UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2019:35

Recent development in conditioned Galton-Watson trees

Anton Falk

Examensarbete i matematik, 15 hp
Handledare: Xing Shi Cai
Examinator: Martin Herschend
Juni 2019

Department of Mathematics
Uppsala University

Recent development in conditioned Galton-Watson trees

Anton Falk

June 23, 2019

Contents

1 Introduction

1.1 Background

In the 19th century England, the number of aristocratic surnames kept decreasing whereas the population of the country exploded. Francis Galton thus asked what the probability of the disappearance of a surname is. The question was answered by Henry William Watson. Together, the two wrote a paper titled *On the probability of the extinction of families* [10], where they introduced a mathematical model describing the evolution of family names.

Nowadays this model is called the *Galton-Watson process*. Roughly speaking, the model is defined as follows: A family starts with one ancestor. Then this person gets a random number of children. Each of these children also gets a random number of children independently, and so on.

As one would usually draw a family tree, the same can be done for a Galton-Watson process. Such a tree is referred to as a *Galton-Watson tree*. This will be made precise in the next subsection.

As a simplification of reality, each person is assumed to get a random number of children independently and also that the probability that a person gets a specific number of children is the same for all the persons in the family.

For such a family tree there are two possible cases. The first is that the number of generations is infinite. The other is that the family eventually becomes extinct. Interestingly, if the expected number of children that each person gets is not greater than 1, the family dies out with probability 1. However, if it is greater than 1, there is a positive probability that the family never dies out.

Let us now restrict these Galton-Watson trees by fixing the number of persons in the tree. This is called conditioning, hence the name of the title. We may now ask questions like: What is the probability that such a tree looks a certain way? Are there any patterns in such a tree if this number of nodes is very large? This review will focus on *conditioned Galton-Watson trees* and some of their properties, in particular recent development within this field.

1.2 Simply generated trees

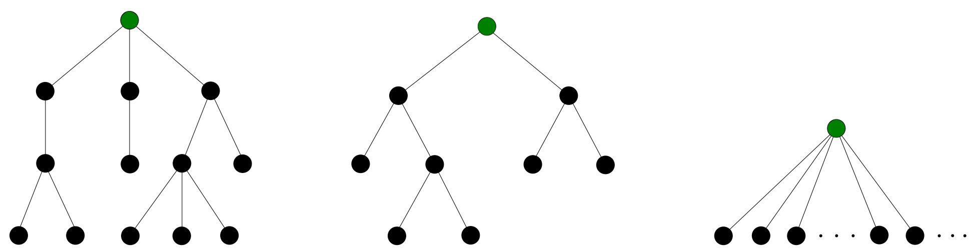


Figure 1: Examples of plane trees. The roots are colored green.

We begin this review by defining simply generated trees which are more general than Galton-Watson trees. A *rooted tree* is a tree in which one node is distinguished as a root. The root is usually denoted o in this text. A *rooted and ordered tree*, also called a *plane tree*, is a rooted tree where the children of each node v, denoted $v_1, v_2, ..., v_{d^+(v)}$, are ordered. Here $d^+(v)$ denotes the outdegree of v, i.e., the number of children of v.

We let $\mathfrak{T}$ denote the set of rooted and ordered trees. Let $\mathfrak{T}_{\text{lf}} \subset \mathfrak{T}$ denote such trees that are locally finite i.e., all nodes have finite outdegrees, and let $\mathfrak{T}_{\text{f}} \subset \mathfrak{T}_{\text{lf}}$ denote such trees that are finite. Lastly, for a positive integer n, let $\mathfrak{T}_n \subset \mathfrak{T}_{\text{f}}$ denote the trees of size n.

A *weight sequence* $(w_k)_{k\geq 0}$, sometimes denoted $\mathbf{w}$, is a sequence of non-negative real numbers. The *weight* of a tree $T \in \mathfrak{T}_{\text{f}}$ is defined as the product $w(T) = \Pi_{v\in T} w_{d^+(v)}$. Given a weight sequence $\mathbf{w}$, the *simply generated tree of size* n, T_n, is a random tree taking values in the set $\mathfrak{T}_n$. The probability of $T_n = T$, for $T \in \mathfrak{T}_n$, is

$$\mathbb{P}(T_n = T) = \frac{w(T)}{Z_n}, \tag{1.1}$$

where $Z_n = \Sigma_{T\in\mathfrak{T}_n} w(T)$ is called the *partition function*. We only consider cases where $Z_n > 0$.

When analyzing simply generated trees, it is convenient to make use of generating functions. We start by defining $\phi(t) := \Sigma_{k=0}^{\infty} w_k t^k$. Let $\rho \in [0, \infty]$

be its radius of convergence. We also define

$$\psi(t) := \frac{t\phi'(t)}{\phi(t)}. \tag{1.2}$$

Furthermore, we define the constant

$$\nu := \psi(\rho), \tag{1.3}$$

where $\psi(\infty) := \lim_{t\to\infty}\psi(t)$. This always exists, by [5, lemma 3.1].

Let ξ be a random variable taking values in $\mathbb{N}_0$. A *Galton-Watson tree* with offspring distribution ξ is a random object taking its values in the set $\mathfrak{T}_{\mathrm{lf}}$, which we denote by $\mathcal{T}$. It is recursively constructed by starting with a root and let this root have ξ children. Then each node v is given an independent copy ξ_v, and spawns ξ_v new children.

The case $\mathbb{E}\xi < 1$ is referred to as subcritical, the case $\mathbb{E}\xi = 1$ is called critical and $\mathbb{E}\xi > 1$ is called supercritical. Interestingly, one may show that in the subcritical and the critical cases $\mathcal{T}$ is finite almost surely (meaning that the event has probability 1 of happening). In the supercritical case, the probability that the tree is infinite is always positive.

This review will mainly focus on *conditioned* Galton-Watson trees, i.e., Galton Watson-trees conditioned on the number of vertices.

On the one hand, such a tree can be described as a special case of simply generated trees. Assume $\mathbf{w}$ is a weight sequence which is also a probability distribution, i.e., $\Sigma_{k=0}^{\infty} w_k = 1$. Let $\mathbb{P}(\xi = k) = w_k$. Then, for $T \in \mathfrak{T}_f$, we have $\mathbb{P}(\mathcal{T} = T) = w(T)$. As a result, $Z_n = \mathbb{P}(\mathcal{T} \in \mathfrak{T}_n)$. We now condition the Galton-Watson tree generated by ξ to have n number of nodes. Then, for $T \in \mathfrak{T}_n$,

$$\mathbb{P}(\mathcal{T} = T | \mathcal{T} \in \mathfrak{T}_n) = \frac{\mathbb{P}(\mathcal{T} = T)}{Z_n} = \frac{w(T)}{Z_n}. \tag{1.4}$$

On the other hand, a simply generated tree is equivalent to a conditioned Galton Watson-tree if and only if the radius of convergence ρ of its corresponding weight sequence is positive. We define an equivalence relation $\sim$ on the class of weight sequences by

$$\mathbf{w} \sim \mathbf{w}' \iff \text{there exist } a > 0,\ b > 0 \text{ such that } w'_k = ab^k w_k \text{ for every } k \geq 0. \tag{1.5}$$

In fact, any weight sequence $\mathbf{w}$ is equivalent to a weight sequence that is a probability distribution if and only if its radius of convergence ρ is greater than 0. Thus, for such a weight sequence, all its equivalent sequences induce the same probability distribution as a conditioned Galton-Watson tree.

2 Convergence of simply generated trees

In this section, we investigate the convergence of sequences of simply generated trees in the subcritical and critical cases. As we will see, such a sequence converges to the *modified Galton-Watson tree*, which has two possible distinct types depending on the underlying distribution. In the critical case the tree has an infinite spine, and in the other case the spine is finite and ends with an explosion, i.e., a vertex with infinite outdegree. These limits will be precisely stated in the end of this section, in the form of an important theorem. We start this section by investigating the notion of convergence of general plane trees. Afterwards, we will describe the construction of the modified Galton-Watson trees and subsequently present the aforementioned theorem.

2.1 Convergence of trees

We start with defining the convergence of a sequence of deterministic trees and then proceed to do this for random trees.

Let $\mathcal{V}_\infty := \bigcup_{k=0}^{\infty} \mathbb{N}_1^k$, the set of finite strings of positive integers, where $\mathbb{N}_1^0 = \emptyset$. The interpretation is that this is an universal set of potential nodes for a plane tree, where $\emptyset$ is the root, $1, 2, 3, ...$ are the possible children of the root, $v1, v2, v3, ...$ are the possible children of v and so on. Let $\mathcal{U}_\infty$ denote the tree obtained by connecting all the potential parent-child relations in $\mathcal{V}_\infty$ by an edge. This tree is called the *Ulam-Harris tree*. For a plane tree $T \in \mathfrak{T}$, its embedding into $\mathcal{U}_\infty$, as defined in Janson, S. [5], is the identification of each T with the subset $V \subset \mathcal{V}_\infty$ satisfying

$$\begin{aligned}
&\emptyset \in \mathcal{V}, \\
&i_1...i_{k+1} \in \mathcal{V} \implies i_1...i_k \in \mathcal{V}, \\
&i_1...i_k i \in \mathcal{V} \implies i_1...i_k j \in \mathcal{V} \text{ for all } j \leq i.
\end{aligned} \tag{2.1}$$

Note that a plane tree $T \in \mathfrak{T}$ is uniquely determined by its degree sequence $(d_T^+(v))_{v \in \mathcal{V}_\infty}$, where we require $d_T^+(v) = 0$ if $v \notin V(T)$. Let $\overline{\mathbb{N}}_0 := \mathbb{N}_0 \cup \{\infty\}$, and consider the set $\overline{\mathbb{N}}_0^{\mathcal{V}_\infty}$, and a sequence (d_v) where $v \in \overline{\mathbb{N}}_0^{\mathcal{V}_\infty}$, with the condition $d_{i_1 \ldots i_k i} = 0$ when $i > d_{i_1 \ldots d_k}$. This gives a natural bijection between $\mathfrak{T}$ and the sequences fulfilling this condition. Thus we may regard the tree T as an element in $\overline{\mathbb{N}}_0^{\mathcal{V}_\infty}$.

Consider $\mathbb{N}_0$ and give it the discrete topology. Now consider the one-point compactification into $\overline{\mathbb{N}}_0$. The point here is that this compactification gives a metrizable topological space. Give $\overline{\mathbb{N}}_0^{V\infty}$ the product topology and $\mathfrak{T}$ its subspace topology. Note that this subspace is compact, since it is closed in $\overline{\mathbb{N}}_0^{V\infty}$. In this metric space, it is easy to see that if T_n, T are trees in $\mathfrak{T}$, then $T_n \to T$ if and only if the outdegrees converge pointwise, i.e.,

$$d^+_{T_n}(v) \to d_{T^+}(v) \qquad \text{for each } v \in \mathcal{V}_\infty, \tag{2.2}$$

or equivalently,

$$d^+_{T_n}(v) \to d_{T^+}(v) \qquad \text{for each } v \in V(T). \tag{2.3}$$

There is an equivalent formulation of convergence in $\mathfrak{T}$, which is visually easy to grasp. First, we consider only locally finite trees, i.e., trees $T \in \mathfrak{T}_{\mathrm{lf}}$. A *truncation at level m of a tree T*, denoted $T^{(m)}$, is the tree obtained by taking T and then removing all generations of height greater than m. We have the following.

Lemma 2.1.1 (Found in Janson [5], Lemma 6.2) *If T is locally finite, then for any sequence of trees $T_n \in \mathfrak{T}$,*

$$T_n \to T \iff T_n^{(m)} \to T^{(m)} \text{for each } m \iff$$
$$T_n^{(m)} = T^{(m)} \text{for each } m \text{ and large } n.$$

However, in order to define convergence for a tree $T \in \mathfrak{T}$, we need to allow for it to have nodes of infinite outdegree. As an example of why the above is not sufficient, let S_i, $i \in \overline{\mathbb{N}}_1$, denote the star of degree i (possibly infinite). We want the sequence $S_1, S_2, \ldots$ to converge to S_∞, but the above definition does not cover this case, for there is no $i \in \mathbb{N}_1$ such that $S_\infty = S_i$. Therefore we define a *left ball* of a tree $T \in \mathfrak{T}$, denoted $T^{[m]}$, as a tree truncated at height m but also *pruned* so that only the first m children of each node is kept. We have the following equivalences.

Lemma 2.1.2 (Found in Janson [5], Lemma 6.3) *For any trees $T \in \mathfrak{T}$, and any sequence $T_n \in \mathfrak{T}$,*

$$T_n \to T \iff T_n^{[m]} \to T^{[m]} \text{for each } m \iff$$
$$T_n^{[m]} = T^{[m]} \text{for each } m \text{ and all large } n.$$

Weak convergence, or convergence in distribution, is well defined in the above mentioned topological space $\mathfrak{T}$. Analogously to the deterministic case, we have the following equivalence for this type of convergence.

Lemma 2.1.3 (Found in Janson [5], Remark 6.4) *For any random trees $T, T_n \in \mathfrak{T}$,*

$$T_n \xrightarrow{d} T \iff T_n^{[m]} \xrightarrow{d} T^{[m]} \text{ for each } m.$$

If $T \in \mathfrak{T}_{lf}$ almost surely (which is the case when T is a subcritical or critical Galton Watson-tree), this is equivalent to

$$T_n \xrightarrow{d} T \iff T_n^{(m)} \xrightarrow{d} T^{(m)} \text{ for each } m.$$

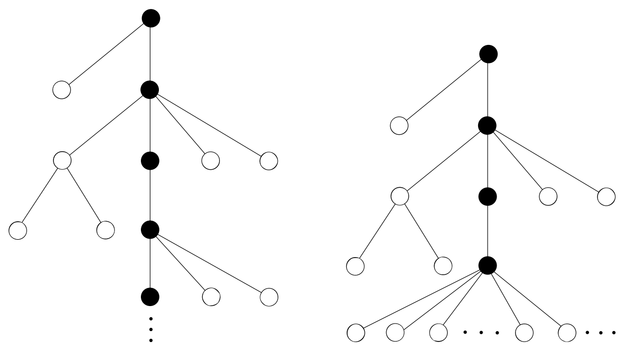
2.2 A modified Galton Watson-tree

We will now describe the modified Galton Watson-tree. Given a random variable ξ with expected value $\mu \leq 1$, We define a new random variable $\hat{\xi}$, referred to as the *size-biased* variable, with distribution:

$$\mathbb{P}(\hat{\xi} = k) = \begin{cases} k\pi_k, & k = 0, 1, 2, ... \\ 1 - \mu, & k = \infty \end{cases} \tag{2.4}$$

Note that this random variable attains values at least 1. We call a vertex *special* if it has offspring $\hat{\xi}$ and *normal* if it has offspring ξ. Now we construct a modified Galton-Watson tree as follows. The root is a special node. If a special node v has $d^+(v) < \infty$, exactly one of its children is selected, uniformly at random, to be special. If $d^+(v) = \infty$ then every child is normal. All normal nodes have normal children.

There are two main cases for the appearance of a modified Galton Watson-tree. One is the subcritical case when $\mu < 1$, originally constructed by Jonsson and Stefánsson [7]. In this case there is a positive probability that a special node has infinite degree. Such a tree may be described as consisting of a finite spine of special nodes ending with an explosion, i.e., a node of infinite outdegree. The other case is the critical case, originally defined by Kesten [8], where the probability of a special node getting an infinite outdegree is 0. Such a tree consists of an infinite spine of special nodes. See Figure 2 for an illustration of these cases.



(a) Case 1: Infinite spine. (b) Case 2: Finite spine ending with explosion.

Figure 2: Modified Galton Watson-tree, the two cases. The black and white nodes are special and normal, respectively.

2.3 Limits of conditioned Galton-Watson trees

The following theorem is due to Janson [5, theorem 7.1]. Roughly, it says that the limit, as n goes to infinity, of a sequence of simply generated trees is the modified Galton Watson-tree. It also, explicitly, gives the offspring distributions of this limit.

Theorem 2.3.1 (Janson [5], Theorem 7.1) *Let $\mathbf{w} = (w_k)_{k\geq 0}$ be any weight sequence with $w_0 > 0$ and $w_k > 0$ for some $k \geq 2$.*

(i) If $\nu \geq 1$, let τ be the unique number in $[0, \rho]$ such that $\psi(\tau) = 1$.

(ii) If $\nu < 1$, let $\tau := \rho$.

In both cases, $0 \leq \tau < \infty$ and $0 < \phi(\tau) < \infty$. Let

$$\pi_k := \frac{\tau^k w_k}{\phi(\tau)}, \qquad k \geq 0; \tag{2.5}$$

then $(\pi_k)_{k\geq 0}$ is a probability distribution, with expectation

$$\mu = \psi(\tau) = \min(\nu, 1) \leq 1 \tag{2.6}$$

and variance $\sigma^2 = \tau\psi'(\tau) \leq \infty$. Let $\hat{\mathcal{T}}$ be the infinite modified Galton-Watson tree for the distribution $(\pi_k)_{k\geq 0}$. Then $\mathcal{T}_n \xrightarrow{d} \hat{\mathcal{T}}_n$ as $n \to \infty$, in the topology described in section 2.1.

Furthermore, in case (i), $\mu = 1$ (the critical case) and $\hat{\mathcal{T}}$ is locally finite with an infinite spine; in case (ii) $\mu = \nu < 1$ (the subcritical case) and $\hat{\mathcal{T}}$ has a finite spine ending with an explosion.

Given a weight sequence, we call τ, as defined in Theorem 2.3.1, the *fundamental constant.*

3 Local limits of large Galton-Watson trees rerooted at a random vertex

In this section, we investigate a type of local convergence, namely convergence of the vicinity of a randomly chosen vertex in a simply generated tree, as the number of nodes tends to infinity. The constructions and the theorems are due to Stufler [9].

In order to describe the convergence of pointed plane trees, we introduce a suitable topological space. We do this in a manner similar to the discussions in section 2.

The nature of a limit, in this sense, is dependent on the underlying weight sequence. For a weight sequence, recall (equation 1.3) the constant ν. If $\nu \geq 1$, then it is of type I, if $0 < \nu < 1$ then it is of type II and if $\nu = 0$ then it is of type III.

A *pointed plane tree*, denoted by (T, v_0) or $T^\bullet$, is a plane tree $T \in \mathfrak{T}$ where one of its nodes, denoted by v_0, is distinguished. Let $\mathfrak{T}^\bullet$ denote the set of pointed plane trees. Note that there is an unique path $v_0v_1...v_h$ of length h from v_0 to the root $o = v_h$, where h denotes the height of v_0. For convenience, for a node v, we refer to a sibling being earlier (later) in the sibling-order as being to the left (right) of v.

We construct a tree in which all pointed plane trees can be embedded. Let u_i, for each $i \geq 0$, be vertices such that u_{i+1} is the parent of u_i. For each $i \geq 0$, we spawn an infinite number of descendants of the vertex u_{i+1} both to the left and to the right of u_i. We let each of these spawned descendants be the root of a copy of the Ulam-Harris tree. We denote this constructed tree $\mathcal{U}_\infty^\bullet$, and its vertex set $\mathcal{V}_\infty^\bullet$. The path $u_0u_1u_2...$ is referred to as the *spine* of $\mathcal{U}_\infty^\bullet$.

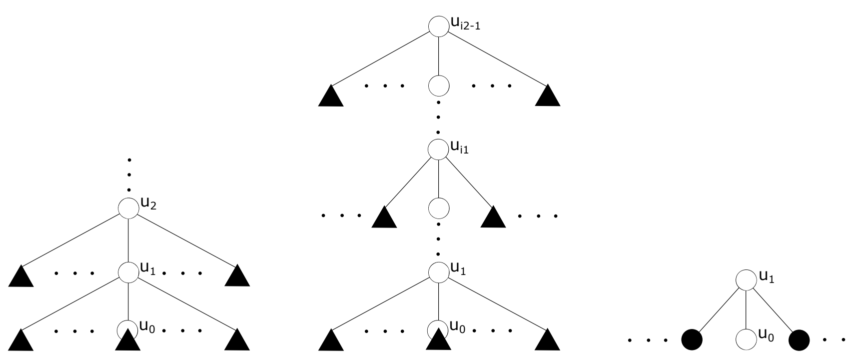
Each pointed plane tree $T^\bullet$ can now be embedded into $\mathcal{U}_\infty^\bullet$ in a natural way, by mapping v_i to u_i for each $i \geq 0$, and then mapping the remaining vertices so that the order and the outdegrees are preserved.

Consider $\mathbb{N}_0$ and give it the discrete topology. We endow $\overline{\mathbb{N}}_0$ with the corresponding one-point compactification topology. This is a Polish space, i.e., separable and completely metrizable. Thus, so are the product topology $\overline{\mathbb{N}}_0 \times \overline{\mathbb{N}}_0$, and consequently, the same is true for the disjoint union topology $\{*\} \cup (\overline{\mathbb{N}}_0 \times \overline{\mathbb{N}}_0)$.

Note that a pointed plane tree $T^\bullet \in \mathfrak{T}^\bullet$ is uniquely determined by its degree sequence $(d^+_{T^\bullet}(v))_{v \in \mathcal{V}^\bullet_\infty}$, where $d^+_{T^\bullet}(v) \in \overline{\mathbb{N}}_0$ for $v \notin \{u_{i+1}, i \geq 0\}$, and $d^+_{T^\bullet}(u_{i+1}) \in \{*\} \cup (\overline{\mathbb{N}}_0 \times \overline{\mathbb{N}}_0)$ for $i \geq 0$. For a vertex u_{i+1}, where $i \geq 0$, $d^+(u_{i+1}) = *$ is interpreted as the vertex being childless, except for the child u_i. If $d^+(u_i) = (m, n)$, then the numbers m and n represent that there are m descendants to the left of u_i and n descendants to its right, respectively. Thus we may regard the tree $T^\bullet$ as an element in the space consisting of all sequences $(d^+_{T^\bullet}(v))_{v \in \mathcal{V}^\bullet_\infty}$. It is a Polish space. Moreover, it is compact. The subspace $\mathfrak{T}^\bullet$ is closed, and thus also compact. Thus, for the metrizable space $\mathfrak{T}^\bullet$, convergence in distribution is well defined.

3.1 The limit theorems

We now have the prerequisites to state the limit theorems for each of the three aforementioned types of weight sequences. We denote the limit by $\mathcal{T}^*$, and describe its construction for each of these types. See Figure 3.



(a) Type I (b) Type II (c) Type III

Figure 3: The three types of the limit tree $\mathcal{T}^*$. The white and black circles represent the spine nodes and the leaves, respectively. The triangles represent the nodes which have an unconditioned Galton-Watson tree $\mathcal{T}$ attached.

Type I

For the type I setting, the tree $\mathcal{T}^*$ is constructed as follows. Let u_0 be the root of an independent copy of the Galton-Watson tree $\mathcal{T}$. For each $i \geq 0$, we let each u_{i+1} receive an independent copy of the size-biased random variable $\hat{\xi}$ (see equation 2.4) and let it have $\hat{\xi}$ offspring. Then, u_i is identified with a child of this vertex, chosen uniformly at random. For each $i = 1, 2, 3, ...$, let each offspring of u_i, except u_{i-1}, be the root of an independent copy of the Galton-Watson tree $\mathcal{T}$. Also, let u_0 be the root of such a copy.

We now state the limit theorem.

Theorem 3.1.1 (Stufler [9], Theorem 5.1) *If the weight-sequence $\boldsymbol{w}$ has type I, then*

$$(\mathcal{T}_n, v_0) \xrightarrow{d} \mathcal{T}^*$$

in the space $\mathfrak{T}^\bullet$.

For the next theorem, we need some definitions. We define the *total variation distance* between two random variables X and Y, in a countable sample space S, as

$$d_{TV}(X, Y) = \sup_{\mathcal{E} \subset S} |\mathbb{P}(X \in \mathcal{E}) - \mathbb{P}(Y \in \mathcal{E})|. \tag{3.1}$$

Also, let $T \in \mathfrak{T}$ be a plane tree, with a vertex $v \in T$. For some $k \geq 0$, let v_k be the kth ancestor of v. We define $H_k(T, v)$ as the fringe subtree of T, rooted at v_k, with the distinguished root v.

Theorem 3.1.2 (Stufler [9], Theorem 5.2) *Suppose that the weight-sequence has type I and the offspring distribution ξ has finite variance. Let k_n be an arbitrary sequence of non-negative integers that satisfies $k_n/\sqrt{n} \to 0$. Then*

$$d_{TV}(H_{k_n}(\mathcal{T}_n, v_0), H_{k_n}(\mathcal{T}^*, u_0)) \to 0$$

as n becomes large.

Here, the redundant notation $(\mathcal{T}^*, u_0)$ is used to emphasize that u_0 is the distinguished node of $\mathcal{T}^*$.

Type II

For the type II setting, the tree $\mathcal{T}^*$ is constructed as follows: Start with u_0. For each $i = 1, 2, ..., i_1$, let $\hat{\xi}_i$ be independent copies of $\hat{\xi}$, where i_1 denotes the first index j such that $\hat{\xi}_j = \infty$. For each $i = 1, 2, ..., i_1 - 1$, spawn u_i

with $\hat{\xi}_i$ offspring and let u_{i-1} be identified with a uniformly at random chosen vertex among these offspring. Let u_{i_1-1} be a descendant of u_{i_1} with an infinite number of siblings to the left and to the right.

Next, for each $i = i_1 + 1, ..., i_2 - 1$, let $\hat{\xi}_i$ be independent copies of $\hat{\xi}$, where i_2 denotes the first index $j > i_1$ such that $\hat{\xi}_j = \infty$. For each $i = i_1 + 1, i_1 + 2, ..., i_2 - 1$, spawn u_i with $\hat{\xi}_i$ offspring and let u_{i-1} be identified with a uniformly at random chosen vertex among these offspring. For each $i = 1, 2, 3, ..., i_2 - 1$, let each offspring of u_i, except u_{i-1}, be the root of an independent copy of the Galton-Watson tree $\mathcal{T}$. Also, let u_0 be the root of such a copy. Note that we do not include a vertex corresponding to i_2 in the construction.

Before we state the next theorem, we need some definitions. A function $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is *slowly varying*, if for any fixed $t > 0$, it holds that $\lim_{x\to\infty} \frac{h(tx)}{h(x)} = 1$.

We also define $o_p(a_n)$ as an unspecified sequence of random variables X_n, such that that $X_n/a_n \xrightarrow{p} 0$ as $n \to \infty$, where a_n is a sequence of numbers.

Theorem 3.1.3 (Stufler [9], Theorem 5.3) *Suppose that the weight-sequence* $\mathbf{w}$ *has type II. Let* μ *be the mean of the corresponding canonical weight sequence. If the maximum degree* $\Delta(\mathcal{T}_n)$ *satisfies*

$$\Delta(\mathcal{T}_n) = (1 - \mu)n + o_p(n),$$

then it holds that

$$(\mathcal{T}_n, v_0) \xrightarrow{d} \mathcal{T}^*$$

in the space $\mathfrak{T}^\bullet$. *In particular, this is the case when there is a constant* $\alpha > 2$ *and a slowly varying function* f *such that for all* k,

$$\mathbb{P}(\xi = k) = f(k)k^{-\alpha}.$$

Type III

In the type III setting, the tree $\mathcal{T}^*$ consists of two marked vertices u_0 and u_1 such that u_0 is a child of u_1, with infinitely many siblings to the left and to the right. Every child of u_1, including u_0, is a leaf. We have the following.

Theorem 3.1.4 (Stufler [9], Proposition 5.5) *If the weight-sequence* $\mathbf{w}$ *has type III, then the following claims are equivalent.*

1. $(\mathcal{T}_n, v_0) \xrightarrow{d} \mathcal{T}^*$ *in* $\mathfrak{T}^\bullet$.

2. $h_{\mathcal{T}_n}(v_0) \xrightarrow{p} 1$.

3. The maximum degree $\Delta(\mathcal{T}_n)$ *satisfies* $\Delta(\mathcal{T}_n) = n + o_p(n)$.

A general class of weight-sequences that demonstrate this behaviour is given by

$$\omega_k = k!^\alpha$$

with $\alpha > 0$ *a constant.*

4 Number of subtrees

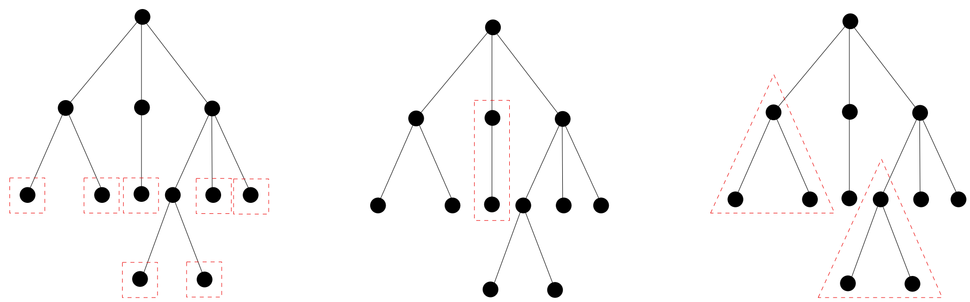
In this section, we investigate the number of subtrees of conditioned Galton-Watson trees. These subtrees are of two different types: fringe, and non-fringe (also called general). We mainly focus on the distributions of the number of such subtrees as the number of vertices goes to infinity. A recursive characterization of trees is introduced, which is a useful tool in the investigation of subtrees.

We consider plane trees, and also rooted trees. For a rooted tree T, a *non-fringe*, or *general*, subtree T' is a subtree of $T' \subset T$ with no other requirement except for it to be a proper tree. Note that such a subtree has a unique vertex $o' \in T'$ of minimal distance from the root o of T. We choose o' to be the root of T'. As a special case of a non-fringe subtree, we define a *fringe subtree* $T_v \subset T$ as a tree rooted at some vertex $v \in T$, such that it contains all descendants of v in T. See Figure 4 for examples.

A *toll function*, also called a *functional*, is a function $f : \mathfrak{T} \to \mathbb{R}$ assigning to each plane tree $T \in \mathfrak{T}$ a real number. With this as a building block, we define an *additive functional* F as a function $F : \mathfrak{T} \to \mathbb{R}$ which satisfies the condition $F(T) = f(T) + \Sigma_{i=1}^{d^+(o)} F(T_i)$, where T_i is the fringe subtree of T rooted at the descendant i of the root o. This gives a useful recursive characterization of trees.

As an example, let the toll function $f_\bullet$ be defined as $f_\bullet(T) = \mathbb{1}\{T \cong \bullet\}$, where $\bullet$ denotes the single-vertex tree. The function $f_\bullet$ is the indicator function which outputs 1 if T is isomorphic to $\bullet$, and 0 otherwise. Let $F_\bullet$ be the induced additive functional. It may be interpreted as the function which counts the number of leaves in an input tree T.

4.1 Law of large number and a Central limit theorem



(a) Copies of a leaf. (b) Copies of an edge. (c) Copies of a cherry.

Figure 4: Examples of the number of fringe subtree-copies of a given tree.

Theorem 4.1.1 is a law of large numbers for $F(\mathcal{T}_n)$. It is stated for two versions of random fringe subtrees. Let T_* be the random finite plane tree obtained by first choosing $T \in \mathfrak{T}_{\mathrm{f}}$ according to some distribution, followed by selecting a fringe subtree of T by choosing a root among its vertices uniformly at random. T_* is referred to as the *annealed* version of fringe subtrees. In contrast, we refer to the conditioned random tree $T_* \mid T$ as the *quenched* version of fringe subtrees, i.e., where T is fixed, and a fringe subtree of T is chosen uniformly at random.

For a finite plane tree $T \in \mathfrak{T}_{\mathrm{f}}$, let $n_T(\cdot)$ denote the additive functional counting the number of fringe subtrees isomorphic to T. See Figure 4. It is induced by the indicator function $f_T(T') := \mathbb{1}\{T' \cong T\}$. Note that the previous example is the case when T is the one-vertex tree.

Theorem 4.1.1 (Aldous [1] et al., also found in [6] as Theorem 1.3)
Let $\mathcal{T}_n$ be a conditioned Galton-Watson tree with n nodes, defined by an offspring distribution ξ with $\mathbb{E}\xi = 1$, and let $\mathcal{T}$ be the corresponding unconditioned Galton-Watson tree. Then, as $n \to \infty$:

(i) (Annealed version) The fringe subtree $\mathcal{T}_{n,*}$ converges in distribution to the Galton-Watson tree $\mathcal{T}$. I.e., for every fixed tree T,

$$\frac{\mathbb{E}n_T(\mathcal{T}_n)}{n} = \mathbb{P}(\mathcal{T}_{n,*} = T) \to \mathbb{P}(\mathcal{T} = T).$$

Equivalently, for any bounded functional f on $\mathfrak{T}$,

$$\frac{\mathbb{E}F(\mathcal{T}_n)}{n} = \mathbb{E}f(\mathcal{T}_{n,*}) \to \mathbb{E}f(\mathcal{T}).$$

(ii) *(Quenched version) The conditional distributions $\mathcal{L}(\mathcal{T}_{n,*}|\mathcal{T}_n)$ converge to the distribution of $\mathcal{T}$ in probability. I.e., for every fixed tree T,*

$$\frac{n_T(\mathcal{T}_n)}{n} = \mathbb{P}(\mathcal{T}_{n,*} = T|\mathcal{T}_n) \xrightarrow{p} \mathbb{P}(\mathcal{T} = T).$$

Equivalently, for any bounded functional f on $\mathfrak{T}$,

$$\frac{F(\mathcal{T}_n)}{n} = \mathbb{E}f(\mathcal{T}_{n,*}|\mathcal{T}_n) \xrightarrow{p} \mathbb{E}f(\mathcal{T}).$$

The following theorem is a central limit theorem for $F(\mathcal{T}_n)$. It says that $F(\mathcal{T}_n)$ is asymptotically normally distributed, and specifies the mean and variance of this random variable.

Theorem 4.1.2 (Janson [6], Theorem 1.5) *Let $\mathcal{T}_n$ be a conditioned Galton-Watson tree of order n with offspring distribution ξ, where $\mathbb{E}\xi = 1$ and $0 < \sigma^2 := \text{Var }\xi < \infty$, and let $\mathcal{T}$ be the corresponding unconditioned Galton-Watson tree. Suppose that $f : \mathfrak{T} \to \mathbb{R}$ is a functional of rooted trees such that $\mathbb{E}|f(\mathcal{T})| < \infty$, and let $\mu := \mathbb{E}f(\mathcal{T})$.*

(i) *If $\mathbb{E}f(\mathcal{T}) \to 0$ as $n \to \infty$, then*

$$\mathbb{E}F(\mathcal{T}_n) = n\mu + o(\sqrt{n}).$$

(ii) *If*

$$\mathbb{E}f(\mathcal{T}_n)^2 \to 0$$

as $n \to \infty$, and

$$\Sigma_{n=0}^{\infty} \frac{\sqrt{\mathbb{E}(f(\mathcal{T}_n)^2)}}{n} < \infty,$$

then

$$\text{Var } F(\mathcal{T}_n) = n\gamma^2 + o(n)$$

where

$$\gamma^2 := 2\mathbb{E}\big(f(\mathcal{T})(F(\mathcal{T}) - |\mathcal{T}|\mu)\big) - \text{Var } f(\mathcal{T}) - \mu^2/\sigma^2$$

is finite; moreover,

$$\frac{F(\mathcal{T}_n) - n\mu}{\sqrt{n}} \xrightarrow{d} N(0, \gamma^2).$$

Theorem 4.1.2 implies that, roughly, $n_T(\mathcal{T}_n)$ has the mean $n\mathbb{P}(\mathcal{T} \cong T)$, and the variance $n\gamma^2$, for the constant γ^2 as defined in the theorem.

4.2 Non-fringe subtrees in conditioned Galton-Watson trees

For a plane tree T, we let $S(T)$ be the number of non-fringe subtrees of T. Also, let $R(T)$ be the number of non-fringe subtrees which contain the root o of T. Theorem 4.2.1 says that, as the number of vertices n in a sequence of conditioned, critical Galton Watson trees $\mathcal{T}_n$ goes to infinity, $S(\mathcal{T}_n)$ and $R(\mathcal{T}_n)$ are both lognormally distributed. Furthermore, for large n, these two random variables have roughly the same distribution, with the relationship between these two variables's mean and variance stated.

Theorem 4.2.1 (Cai [3] et al., Theorem 2.1) *Let $\mathcal{T}_n$ be a random conditioned Galton-Watson tree of order n, defined by some offspring distribution ξ with $\mathbb{E}\xi = 1$ and $0 < \operatorname{Var}\xi < \infty$. Then there exist constants $\mu, \sigma^2 > 0$ such that, as $n \to \infty$,*

$$\frac{\log R(\mathcal{T}_n) - \mu n}{\sqrt{n}} \xrightarrow{d} N(\mu, \sigma^2),$$
$$\frac{\log S(\mathcal{T}_n) - \mu n}{\sqrt{n}} \xrightarrow{d} N(\mu, \sigma^2),$$

where $N(\mu, \sigma^2)$ denotes the normal distribution with mean 0 and variance σ^2. Furthermore,

$$\mathbb{E}[\log R(\mathcal{T}_n)] = \mathbb{E}[\log S(\mathcal{T}_n)] + O(\log n) = n\mu + o(\sqrt{n}),$$
$$\operatorname{Var}[\log R(\mathcal{T}_n)] = \operatorname{Var}[\log S(\mathcal{T}_n)] + o(n) = n\sigma^2 + o(n).$$

The next theorem gives the moments of the random variable $R(\mathcal{T}_n)$, under some assumptions. Note that, under these assumptions, all moments exist.

Theorem 4.2.2 (Cai [3] et al., Theorem 2.2) *Let $\mathcal{T}_n$ be as in Theorem 4.2.1, and assume further that $\mathbb{E}e^{t\xi} < \infty$ for some $t > 0$. Assume further that if $R \leq \infty$ is the radius of convergence of the probability generating function $\phi(z) := \mathbb{E}z^{\xi}$, then $\phi'(R) := \lim_{z \to R} \phi'(z) = \infty$. Then there exist sequences of numbers $\gamma_m > 0$ and $1 < \tau_1 < \tau_2 < \dots$ such that for any fixed $m \geq 1$,*

$$\mathbb{E}R(\mathcal{T}_n)^m = (1 + O(n^{-1}))\gamma_m \tau_m^n. \tag{4.1}$$

The next theorem gives the moments of the random variable $S(\mathcal{T}_n)$, under some assumptions. It also gives the covariance between $R(\mathcal{T}_n)$ and $S(\mathcal{T}_n)$. Note that, under the assumptions stated in the theorem, all the moments and covariances exist.

Theorem 4.2.3 (Cai [3] et al., Theorem 2.4) *Let $\mathcal{T}_n$ be as in Theorem 4.2.2. Then, for any $m \geq 1$,*

$$\mathbb{E}S(\mathcal{T}_n)^m = (1 + O(n^{-1}))\gamma'_m \tau_m^n,$$

where τ_m is as in 4.1, and $\gamma'_m > 0$.
More generally, for $m, l \geq 0$,

$$\mathbb{E}[R(\mathcal{T}_n)^l S(\mathcal{T}_n)^m] = (1 + O(n^{-1}))\gamma'_{m,l} \tau_{l+m}^n,$$

for some $\gamma'_{m,l} > 0$.

4.3 Large fringe and non-fringe subtrees in conditioned Galton-Watson trees

As pointed out in [2], the additive functional $n_T(\cdot)$, which counts the number of fringe subtree copies of T in an input tree, may be generalized as follows. Let $\mathcal{T}$ be a critical Galton-Watson tree. Instead of considering a fixed tree T, we consider a sequence $(T_n)_{n\geq 0}$ of trees, and the corresponding random sequence $(n_{T_n}(\mathcal{T}_n))_{n\geq 0}$.

We restrict ourselves to tree sequences where its corresponding size sequence $(|T_n|)_{n\geq 0} = o(n)$, i.e., grows slower than a linear function of n. Roughly, theorem 4.3.1 says that, as n grows large, the sequence $(n_{T_n}(\mathcal{T}_n))_{n\geq 0}$ behaves as a sequence of Poisson distributed random variables, and also that it has different types of asymptotic distributions, depending on how fast the sequence $(\mathbb{P}(\mathcal{T} = T_n))_{n\geq 0}$ decreases in relation to a linear function of n.

Theorem 4.3.1 (Cai [2] et al., Theorem 2) *Let ξ be the offspring distribution of the Galton-Watson tree $\mathcal{T}$, with $\mathbb{E}\xi = 1$ and $0 < \operatorname{Var} \xi < \infty$. Let $Po(\lambda)$ denote a Poisson-distributed random variable with parameter λ and let $\pi(T) := \mathbb{P}(\mathcal{T} = T)$. Also, let $k_n \to \infty$, $k_n = o(n)$. Then*

$$\lim_{n\to\infty} \sup_{T:|T|=k_n} d_{TV}(n_T(\mathcal{T}_n), Po(n\pi(T))) = 0,$$

where $d_{TV}(\cdot,\cdot)$ denotes the total variation distance. Therefore, letting T_n be a sequence of trees with $|T_n| = k_n$, we have as $n \to \infty$:

(i) If $n\pi(T_n) \to 0$, then $n_{T_n}(\mathcal{T}_n) = 0$ whp.

(ii) If $n\pi(T_n) \to \mu \in (0, \infty)$, then $n_{T_n}(\mathcal{T}_n) \xrightarrow{d} Po(\mu)$.

(iii) If $n\pi(T_n) \to \infty$*, then*

$$\frac{n_{T_n}(\mathcal{T}_n) - n\pi(T_n)}{\sqrt{n\pi(T_n)}} \xrightarrow{d} N(0,1),$$

where $N(0,1)$ *denotes the standard normal distribution, and* $\xrightarrow{d}$ *denoted convergence in distribution.*

We also consider counting the number of *non-fringe* subtree copies in a random tree sequence, of a fixed tree. Let $n_T^{\mathrm{nf}}(\cdot)$ be the additive functional which counts the number of non-fringe subtree copies of T in an input tree. We define a partial order $\prec$ on the the set of plane trees $\mathfrak{T}$ as

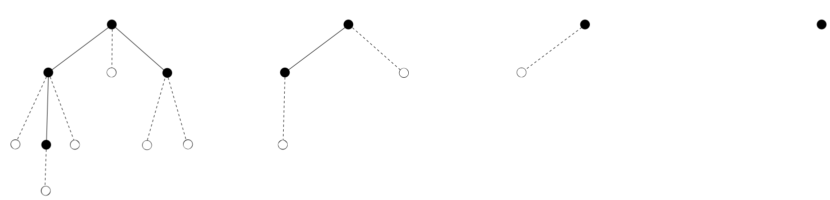
$$\text{For } T, T' \in \mathfrak{T}, T \prec T' \iff T \text{ is a general subtree of } T'.$$

Theorem 4.3.2 (Cai [2] et al., Theorem 4) *Let* ξ *be the offspring distribution of the Galton-Watson tree* $\mathcal{T}$*, with* $\mathbb{E}\xi = 1$ *and* $0 < \operatorname{Var} \xi < \infty$*. Let* $Po(\lambda)$ *denote a Poisson-distributed random variable with parameter* λ*, and let* $\pi^{\mathrm{nf}}(T) := \mathbb{P}(T \prec \mathcal{T})$*. Let* $n_T^{\mathrm{nf}}(T')$ *be the number of non-fringe subtree copies of* T *in* T'*. Let* T_n *be a sequence of trees with* $|T_n| = k_n$*, where* $k_n \to \infty$ *and* $k_n = o(n)$*. We have, as* $n \to \infty$*,*

(i) If $\pi^{\mathrm{nf}}(T_n) \to 0$*, then* $n_{T_n}^{\mathrm{nf}}(\mathcal{T}_n) \xrightarrow{p} 0$*.*

(ii) If $\pi^{\mathrm{nf}}(T_n) \to \infty$*, then* $n_{T_n}^{\mathrm{nf}}(\mathcal{T}_n)/(n\pi^{\mathrm{nf}}(T_n)) \xrightarrow{p} 1$*.*

5 Iterative leaf cutting



(a) The original tree. The white nodes are leaves to be cut. (b) After one cut. (c) After two cuts. (d) After three cuts only the root remains.

Figure 5: Example of a cutting procedure.

Consider the following procedure. Given a deterministic tree T, we remove all the leaves from it. From the resulting tree, we once more cut away the leaves. We repeat this cutting an arbitrary number of times, presuming there always are leaves left to cut away. One might now ask how many nodes, in total, that have been removed after a certain number of iterations. See Figure 5.

Instead of considering a cutting procedure for a deterministic tree, we consider such a procedure for random trees, where we restrict ourselves to the class of simply generated trees. We first define a cutting procedure in the context of toll functions and additive functionals, which are introduced in section 4.

Suppose we want to carry out r iterations of the cutting process on a fixed tree T. We define a toll function $f_r(\cdot)$, where r is a parameter, as follows:

$$f_r(T) = \begin{cases} 1, & H(T) < r \\ 0, & H(T) \geq r \end{cases}$$

where H denotes the height of the tree t. Now consider the induced additive functional $F_r(T) = f_r(T) + \Sigma_{i=1}^{d^+(o)} F_r(T_i)$, where T_i is the ith child of o, the root of T.

Note that the number of fringe subtrees of T is exactly equal to the number of nodes in T. Thus, as the proof in [4] shows, a fringe subtree T_v of T contributes 1 to the sum $F_r(T)$ if and only if it is completely removed during the cutting procedure, and 0 otherwise, or equivalently, that its root v is removed at some cutting step. Consequently, $F_r(T)$ counts the number of nodes removed in the first r iterations of the cutting procedure.

5.1 Main result for iterative leaf cutting

The following theorem shows that $f_r(T)$ is asymptotically normally distributed, and also specifies the expectation and the variance.

Theorem 5.1.1 (Hackl [4] et al., Theorem 3.1) *Let $r \in \mathbb{N}_1$ be fixed and let $\mathbb{T}$ be a family of simply generated trees with weight generating function ϕ and fundamental constant τ. Let ρ be the radius of convergence of ϕ. Furthermore, let $\mathbb{T}_r \subset \mathbb{T}$ denote the set of trees with height less than r and let $g_r(x) := \Sigma_{T \in \mathbb{T}_r} w(T) x^{|T|}$. Also, let $F_r(T)$ denote the additive func-*

tional which outputs the number of nodes removed after applying the "cutting leaves" procedure r times to $T \in \mathbb{T}$.

1. *If $\mathcal{T}_n$ denotes a random tree from $\mathbb{T}$ of size n, then, as $n \to \infty$:*

$$\mathbb{E}F_r(\mathcal{T}_n) = \mu_r n + \frac{\rho\tau^2 g_r'(\rho) + 3\beta_\tau g_r(\rho) - \alpha^2 g_r(\rho)}{3\tau^3} + O(n^{-1})$$

and

$$\text{Var } F_r(\mathcal{T}_n) = \sigma_r^2 n + O(1).$$

The constants μ_r and σ_r^2 are given by

$$\mu_r = \frac{g_r(\rho)}{\tau}$$

and

$$\sigma_r^2 = \frac{1}{2\tau^4}(4\rho\tau^3 g_r'(\rho) - 4\rho\tau^2 g_r(\rho) g_r'(\rho) + (2\tau^2 - \alpha^2) g_r(\rho)^2 - r\tau^3 g_r(\rho)),$$

respectively. The constants α and β are given by

$$\alpha = \sqrt{\frac{2\tau}{\rho\phi''(\tau)}}, \beta = \frac{1}{\rho\phi''(\tau)} - \frac{\tau\phi'''(\tau)}{3\rho\phi''(\tau)^2}.$$

2. *For $r \to \infty$ the constants μ_r and σ_r^2 behave as follows:*

$$\mu_r = 1 - \frac{2}{\rho\tau\phi''} r^{-1} + o(r^{-1})$$

and

$$\sigma_r^2 = \frac{1}{3\rho\tau\phi''(\tau)} + o(1).$$

3. *Finally, if $r \geq 2$ or $\mathbb{T}$ is not a family of d-ary trees, then $F_r(\mathcal{T}_n)$ is asymptotically normally distributed, meaning that for $x \in \mathbb{R}$ we have*

$$\mathbb{P}\left(\frac{F_r(\mathcal{T}_n) - \mu_r n}{\sqrt{\sigma_r^2 n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt + O(n^{-1/2}).$$

If $r = 1$ and $\mathbb{T}$ is a family of d-ary trees, then $F_r(\mathcal{T}_n) = \frac{n(d-1)+1}{d}$.

References

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

[2] Xing Shi Cai and Luc Devroye. A study of large fringe and non-fringe subtrees in conditional galton-watson trees. *ALEA*, 14:579–611, 2017.

[3] Xing Shi Cai and Svante Janson. Non-fringe subtrees in conditioned galton—watson trees. *The Electronic Journal of Combinatorics*, 25(3):3–40, 2018.

[4] Benjamin Hackl, Clemens Heuberger, and Stephan Wagner. Reducing simply generated trees by iterative leaf cutting. In *2019 Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 36–44. SIAM, Philadelphia, PA, 2019.

[5] Svante Janson. Simply generated trees, conditioned galton–watson trees, random allocations and condensation. *Probab. Surveys*, 9:103–252, 2012.

[6] Svante Janson. Asymptotic normality of fringe subtrees and additive functionals in conditioned galton–watson trees. *Random Structures & Algorithms*, 48(1):57–101, 2016.

[7] Thordur Jonsson and Sigurdur Örn Stefánsson. Condensation in nongeneric trees. *J. Stat. Phys.*, 142(2):277–313, 2011.

[8] Harry Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(4):425–487, 1986.

[9] Benedikt Stufler. Local limits of large Galton-Watson trees rerooted at a random vertex. *arXiv e-prints*, page arXiv:1611.01048, Nov 2016.

[10] Henry William Watson and Francis Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.