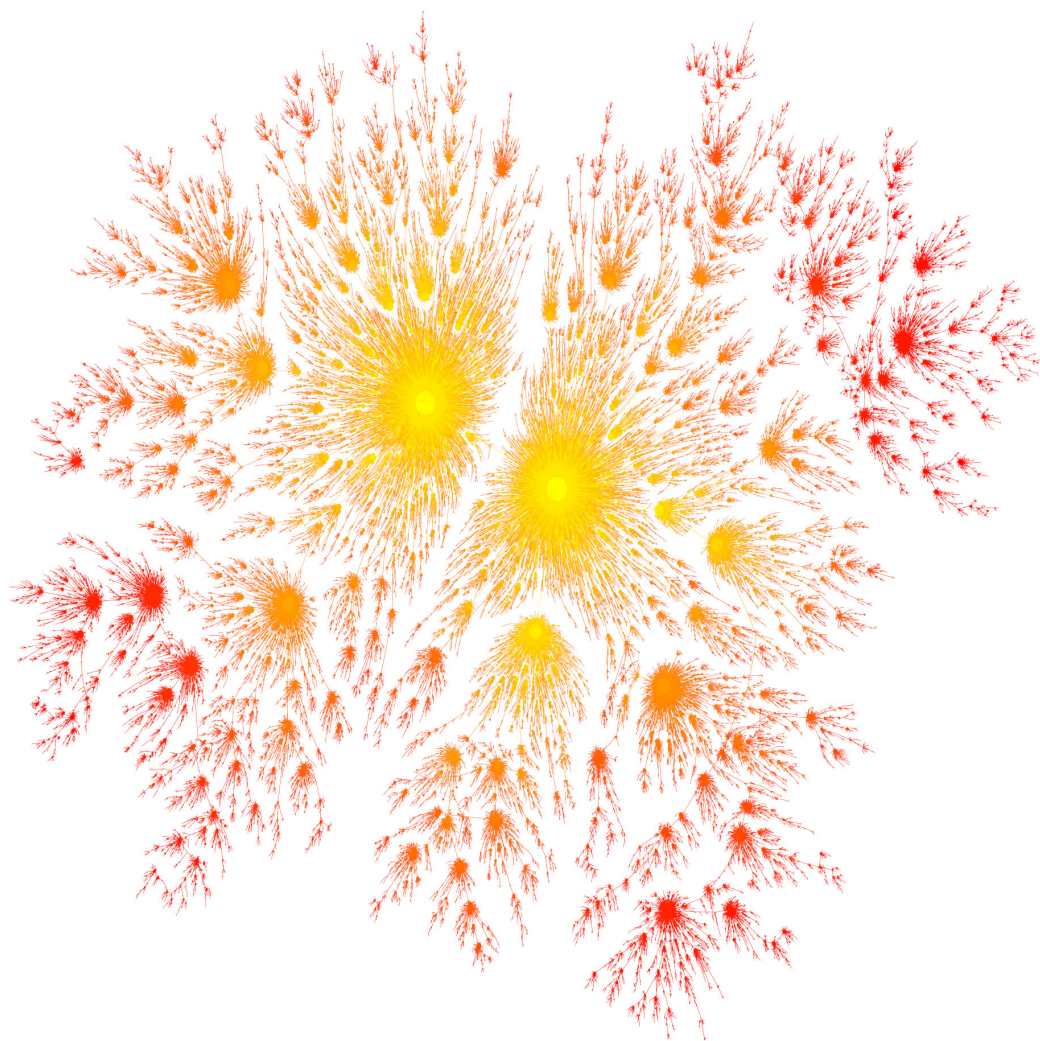

RANDOM TREES

by

Benedikt Stuffer



1. Introduction

The study of randomly generated trees and tree-like structures is a growing field with connections to stochastic processes, combinatorics, and computer science. The present notes summarize the content of a course on the topic given by the author at LMU Munich.

2. Basic vocabulary

2.1. Notation. — We let

$$\mathbb{N} = \{1, 2, 3, \dots\}$$

denote the positive integers and $\mathbb{N}_0$ the non-negative integers. We also use the notation

$$\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\} \quad \text{and} \quad \overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}.$$

For any integer $n \geq 0$ we set

$$[n] = \{1, \dots, n\}.$$

This includes the case $n = 0$, for which $[0] = \emptyset$. Summation and products of numbers are also subject to the usual conventions. That is, the sum of an empty collection (of numbers) equals 0, the product of an empty collection equals 1. Given a power series $f(z) = \sum_{n \geq 0} a_n z^n$ we use the notation $[z^n]f(z) = a_n$ to denote its n th coefficient.

All unspecified limits are as $n \rightarrow \infty$. We say an event (that depends on n) holds with high probability if its probability tends to 1 as n becomes large. Convergence in probability and distribution are denoted by $\xrightarrow{p}$ and $\xrightarrow{d}$. For any sequence $a_n > 0$ we let $o_p(a_n)$ denote an unspecified random variable Z_n such that $Z_n/a_n \xrightarrow{p} 0$. Likewise $O_p(a_n)$ is a random variable Z_n such that Z_n/a_n is stochastically bounded. In other words, for any $\epsilon > 0$ we may select a number M such that

$$\mathbb{P}(|Z_n| > Ma_n) < \epsilon$$

for all large enough n .

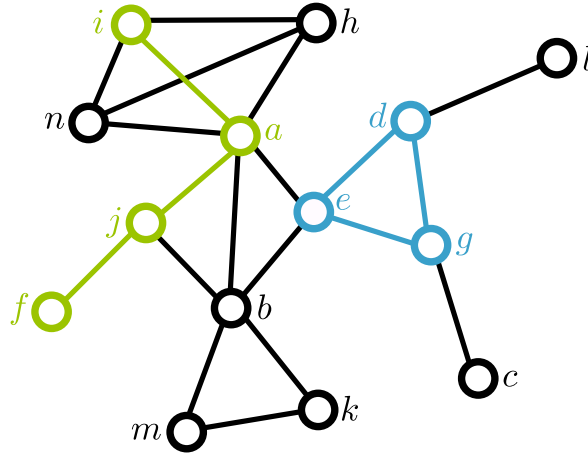


FIGURE 1. A connected graph with vertex set $\{a, b, \dots, n\}$. There is a green path from i to f and a blue cycle with vertices e , d , and g .

2.2. Graphs. —

DEFINITION 2.1 (GRAPH). — A graph $G = (V, E)$ is a pair of a set V and a collection E of 2-element subsets of V . Elements of V are called vertices. Elements of E are called edges. We say an edge $e \in E$ is incident to a vertex $v \in V$ if $v \in e$.

Figure 1 depicts an example of a graph. Vertices are displayed as dots and edges by lines.

REMARK 2.2. — There are many variants of the notion of a graph. For example, edges may be directed. Such a graph is called directed, or a digraph. Multigraphs are graphs where we allow multiple edges between the same pair of vertices, and edges with identical ends.

REMARK 2.3. — By abuse of notation, we will sometimes write $v \in G$ instead of $v \in V$. We will also let $|G| \in \overline{\mathbb{N}}_0$ denote the number of vertices in G .

DEFINITION 2.4 (VERTEX DEGREE). — Let $G = (V, E)$ be a graph and $v \in V$ a vertex. A vertex $u \in V$ is called a neighbour of v if $\{v, u\}$ is an edge of E . We say u and v are adjacent. The degree of v , denoted by $d_G(v) \in \overline{\mathbb{N}}_0$, is the number of neighbours of v .

DEFINITION 2.5 (PATH). — Let $n \geq 0$ be an integer. A path of length n is an $(n + 1)$ -vertex graph (V, E) of the form

$$V = \{v_i \mid 0 \leq i \leq n\}, \quad E = \{\{v_i, v_{i+1}\} \mid 0 \leq i < n\}.$$

The vertices v_0 and v_n are called endvertices or ends. All other vertices of the path are called inner vertices. A directed path is a path with an additional direction - from v_0 to v_n or from v_n to v_0 .

DEFINITION 2.6 (CYCLE). — *A cycle is a graph obtained by joining the end-points of a path of length at least 2 by an additional edge. The length of a cycle is its number of edges.*

DEFINITION 2.7 (SUBGRAPHS). — *A graph $G' = (V', E')$ is a subgraph of a graph $G = (V, E)$ if $V' \subset V$ and $E' \subset E$. We say G contains G' . We say G' is an induced subgraph, if any edge of G with both ends in G' is also an edge of G' . For any subset $U \subset V$ we may form the subgraph $G[U]$ induced by U .*

DEFINITION 2.8 (CONNECTED GRAPH). — *Two vertices $x, y \in V$ in a graph $G = (V, E)$ are joined by a path P in G if P is a subgraph of G and x, y are its endvertices. We say G is connected if any two vertices of G may be joined by some path in G .*

DEFINITION 2.9 (BRIDGES AND CUT VERTICES). — *Let G be a connected graph. Any vertex of G whose removal disconnects G is called a cut vertex. Any edge of G whose removal disconnects G is called a bridge.*

DEFINITION 2.10 (ROOTED GRAPH). — *A rooted graph is a pair (G, o) of a graph G and a vertex o of G . We say o is the root vertex of G , and G is rooted at o .*

DEFINITION / PROPOSITION 2.11 (GRAPH DISTANCE). — *Let $G = (V, E)$ denote a connected graph. The graph distance $d_G(x, y)$ of two vertices $x, y \in V$ is the length of a shortest path from x to y . The pair (V, d_G) is a metric space.*

DEFINITION 2.12 (HEIGHT AND DIAMETER). — *Let G denote a connected graph.*

1) *The diameter of G is given by*

$$D(G) = \sup_{x, y \in V} d_G(x, y).$$

2) *Let $v \in V$ be a vertex so that $G^\bullet = (G, v)$ is a rooted connected graph. The height of a vertex $u \in V$ is given by*

$$h_{G^\bullet}(u) = d_G(u, v).$$

3) *The height of $G^\bullet$ is defined by*

$$H(G^\bullet) = \sup_{u \in V} h_G(u).$$

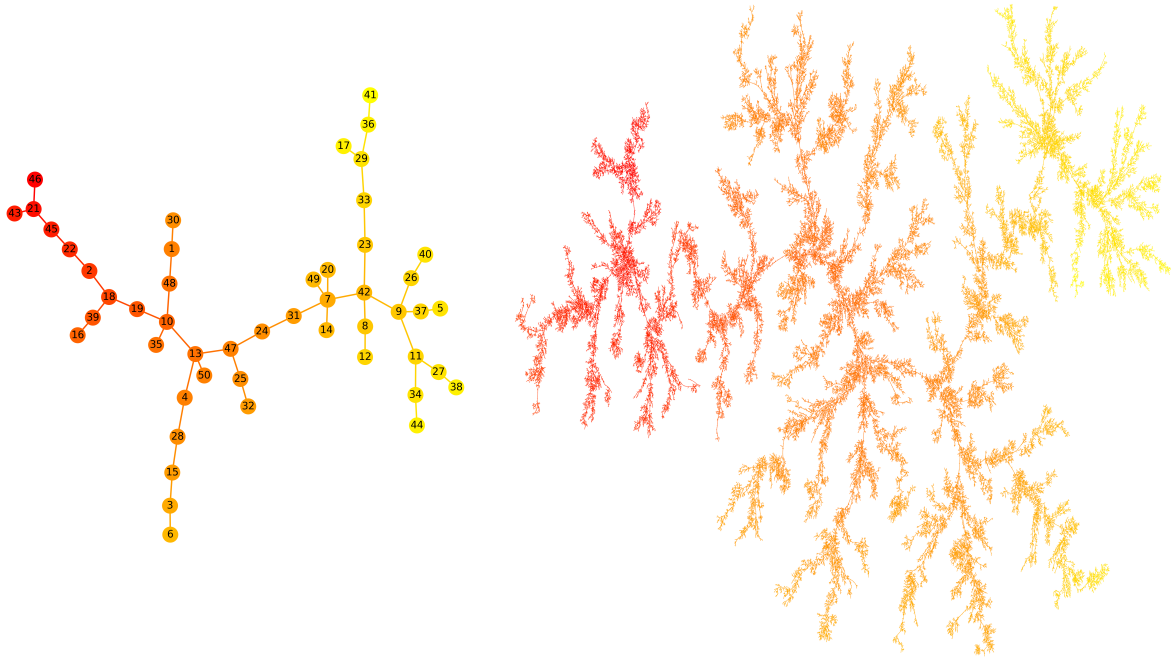


FIGURE 2. To the left: a uniformly generated tree with vertex set $[50] = \{1, \dots, 50\}$. The colours correspond to the graph distance from a uniformly selected root vertex, in this case the vertex 21. To the right: the same, but with half a million vertices.

2.3. Trees. —

DEFINITION 2.13 (TREE). — *A tree is a connected graph with no cycles.*

Figure 2 illustrates two randomly generated trees.

PROPOSITION 2.14. — *The following assertions are equivalent for a graph T .*

- i) T is a tree.
- ii) Any two vertices of T are linked by a unique path.
- iii) T is connected and removing any single edge of T disconnects the graph.
- iv) T has no cycles and adding any single edge that is not already present produces a cycle.

Proof. — See exercises. □

COROLLARY 2.15. — *A finite graph T with n vertices is a tree if and only if it is connected and has $n - 1$ edges.*

Proof. — See exercises. □

We will need some vocabulary for referring to vertices in a tree:

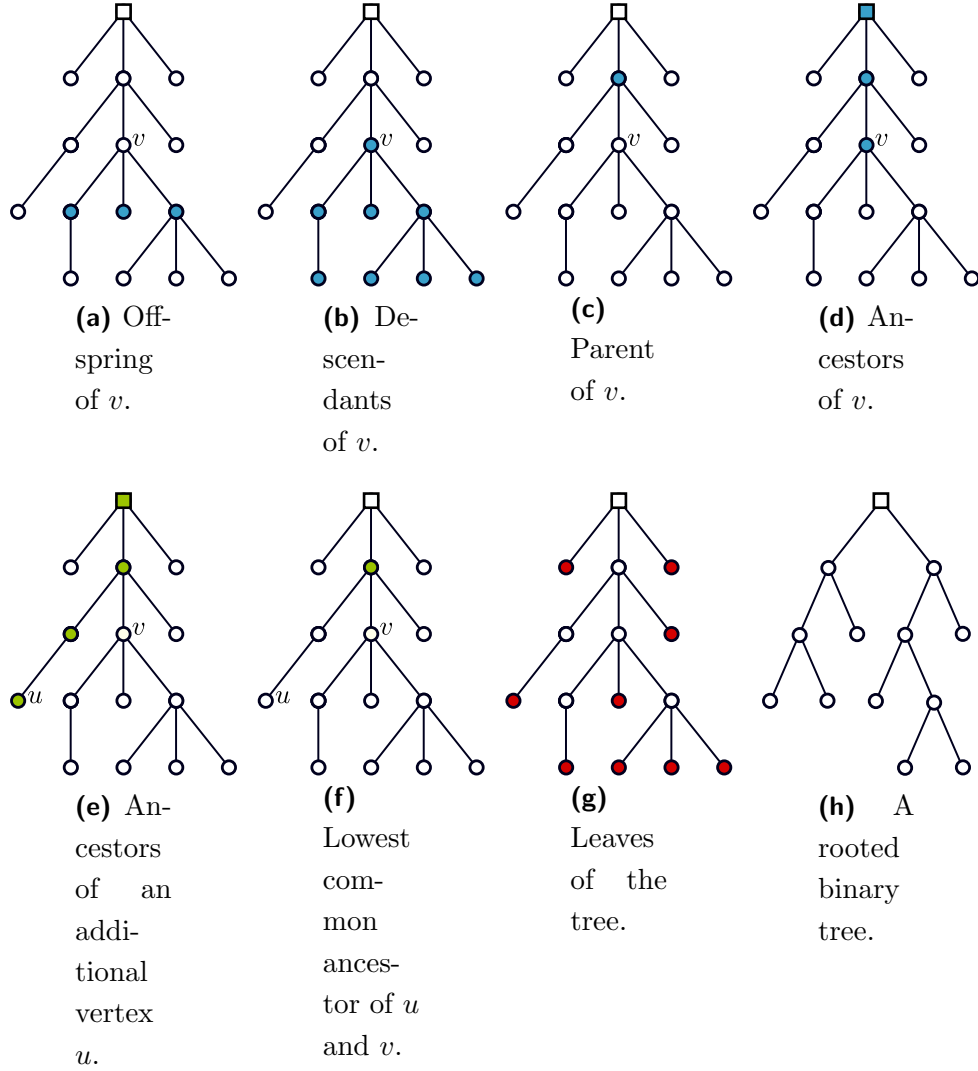


FIGURE 3. Notation for vertices in rooted trees. The root is marked with a square.

DEFINITION 2.16 (NOTATION FOR ROOTED TREES). — Let T be a rooted tree with root vertex o . Let u and v be vertices of T .

- 1) The ancestors of v are all vertices of T that lie on the path from v to the root o . In particular, o and v are both ancestors of v .
- 2) A vertex is called a descendant of v , if v is one of its ancestors.
- 3) The directed path from u to o and the directed path from v to o meet for the first time at a vertex w and then proceed in the same way all the way to the root o . We say w is the lowest (or youngest) common ancestor of u and v .
- 4) If $v \neq o$, then the second vertex x on the directed path from v to the root o is called the parent of v . Conversely, we say v is an offspring vertex or child of x .

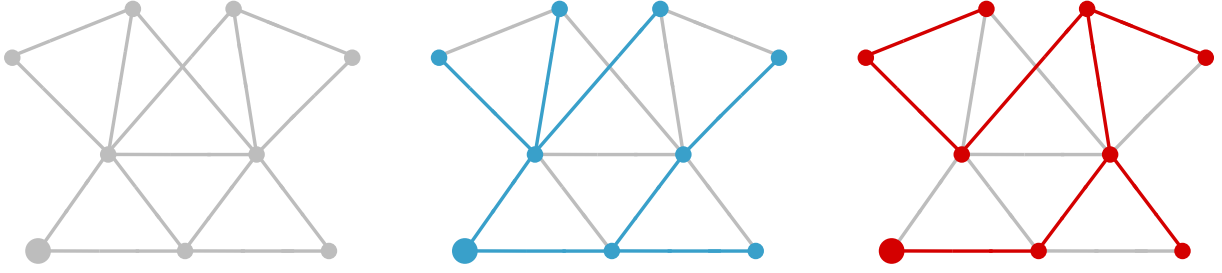


FIGURE 4. Spanning trees of a connected graph. The one in the middle was obtained via a breadth-first-search started at the vertex marked by a big disc. The one on the right by a depth-first-search started at the same vertex.

- 5) The number of offspring vertices of v is called its outdegree and denoted by $d_T^+(v) \in \overline{\mathbb{N}}_0$. The collection of all offspring vertices of v is called its offspring set.
- 6) A vertex is called a leaf, if it has outdegree 0.
- 7) A rooted tree is called d -ary for $d \geq 2$ an integer, if all of its vertices have outdegree d .

See Figure 3 for an illustration of these concepts.

REMARK 2.17. — The outdegree of a vertex makes only sense for rooted trees. For unrooted trees, different definition are often used, that may be at odds with the rooted world. For example, a leaf in an unrooted tree is a vertex with degree 1.

DEFINITION 2.18 (SPANNING TREE). — Let G be a connected graph. A spanning tree of G is a subtree containing all the vertices of G .

DEFINITION / PROPOSITION 2.19 (BREADTH-FIRST-SEARCH)

The breadth-first-search (BFS) of a rooted connected finite graph (G, o) maintains a queue Q of queued vertices, a list L of visited vertices, and a tree T .

1. In the beginning we mark the root o as queued, adding it to Q , and let T be a rooted tree consisting only of o .
2. In each step the first vertex v of Q is moved to the end of L , and all its neighbours in G that haven't been visited or queued so far are added to the end of Q and also added as offspring of v in T .
3. We proceed in this way until the queue Q is empty.

After the procedure terminates, the rooted tree T is a spanning tree and L is an ordering of the vertices of G .

See Figure 4 for an illustration.

PROPOSITION 2.20. — *Let $G^\bullet = (G, o)$ be rooted connected finite graph and T a rooted spanning tree obtained via a breadth-first-search of $G^\bullet$. Then*

$$h_{G^\bullet}(v) = h_T(v)$$

for all vertices v of G .

DEFINITION / PROPOSITION 2.21 (DEPTH-FIRST-SEARCH)

The depth-first-search (DFS) of a rooted connected finite graph (G, o) is an algorithm that produces an ordering of the vertices of G together with a spanning tree. Starting from o , we move along the edges of G , going only to vertices not visited so far. If there is no such vertex there are two cases. If our current vertex v satisfies $v \neq o$, then we traverse back along the edge by which the current vertex was first reached. If our current vertex is the root $v = o$, then we stop.

The edges traversed form a spanning tree, and the order in which each vertex is visited for the first time gives an ordering of the vertices of G .

Note that we may describe depth-first-search in the same way as we described breadth-first-search. The only difference is that in the second step, when adding the so far not visited and not queued neighbours of v to the queue Q , we add them at the beginning instead of the end.

Depth-first-search may be used, for example, to find all bridges of a connected finite graph G . See the exercises.

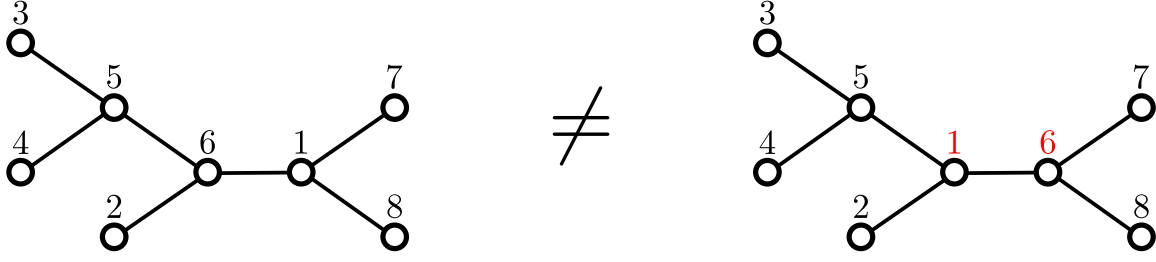


FIGURE 5. The two trees are not identical, but isomorphic.

2.4. Symmetries. —

DEFINITION 2.22 (GRAPH ISOMORPHISMS). — Let (V, E) and (V', E') be graphs. A graph isomorphism between the two is a bijective map

$$\varphi : V \rightarrow V'$$

such that for all $x, y \in V$ we have $\{x, y\} \in E$ if and only if $\{\varphi(x), \varphi(y)\} \in E'$. In this case we say the two graphs are isomorphic. An unlabelled graph is an isomorphism class of graphs.

Figure 5 illustrates an example of two isomorphic graphs.

DEFINITION 2.23 (GRAPH ISOMORPHISMS OF ROOTED GRAPHS)

Two rooted graphs are considered isomorphic, if there exists an isomorphism between the two that maps the root vertices to each other. An unlabelled rooted graph is an isomorphism class of rooted graphs.

DEFINITION 2.24 (SYMMETRIC GROUP). — Given a set X we let

$$\mathfrak{S}(X) = \{f : X \rightarrow X \mid f \text{ is bijective}\}$$

denote the corresponding symmetric group on X . Its unit element is the identity map id_X .

DEFINITION 2.25 (FIXED POINTS AND SUPPORT). — Let $f \in \mathfrak{S}(X)$ be a permutations. Any element $x \in X$ with $f(x) = x$ is called a fixed point of f . The subset

$$\{y \in X \mid f(y) \neq y\}$$

is called the support of f .

DEFINITION 2.26 (CYCLES). — An element $\tau \in \mathfrak{S}(X)$ is called a cycle, if its support may be ordered in a way $x_1, \dots, x_k$ with $k \geq 2$ such that

$$\tau(x_i) = \begin{cases} x_{i+1}, & 1 \leq i < k \\ x_1, & i = k \end{cases}.$$

In this case we say τ is a k -cycle. Two cycles in $\mathfrak{S}(X)$ are disjoint if they have disjoint support sets.

DEFINITION 2.27 (GROUP OPERATIONS). — Let M be a group with unit element e_M and let X denote a set. A (left-)group operation of M on X is a map

$$\mu : M \times X \rightarrow X, \quad (m, x) \mapsto m.x$$

such that for all $m, m' \in M$ and $x \in X$ we have $e_M.x = x$ and $m.(m'.x) = (mm').x$.

PROPOSITION 2.28. — Let M be a group and X a set. Any group operations

$$\mu : M \times X \rightarrow X$$

corresponds to a group homomorphism

$$\varphi_\mu : M \rightarrow \mathfrak{S}(X), m \mapsto (x \mapsto m.x).$$

Conversely, any group homomorphism

$$\varphi : M \rightarrow \mathfrak{S}(X)$$

corresponds to a group operation

$$\mu_\varphi : M \times X \rightarrow X, (m, x) \mapsto \varphi(m)(x).$$

Moreover, these two constructions are inverse to each other.

DEFINITION / PROPOSITION 2.29 (ORBITS). — Suppose that a group M operates on a set X . This defines an equivalence relation $\sim$ on X with $x \sim y$ if there is a group element $m \in M$ with $m.x = y$. The equivalence classes are called orbits of the group operation. We denote the orbit of an element $x \in X$ by $\text{Orbit}(x)$.

DEFINITION / PROPOSITION 2.30 (STABILIZERS). — Let a group M operate on a set X . For any $x \in X$ the stabilizer

$$\text{Stab}(x) = \{m \in M \mid m.x = x\}$$

is a subgroup of M .

LEMMA 2.31. — Let a group M operate on a set X . For any element $x \in X$, the map

$$M/\text{Stab}(x) \rightarrow \text{Orbit}(x), \quad \bar{m} \mapsto m.x$$

is well-defined and bijective.

Proof. — The map is well-defined: let $m, m' \in M$ such that $m \equiv m' \pmod{\text{Stab}(x)}$. That is, there is a group element $g \in \text{Stab}(x)$ with $m' = mg$. Hence

$$m'.x = (mg).x = m.(g.x) = m.x.$$

It is clearly surjective. To see that it is injective suppose that $m, m' \in M$ with $m.x = m'.x$. Then $m^{-1}m' \in \text{Stab}(x)$ and hence $m \equiv m' \pmod{\text{Stab}(x)}$. $\square$

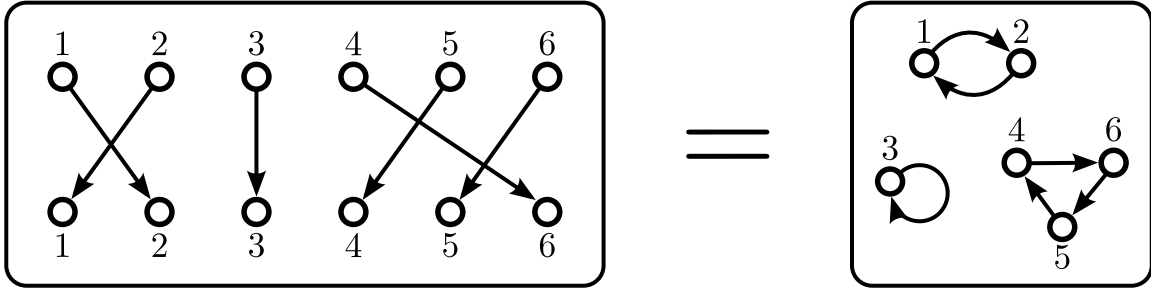


FIGURE 6. Decomposition of a permutation into disjoint cycles.

LEMMA 2.32. — *Let a group M operate on a set X . For any element $x \in X$ and any group element $m \in M$ it holds that*

$$\text{Stab}(g.x) = g\text{Stab}(x)g^{-1}.$$

Here we define $g\text{Stab}(x)g^{-1} := \{ghg^{-1} \mid h \in \text{Stab}(x)\}$.

Proof. — This easy observation is left as an exercise to the reader. $\square$

EXAMPLE 2.33. — *Let V be a fixed set and $\mathfrak{G}(V)$ the collection of all graphs with vertex set V . The symmetric group $\mathfrak{S}(V)$ operates on $\mathfrak{G}(V)$ via relabelling of vertices. The orbits of this operation correspond bijectively to unlabelled graphs whose vertex set has the same cardinality as V . The stabilizer of a graph is called its automorphism group. If V has $n \geq 1$ elements, then*

$$n! = |\text{Orbit}(G)| |\text{Stab}(G)|$$

for any graph $G \in \mathfrak{G}(V)$.

THEOREM 2.34. — *Let X be a finite set. Then any permutation on X may be represented as a product of disjoint cycles. Moreover, this product representation is unique up to the order of the factors.*

Proof. — For any permutation $\sigma \in \mathfrak{S}(X)$ the generated subgroup $\langle \sigma \rangle$ operates on the set X in a canonical way. The orbits with at least 2 elements correspond precisely to the disjoint cycles in the product composition of σ . Compare with Figure 6. $\square$

THEOREM 2.35 (CAYLEY). — *Let V be a fixed set with $n \geq 1$ elements. There are precisely n^{n-2} edge-configurations E such that (V, E) is a tree.*

Proof. — Consider the collection $\text{End}(V)$ of all functions $f : V \rightarrow V$. Clearly $\text{End}(V)$ has n^n elements. For any function $f : V \rightarrow V$ we may consider its recurrent points $v \in V$ that satisfy $f^k(v) = v$ for some $k \geq 1$, and its non-recurrent points $u \in V$, that satisfy $f^k(u) \neq u$ for all $k \geq 1$. Any non-recurrent point corresponds to a unique recurrent point. Conversely, the collection of all non-recurrent points corresponding to a given recurrent point u may be viewed as a tree rooted at u , with f acting like the parent function. The

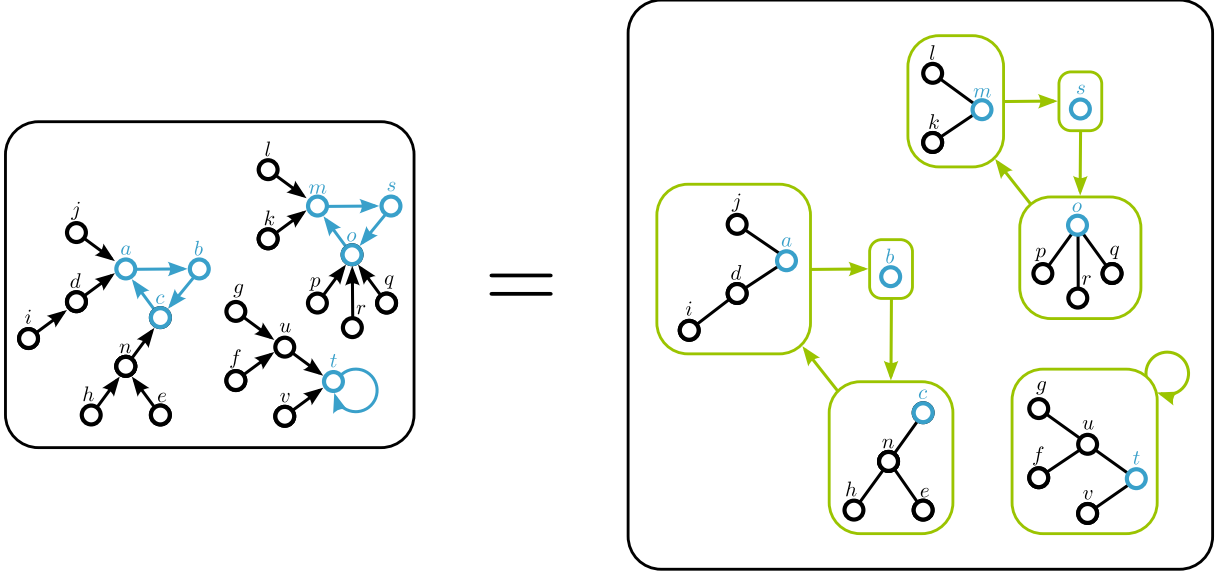


FIGURE 7. Decomposition of an endofunction into a composite structure consisting of a partition with a rooted tree structure on each partition class and a permutation structure on the collection of classes.

recurrent points may be grouped into cycles and fixed-points, in the same way as any permutation may be decomposed into cycles and fixed-points in a unique way. Summing up, there is a bijection between functions from $\text{End}(V)$ and triples $(\pi, \sigma, (A_Q)_{Q \in \pi})$ with π a partition of V , $\sigma \in \mathfrak{S}(V)$ a permutation, and A_Q a rooted tree with vertex set Q for each class $Q \in \pi$. Compare with Figure 7.

Now let us consider the collection $\mathfrak{A}^\bullet(V)$ of all doubly rooted trees with vertex set V . Any such tree has a unique directed path from the first root to the second. Any vertex on that path may be viewed as the root of a subtree consisting of all the non-path vertices we may reach without passing through another path-vertex. Summing up, there is a bijection between doubly rooted trees from $\mathfrak{A}^\bullet(V)$ and triples $(\pi, \omega, (A_Q)_{Q \in \pi})$ with π a partition of V , ω a linear order on V , and A_Q a rooted tree with vertex set Q for each partition class $Q \in \pi$. Compare with Figure 8.

The number of linear orders on a given set is equal to the number of permutations on that set. Hence there is a bijection between $\text{End}(V)$ and $\mathfrak{A}^\bullet(V)$, yielding that there are precisely n^n doubly rooted trees on a given n -element vertex set. Hence there are n^{n-1} rooted trees and n^{n-2} unrooted trees on that set. $\square$

REMARK 2.36. — *There many different proofs of Cayley's theorem. The one given here illustrates two important concepts. First, introducing roots serves as a starting point for decompositions. Second, we used a very important tool in enumerative combinatorics: partitions of sets where each partition class is endowed with an additional structure of*

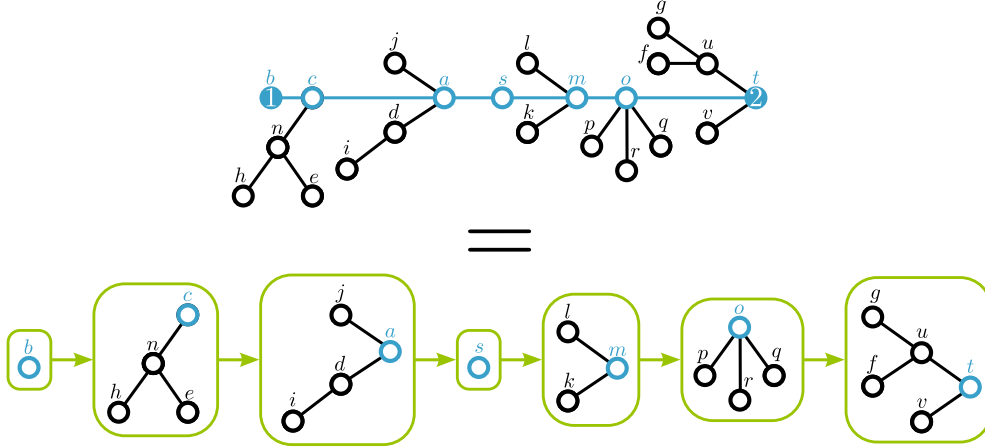


FIGURE 8. Decomposition of a doubly rooted tree into a linearly ordered of rooted trees.

some kind (such as graphs or trees) and where the collection of classes is endowed with an additional structure of another kind (such as linear orders or permutations).

There is also a third remark to be made here. The proof does not work for unlabelled graphs. The bijection between $\text{End}(V)$ and $\mathfrak{A}^\bullet(V)$ is not compatible with relabelling operations. Moreover, although any labelled graph with n vertices has n rooted versions, this is no longer true (in general) for unlabelled graphs.

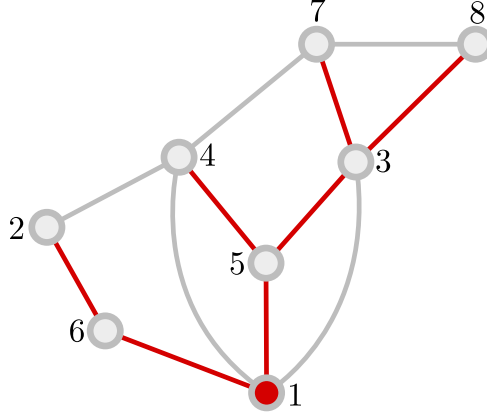


FIGURE 9. The spanning tree in the case the random walk starts like $(1, 6, 1, 5, 4, 5, 1, 6, 2, 4, 5, 3, 8, 3, 1, 3, 7, \dots)$.

3. The random walk construction of uniform spanning trees

3.1. Generating uniform spanning trees. — Let $G = (V, E)$ denote a finite (non-empty) connected graph. A simple random walk $(X_n)_{n \geq 0}$ starts at some (random or deterministic) vertex X_0 and walks in each step to a uniformly selected neighbour of the current vertex. We may view $(X_n)_{n \geq 0}$ as a discrete time Markov chain with transition matrix P of the form

$$P(v, w) = \begin{cases} \frac{1}{d_G(v)}, & \{v, w\} \in E \\ 0, & \text{otherwise} \end{cases} \quad \text{for } v, w \in V.$$

For each vertex v of G we let

$$T_v = \min\{n \geq 0 : X_n = v\}$$

denote the first hitting time.

PROPOSITION 3.1. — *The stationary distribution $(\pi(v))_{v \in V}$ of the simple random walk $(X_n)_{n \geq 0}$ on the finite connected graph G is given by*

$$\pi(v) = \frac{d_G(v)}{2|E|}, \quad v \in V.$$

Proof. — First, note that π defined as above is a probability distribution: In the sum $\sum_{v \in V} d_G(v)$ we count each edge twice (once for each end), and hence

$$\sum_{v \in V} d_G(v) = 2|E|.$$

Moreover, π is the stationary distribution, as for each vertex $v \in V$

$$\sum_{w \in V} \pi(w)P(w, v) = \frac{1}{2|E|} \sum_{w \in V} \mathbb{1}_{\{v, w\} \in E} = \frac{d_G(v)}{2|E|}.$$

□

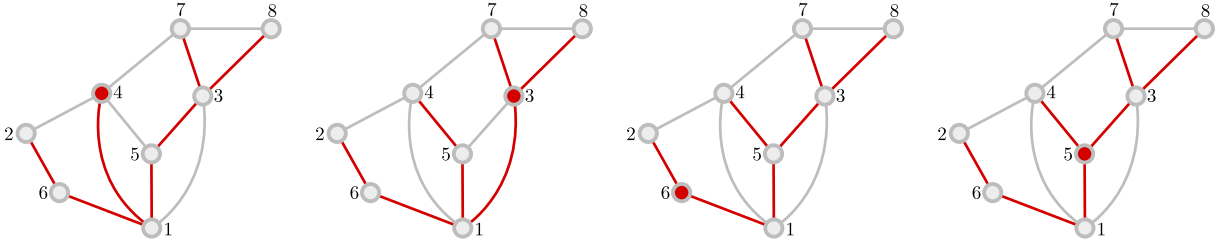


FIGURE 10. The possible predecessors of the spanning tree in Figure 9.

DEFINITION / PROPOSITION 3.2. — *The graph $\mathcal{T}$ with vertex set V and edges $\{X_{T_v-1}, v\}$ for $v \in V \setminus \{X_0\}$ is a spanning tree of G that we view as rooted at X_0 .*

Proof. — It is clear that the covering time is almost surely finite, so we eventually reach each vertex. The construction means that we start at X_0 and whenever we walk across an edge e to arrive at a vertex $v \in V \setminus \{X_0\}$ for the first time, we keep e . This way, we keep $|V| - 1$ edges in total, and we construct $\mathcal{T}$ via a growth procedure, attaching always one edge to a pre-existing tree to form a new tree in each step. $\square$

See Figure 9 for an illustration.

THEOREM 3.3 (ALDOUS, BRODER). — *If we view $\mathcal{T}$ as an unrooted tree, then it is uniformly distributed among all spanning trees of G . Conditional on $X_0 = v$ for some vertex v of G , the tree $\mathcal{T}$ is uniform among all rooted spanning trees of G with root v . If X_0 is uniform, then $\mathcal{T}$ is uniform among all rooted spanning trees of G .*

Proof. — For now, suppose that X_0 follows the stationary distribution π . Let us extend the random walk in negative time so that $(X_n)_{n \in \mathbb{Z}}$ forms a stationary process.

We let $\mathcal{S}$ denote the collection of rooted spanning subtrees of G . For each $n \in \mathbb{Z}$ and $v \in V$ we set

$$T_{v,n} = \min\{k \geq n \mid X_k = v\}.$$

That is, $T_{v,n}$ is the first time starting from n that we encounter the vertex v . We define the spanning subtree S_n rooted at X_n with vertex set V and edges $\{X_{T_{v,n}-1}, v\}$ for $v \in V \setminus \{X_n\}$. As $(X_n)_{n \in \mathbb{Z}}$ is stationary, it follows that $(S_n)_{n \in \mathbb{Z}}$ is an $\mathcal{S}$ -valued stationary process.

For each spanning tree $t \in \mathcal{S}$ with root vertex v we let $D(t) = d_G(v)$ denote the number of neighbours of v in G . Note that if $S_0 = t$, then X_{-1} must be one of the neighbours of v . Each neighbour w has chance $1/D(t)$, and each possibility leads to S_{-1} being equal to some tree t' that is determined by t and w . Specifically, t' is obtained from t by deleting the edge from w to its parent and adding an edge from w to v . See Figure 10 for an example.

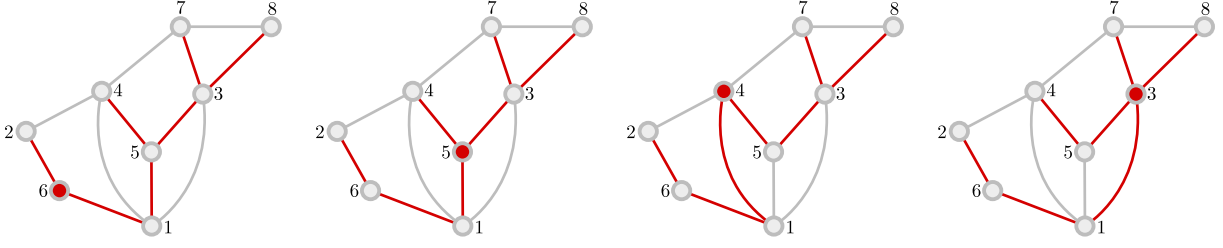


FIGURE 11. The possible successors of the spanning tree in Figure 9. For example, we would obtain the first if the random walk starts with $(1, 6, 1, 5, 4, 5, 1, 6, 2, 4, 5, 3, 8, 3, 1, 3, 7, \dots)$ as in the caption of Figure 9.

It follows that $(S_{-n})_{n \in \mathbb{Z}}$ is a stationary $\mathcal{S}$ -valued Markov chain. Its transition matrix Q with

$$(3.1) \quad Q(t, t') = \mathbb{P}(S_{-1} = t' \mid S_0 = t)$$

satisfies the property, that for each $t \in \mathcal{S}$ there are $D(t)$ trees $t' \in \mathcal{S}$ for which

$$(3.2) \quad Q(t, t') = 1/D(t),$$

and of course

$$(3.3) \quad Q(t, t') = 0$$

for all other t' . Note that for (3.1) we have used that X_0 follows the stationary distribution π .

Now, conversely, suppose that $t' \in \mathcal{S}$ is given and let w denote its root. We know that for each $t \in \mathcal{S}$ it holds that $Q(t, t') \in \{1/D(t), 0\}$. The question is, how many rooted spanning trees t with $Q(t, t') > 0$ are there exactly? To answer this, note that removing w from t' produces $d_{t'}(w)$ connected components (or fringe subtrees) $T_1, \dots, T_{d_{t'}(w)}$, each rooted at the corresponding t' -neighbour of w . We know that if $S_{-1} = t'$, then the random walk beginning at time -1 sets foot on one of these components first and it visits w again before arriving at any of the other component. Hence t must be obtained from t' by selecting some index $1 \leq i \leq d_{t'}(w)$, removing the edge from w to the root of T_i , and adding an arbitrary G -edge that links T_i with w . The choice of component and edge of course depends on the precise route of the random walk. But, in total, we have $D(t) = d_G(w)$ possibilities for t . Compare with Figure 11.

Thus, for each rooted spanning tree $t' \in \mathcal{S}$ there are $D(t')$ trees $t \in \mathcal{S}$ for which

$$Q(t, t') = 1/D(t)$$

and

$$Q(t, t') = 0$$

for all other t . This implies

$$(3.4) \quad \sum_{t \in \mathcal{S}} D(t) Q(t, t') = D(t').$$

Note that the Markov chain $(S_{-n})_{n \in \mathbb{Z}}$ is irreducible: Indeed, given any rooted spanning tree $t \in \mathcal{S}$ we may consider a sequence $v_0, \dots, v_s$ of its vertices obtained from a depth-first-search walk, including the steps we have to backtrack towards the root. (This way, s equals twice the number of edges of t .) The probability for $(X_0, \dots, X_s)$ to equal $(v_0, \dots, v_s)$ is positive and a lower bound for the probability that $S_0 = t$.

It follows from Equation (3.4) that the stationary distribution of $(S_{-n})_{n \in \mathbb{Z}}$ is proportional to $(D(t))_{t \in \mathcal{S}}$.

We have shown that when X_0 follows the stationary distribution π from Proposition 3.1, then $\mathbb{P}(\mathcal{T} = t)$ is proportional to $D(t)$. Hence, conditional on $X_0 = v$ for some fixed vertex v of G , $\mathcal{T}$ is uniformly distributed among all rooted spanning trees of G with root v . If we regard $\mathcal{T}$ as an unrooted tree, then it is uniformly distributed among all spanning trees of T . Likewise, if X_0 is assumed to have uniform distribution, it follows that $\mathcal{T}$ is uniform among all rooted spanning trees of G . $\square$

REMARK 3.4. — *Suppose that $G = K_n$ is the complete graph with vertex set $[n]$ (such that any two distinct vertices are connected). The irreducible Markov chain on the set of all n^{n-1} rooted trees on $[n]$ with transition matrix given in (3.1) takes the following shape: In each step, choose a uniform non-root vertex v . Delete the edge to its parent, add a new edge linking v and the old root, and declare v to be the new root. The stationary distribution of this chain is the uniform distribution on all rooted trees.*

3.2. Generating uniform labelled trees. —

THEOREM 3.5 (ALDOUS). — *For each integer $n \geq 2$ consider the following procedure that generates a random tree $\mathcal{T}_n$ with vertex set $[n] = \{1, \dots, n\}$.*

- 1) *Let $U_2, \dots, U_n$ be random elements that are independent and uniform on $[n]$.*
- 2) *Let $\sigma \in \mathfrak{S}_n$ be a uniformly selected permutation of degree n .*
- 3) *For $2 \leq j \leq n$ add an edge between $\sigma(j)$ and $\sigma(\min(j-1, U_j))$.*

Then $\mathcal{T}_n$ is uniform among all labelled trees with vertex set $[n]$.

Note that there is always an edge between $\sigma(1)$ and $\sigma(2)$. The random variable U_2 is not used, we just keep it for ease of notation.

Proof of Theorem 3.5. — Let $(Z_i)_{i \geq 0}$ be a sequence of independent random variables, each uniformly distributed on $[n]$. Let $\pi_1, \pi_2, \dots, \pi_n \in [n]$ denote the distinct states hit first, second, etc. by $(Z_i)_{i \geq 0}$, and let $0 = \xi_1 < \xi_2 < \dots < \xi_n$ denote the corresponding hitting times. For $2 \leq j \leq n$, we let $P_j \in [n]$ denote the last state right before hitting π_j for the first time. That is,

$$(3.5) \quad P_j = Z_{\xi_j-1}.$$

Consider the tree $\mathcal{R}$ with vertex set $[n]$ in which π_j is connected to P_j for all $2 \leq j \leq n$.

The construction of $\mathcal{R}$ does not change if we replace $(Z_i)_{i \geq 0}$ by the sequence $(Z'_i)_{i \geq 0}$ where terms identical to their predecessor are deleted. Clearly $(Z'_i)_{i \geq 0}$ is the simple random walk on the complete graph with n vertices, and the construction of $\mathcal{R}$ from $(Z'_i)_{i \geq 0}$ is identical to the construction of the tree $\mathcal{T}$. By Theorem 3.3 it follows that $\mathcal{R}$ is uniformly distributed among all trees with vertex set $[n]$.

Let us analyse the distribution of P_j conditionally on the ordered sequence $(\pi_1, \dots, \pi_{j-1})$. Conditionally on $(\pi_1, \dots, \pi_{j-1})$, the last state π_{j-1} has a higher chance than the others: At time $\xi_{j-1} + 1$, that is, right after hitting π_{j-1} for the first time, the probability for $Z_{\xi_{j-1}+1}$ to belong to one of the $n - (j-1)$ not previously visited states is $1 - \frac{j-1}{n}$. In this case we have $\xi_j = \xi_{j-1} + 1$ and P_j is the state at time ξ_{j-1} , that is $P_j = Z_{\xi_{j-1}} = \pi(j-1)$. Now consider the other case, assumed with probability $\frac{j-1}{n}$, where $Z_{\xi_{j-1}+1}$ belongs to one of the $j-1$ previously visited states. Here $Z_{\xi_{j-1}+1}$ assumes any of these previously visited states with equal conditional probability. Hence, regardless how many additional steps it takes to hit a state that has not been previously visited, any of the states $\pi_1, \dots, \pi_{j-1}$ is equally likely to be the last before hitting a new state. Hence in the second case, P_j is conditionally uniformly distributed among $\{\pi_1, \dots, \pi_{j-1}\}$. Combining the two cases, we obtain the conditional probability

$$(3.6) \quad \mathbb{P}(P_j = \pi_i \mid \pi_1, \dots, \pi_{j-1}) = \begin{cases} \frac{1}{n}, & 1 \leq i < j-1 \\ 1 - \frac{j-2}{n}, & i = j-1 \end{cases}.$$

Note that this conditional probability only depends on i , j , and n . Moreover, interpreting $(\pi_i)_{1 \leq i \leq n}$ as permutation, it holds by symmetry that

$$(3.7) \quad \mathbb{P}((\pi_i)_{1 \leq i \leq n} = \gamma) = \frac{1}{n!}$$

for all permutations $\gamma \in \mathfrak{S}_n$.

Recall that in the construction of $\mathcal{R}$ we start with vertex set $[n]$ and connect π_j to P_j for all $2 \leq j \leq n$. This means the construction of $\mathcal{R}$ is equivalent to first selecting a uniform permutation σ from $\mathfrak{S}_n$, and then adding for each $2 \leq j \leq n$ an edge between $\sigma(j)$ and $\sigma(V_j)$ for an independent random index $1 \leq V_j \leq j-1$ with distribution

$$(3.8) \quad \mathbb{P}(V_j = i) = \begin{cases} \frac{1}{n}, & 1 \leq i < j-1 \\ 1 - \frac{j-2}{n}, & i = j-1 \end{cases}.$$

It is elementary that

$$(3.9) \quad V_j \stackrel{d}{=} \min(i-1, U)$$

for an uniform random element U of $[n]$. Hence this sampling procedure is equivalent to the construction of $\mathcal{T}_n$. This completes the proof. $\square$

REMARK 3.6. — *Theorem 3.3 also entails that the random tree $\mathcal{R}$ from the proof rooted at π_1 is uniform among all rooted trees with vertex set $[n]$. Thus, $\mathcal{T}_n$ rooted at $\sigma(1)$ is also uniform among all rooted trees on $[n]$.*

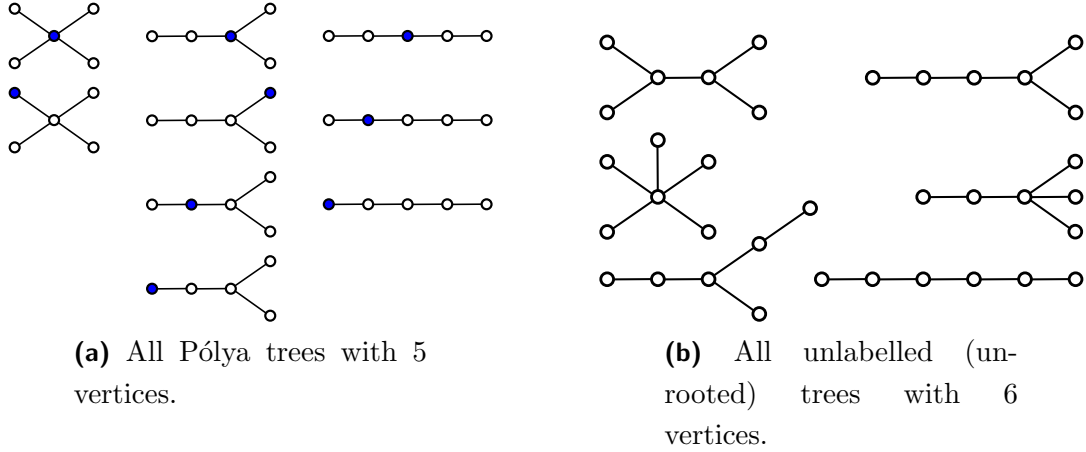


FIGURE 12. Notation for vertices in rooted trees. The root is marked with a square.

4. Plane trees

Trees in the graph-theoretical sense are also called labelled trees, and the vertex set is often referred to as the set of labels. Unlabelled rooted trees are often called Pólya trees, in honour of the Hungarian mathematician George Pólya (1887 – 1985). Note that unlike the labelled case, the number root locations in an unlabelled (unrooted) tree depends on the tree and not only on the number of vertices. See Figure 12 for an example.

These three tree models - labelled, unlabelled rooted, and unlabelled unrooted - are also called *unordered trees*. There are also various kinds of *ordered trees*, the most important and common one being *planted plane trees*. They are the most common type of trees in the probabilistic literature. Often, the shorter term *plane tree* is used synonymously with planted plane trees when there is no risk of confusion with other types. We will abide by this convention.

DEFINITION 4.1 (FINITE PLANE TREE). — *A finite plane tree is an unlabelled rooted tree where each offspring set is endowed with a linear order.*

That is, the offspring of any vertex v of a finite plane tree T is enumerated from 1 to $d_T^+(v)$. This allows us to refer to the first, second, etc. offspring vertex of v .

NOTATION 4.2. — *The set of all plane trees with $n \geq 1$ vertices is denoted by $\mathfrak{T}_n$. We let*

$$\mathfrak{T}_f = \bigcup_{n \geq 1} \mathfrak{T}_n$$

denote the collection of all finite plane trees.

See Figure 13 for the collection of all plane trees with 5 vertices. We may also define plane trees that admit vertices with infinite degree.

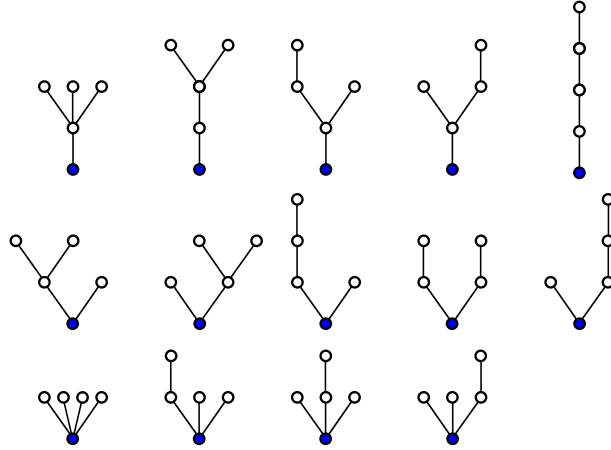


FIGURE 13. All 14 plane trees with 5 vertices.

DEFINITION 4.3 (INFINITE PLANE TREES). — *An infinite plane tree T may have nodes v with outdegree $d_T^+(v) = \infty$. In this case we require the children to be ordered $v_1, v_2, \dots$. That is, the order type on the offspring set is $\mathbb{N}$. We let $\mathfrak{T}$ denote the collection of all plane trees, finite and infinite.*

DEFINITION 4.4 (LOCALLY FINITE PLANE TREES). — *A plane tree T is called locally finite, if $d_T^+(v) < \infty$ for all vertices v of T . We let $\mathfrak{T}_{\text{lf}}$ denote the collection of all locally finite plane trees.*

Note that

$$\mathfrak{T}_{\text{f}} \subset \mathfrak{T}_{\text{lf}} \subset \mathfrak{T}.$$

4.1. Ulam–Harris tree encoding. — Plane trees may be realized as canonical subtrees of the Ulam–Harris tree:

DEFINITION 4.5 (ULAM–HARRIS TREE). — *The Ulam–Harris tree U_∞ is a tree whose vertex set is the collection*

$$V_\infty = \{\emptyset\} \cup \bigcup_{n \geq 1} \mathbb{N}^n$$

of finite sequences of positive integers. Here any such sequence $(i_1, \dots, i_k) \in \mathbb{N}^k$, $k \geq 0$ has the linearly ordered offspring

$$(i_1, \dots, i_k, 1), (i_1, \dots, i_k, 2), \dots$$

The empty sequence $\emptyset$ is the root of the Ulam–Harris tree.

PROPOSITION 4.6 (ENCODING AS SUBTREES OF THE ULAM–HARRIS TREE)

A finite plane tree T may be canonically realized as a subtree of the Ulam–Harris tree U_∞ by labelling the root of T with the empty sequence $\emptyset$. The linearly ordered offspring

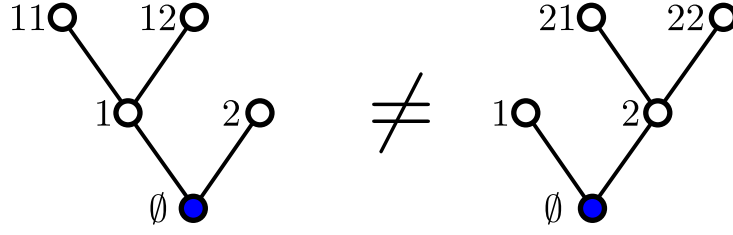


FIGURE 14. The two depicted trees are not identical as planted plane trees.

set $v_1, \dots, v_{d_T^+(v)}$ of a vertex of T that received label $(i_1, \dots, i_k) \in \mathbb{N}^k$, $k \geq 0$, receives the labels

$$(i_1, \dots, i_k, 1), (i_1, \dots, i_k, 2), \dots, (i_1, \dots, i_k, d_T^+(v)).$$

Compare with Figure 14. This embedding allows us to identify plane trees with families of outdegrees, indexed by V_∞ :

COROLLARY 4.7 (ENCODING BY OUTDEGREES). — *There is a bijection between the collection of plane trees $\mathfrak{T}$ and the collection of all families*

$$(d_v)_{v \in V_\infty} \in \overline{\mathbb{N}}_0^{V_\infty}$$

satisfying

$$d_{i_1 \dots i_k i} = 0 \quad \text{when} \quad i > d_{i_1 \dots i_k}.$$

The bijection maps a plane tree T to the family of outdegrees $(d_T^+(v))_{v \in V_\infty}$. Here we identify the vertex set of T in the canonical way as a subset of V_∞ , and set

$$d_T^+(v) = 0 \quad \text{when} \quad v \notin T.$$

Be careful that $d_T^+(v) = 0$ may also hold for some vertices $v \in T$.

4.2. Enumeration of plane trees. — In general, a depth-first-search of a rooted tree may yield different outcomes, since there is no general rule along which previously unvisited neighbour to proceed first. For plane trees, there is fortunately a canonical way:

DEFINITION 4.8 (CANONICAL DFS ORDERING). — *Given a finite plane tree with we may order its vertices in a canonical way via depth-first-search. Here we start at the root and always try to proceed along the left-most unvisited offspring. This ordering is identical to the lexicographic ordering when interpreting the vertex set of T as a subset of the vertex set V_∞ of the Ulam–Harris tree.*

DEFINITION 4.9 (DEGREE SEQUENCE AND ŁUKASIEWICZ PATH)

Let T be a finite plane tree and let $v_1, \dots, v_n$ be its depth-first-search ordered list of

vertices. We say $(d_T^+(v_1), \dots, d_T^+(v_n))$ is the degree sequence of T . The corresponding sequence $(W_k)_{0 \leq k \leq n}$ with

$$W_k = \sum_{i=1}^k (d_T^+(v_i) - 1)$$

is called the *Lukasiewicz path* of T .

LEMMA 4.10. — A sequence $d_1, \dots, d_n$ of $n \geq 1$ non-negative integers is a degree sequence of some plane tree if and only if

$$\sum_{i=1}^n (d_i - 1) = -1$$

and

$$\sum_{i=1}^k (d_i - 1) \geq 0 \quad \text{for all} \quad 1 \leq k < n.$$

Proof. — See exercises. □

The following combinatorial result ensures that any sequence $(d_1, \dots, d_n)$ of $n \geq 1$ non-negative integers satisfying $\sum_{i=1}^n d_i = n - 1$ corresponds via a unique cyclic shift to a degree sequence of a plane tree:

LEMMA 4.11 (CYCLE LEMMA). — Let $x_1, \dots, x_n \geq -1$ and $r \geq 1$ be integers such that

$$\sum_{i=1}^n x_i = -r.$$

For all $i, j \in \mathbb{Z}$ let $x_i^{(j)} = x_{i+j}$ with the index taken modulo n . For each integer $k \geq 0$ consider the partial sums

$$S_k^{(j)} = \sum_{i=1}^k x_i^{(j)}.$$

Show that there are precisely r values $j \in \{1, \dots, n\}$ such that

$$S_k^{(j)} > -r$$

for all $1 \leq k < n$.

Proof. — See exercises. □

THEOREM 4.12. — The number of planted plane trees with $n \geq 1$ vertices equals

$$|\mathfrak{T}_n| = \frac{1}{n} \binom{2n-2}{n-1}.$$

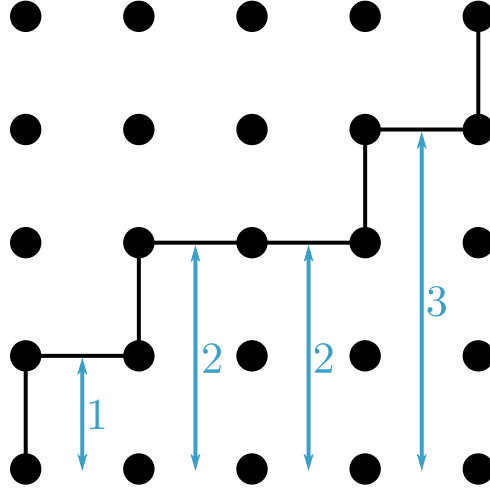


FIGURE 15. The lattice walk corresponding to the monotonically increasing sequence $(1, 2, 2, 3)$.

Proof. — Let $Z(n-1, n)$ denote the number of sequences $(d_1, \dots, d_n)$ of non-negative integers satisfying $\sum_{i=1}^n d_i = n-1$. By Lemmas 4.10 and 4.11 it follows that

$$|\mathfrak{T}_n| = \frac{1}{n} Z(n-1, n).$$

Consider the grid $G = \{(i, j) \mid i, j \in \{0, \dots, n-1\}\}$. Suppose we want to walk in G from $(0, 0)$ to $(n-1, n-1)$ in such a way that in each step we either move distance 1 upwards or distance 1 to the right. In total we have to take $2n-2$ steps, precisely $n-1$ of which are upwards and $n-1$ are to the right. The total number of such walks is hence $\binom{2n-2}{n-1}$.

Each walk corresponds in a bijective way to a sequence $(D_1, \dots, D_{n-1})$ of integers satisfying

$$0 \leq D_1 \leq D_2 \leq \dots \leq D_{n-1} \leq n-1.$$

Here for each $1 \leq k \leq n-1$ the number D_k corresponds to the height of the k th horizontal segment of the walk. Compare with Figure 15.

Each such sequence $(D_1, \dots, D_{n-1})$ corresponds bijectively to a sequence $(d_1, \dots, d_n)$ of non-negative integers satisfying $\sum_{i=1}^n d_i = n-1$. Here for each $1 \leq k \leq n-1$ the number D_k corresponds to the sum $\sum_{i=1}^k d_i$.

Thus,

$$Z(n-1, n) = \binom{2n-2}{n-1}.$$

□

5. Galton–Watson trees

5.1. Basic definitions. — Throughout this section we let ξ denote a random non-negative integer satisfying

$$(5.1) \quad \mathbb{P}(\xi = 0) > 0 \quad \text{and} \quad \mathbb{P}(\xi \geq 2) > 0.$$

We let

$$(5.2) \quad \phi(z) = \mathbb{E}[z^\xi]$$

denote its probability generating series, and

$$(5.3) \quad \mu = \mathbb{E}[\xi] \in [0, \infty]$$

its first moment.

DEFINITION 5.1 (GALTON–WATSON PROCESS). — Let $(\xi_i^{(k)})_{i,k \geq 1}$ denote a family of independent copies of ξ . A ξ -Galton–Watson process is a stochastic process $(Z_n)_{n \geq 0}$ in discrete time, defined by

$$(5.4) \quad Z_0 = 1, \quad Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n)}.$$

The process models the genealogical structure of a population that starts with a single individual and reproduces asexually. The number Z_n represents the size of the n th generation, and the offspring of the i th individual in this generation is represented by $\xi_i^{(n)}$ for each $1 \leq i \leq Z_n$.

The random variable ξ and its distribution are called the *branching law* or the *offspring distribution* of the process.

DEFINITION 5.2 (GALTON–WATSON TREE). — The family tree $\mathcal{T}$ corresponding to a ξ -Galton–Watson process $(Z_n)_{n \geq 0}$ is called a ξ -Galton–Watson tree. We interpret $\mathcal{T}$ as a random element of the collection $\mathfrak{T}$ of plane trees.

That is, the root of $\mathcal{T}$ has $\xi_1^{(1)}$ children. For each $1 \leq i \leq \xi_1^{(1)}$, the i th child of the root has $\xi_i^{(2)}$ children of its own. This process continues on and on, producing a tree that a priori may be infinite.

PROPOSITION 5.3. — 1) For each $n \geq 1$ the number Z_n corresponds to the total number of vertices in $\mathcal{T}$ with height n .

2) The total number of vertices of $\mathcal{T}$ equals $|\mathcal{T}| = \sum_{n \geq 0} Z_n$.

3) The height of $\mathcal{T}$ is given by $H(\mathcal{T}) = \sup\{k \geq 0 \mid Z_k > 0\}$.

4) The maximal outdegree of $\mathcal{T}$ is given by $\Delta(\mathcal{T}) := \sup\{\xi_i^{(n)} \mid n \geq 0, 1 \leq i \leq Z_n\}$.

DEFINITION 5.4 (CRITICALITY). — We say the Galton–Watson tree $\mathcal{T}$ is *critical*, *subcritical*, or *supercritical*, if $\mu = 1$, $\mu < 1$, or $\mu > 1$.

5.2. The probability of extinction. — Depending on the offspring distribution, it is possible that the tree $\mathcal{T}$ is infinite. We will clarify this situation.

LEMMA 5.5. — 1) Let $n \geq 1$ be an integer. The population Z_n at time n has probability generating function

$$\mathbb{E}[z^{Z_n}] = \underbrace{(\phi \circ \dots \circ \phi)}_{n \text{ times}}(z).$$

and first moment

$$\mathbb{E}[Z_n] = \mu^n.$$

2) It holds that

$$\mathbb{P}(|\mathcal{T}| < \infty) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0)$$

Proof. — By (5.4) it follows that

$$\mathbb{E}[z^{Z_1}] = \phi(z)$$

and

$$\mathbb{E}[z^{Z_{n+1}}] = \phi(\mathbb{E}[z^{Z_n}]).$$

By induction on n , it follows that the probability generating series for Z_n is the n th iterate of the probability generating series $\phi(z)$ of the offspring distribution. Using the chain rule, it follows again by induction on n that

$$\mathbb{E}[Z_n] = \frac{d}{dz} \mathbb{E}[z^{Z_n}] \Big|_{z=1} = \frac{d}{dz} \underbrace{(\phi \circ \dots \circ \phi)}_{n \text{ times}}(z) \Big|_{z=1} = \mu^n.$$

The Galton–Watson tree $\mathcal{T}$ is finite if and only if $Z_n = 0$ for some integer n . Letting $\mathcal{E}_n$ denote the event $Z_n = 0$, it follows that

$$\mathbb{P}(|\mathcal{T}| < \infty) = \mathbb{P}\left(\bigcup_{n \geq 0} \mathcal{E}_n\right).$$

Of course, if $Z_n = 0$, then $Z_{n+1} = 0$ holds as well. That is, $\mathcal{E}_n \subset \mathcal{E}_{n+1}$. Consequently,

$$\mathbb{P}(\mathcal{E}_n) \rightarrow \mathbb{P}\left(\bigcup_{n \geq 0} \mathcal{E}_n\right)$$

as $n \rightarrow \infty$. This shows $\mathbb{P}(|\mathcal{T}| < \infty) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0)$. □

THEOREM 5.6 (THE PROBABILITY OF EXTINCTION). —

- 1) If $\mu \leq 1$, then $\mathbb{P}(|\mathcal{T}| < \infty) = 1$.
- 2) If $\mu > 1$, then $0 < \mathbb{P}(|\mathcal{T}| < \infty) < 1$.

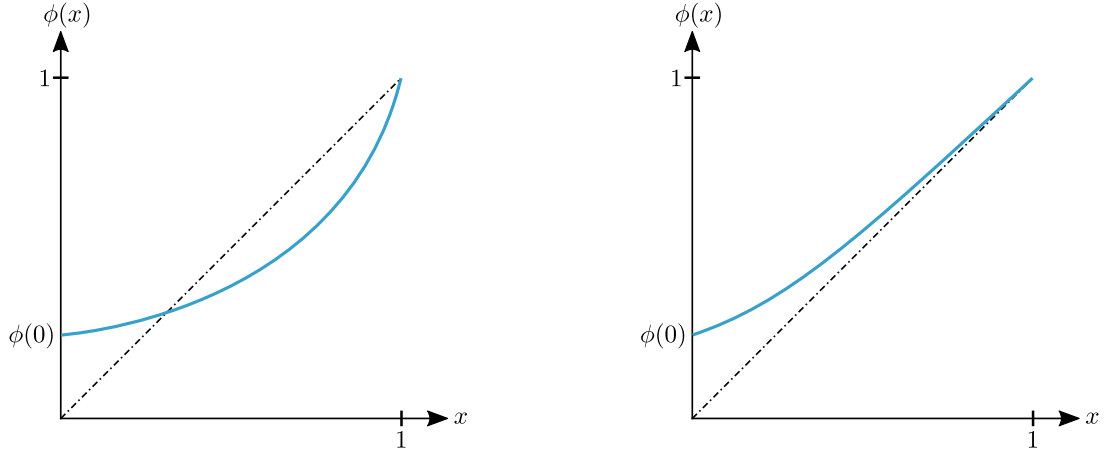


FIGURE 16. The probability generating function $\phi(x) = \mathbb{E}[x^\xi]$ has either one or two fixed points in the unit interval $[0, 1]$, depending on whether $\mathbb{E}[\xi] > 1$ or $\mathbb{E}[\xi] \leq 1$.

Proof. — For ease of notation, let us set

$$p := \mathbb{P}(|\mathcal{T}| < \infty) \geq \mathbb{P}(\xi = 0) > 0.$$

The Galton–Watson tree $\mathcal{T}$ is finite if and only if the $\xi_1^{(1)}$ fringes subtrees dangling from the root are all finite. Conditionally on $\xi_1^{(1)}$, each of these is distributed like $\mathcal{T}$. This readily yields

$$p = \sum_{n \geq 0} \mathbb{P}(\xi = n) p^n = \phi(p).$$

Assumption (5.1) asserts that $\mathbb{P}(\xi = 0) > 0$ and $\mathbb{P}(\xi \geq 2) > 0$. Hence ϕ is strictly increasing and strictly convex on the unit interval. Furthermore, $\phi(0) = \mathbb{P}(\xi = 0) > 0$ and $\phi(1) = 1$. By basic properties of convex functions it follows that 1 is the only fixed point of ϕ in the unit interval if

$$\mu = \lim_{t \nearrow 1} \phi'(t) \leq 1.$$

Hence in this case, p must be equal to 1. In the other case, when

$$\mu = \lim_{t \nearrow 1} \phi'(t) > 1,$$

then there is some number $0 < q < 1$ so that q and 1 are the only fixed points of ϕ in the unit interval. Compare with Figure 16. Hence $p \in \{q, 1\}$. In order to see that $p = q$, note that $0 < q$ implies $\phi(0) < \phi(q) = q$. More generally, it follows from Lemma 5.5 that

$$\mathbb{P}(Z_n = 0) = \underbrace{(\phi \circ \dots \circ \phi)}_{n \text{ times}}(0) < q.$$

As, again by Lemma 5.5, $p = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0)$, it follows that $p \leq q$ and hence $p = q \in]0, 1[$. $\square$

5.3. Topological properties of the collection of plane trees. —

5.3.1. Tychonoff's theorem. — In Corollary 4.7, we learned that we may interpret the collection $\mathfrak{T}$ as a subspace of the product $\overline{\mathbb{N}}_0^{V_\infty}$. Note that V_∞ is countably infinite. We may endow each factor

$$\overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}$$

with the usual compact topology given by the one-point compactification of the discrete space $\mathbb{N}_0$. That is, a subset $O \subset \overline{\mathbb{N}}_0$ is open if and only if $O \subset \mathbb{N}_0$, or $\infty \in O$ and $\overline{\mathbb{N}}_0 \setminus O \subset \mathbb{N}_0$ is finite. This makes the bijection

$$\overline{\mathbb{N}}_0 \rightarrow \{0\} \cup \{1/n \mid n \geq 1\}, \quad n \mapsto 1/(n+1)$$

a homeomorphism, if we endow $\{0\} \cup \{1/n \mid n \geq 1\} \subset \mathbb{R}$ with the subspace topology. In particular, $\overline{\mathbb{N}}_0$ is metrizable by a metric $d_{\overline{\mathbb{N}}_0}$ induced from this bijection.

A famous theorem in topology states that the product of compact topological spaces is compact:

THEOREM 5.7 (TYCHONOFF). — *If $(X_i, \tau_i)_{i \in I}$ is a family of compact topological spaces, then their product $\prod_{i \in I} X_i$ is compact with respect to the product topology.*

This makes $\overline{\mathbb{N}}_0^{V_\infty}$ a compact space. A direct proof of Tychonoff's theorem is sufficiently straightforward that it seems worth giving before proceeding.

LEMMA 5.8. — *Let $(X_i, \tau_i)_{i \in I}$ be a family of compact topological spaces and let $X = \prod_{i \in I} X_i$ be endowed with the product topology. For any $i \in I$ let $\pi_i : X \rightarrow X_i$ denote the projection to the i th coordinate. Then any open cover C of X satisfying*

$$C \subset \{\pi_i^{-1}(O) \mid i \in I, O \in \tau_i\}$$

has a finite subcover.

Proof. — For each $i \in I$ define

$$C_i = \{O \in \tau_i \mid \pi_i^{-1}(O) \in C\}.$$

Suppose that there is no index $i \in I$ such that C_i covers X_i . Then for each $i \in I$ we may select an element x_i that is not covered by C_i . Hence $x := (x_i)_{i \in I} \in X$ would not be covered by C , a contradiction as we assumed C to be a cover of X .

Hence there is an index $i \in I$ such that C_i covers X_i . As X_i is compact, it follows that C_i admits a finite subcover $C'_i \subset C_i$ of X_i . Hence

$$\{\pi^{-1}(O) \mid O \in C'_i\} \subset C$$

is a finite subcover of X . □

Recall that a *basis* of a topological space (X, τ) is a subset $B \subset \tau$ such that each open subset of X is a union of elements from B . A *subbasis* $S \subset \tau$ is a collection of open sets such that the collection of all finite intersections of elements from S forms a basis of τ . Care has to be taken that “finite” is allowed to mean zero, using the convention $\bigcap_{O \in \emptyset} O := X$. (Without this convention, only proper subsets of X would be required to admit expressions as unions of finite intersections of elements from S .) Equivalently, S is a subbasis if τ is the smallest topology on X containing S .

For example, if $X = \prod_{i \in I} X_i$ is equipped with the product topology as in Lemma 5.8, then

$$\{\pi_i^{-1}(O) \mid i \in I, O \in \tau_i\}$$

is a subbasis. In Lemma 5.8 we showed that if X_i is compact for all $i \in I$, then any collection of sets from this subbasis that covers X has a finite subcover. Hence Tychonoff’s theorem may be readily deduced from the following result:

LEMMA 5.9 (ALEXANDER’S SUBBASIS THEOREM). — *Let S be a subbasis of a topological space (X, τ) . If every collection of sets from S that covers X has a finite subcover, then X is compact.*

Proof. — Consider the collection E of all open covers of X that do not have a finite subcover. E is partially ordered by set inclusion. Any totally ordered subset $T \subset E$ has $C = \bigcup_{C' \in T} C'$ as an upper bound. To see that C has no finite subcover, note that any finite subset of C is necessarily a subset of some element of T .

Now, if X is not compact, then E is nonempty and hence has a maximal element M by Zorn’s Lemma. We claim that $M \cap S$ is a cover of X . To see this, let $x \in X$ be an element. M is a cover of X , hence there is an element $O \in M$ with $x \in O$. As M has no finite subcover, it must hold that $O \subsetneq X$. As S is a subbasis, it follows that

$$x \in O_1 \cap \dots \cap O_n \subset O$$

for some $n \geq 1$ and elements $O_1, \dots, O_n \in S$. If $O_j \in M$ for some $1 \leq j \leq n$, then x is covered by $M \cap S$. If $O_1, \dots, O_n \notin M$, then, by maximality of M , $M \cup \{O_j\}$ has a finite subcover containing O_j for each $1 \leq j \leq n$. That is, we may write

$$X = O_j \cup U_j$$

for some open set $U_j \subset X$ that is the union of finitely many elements of M . This means that $O \cup \bigcup_{j=1}^n U_j$ is a union of finitely many elements from M . But

$$O \cup \bigcup_{j=1}^n U_j \supset \left(\bigcap_{j=1}^n O_j \right) \cup \bigcup_{j=1}^n U_j \supset \bigcap_{j=1}^n (O_j \cup U_j) = X.$$

This contradicts the fact that M has by construction no finite subcover. Thus $M \cap S$ is a cover of X .

By assumption on S , it follows that $M \cap S$ has a finite subcover. This contradicts the fact that M has no finite subcover. This means X must be compact. $\square$

This concludes the proof of Tychonoff's theorem.

5.3.2. The space $\mathfrak{T}$ is Polish. — Going back to $\overline{\mathbb{N}}_0^{V_\infty}$, we have now verified that $\overline{\mathbb{N}}_0^{V_\infty}$ is a (countable) product of compact spaces and hence compact with respect to the product topology.

PROPOSITION 5.10. — *A sequence of points in $\overline{\mathbb{N}}_0^{V_\infty}$ converges if and only if each coordinate converges.*

Proof. — Indeed, the projections $\pi_v : \overline{\mathbb{N}}_0^{V_\infty} \rightarrow \overline{\mathbb{N}}_0$ (with $v \in V_\infty$) are continuous, hence convergence in $\overline{\mathbb{N}}_0^{V_\infty}$ implies convergence of each coordinate. Conversely, suppose that $\mathbf{x}_n$, $n \geq 1$ is a sequence in $\overline{\mathbb{N}}_0^{V_\infty}$ and $\mathbf{x} \in \overline{\mathbb{N}}_0^{V_\infty}$ so that $\pi_v(\mathbf{x}_n) \rightarrow \pi_v(\mathbf{x})$ for each $v \in V_\infty$. We know that

$$\{\pi_v^{-1}(O) \mid v \in V_\infty, O \subset \overline{\mathbb{N}}_0 \text{ open}\}$$

is a subbasis. Hence if $O \subsetneq \overline{\mathbb{N}}_0^{V_\infty}$ is an open neighbourhood of $\mathbf{x}$, then there exists an open set O' that is the intersection of finitely many elements of the subbasis such that $\mathbf{x} \in O' \subset O$. Convergence of each coordinate of $\mathbf{x}_n$ ensures that for n large enough $\mathbf{x}_n \in O'$ and hence $\mathbf{x}_n \in O$. Thus, $\mathbf{x}_n$ converges to $\mathbf{x}$ in the product topology. $\square$

The space $\overline{\mathbb{N}}_0^{V_\infty}$ has a countable basis, given by the intersections of finitely many elements of its countable subbasis

$$\{\pi_v^{-1}(\{n\}) \mid v \in V_\infty, n \in \mathbb{N}_0\} \cup \{\pi_v^{-1}(\{\infty\} \cup \{n, n+1, \dots\}) \mid v \in V_\infty, n \in \mathbb{N}_0\}.$$

Hence $\overline{\mathbb{N}}_0^{V_\infty}$ is a *second-countable space*. Consequently, it also admits a countable dense subset, just pick an element of each basis element. This means $\overline{\mathbb{N}}_0^{V_\infty}$ is *separable*.

The space $\overline{\mathbb{N}}_0$ is *metrizable*, for example by the metric $d_{\overline{\mathbb{N}}_0}$ induced from the homeomorphism

$$\overline{\mathbb{N}}_0 \rightarrow \{0\} \cup \{1/n \mid n \geq 1\}$$

and the subspace metric on the right-hand side. Consequently, the product $\overline{\mathbb{N}}_0^{V_\infty}$ is metrizable as well: if we pick any bijection $\iota : V_\infty \rightarrow \mathbb{N}$, then the metric

$$d_{\overline{\mathbb{N}}_0^{V_\infty}}((x_v)_{v \in V_\infty}, (y_v)_{v \in V_\infty}) = \sup_{n \geq 1} \frac{1}{n} d_{\overline{\mathbb{N}}_0}(x_{\iota(n)}, y_{\iota(n)})$$

induces the product topology on $\overline{\mathbb{N}}_0^{V_\infty}$. One way to see this is to first observe that convergence with respect to $d_{\overline{\mathbb{N}}_0^{V_\infty}}$ is equivalent to convergence of each coordinate, just as with the product topology. Hence a set is closed in the product topology if and only if it is closed with respect to $d_{\overline{\mathbb{N}}_0^{V_\infty}}$.

A topological space is called *Polish* if it is homeomorphic to a separable complete metric space. We have shown that the metric space $(\bar{\mathbb{N}}_0^{V_\infty}, d_{\bar{\mathbb{N}}_0^{V_\infty}})$ is compact, hence it is also complete. Since $\bar{\mathbb{N}}_0^{V_\infty}$ is separable, it follows that $\bar{\mathbb{N}}_0^{V_\infty}$ is Polish.

Recall that by Corollary 4.7 we may identify $\mathfrak{T}$ with a subset of $\bar{\mathbb{N}}_0^{V_\infty}$.

PROPOSITION 5.11. — *The subset $\mathfrak{T} \subset \bar{\mathbb{N}}_0^{V_\infty}$ is closed.*

Proof. — Recall that a point $(x_v)_{v \in V_\infty} \in \bar{\mathbb{N}}_0^{V_\infty}$ belongs to $\mathfrak{T}$ if and only if for each $v \in V_\infty$ and each $i \in \mathbb{N}$ with $i > x_v$ it holds that

$$x_{vi} = 0.$$

Hence we may write

$$\mathfrak{T} = \bar{\mathbb{N}}_0^{V_\infty} \setminus \bigcup_{v \in V_\infty, x \in \mathbb{N}_0, i \in \mathbb{N}, i > x} U_{v,x,i}$$

for

$$\begin{aligned} U_{v,x,i} &= \{\mathbf{x} \in \bar{\mathbb{N}}_0^{V_\infty} \mid \pi_v(\mathbf{x}) = x, \pi_{vi}(\mathbf{x}) \neq 0\} \\ &= \pi_v^{-1}(\{x\}) \cap \pi_{vi}^{-1}(\bar{\mathbb{N}}_0 \setminus \{0\}). \end{aligned}$$

Each such set $U_{v,x,i}$ is open, making $\mathfrak{T}$ a closed subset. □

COROLLARY 5.12. — *$\mathfrak{T}$ is a compact Polish space.*

Proof. — Closed subsets of compact spaces are compact, hence $\mathfrak{T}$ is compact. As a compact metric space, $\mathfrak{T}$ is of course complete. Subspaces of second-countable space are second-countable, making $\mathfrak{T}$ second-countable and hence separable. □

Note that $\mathfrak{T}_{\text{lf}}$ and $\mathfrak{T}_{\text{f}}$ are *not* closed subsets of $\mathfrak{T}$.

5.3.3. Truncated trees and distributional convergence. —

DEFINITION 5.13 (TRUNCATIONS OF TREES). — *For any tree $T \in \mathfrak{T}$ and any integer $m \geq 1$ we let $T^{(m)}$ denote the tree obtained by removing all vertices with height larger than m . We also define the subset*

$$V^{[m]} = \bigcup_{k=0}^m \{1, \dots, m\}^k \subset V_\infty$$

and let $T^{[m]}$ denote the tree with vertex set restricted to $V(T) \cap V^{[m]}$.

That is, the tree $T^{(m)}$ is obtained from T by truncating at height m . The tree $T^{[m]}$ is obtained by additionally pruning so that all outdegrees are at most m . The following result is an immediate consequence from the characterization of convergence in $\bar{\mathbb{N}}_0^{V_\infty}$:

PROPOSITION 5.14. — *Let $(T_n)_{n \geq 1}$ be a sequence of trees in $\mathfrak{T}$ and let $T \in \mathfrak{T}$.*

- 1) The sequence converges to T if and only if for each $m \geq 1$ there is a constant $n(m) \geq 1$ such that for all $n \geq n(m)$

$$T_n^{[m]} = T^{[m]}.$$

- 2) Suppose that $T \in \mathfrak{T}_{\text{lf}}$. Then the sequence converges to T if and only if for each $m \geq 1$ there is a constant $n(m) \geq 1$ such that for all $n \geq n(m)$

$$T_n^{<m>} = T^{<m>}.$$

PROPOSITION 5.15. — We equip $\overline{\mathbb{N}}_0^{V_\infty}$ with its Borel σ -algebra. A sequence of random points $\mathbf{x}_1, \mathbf{x}_2, \dots \in \overline{\mathbb{N}}_0^{V_\infty}$ converges in distribution to a random point $\mathbf{x} \in \overline{\mathbb{N}}_0^{V_\infty}$, if and only if for each integer $m \geq 1$

$$(\pi_v(\mathbf{x}_n))_{v \in V^{[m]}} \xrightarrow{d} (\pi_v(\mathbf{x}))_{v \in V^{[m]}}.$$

Proof. — Necessity follows from the mapping theorem. Turning to sufficiency, define for each $m \geq 1$ the projection

$$p_m : \overline{\mathbb{N}}_0^{V_\infty} \rightarrow \overline{\mathbb{N}}_0^{V^{[m]}}, \quad (x_v)_{v \in V_\infty} \mapsto (x_v)_{v \in V^{[m]}}.$$

The finite dimensional marginals are measure-determining: If $\mathbf{x}$ and $\mathbf{y}$ are two random elements with $p_m(\mathbf{x}) \stackrel{d}{=} p_m(\mathbf{y})$ for all $m \geq 1$, then $\mathbf{x} \stackrel{d}{=} \mathbf{y}$. This follows from the usual π - λ -Lemma.

Now, suppose that $\mathbf{x}_n$ is a sequence of random elements of $\overline{\mathbb{N}}_0^{V_\infty}$ such that the finite dimensional marginals convergence in distribution. Our space is a compact Polish space, hence so this the corresponding space of Borel probability measures. Hence any sequence of random elements is tight and thus has a convergent subsequence. In order to verify distributional convergence of $\mathbf{x}_n$ it suffices to show that the distribution of the limit along a subsequence does not depend on the choice of subsequence. But this is clear, since any two distributional limits along subsequences must have the same finite dimensional marginal distributions. $\square$

Proposition 5.15 allows us to verify the following convergence criteria in a straightforward way.

PROPOSITION 5.16. — Let $(T_n)_{n \geq 1}$ be a sequence of random trees in $\mathfrak{T}$.

- 1) For a random tree T in $\mathfrak{T}$ it holds that

$$T_n \xrightarrow{d} T$$

if and only if

$$T_n^{[m]} \xrightarrow{d} T^{[m]}$$

for each $m \geq 1$.

2) For a random tree T that almost surely lies in $\mathfrak{T}_{\text{lf}}$ it holds that

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

if and only if

$$\mathsf{T}_n^{(m)} \xrightarrow{d} \mathsf{T}^{(m)}$$

for each $m \geq 1$.

5.4. Size-biased Galton–Watson trees. — Recall that our assumptions on the offspring distribution ξ entail $\mu > 0$. Suppose that $\mu < \infty$. Let us define the size-biased random variable ξ^* with distribution given by

$$(5.5) \quad \mathbb{P}(\xi^* = k) = \frac{1}{\mu} k \mathbb{P}(\xi = k), \quad k \geq 0.$$

DEFINITION 5.17 (SIZE-BIASED GALTON–WATSON TREE). — Suppose that $\mu < \infty$. We let $\mathcal{T}^*$ denote a size-biased Galton–Watson tree where there are two types of vertices, normal and special. Normal vertices receive offspring according to independent copies of ξ , all of which are declared normal. The root is declared special. Any special vertex receives offspring according to an independent copy of ξ^* , and among these a uniformly selected vertex is declared special again, whereas the rest is declared normal.

DEFINITION / PROPOSITION 5.18 (SPINE OF THE SIZE-BIASED GALTON–WATSON TREE)

The special vertices of the size-biased Galton–Watson tree $\mathcal{T}^$ form a one-sided infinite path that starts at the root vertex. This path is called the spine of $\mathcal{T}^*$.*

DEFINITION 5.19 (LEVEL SETS). — For any rooted tree T and any integer $n \geq 0$ we let $L_n(T)$ the number of vertices with height n .

For example, $L_n(\mathcal{T}) = Z_n$ for all $n \geq 0$.

PROPOSITION 5.20. — For each integer $k \geq 0$,

$$\mathbb{P}(L_n(\mathcal{T}^*) = k) = k \mathbb{P}(Z_n = k) \mu^{-n}.$$

For each finite plane tree T ,

$$\mathbb{P}(\mathcal{T}^{*(n)} = T) = \mu^{-n} L_n(T) \mathbb{P}(\mathcal{T}^{(n)} = T).$$

Proof. — We may mark the finite trimmed tree $\mathcal{T}^{*(n)}$ at the unique special vertex u_n with height n . For finite plane tree T with height n and any vertex v of T with height $h_T(v) = n$ it holds that

$$(5.6) \quad \mathbb{P}((\mathcal{T}^{*(n)}, u_n) = (T, v)) = \mu^{-n} \mathbb{P}(\mathcal{T}^{(n)} = T).$$

To see this, note that the event $(\mathcal{T}^{*(n)}, u_n) = (T, v)$ corresponds to a unique choice of outdegrees and selections of special vertices. The probability that an independent copy of ξ^* assumes a given outdegree $d \geq 1$ and that a given child is declared special is given by

$$\mathbb{P}(\xi^* = d) \frac{1}{d} = \mathbb{P}(\xi = d) \frac{1}{\mu}.$$

As there are n choices to be made for selecting special children in $\mathcal{T}^{*\langle n \rangle}$, Equation (5.6) follows. The event $\mathcal{T}^{*\langle n \rangle} = T$ corresponds to $L_n(T)$ equally likely outcomes for $(\mathcal{T}^{*\langle n \rangle}, u_n)$, hence

$$(5.7) \quad \mathbb{P}(\mathcal{T}^{*\langle n \rangle} = T) = \mu^{-n} L_n(T) \mathbb{P}(\mathcal{T}^{\langle n \rangle} = T).$$

Summing over all T with $L_n(T) = k$, this readily implies

$$\mathbb{P}(L_n(\mathcal{T}^*) = k) = \mu^{-n} k \mathbb{P}(Z_n = k).$$

□

DEFINITION / PROPOSITION 5.21. — *For any $\ell \geq 0$ we may consider the size-biased tree $\mathcal{T}^{*(\ell)}$ that is constructed like $\mathcal{T}^*$, with the only difference being that the ℓ th special vertex u_ℓ is declared normal instead, and hence receives offspring according to an independent copy of ξ , all of which are declared normal again. Thus, $\mathcal{T}^{*(\ell)}$ has a spine of length ℓ .*

PROPOSITION 5.22. — *For any finite plane tree T and any vertex v of T with height ℓ it holds that*

$$\mathbb{P}((\mathcal{T}^{*(\ell)}, u_\ell) = (T, v)) = \mu^{-\ell} \mathbb{P}(\mathcal{T} = T).$$

The proof is by identical arguments as for Equation (5.6). Note that if $\mu > 1$, then $\mathcal{T}^{*(\ell)}$ and $\mathcal{T}$ may be infinite. The Equation in Proposition 5.22 still holds for infinite trees T , but in this case both sides of the Equation equal zero.

5.5. The Kesten–Stigum theorem. — We observed that the population of a super-critical Galton–Watson process has a positive chance of survival. Our aim in this section is to describe the growth of population in case of survival.

LEMMA 5.23. — *Suppose that $\mu < \infty$. Then*

$$(5.8) \quad W_n = \frac{Z_n}{\mu^n}.$$

is a martingale with respect to the natural filtration.

Proof. — It holds that $\mathbb{E}[W_n] = 1$ for all $n \geq 1$ and

$$\begin{aligned} \mathbb{E}[W_{n+1} \mid W_n] &= \frac{1}{\mu^{n+1}} \mathbb{E} \left[\sum_{i=1}^{Z_n} \xi_i^{(n)} \mid W_n \right] \\ &= \frac{1}{\mu^{n+1}} \sum_{i=1}^{Z_n} \mathbb{E} \left[\xi_i^{(n)} \mid W_n \right] \\ &= \frac{Z_n}{\mu^n}. \end{aligned}$$

□

DEFINITION / PROPOSITION 5.24. — *Suppose that $\mu < \infty$. As $\mathbb{E}[W_n] = 1$ for all $n \geq 1$, it follows from the martingale convergence theorem that*

$$W := \lim_{n \rightarrow \infty} W_n$$

exists almost surely. Fatou's lemma implies

$$\mathbb{E}[W] \leq 1.$$

PROPOSITION 5.25. — *Suppose that $\mu < \infty$. Then*

$$\mathbb{P}(W = 0) \in \{1, \mathbb{P}(|\mathcal{T}| < \infty)\}.$$

Proof. — The $(n+1)$ th generation of $\mathcal{T}$ may be partitioned into the n th generations $Z_n^{(1)}, \dots, Z_n^{(Z_1)}$ of the Z_1 fringe subtrees dangling from the root of $\mathcal{T}$. Thus

$$Z_{n+1} = \sum_{i=1}^{Z_1} Z_i^{(n)}.$$

The generations $(Z_i^{(n)})_{i \geq 1}$ are i.i.d., each distributed and constructed like the n th generation of a ξ -Galton–Watson process. Furthermore, they are independent from Z_1 .

Dividing both sides of the equation by μ^n and letting $n \rightarrow \infty$, it follows that

$$\mu W = \sum_{i=1}^{Z_1} W^{(i)},$$

with $(W^{(i)})_{i \geq 1}$ representing i.i.d. random variables, each distributed like W , that are also independent from Z_1 . This entails

$$\mathbb{P}(W = 0) = \sum_{i \geq 0} \mathbb{P}(\xi = i) \mathbb{P}(W = 0)^i = \phi(\mathbb{P}(W = 0)).$$

Consequently,

$$\mathbb{P}(W = 0) \in \{1, \mathbb{P}(|\mathcal{T}| < \infty)\}.$$

□

Of course, in the case $\mu \leq 1$, we have $W = 0$ almost surely. The question is whether W is also degenerate in the case $\mu > 1$. The Kesten–Stigum theorem provides an answer:

THEOREM 5.26 (KESTEN, STIGUM). — *Suppose that $1 < \mu < \infty$. The following claims are equivalent.*

- 1) $\mathbb{E}[W] = 1$
- 2) $\mathbb{P}(W = 0) = \mathbb{P}(|\mathcal{T}| < \infty)$
- 3) $\mathbb{E}[\xi \log \max(1, \xi)] < \infty$.

We are going to need some preliminary observations before we can proceed to prove this.

5.5.1. Preliminaries 1: Properties of Radon–Nikodym derivatives. — Let $(\Omega, \mathcal{F})$ be a measurable space. Let Q be a finite measure on $\mathcal{F}$ (with *finite* meaning $Q(\Omega) < \infty$), and let P be a probability measure on $\mathcal{F}$. Let $(\mathcal{F}_n)_{n \geq 1}$ be a filtration of $\mathcal{F}$ such that $\mathcal{F} = \sigma(\bigcup_{n \geq 1} \mathcal{F}_n)$. Let Q_n and P_n denote the restrictions of Q and P to $\mathcal{F}_n$.

Suppose that $Q_n \ll P_n$ for all $n \geq 1$. Set $X_n = \frac{dQ_n}{dP_n}$ and $X = \limsup_{n \rightarrow \infty} X_n$. We are going to show:

THEOREM 5.27. — *The Lebesgue decomposition $Q = Q_r + Q_s$ with $Q_r \ll P$ and $Q_s \perp P$ is given by*

$$Q_r(A) = \int_A X \, dP, \quad \text{and} \quad Q_s(A) = Q(A \cap \{X = \infty\})$$

for all $A \in \mathcal{F}$.

The first part of the proof is the following observation:

LEMMA 5.28. — *$(X_n)_n$ defined on $(\Omega, \mathcal{F}, P)$ is a martingale with respect to the filtration $(\mathcal{F}_n)_n$.*

Proof. — By definition, $X_n = \frac{dQ_n}{dP_n}$ is $\mathcal{F}_n$ -measurable. As P_n is the restriction of P to $\mathcal{F}_n$, it follows that for any $A \in \mathcal{F}_n$

$$\int_A X_n \, dP = \int_A X_n \, dP_n = Q_n(A) = Q(A).$$

Since this holds for all $n \geq 1$ and since $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, it follows that

$$\int_A X_{n+1} dP = Q(A) = \int_A X_n dP.$$

As this holds for all $A \in \mathcal{F}_n$, it follows that

$$X_n = \mathbb{E}[X_{n+1} \mid \mathcal{F}_n].$$

□

Lemma 5.28 and the martingale convergence theorem entail that $X = \lim_{n \rightarrow \infty} X_n < \infty$ holds P -a.s.. In particular, the measure $A \mapsto Q(A \cap \{X = \infty\})$ is singular with respect to P . Thus, verifying Theorem 5.27 reduces to showing

$$(5.9) \quad Q(A) = \int_A X dP + Q(A \cap \{X = \infty\})$$

for all $A \in \mathcal{F}$.

Proof of Equation (5.9) and hence Theorem 5.27. — Dividing Q by $Q(\Omega)$ allows us assume without loss of generality that Q is a probability measure.

We define the probability measures $R = (P + Q)/2$, so that $R_n = (P_n + Q_n)/2$ is the restriction of R to $\mathcal{F}_n$. By standard properties of Radon–Nikodym derivatives,

$$\frac{dP_n}{dR_n} + \frac{dQ_n}{dR_n} = 2 \frac{dR_n}{dR_n} = 2.$$

Using Lemma 5.28 and the martingale convergence theorem, it follows that the martingales $\frac{dP_n}{dR_n}$ and $\frac{dQ_n}{dR_n}$ are R -a.s. convergent with limits Y and Z . It holds that

$$Y = \frac{dP}{dR}, \quad \text{and} \quad Z = \frac{dQ}{dR}.$$

To see this, it suffices to show the first equality. By dominated convergence, for any $A \in \mathcal{F}_n$

$$P(A) = P_n(A) = \int_A \frac{dP_n}{dR_n} dR_n = \int_A \frac{dP_n}{dR_n} dR \rightarrow \int_A Y dR.$$

Hence,

$$P(A) = \int_A Y dR$$

for all $A \in \bigcup_{n \geq 1} \mathcal{F}_n$. As we assumed that $\mathcal{F} = \sigma(\bigcup_{n \geq 1} \mathcal{F}_n)$, and since $\bigcup_{n \geq 1} \mathcal{F}_n$ is a π -system, it follows from Dynkin's π - λ -Lemma that $P(A) = \int_A Y dR$ for all $A \in \mathcal{F}$. Thus, $Y = \frac{dP}{dR}$.

Since $Q_n \ll R_n \ll P_n$ and $P_n \ll R_n$, it holds that

$$X_n = \frac{dQ_n}{dP_n} = \frac{dQ_n}{dR_n} \frac{dR_n}{dP_n} = \frac{dQ_n}{dR_n} / \frac{dP_n}{dR_n}.$$

Since $\frac{dP_n}{dR_n} + \frac{dQ_n}{dR_n} = 2$, it follows that $Y + Z = 2$ holds R -a.s., and hence

$$R(Y = 0, Z = 0) = 0.$$

Thus,

$$X_n = \frac{dQ_n}{dR_n} / \frac{dP_n}{dR_n} \rightarrow Z/Y$$

holds R -a.s.. As we defined $X = \limsup_{n \rightarrow \infty} X_n$, it follows that

$$X = Z/Y$$

holds R -a.s.. Now, for any $A \in \mathcal{F}$ we have

$$\begin{aligned} Q(A) &= \int_A Z \, dR \\ &= \int_A Y Z/Y \mathbb{1}_{Y>0} dR + \int_A Z \mathbb{1}_{Y=0} dR \\ &= \int_A XY \, dR + \int_A Z \mathbb{1}_{X=\infty} dR \\ &= \int_A X \, dP + Q(A \cap \{X = \infty\}). \end{aligned}$$

□

COROLLARY 5.29. — *In this setting,*

$$Q \ll P \quad \Leftrightarrow \quad X < \infty \text{ } Q\text{-a.s.} \quad \Leftrightarrow \quad \mathbb{E}_P[X] = 1$$

and

$$Q \perp P \quad \Leftrightarrow \quad X = \infty \text{ } Q\text{-a.s.} \quad \Leftrightarrow \quad \mathbb{E}_P[X] = 0.$$

5.5.2. Preliminaries 2: Seneta's theorem. — Seneta's theorem is stated in the general context of Galton–Watson processes with immigration:

DEFINITION 5.30 (GALTON–WATSON PROCESS WITH IMMIGRATION)

Let ξ be an offspring distribution subject to our requirements (5.1). Let ζ be a random non-negative integer with $\mathbb{P}(\zeta > 0) > 0$, which in this context will be called the immigration distribution. The processes starts with no particles, say, and in each generation $n \geq 1$ there is an immigration of Y_n individuals, with $(Y_n)_{n \geq 1}$ being independent copies of the immigration law. Each individual has, independently, a ξ -Galton–Watson descendant tree.

For example, we may view the non-spine vertices of the size-biased tree $\mathcal{T}^*$ as a Galton–Watson process with offspring distribution ξ and immigration distribution $\xi^* - 1$.

THEOREM 5.31 (SENETA). — Let $(Z_n^*)_{n \geq 1}$ denote the generation sizes of a Galton–Watson process with offspring distribution ξ and immigration distribution ζ . Suppose that $\mu = \mathbb{E}[\xi] > 1$. If $\mathbb{E}[\log \max(\zeta, 1)] < \infty$, then $\lim_{n \rightarrow \infty} Z_n^*/\mu^n$ exists and is finite almost surely. If $\mathbb{E}[\log \max(\zeta, 1)] = \infty$, then $\limsup_{n \rightarrow \infty} Z_n^*/c^n = \infty$ for every constant $c > 0$.

The first step of the proof is a small general observation that follows from the Borel–Cantelli Lemma:

LEMMA 5.32. — *Let $X, X_1, X_2, \dots$ be non-negative i.i.d. random variables. Then*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} X_n = \begin{cases} 0, & \mathbb{E}[X] < \infty \\ \infty, & \mathbb{E}[X] = \infty \end{cases}.$$

Proof. — Suppose that $\mathbb{E}[X] < \infty$. Then for any $\epsilon > 0$

$$\sum_{n \geq 1} \mathbb{P}(X > \epsilon n) \leq \int_0^\infty \mathbb{P}(X/\epsilon > x) dx = \epsilon^{-1} \mathbb{E}[X] < \infty.$$

By the first Borel–Cantelli Lemma it follows that almost surely there is an $N > 0$ with $X_n < \epsilon n$ for all $n > N$. Since this holds for any $\epsilon > 0$, it follows that $\limsup_{n \rightarrow \infty} \frac{1}{n} X_n = 0$.

Now, suppose that $\mathbb{E}[X] = \infty$. Then for any $\epsilon > 0$

$$\sum_{n \geq 1} \mathbb{P}(X > \epsilon n) \geq \int_1^\infty \mathbb{P}(X/\epsilon > x) dx \geq \epsilon^{-1} \mathbb{E}[X] - 1 = \infty.$$

Since $(X_n)_{n \geq 1}$ are i.i.d., it follows from the second Borel–Cantelli Lemma that almost surely $X_n > \epsilon n$ for infinitely many n . Since this holds for each $\epsilon > 0$, it follows that $\limsup_{n \rightarrow \infty} \frac{1}{n} X_n = \infty$. $\square$

Proof of Theorem 5.31. — Assume first that $\mathbb{E}[\log \max(\zeta, 1)] = \infty$. Then, by Lemma 5.32, the number Y_n of immigrants in generation $n \geq 1$ satisfies

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \max(Y_n, 1) = \infty$$

almost surely. As $Z_n^* \geq Y_n$, it follows that $\limsup_{n \rightarrow \infty} Z_n^*/c^n = \infty$ for every constant $c > 0$.

Now, assume instead that $\mathbb{E}[\log \max(\zeta, 1)] < \infty$. Let $\mathcal{Y}$ denote the σ -field generated by $(Y_n)_{n \geq 1}$. For all integers $n \geq k \geq 1$ let $Z_{n,k}^*$ denote the number of descendants at level n of the individuals who immigrated in generation k . Thus, the total population at level n is given by

$$Z_n^* = \sum_{k=1}^n Z_{n,k}^*.$$

Thus,

$$\mathbb{E}[\mu^{-n} Z_n^* \mid \mathcal{Y}] = \mathbb{E}\left[\mu^{-n} \sum_{k=1}^n Z_{n,k}^* \mid \mathcal{Y}\right] = \sum_{k=1}^n \mu^{-k} \mathbb{E}[Z_{n,k}^* \mu^{-(n-k)} \mid \mathcal{Y}].$$

Given Y_k , the size $Z_{n,k}^*$ is distributed like the population in generation $n - k$ of Y_k independent ξ -Galton–Watson processes. Thus, $\mathbb{E}[Z_{n,k}^* \mid Y_k] = Y_k \mu^{n-k}$ and hence

$$\mathbb{E}[\mu^{-n} Z_n^* \mid \mathcal{Y}] = \sum_{k=1}^n \mu^{-k} Y_k.$$

Since we assumed $\mathbb{E}[\log \max(\zeta, 1)] < \infty$, it follows from Lemma 5.32 that almost surely

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \max(Y_n, 1) = 0.$$

Hence $\sum_{k=1}^{\infty} \mu^{-k} Y_k$ converges almost surely, providing a bound

$$\mathbb{E} [\mu^{-n} Z_n^* \mid \mathcal{Y}] \leq \sum_{k=1}^{\infty} \mu^{-k} Y_k$$

that is uniform in $n \geq 1$. Furthermore, note that

$$\mathbb{E} [\mu^{-(n+1)} Z_{n+1}^* \mid \mathcal{Y}, Z_n^*] = \mu^{-n} Z_n^* + \mu^{-(n+1)} Y_{n+1} \geq \mu^{-n} Z_n^*.$$

Thus, $\mu^{-n} Z_n^*$ is a submartingale when conditioned on $\mathcal{Y}$ with bounded expectation (given $\mathcal{Y}$). Hence $\mu^{-n} Z_n^*$ converges almost surely to a finite limit. $\square$

5.5.3. Proof of the Kesten–Stigum theorem. — Having done all preparations, the Kesten–Stigum theorem may now be verified in a short and elegant way.

Proof of Theorem 5.26. — Consider the space $\mathfrak{T}_{\text{lf}}$ of locally finite plane trees, equipped with its Borel σ -algebra $\mathcal{F}$. For each integer $n \geq 1$ let $\mathcal{F}_n \subset \mathcal{F}$ denote the σ -algebra generated by the projection $T \mapsto T^{(n)}$ to the first n levels. Then $\mathcal{F} = \sigma(\bigcup_{n \geq 1} \mathcal{F}_n)$. Let Q and P be the probability measures on $\mathcal{F}$ corresponding to the laws of the size-biased tree $\mathcal{T}^*$ and the Galton–Watson tree $\mathcal{T}$. We let Q_n and P_n denote the corresponding restrictions to $\mathcal{F}_n$. Recall that by Proposition 5.20

$$\mathbb{P}(\mathcal{T}^{*(n)} = T) = \mu^{-n} L_n(T) \mathbb{P}(\mathcal{T}^{(n)} = T).$$

for any finite plane tree T and any integer $n \geq 1$. In other words, $Q_n \ll P_n$ with

$$\frac{dQ_n}{dP_n} = \frac{L_n(\cdot)}{\mu^n}.$$

We know that $X = \limsup_{n \rightarrow \infty} \frac{L_n(\cdot)}{\mu^n}$ is P -a.s. finite by Lemma 5.28. By Corollary 5.29,

$$Q \ll P \quad \Leftrightarrow \quad X < \infty \text{ } Q\text{-a.s.} \quad \Leftrightarrow \quad \mathbb{E}_P[X] = 1$$

and

$$Q \perp P \quad \Leftrightarrow \quad X = \infty \text{ } Q\text{-a.s.} \quad \Leftrightarrow \quad \mathbb{E}_P[X] = 0.$$

The generation size $L_n(\mathcal{T}^*)$ may be interpreted as 1 plus the size of the n -th generation in a Galton–Watson process with immigration, with the immigration law given by $\xi^* - 1$. Hence by Theorem 5.31,

$$X < \infty \text{ } Q\text{-a.s.} \quad \Leftrightarrow \quad \mathbb{E}[\log \max(\xi^* - 1, 1)] < \infty.$$

Note that the definition of ξ^* implies

$$\mathbb{E}[\log \max(\xi^* - 1, 1)] = \mu^{-1} \sum_{k \geq 1} \mathbb{P}(\xi = k) k \log k.$$

Hence the previous condition is also equivalent to

$$\mathbb{E}[\xi \log \xi, \xi \geq 1] < \infty.$$

Summing up,

$$\mathbb{E}_P[X] = 1 \quad \Leftrightarrow \quad \mathbb{E}[\xi \log \xi, \xi \geq 1] < \infty.$$

We know that $W = X$ P -a.s. and $\mathbb{P}(W = 0) \in \{1, \mathbb{P}(|\mathcal{T}| < \infty)\}$ by Proposition 5.25. As $E_P[X] = 0$ if and only if $X = \infty$ Q -a.s., it follows that

$$\mathbb{E}_P[X] = 1 \quad \Leftrightarrow \quad \mathbb{P}(W = 0) = \mathbb{P}(|\mathcal{T}| < \infty).$$

This completes the proof of the Kesten–Stigum theorem. □

5.6. The total population. — The total population of the Galton–Watson process is given by the number of vertices

$$|\mathcal{T}| = \sum_{n \geq 0} Z_n$$

of the associated Galton–Watson tree. We observed in (5.5) that $\mathbb{E}[Z_n] = \mu^n$. Hence:

PROPOSITION 5.33. —

- 1) If $\mu = 1$, then $\mathbb{E}[|\mathcal{T}|] = \infty$.
- 2) If $\mu < 1$, then $\mathbb{E}[|\mathcal{T}|] = \frac{1}{1 - \mathbb{E}[\xi]}$.

PROPOSITION 5.34. — Suppose that $\mu \leq 1$. Then the probability generating series $\mathcal{Z}(z) = \mathbb{E}[z^{|\mathcal{T}|}]$ of the total population and $\phi(z) = \mathbb{E}[z^\xi]$ are related by the equation

$$\mathcal{Z}(z) = z\phi(\mathcal{Z}(z)).$$

Proof. — Letting $(\mathcal{T}^i)_{i \geq 1}$ denote independent copies of $\mathcal{T}$, it holds that

$$|\mathcal{T}| \stackrel{d}{=} 1 + \sum_{i=1}^{\xi} |\mathcal{T}^i|.$$

This readily implies $\mathcal{Z}(z) = z\phi(\mathcal{Z}(z))$. □

Given integers $k, n \geq 1$, let $\mathfrak{T}_{k,n}$ denote the collection of sequences of k finite plane trees, such that the total number of vertices of all trees equals n . Let $\mathfrak{B}_{n-k,n}$ denote the collection of sequences $(d_i)_{1 \leq i \leq n}$ of n non-negative integers satisfying

$$\sum_{i=1}^n d_i = n - k.$$

Recall that any finite plane tree has an outdegree sequence, obtained by listing the outdegrees of its vertices in a depth-first-search order (starting with the root). The outdegree sequence of a forest $(T_i)_{1 \leq i \leq k} \in \mathfrak{T}_{k,n}$ is formed by concatenating the outdegree sequences of the individual trees.

LEMMA 5.35. — Consider the map

$$\mathfrak{T}_{k,n} \times [n] \rightarrow \mathfrak{B}_{n-k,n}$$

that maps a forest $(T_i)_{1 \leq i \leq k}$ with outdegree sequence $(d_i)_{1 \leq i \leq n}$ and an integer $j \in [n]$ to the cyclically shifted list

$$(d_{i+j})_{1 \leq i \leq n}$$

with the index taken modulo n . Then any element of $\mathfrak{B}_{n-k,n}$ has precisely k preimages.

Proof. — By Lemma 4.10, for any integer $\ell \geq 1$ there is a bijection between the collection $\mathfrak{T}_\ell$ of plane trees with ℓ vertices and the collection of all sequences $d'_1, \dots, d'_\ell$ of non-negative integers satisfying

$$\sum_{i=1}^{\ell} (d'_i - 1) = -1$$

and

$$\sum_{i=1}^s (d'_i - 1) \geq 0 \quad \text{for all} \quad 1 \leq s < \ell.$$

The bijection maps a tree to its outdegree sequence.

Consequently, mapping a forest from $\mathfrak{T}_{k,n}$ to its outdegree sequence yields a bijection from $\mathfrak{T}_{k,n}$ to the collection of all sequences $(d_i)_{1 \leq i \leq n}$ of non-negative integers satisfying

$$(5.10) \quad \sum_{i=1}^n (d_i - 1) = -k$$

and

$$(5.11) \quad \sum_{i=1}^s (d_i - 1) > -k \quad \text{for all} \quad 1 \leq s < n.$$

By Lemma 5.10, any sequence $(d_i)_{1 \leq i \leq n}$ of n non-negative integers satisfying (5.10) admits precisely k integers $j \in [n]$ such that shifting $(d_i)_{1 \leq i \leq n}$ cyclically by j yields a sequence satisfying (additionally) Equation (5.11). Hence the k pairs $((d_{i+j})_{1 \leq i \leq n}, -j)$ for these k choices of j , with the index $i+j$ and the shift $-j$ both taken modulo n , correspond precisely to the preimages of $(d_i)_{1 \leq i \leq n}$ under the mapping $\mathfrak{T}_{k,n} \times [n] \rightarrow \mathfrak{B}_{n-k,n}$. $\square$

THEOREM 5.36 (OTTER–DWASS FORMULA). — *Let $(\mathcal{T}^i)_{i \geq 1}$ denote independent copies of $\mathcal{T}$, and let $(\xi_i)_{i \geq 1}$ denote independent copies of ξ . Then, for all integers $k, n \geq 1$,*

$$\mathbb{P} \left(\sum_{i=1}^k |\mathcal{T}^i| = n \right) = \frac{k}{n} \mathbb{P} \left(\sum_{i=1}^n \xi_i = n - k \right).$$

Proof. — If a forest $(T^i)_{1 \leq i \leq k} \in \mathfrak{T}_{k,n}$ has outdegree sequences $(d_i)_{1 \leq i \leq n}$, then

$$\mathbb{P}((\mathcal{T}^i)_{1 \leq i \leq k} = (T^i)_{1 \leq i \leq k}) = \prod_{i=1}^n \mathbb{P}(\xi = d_i).$$

Using Equations (5.10) and (5.11), and Lemma 5.35 it follows that

$$\begin{aligned} \mathbb{P} \left(\sum_{i=1}^k |\mathcal{T}^i| = n \right) &= \mathbb{P} \left(\sum_{i=1}^n (\xi_i - 1) = -k, \sum_{i=1}^s (\xi_i - 1) > -k \text{ for all } 1 \leq s < n \right) \\ &= \frac{k}{n} \mathbb{P} \left(\sum_{i=1}^n \xi_i = n - k \right). \end{aligned}$$

□

The probability $\mathbb{P}(\sum_{i=1}^n \xi_i = n)$ may be approximated using many forms of local limit theorems, thus providing precise asymptotic expressions for $\mathbb{P}\left(\sum_{i=1}^k |\mathcal{T}^i| = n\right)$ when k is fixed and $n \rightarrow \infty$.

Recall the classical Gnedenko local limit theorem for the lattice case:

LEMMA 5.37 (GNEDENKO LOCAL LIMIT THEOREM). — *Let $(X_i)_{i \geq 1}$ denote independent copies of a random integer X with finite first moment $\mathbb{E}[X]$ and finite non-zero variance $\mathbb{V}[X]$. Let $d \geq 1$ denote the smallest positive integer such that the support $\{i \in \mathbb{Z} \mid \mathbb{P}(X = i) > 0\}$ is contained in a lattice of the form $a + d\mathbb{Z}$ for some $a \in \mathbb{Z}$. Fix any such integer a . Then the sum $S_n = X_1 + \dots + X_n$ satisfies $\mathbb{P}(S_n = x) = 0$ for all $x \in \mathbb{Z} \setminus (na + d\mathbb{Z})$ and*

$$\lim_{n \rightarrow \infty} \sup_{x \in na + d\mathbb{Z}} \left| \sqrt{n} \mathbb{P}(S_n = x) - \frac{d}{\sqrt{2\pi \mathbb{V}[X]}} \exp\left(-\frac{(n\mathbb{E}[X] - x)^2}{2n\mathbb{V}[X]}\right) \right| = 0.$$

Applying this local limit theorem, we obtain:

COROLLARY 5.38. — *Suppose that the offspring distribution ξ is critical and has finite variance σ^2 . Let $d \geq 1$ denote the greatest common divisor of all integers $i \geq 1$ with $\mathbb{P}(\xi = i) > 0$. Then $\mathbb{P}(|\mathcal{T}| = n) = 0$ when $n \notin 1 + d\mathbb{N}$, and*

$$\mathbb{P}(|\mathcal{T}| = n) \sim \frac{d}{\sqrt{2\pi\sigma^2}} n^{-3/2}$$

as $n \in 1 + d\mathbb{N}$ tends to infinity.

Note that this means $\mathbb{P}(|\mathcal{T}| = n) > 0$ at least for all large enough $n \in 1 + d\mathbb{N}$.

Proof of Corollary 5.38. — By the Otter–Dwass formula, it holds that

$$\mathbb{P}(|\mathcal{T}| = n) = \frac{1}{n} \mathbb{P}\left(\sum_{i=1}^n (\xi_i - 1) = -1\right).$$

The support of the shifted integer $\xi - 1$ is contained in $-1 + d\mathbb{Z}$ and in no lattice of the form $a + d'\mathbb{Z}$ for $a \in \mathbb{Z}$ and $1 \leq d' < d$. Hence this probability is equal to zero when $-1 \notin -n + d\mathbb{Z}$, that is, when $n \notin 1 + d\mathbb{N}$. By the Gnedenko local limit theorem for $x = -1$ and $X = \xi - 1$, we obtain

$$\mathbb{P}\left(\sum_{i=1}^n (\xi_i - 1) = -1\right) = (1 + o(1)) \frac{d}{\sqrt{2\pi\sigma^2}} n^{-1/2}$$

when $n \in 1 + d\mathbb{N}$ tends to infinity. □

5.7. Local convergence. — For positive integers n the *simply generated tree* $\mathcal{T}_n$ is obtained by conditioning the ξ -Galton-Watson tree $\mathcal{T}$ on having n vertices. That is, the process is restarted over and over again until a realization with total population

$$\sum_{k \geq 0} Z_k = n$$

is generated.

The asymptotic behaviour of the tree $\mathcal{T}_n$ as n tends to infinity forms a vital part of the asymptotic analysis of random trees.

THEOREM 5.39. — *Let $d \geq 1$ denote the greatest common divisor of all integers $i \geq 1$ with $\mathbb{P}(\xi = i) > 0$. If $\mathcal{T}$ is critical and its offspring distribution ξ has finite variance, then*

$$\mathcal{T}_n \xrightarrow{d} \mathcal{T}^*$$

as $n \in 1 + d\mathbb{N}$ tends to infinity.

Proof. — Since $\mathcal{T}^*$ is locally finite and has infinite height, it suffices by Proposition 5.16 to show that for each integer $h \geq 1$ and each finite plane tree T with height h and we have

$$\mathbb{P}(\mathcal{T}_n^{(h)} = T) \rightarrow \mathbb{P}(\mathcal{T}^{*(h)} = T)$$

as n becomes large.

We let $m \geq 1$ denote the number of vertices of T with height less than h . Let $k \geq 1$ denote the number of vertices of T with height h . If $\mathcal{T}^{(h)} = T$ then the total size of $\mathcal{T}$ is given by m plus the sum of the sizes of the k fringe subtrees rooted at its vertices of height h . Each of these is distributed like an independent copy $\mathcal{T}$. Hence, with $(\mathcal{T}^i)_{i \geq 1}$ denoting independent copies of $\mathcal{T}$ we have

$$\mathbb{P}(|\mathcal{T}| = n \mid \mathcal{T}^{(h)} = T) = \mathbb{P}\left(\sum_{i=1}^k |\mathcal{T}^i| = m\right).$$

By the Otter-Dwass formula and the local limit theorem, this simplifies to

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^k |\mathcal{T}^i| = m\right) &= \frac{k}{n} \mathbb{P}\left(\sum_{i=1}^n \xi_i = n - k\right) \\ &\sim \frac{kd}{\sqrt{2\pi\sigma^2}} n^{-3/2}. \end{aligned}$$

This is precisely k times the asymptotic expression for $\mathbb{P}(|\mathcal{T}| = n)$ we obtained in Corollary 5.38. Hence

$$\mathbb{P}(\mathcal{T}_n^{(h)} = T) = \frac{\mathbb{P}(|\mathcal{T}| = n, \mathcal{T}^{(h)} = T)}{\mathbb{P}(|\mathcal{T}| = n)} \rightarrow k \mathbb{P}(\mathcal{T}_n^{(h)} = T).$$

By Proposition 5.20 this equals $\mathbb{P}(\mathcal{T}^{*(h)} = T)$ and the proof is complete. $\square$

References

- [1] F. BERGERON, G. LABELLE, AND P. LEROUX, *Combinatorial species and tree-like structures*, vol. 67 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1998. Translated from the 1994 French original by Margaret Readdy, With a foreword by Gian-Carlo Rota.
- [2] M. BODIRSKY, É. FUSY, M. KANG, AND S. VIGERSKE, *Boltzmann samplers, Pólya theory, and cycle pointing*, SIAM J. Comput., 40 (2011), pp. 721–769.
- [3] R. DIESTEL, *Graph theory*, vol. 173 of Graduate Texts in Mathematics, Springer, Heidelberg, fourth ed., 2010.
- [4] M. DRMOTA, *Random trees*, SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.
- [5] P. FLAJOLET AND R. SEDGEWICK, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.
- [6] S. JANSON, *Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation*, Probab. Surv., 9 (2012), pp. 103–252.
- [7] R. VAN DER HOFSTAD, *Random graphs and complex networks. Volume 1*, vol. 43 of Camb. Ser. Stat. Probab. Math., Cambridge: Cambridge University Press, 2017.