} \overset{d}{=} \mathsf{T}$ follows the Boltzmann distribution for plane trees. In particular, it is almost surely finite.

Many of the results presented here for the uniform plane tree T_n extend to genealogical trees of branching processes conditioned to have n vertices, as long as the reproduction mechanism satisfies certain properties.

References

[1] D. Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1(2):228–266, 1991.

[2] D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[3] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

[4] D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[5] M. Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.

[6] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[7] G. Kersting. On the height profile of a conditioned galton-watson tree. 2011.

[8] A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian J. Math.*, 30(5):997–1015, 1978.

[9] B. Stufler. Local limits of large Galton-Watson trees rerooted at a random vertex. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(1):155–183, 2019.

Benedikt Stufler, Vienna University of Technology • *E-mail :* `benedikt.stufler at tuwien.ac.at`