

# Local Limits of Large Galton–Watson Trees Rerooted at a Random Vertex

Benedikt Stuffer<sup>1</sup>

University of Zurich, Institute of Mathematics, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland  
benedikt.stuffer@math.uzh.ch

---

## Abstract

We prove limit theorems describing the asymptotic behaviour of a typical vertex in random simply generated trees as their sizes tends to infinity. In the standard case of a critical Galton–Watson tree conditioned to be large, the limit is the invariant random sin-tree constructed by Aldous (1991). Our main contribution lies in the condensation regime where vertices of macroscopic degree appear. Here we describe in complete generality the asymptotic local behaviour from a random vertex up to its first ancestor with “large” degree. Beyond this distinguished ancestor, different behaviours may occur, depending on the branching weights. In a subregime of complete condensation, we obtain convergence toward a novel limit tree, that describes the asymptotic shape of the vicinity of the full path from a random vertex to the root vertex. This includes the important case where the offspring distribution follows a power law up to a factor that varies slowly at infinity.

**2012 ACM Subject Classification** Mathematics of computing → Stochastic processes

**Keywords and phrases** Galton–Watson trees, local weak limits

**Digital Object Identifier** 10.4230/LIPIcs.AofA.2018.34

## 1 Introduction

A Galton–Watson branching process is a classical stochastic model for population evolution. The process starts with a single individual. All individuals reproduce asexually and independently from each other according to the same offspring distribution. The genealogical tree corresponding to such a process is called a Galton–Watson tree. We call such a tree critical if the average number of children of a node equals 1, and subcritical if it is strictly less than one. In this context we call the number of offspring of node its outdegree. The tree obtained by conditioning the total population size to be equal to an integer  $n$  is a special case of a simply generated tree  $\mathcal{T}_n$ . That is, there is a fixed sequence  $(\omega_k)_{k \geq 0}$  of so-called branching weights and the tree  $\mathcal{T}_n$  assumes a plane tree  $T$  with  $n$  vertices with probability proportional to the product  $\prod_{v \in T} \omega_{d_T^+(v)}$  of weights corresponding to the vertex outdegrees  $d_T^+(v)$ . The present work aims to describe the vicinity of a typical vertex in  $\mathcal{T}_n$  as  $n$  tends to infinity. We refer the reader to Section 3 for a brief introduction of this model of random trees and to Drmota’s book [7, Sec. 1.2.7] for a more detailed discussion.

The asymptotic shape of the vicinity of the fixed root vertex in random trees has received considerable attention in recent literature. Jonsson and Stefánsson [11] described a phase transition between an infinite spine case and a condensation setting for large Galton–Watson trees with a power-law offspring distribution. A third regime for random

---

<sup>1</sup> The author gratefully acknowledges support by the Swiss National Science Foundation grant number 200020\_172515.



© Benedikt Stuffer;

licensed under Creative Commons License CC-BY

29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2018).

Editors: James Allen Fill and Mark Daniel Ward; Article No. 34; pp. 34:1–34:11



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

simply generated trees with superexponential branching weights was studied by Janson, Jonsson and Stefánsson [10]. The asymptotic shape of simply generated trees as their size tends to infinity was later described in complete generality by Janson [9]. Abraham and Delmas [3, 4] classified the limits of conditioned Galton–Watson trees as the total number of vertices with outdegree in a given fixed set tends to infinity. Limits of Galton–Watson trees having a large number of protected nodes were established by Abraham, Bouaziz, and Delmas [1]. The asymptotic shape of conditioned multi-type Galton–Watson trees was studied by Stephenson [14], Abraham, Delmas, and Guo [2], and Pénisson [13].

Clearly considerable effort and progress is being made in understanding local limits of random trees that describe the asymptotic behaviour near the fixed root vertex, and for random simply generated trees even a complete classification is available. As for the question of the asymptotic shape of the vicinity of a random vertex, Aldous [5] studied in his pioneering work asymptotic fringe distributions for general families of random trees. For the case of critical Galton–Watson trees, he established, at least when the offspring distribution has finite variance, convergence of the tree obtained by rerooting at a random vertex. Janson [9, Thm. 7.12] described the asymptotic behaviour of the fringe subtree rooted at a random vertex of a simply generated tree. Here a fringe subtree at a vertex  $v$  refers to the subtree formed by  $v$  and all its descendants. A fringe subtree of some ancestor of  $v$  is also called an extended fringe subtree. A recent work by Holmgren and Janson [8] studied fringe subtrees and extended fringe subtrees of models of random trees that may be described by the family tree of a Crump–Mode–Jagers branching process stopped at a suitable time, including random recursive trees, preferential attachment trees, fragmentation trees and  $m$ -ary search trees.

Janson [9] distinguishes three types of simply generated trees, numbered I, II and III, and for each the local limit exhibits a distinguishing characteristic. See Subsection 3.1 for a brief review of this notation. We use this terminology in our study of the vicinity of a random vertex. In the type I setting, the simply generated tree  $\mathcal{T}_n$  is distributed like a critical Galton–Watson tree conditioned on having  $n$  vertices. Thus the height of a random vertex in  $\mathcal{T}_n$  is typically large and extended fringe trees are typically small. In this regime, the limit is given by the random *sin*-tree constructed by Aldous [5]. Here the word *sin* refers to the fact that, like the Kesten tree, this tree has almost surely up to finite initial segments only a single infinite path. When the offspring distribution has finite variance, we may even verify total variational convergence of the extended fringe subtree up to  $o(\sqrt{n})$ -distant ancestors.

While trees in the type I regime usually have small maximum degree, the types II and III are characterized by the appearance of vertices with large degree, which may be viewed as a form of condensation. Specifically, type II simply generated trees correspond to subcritical Galton–Watson trees with a heavy-tailed offspring distribution, and type III simply generated trees have superexponential branching weights such that no equivalent conditioned Galton–Watson tree exists. Our main contribution is in this condensation setting, where contrary to the type I regime a random vertex may be near to the root, and extended fringe trees may have size comparable to the total number of vertices of  $\mathcal{T}_n$ , as we are likely to encounter an ancestor with large degree. This is also a major difference to the settings addressed in the mentioned works by Aldous [5] and Holmgren and Janson [8].

We set up a compact space that encodes rooted plane trees that are centered around a second distinguished vertex, and establish several limit theorems. For arbitrary weight-sequences having type II or III, we establish a limit that describes the vicinity of a random vertex up to and including its first ancestor with large degree. Here *large* means having outdegree bigger than a deterministic sequence that tends to infinity sufficiently slowly. The

asymptotic shape of what lies beyond this ancestor appears to depend on the branching weights. In a way, the vertex with large degree obstructs the view to older generations.

We describe a novel limit object  $\mathcal{T}^*$  given by a random pointed plane tree, in which the pointed vertex has random distance from its first ancestor with infinite degree, and this ancestor again has a random number of ancestors with finite degree before the construction stops. For arbitrary weight-sequences, the asymptotic probability for the vicinity of a random vertex of  $\mathcal{T}_n$  to have a specific shape that admits at most one single ancestor of large degree, but allows ancestors with small degrees afterwards, coincides with the corresponding probability for the tree  $\mathcal{T}^*$ . Our approach is based on a heavily modified depth-first-search to explore the tree  $\mathcal{T}_n$ . This yields information on how parts of a limit tree for the complete vicinity, that is not truncated at the first large ancestor, must look, if the simply generated tree  $\mathcal{T}_n$  pointed at a random vertex converges weakly (along a subsequence). Note also that the compactness of the space, in which we formulate our limits, guarantees the existence of such subsequences. Thus the obstruction by the ancestor with large degree, that prevents us from seeing older generations, is not a complete blockage. However, this is not yet sufficient to deduce convergence in the space of pointed plane trees. In general, the tip of the backwards growing spine, where the construction of  $\mathcal{T}^*$  breaks off, may correspond to the root vertex of  $\mathcal{T}_n$ , but just as well to a second ancestor with large degree.

If the branching weights belong to a general regime of complete condensation, we manage to surpass the blockage and deduce weak convergence toward  $\mathcal{T}^*$ . There are two main steps involved. First, we show that convergence toward  $\mathcal{T}^*$  is in fact equivalent to weak convergence of the height of a random vertex in  $\mathcal{T}_n$  to the height of the pointed vertex in the tree  $\mathcal{T}^*$ , which in the type II regime is distributed like 1 plus the sum of two independent identically distributed geometric random variables, and in the type III regime equals 1. In this case, the root of  $\mathcal{T}^*$  really corresponds to the root of  $\mathcal{T}_n$ . The second step verifies this property in the case of complete condensation, where the maximum degree of  $\mathcal{T}_n$  has the correct order.

In particular, Kortchemski's central limit theorem for the maximum degree [12, Theorem 1] allows us to deduce convergence toward  $\mathcal{T}^*$  in the general case of a subcritical Galton–Watson tree conditioned on having  $n$  vertices, if the offspring distribution  $\xi$  satisfies  $\mathbb{P}(\xi = k) = f(k)k^{-\alpha}$  for a constant  $\alpha > 2$  and a function  $f$  that varies slowly at infinity. In the type III regime where branching weights grow superexponentially fast, we consider the specific case where  $\omega_k = k!^\alpha$  for  $\alpha > 0$ . It is known that for these weights the maximum degree of  $\mathcal{T}_n$  has order  $n + o_p(n)$ , which may also be seen as complete condensation, see Janson, Jonsson, and Stefánsson [10] and Janson [9, Example 19.36]. Thus here the tree  $\mathcal{T}^*$  is also the weak limit of the simply generated tree  $\mathcal{T}_n$  pointed at a random vertex. There are, however, also examples of superexponential branching weights that exhibit a more irregular behaviour [9, Example 19.38], in which weak convergence towards  $\mathcal{T}^*$  does not hold.

The present work is an extended abstract of [15], where detailed justifications of all results are provided.

## 2 The space of pointed plane trees

### 2.1 Centering at a specified vertex

A *plane tree* is a rooted tree in which the offspring set of each vertex is endowed with a linear order. (Such trees are also sometimes referred to as planted plane trees or corner rooted plane trees, in order to distinguish them from related planar structures [7].) In the present work we will also encounter plane trees that have no root, but whose vertex sets are

endowed with a partial order that specifies the ancestry relations, and whose offspring sets are endowed with a linear order that is not required to have a smallest element.

Given a plane tree  $T$  and a vertex  $v \in T$  we let  $d_T^+(v)$  denote its *outdegree*, that is, the number of offspring. Its *height*  $h_T(v)$  is its distance from the root-vertex. Traditionally, plane trees are encoded as subtrees of the *Ulam–Harris tree*. The Ulam–Harris tree  $\mathcal{U}_\infty$  is an infinite plane tree with vertex set  $\mathcal{V}_\infty = \mathbb{N}^{(\mathbb{N})}$  given by the space of finite sequences of non-negative integers. Its root vertex is the unique sequence with length zero, and the ordered offspring of a vertex  $v$  are the concatenations  $(v, i)$  for  $i \geq 1$ . Thus a plane tree is a subtree of the Ulam–Harris tree that contains its root, such that the offspring set of each vertex is an initial segment of the offspring of the corresponding vertex in  $\mathcal{U}_\infty$ . Here we explicitly allow trees with infinitely many vertices, and vertices with countably infinite outdegree. If all outdegrees of a plane tree are finite, we say that it is *locally finite*. The tree is *finite*, if its total number of vertices is. We will usually let  $o$  denote the root-vertex of a plane tree.

Subtrees of the Ulam–Harris tree are however not an adequate form to represent the vicinity of a specified vertex in a plane tree. If this vertex does not coincide with the root of the tree, then it has an ordered sequence of ancestors and possibly also siblings that lie to the left and right of it. If we look at a random vertex of the simply generated  $\mathcal{T}_n$ , then it may happen that the number of siblings to the left and/or right of it is asymptotically large, or that its distance from the root vertex is large. A sensible space in which we may describe the limit of the vicinity of the random vertex in  $\mathcal{T}_n$  must hence contain trees with a center that may have infinitely many ancestors, such that each may have infinitely many siblings to the left and/or right of it, including the center vertex itself.

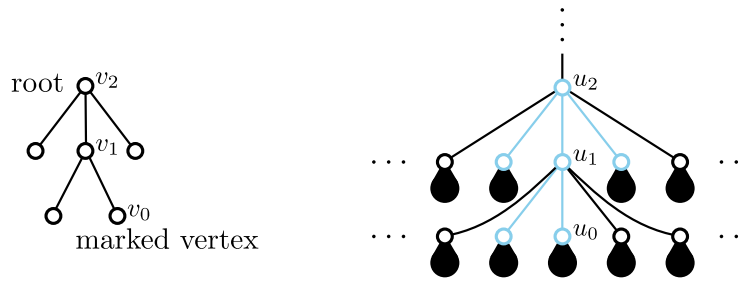
For this reason, we describe the construction of an infinite tree  $\mathcal{U}_\infty^\bullet$  that is embedded in the plane and has a spine  $(u_i)_{i \geq 0}$  that grows "backwards". That is, we construct the tree  $\mathcal{U}_\infty^\bullet$  by starting with an infinite path  $u_0, u_1, \dots, u_i$  of abstract vertices and define  $u_i$  to be a parent of  $u_{i-1}$  for all  $i \geq 1$ . Additionally, any vertex  $u_i$  with  $i \geq 1$  receives an infinite number of vertices to the left and to the right of its distinguished offspring  $u_{i-1}$ , and each of these "non-centered" offspring vertices is the root of a copy of the Ulam–Harris tree  $\mathcal{U}_\infty$ . To conclude the construction, the start-vertex  $u_0$  of the spine also gets identified with the root of a copy of  $\mathcal{U}_\infty$ . We let  $\mathcal{V}_\infty^\bullet$  denote the vertex-set of the tree  $\mathcal{U}_\infty^\bullet$ . The tree  $\mathcal{U}_\infty^\bullet$  is illustrated in Figure 1.

Note that the vertex set  $\mathcal{V}_\infty^\bullet$  carries a natural partial order (given by the transitive hull of the parent-child relations specified in the construction of  $\mathcal{U}_\infty^\bullet$ ), and the offspring set of any given vertex carries a natural linear order. This allows us to continue using the terms *ancestor* and *offspring* in this context.

A plane tree  $T$  together with a distinguished vertex  $v_0$  is called a *pointed* plane tree, and may be interpreted in a canonical way as a subtree of  $\mathcal{U}_\infty^\bullet$ . To do so, let  $v_0, v_1, \dots, v_k$  denote the path from  $v_0$  to the root of  $T$ . This way, any vertex  $v_i$  for  $i \geq 1$  may have offspring to the left and to the right of  $v_{i-1}$ . Thus there is a unique order-preserving and outdegree preserving embedding of  $T$  into  $\mathcal{U}_\infty^\bullet$  such that  $v_i$  corresponds to  $u_i$  for all  $0 \leq i \leq k$ . Compare with Figure 1.

## 2.2 Topological properties

Any plane tree  $T$  may be identified with its family of outdegrees  $(d_T^+(v))_{v \in \mathcal{V}_\infty} \in \overline{\mathbb{N}}_0^{\mathcal{V}_\infty}$ , where we set  $\overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}$ . Here we use the convention  $d_T^+(v) = 0$  if  $v \in \mathcal{V}_\infty$  is not a vertex of the tree  $T$ . We endow  $\overline{\mathbb{N}}_0$  with the one-point compactification topology of the discrete space



■ **Figure 1** Embedding of a pointed plane tree into the tree  $\mathcal{U}_\infty^\bullet$ . Each black blob represents a copy of the Ulam–Harris tree.

$\mathbb{N}_0$ . Thus plane trees are elements of the compact product space  $\overline{\mathbb{N}}_0^{\mathcal{V}_\infty}$ . It is not hard to see that the subspace  $\mathfrak{T} \subset \overline{\mathbb{N}}_0^{\mathcal{V}_\infty}$  of all plane trees is closed.

Similarly, we may identify a pointed plane tree  $T^\bullet = (T, v_0)$  with the corresponding family of outdegrees  $(d_{T^\bullet}^+(v))_{v \in \mathcal{V}_\infty}$ , such that  $d_{T^\bullet}^+(v) \in \overline{\mathbb{N}}_0$  for  $v \notin \{u_1, u_2, \dots\}$ , and  $d_{T^\bullet}^+(u_i) \in \{*\} \sqcup (\overline{\mathbb{N}}_0 \times \overline{\mathbb{N}}_0)$  for  $i \geq 1$ . Here the two numbers represent the number of offspring vertices to the left and right of the distinguished son  $u_{i-1}$ , and the  $*$ -placeholder represents the fact that the vertex does not belong to the tree.

Since  $\overline{\mathbb{N}}_0$  is a compact Polish space, so are the product  $\overline{\mathbb{N}}_0 \times \overline{\mathbb{N}}_0$  and the disjoint union topology on  $\{*\} \sqcup (\overline{\mathbb{N}}_0 \times \overline{\mathbb{N}}_0)$ . Hence the space of all families  $(d^+(v))_{v \in \mathcal{V}_\infty}$  satisfying

$$d^+(v) \in \begin{cases} \overline{\mathbb{N}}_0 & \text{for } v \notin \{u_1, u_2, \dots\} \\ \{*\} \sqcup (\overline{\mathbb{N}}_0 \times \overline{\mathbb{N}}_0) & \text{for } v \in \{u_1, u_2, \dots\} \end{cases}$$

is the product of countably many compact Polish spaces, and hence also compact and Polish. It is easy to verify that the subset  $\mathfrak{T}^\bullet$  of all elements that correspond to trees (that is, connected acyclic graphs) is closed, and hence also a compact Polish space with respect to the subspace topology.

### 3 Simply generated trees

We let  $\mathbf{w} = (\omega_i)_{i \geq 0}$  denote a sequence of non-negative weights satisfying  $\omega_0 > 0$  and  $\omega_k > 0$  for at least one  $k \geq 2$ . The weight of a plane tree  $T$  is defined by  $\omega(T) = \prod_{v \in T} \omega_{d_T^+(v)}$ . The simply generated tree  $\mathcal{T}_n$  with  $n$  vertices gets drawn from the set of all  $n$ -vertex plane trees with probability proportional to its weight. Galton–Watson trees conditioned on having a fixed number of vertices are encompassed by this model of random plane trees. Of course, the tree  $\mathcal{T}_n$  is only well-defined if there is at least one plane tree with  $n$  vertices that has positive weight. We set  $\text{span}(\mathbf{w}) = \text{gcd}\{i \geq 0 \mid \omega_i > 0\}$ . As argued in [9, Corollary 15.6],  $n$ -sized trees with positive weight may only exist for  $n \equiv 1 \pmod{\text{span}(\mathbf{w})}$ , and conversely, they always exist if  $n$  is large enough and belongs to this congruence class. We tacitly only consider such  $n$  throughout this paper.

#### 3.1 Three types of weight-sequences

Janson [9, Chapter 8] distinguishes three types of weight-sequences. The classification is as follows. Let  $\rho_\phi$  denote the radius of convergence of the generating series  $\phi(z) = \sum_{k \geq 0} \omega_k z^k$ . As argued in [9, Lemma 3.1], if  $\rho_\phi > 0$  then the function  $\psi(t) = \phi'(t)t/\phi(t)$  admits a limit  $\nu = \lim_{t \nearrow \rho_\phi} \psi(t) \in ]0, \infty]$  with the following properties. If  $\nu \geq 1$ , then there is a unique

number  $\tau$  with  $\psi(\tau) = 1$  and we say the weight sequence  $\mathbf{w}$  has type I. If  $0 < \nu < 1$ , then we set  $\tau := \rho_\phi < \infty$  and say  $\mathbf{w}$  has type II. If  $\rho_\phi = 0$ , we say  $\mathbf{w}$  has type III and set  $\nu = 0$  and  $\tau = 0$ . The constant  $\nu$  has a natural interpretation as the supremum of the means of all probability weight sequences equivalent to  $\mathbf{w}$ . The inclined reader may see [9, Remark 4.3] for details.

### 3.2 An associated Galton–Watson tree

We define the probability distribution  $(\pi_k)_k$  on  $\mathbb{N}_0$  by  $\pi_k = \tau^k \omega_k / \phi(\tau)$ . The mean and variance of the distribution  $(\pi_k)_k$  are given by  $\mu = \min(\nu, 1)$  and  $\sigma^2 = \tau \psi'(\tau) \leq \infty$ . We let  $\xi$  denote a random non-negative integer with density  $(\pi_k)_k$ , and  $\mathcal{T}$  a Galton–Watson tree with offspring distribution  $\xi$ . Note that if  $\mathbf{w}$  has type III, then  $\xi = 0$  almost surely and the tree  $\mathcal{T}$  consists of a single deterministic vertex. As detailed in [9, Section 4], if  $\mathbf{w}$  has type I or II then the simply generated tree  $\mathcal{T}_n$  is distributed like the Galton–Watson tree  $\mathcal{T}$  conditioned on having  $n$  vertices.

## 4 The limit theorems

As discussed in Section 3 there is a probability distribution  $(\pi_k)_k$  associated with the weight sequence  $\mathbf{w}$ . Let  $\xi$  be distributed according to  $(\pi_k)_k$  and let  $\mathcal{T}$  be a  $\xi$ -Galton–Watson tree. Thus  $\mu := \mathbb{E}[\xi] \leq 1$ . We may consider the *size-biased* random variable  $\hat{\xi}$  with values in  $\overline{\mathbb{N}}_0$  and distribution given by

$$\mathbb{P}(\hat{\xi} = k) = k\pi_k \quad \text{and} \quad \mathbb{P}(\hat{\xi} = \infty) = 1 - \mu.$$

For any tree  $T$  and any vertex  $v \in T$  we let  $f(T, v)$  denote the *fringe-subtree* of  $T$  at  $v$ . That is, the maximal subtree of  $T$  that is rooted at the vertex  $v$ .

Throughout the following, we let  $v_0$  denote a uniformly at random selected vertex of the simply generated plane tree  $\mathcal{T}_n$ , that in the type I and II regime is distributed like the Galton–Watson tree  $\mathcal{T}$  conditioned on having  $n$  vertices.

### 4.1 The type I regime

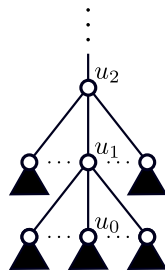
If the weight-sequence  $\mathbf{w}$  has type I, then  $\hat{\xi} < \infty$  almost surely, and we define the random pointed tree  $\mathcal{T}^*$  as follows. Let  $u_0$  be the root of an independent copy of the Galton–Watson tree  $\mathcal{T}$ . For each  $i \geq 1$ , we let  $u_i$  receive offspring according to an independent copy of  $\hat{\xi}$ . The vertex  $u_{i-1}$  gets identified with a uniformly at random chosen offspring of  $u_i$ . All other offspring vertices of  $u_i$  become the root of an independent copy of the Galton–Watson tree  $\mathcal{T}$ . Compare with Figure 2.

► **Theorem 1.** *If the weight-sequence  $\mathbf{w}$  has type I, then*

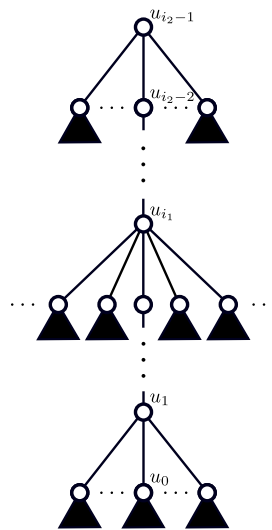
$$(\mathcal{T}_n, v_0) \xrightarrow{d} \mathcal{T}^*$$

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

Let  $T$  be a plane tree,  $v \in T$  a vertex, and  $k \geq 0$  an integer. If the vertex  $v$  has a  $k$ th ancestor  $v_k$ , then we may define the pointed plane tree  $H_k(T, v)$  as the fringe tree  $f(T, v_k)$  that is rooted at the vertex  $v_k$  and pointed at the vertex  $v$ . Here we use the term vertex in the graph-theoretic sense, since the coordinates of the vertex  $v$  as node of the Ulam–Harris tree depend on whether we talk about  $v \in T$  or  $v \in f(T, v_k)$ . If the vertex  $v$  has height  $h_T(v) < k$ , we set  $H_k(T, v) = \diamond$  for some placeholder symbol  $\diamond$ .



■ **Figure 2** The limit tree  $\mathcal{T}^*$  in the type I regime. Each triangle represents an independent copy of the Galton–Watson tree  $\mathcal{T}$ . For each  $i \geq 1$  the vertex  $u_i$  receives offspring according to an independent copy of  $\hat{\xi}$ , and the location of  $u_{i-1}$  within that offspring set is chosen uniformly at random.



■ **Figure 3** The limit tree  $\mathcal{T}^*$  in the complete condensation regime. The vertex  $u_{i_1}$  is the only one having infinite degree, and each triangle represents an independent copy of the Galton–Watson tree  $\mathcal{T}$ .

► **Theorem 2.** *Suppose that weight-sequence has type I and the offspring distribution  $\xi$  has finite variance. Let  $k_n$  be an arbitrary sequence of non-negative integers that satisfies  $k_n/\sqrt{n} \rightarrow 0$ . Then*

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

as  $n$  becomes large.

Here we use the redundant notation  $(\mathcal{T}^*, u_0)$  to emphasize that the tree  $\mathcal{T}^*$  is marked at the vertex  $u_0$ .

## 4.2 Complete condensation in the type II regime

If the weight-sequence  $\mathbf{w}$  has type II or III, then we construct  $\mathcal{T}^*$  similarly as in the type I case, letting  $u_0$  become the root of an independent copy of the Galton–Watson tree  $\mathcal{T}$ , and letting for  $i = 1, 2, \dots$  the vertex  $u_i$  receives offspring according to an independent copy  $\hat{\xi}_i$  of  $\hat{\xi}$ , where a uniformly at random chosen son gets identified with  $u_{i-1}$  (specifying the number

of siblings to the left and right of  $u_{i-1}$ ) and the rest become roots of independent copies of  $\mathcal{T}$ . We proceed in this way for  $i = 1, 2, \dots$  until it occurs for the first time  $i_1$  that  $\hat{\xi}_{i_1} = \infty$ . When  $\hat{\xi}_1, \dots, \hat{\xi}_{i_1-1} < \infty$  and  $\hat{\xi}_{i_1} = \infty$ , then  $u_{i_1}$  receives infinitely many offspring to the left and right of its son  $u_{i_1-1}$ . Each of these vertices (except  $u_{i_1-1}$  of course) gets identified with an independent copy of the Galton–Watson tree  $\mathcal{T}$ . We then proceed as before for  $i = i_1, i_1 + 1, \dots$ , such that  $u_i$  receives offspring according to an independent copy  $\hat{\xi}_i$  of  $\hat{\xi}$ , with a random son being identified with  $u_{i-1}$  and the rest becoming roots of independent copies of  $\mathcal{T}$ , until it happens for the second time  $i_2$  that  $\hat{\xi}_{i_2} = \infty$ . When  $\hat{\xi}_{i_1} = \infty = \hat{\xi}_{i_2}$  for  $i_1 < i_2$  and  $\hat{\xi}_i < \infty$  for all  $i < i_2$  with  $i \neq i_1$ , then we stop the construction. The spine of the resulting tree is then given by the ordered path  $u_0, \dots, u_{i_2-1}$ . Compare with Figure 3.

Note that this construction also works in the type I regime and yields the tree  $\mathcal{T}^*$  as defined in Section 4.1, since  $\mathbf{w}$  having type I implies that almost surely  $\hat{\xi}_i < \infty$  for all  $i \geq 1$ .

► **Theorem 3.** *Suppose that the weight-sequence  $\mathbf{w}$  has type II. 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  $k$   $\mathbb{P}(\xi = k) = f(k)k^{-\alpha}$ .

Here we make use of a result by Kortchemski [12, Theorem 1] who established a central limit theorem for the maximum degree, that ensures that  $\Delta(\mathcal{T}_n)$  has the correct order if the offspring distribution  $\xi$  has a power law up to a slowly varying factor. There are also examples of offspring distributions with a more irregular behaviour. Janson [9, Example 19.37] constructed a weight sequence such that along a subsequence  $n = n_k$  it holds that  $\Delta(\mathcal{T}_n) = o_p(n)$ , and along another subsequence several vertices with degree comparable to  $n$  exist. This may be seen as incomplete condensation.

The proof idea of Theorem 3 is to deduce the asymptotic distribution of the height  $h_{\mathcal{T}_n}(v_0)$  by localizing the vertex of  $\mathcal{T}_n$  having maximum degree at a position, that was also given in [12, Theorem 2] using results by Armendáriz and Loulakis [6] concerning conditioned random walks having a subexponential jump distribution. To do so, we employ results of Janson [9, Chapter 20] that (partially) use

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

but do not assume the offspring distribution to be subexponential. The following main lemma, which characterizes convergence toward the tree  $\mathcal{T}^*$  in terms of weak convergence of the height  $h_{\mathcal{T}_n}(v_0)$ , then finalizes the proof of Theorem 3.

► **Lemma 4.** *If the weight-sequence  $\mathbf{w}$  has type II or III, then the following three conditions are equivalent.*

1.  $(\mathcal{T}_n, v_0) \xrightarrow{d} \mathcal{T}^*$  in  $\mathfrak{T}^\bullet$ .
2.  $h_{\mathcal{T}_n}(v_0) \xrightarrow{d} h_{\mathcal{T}^*}(u_0)$ .
3.  $\limsup_{n \rightarrow \infty} \mathbb{P}(h_{\mathcal{T}_n}(v_0) \geq k) \leq \mu^k + k(1 - \mu)\mu^{k-1}$  for all  $k \geq 1$ .

Note that  $h_{\mathcal{T}^*}(u_0)$  is distributed like 1 plus the sum of two independent identically distributed geometric random variables that assume an integer  $i$  with probability  $\mu^i(1 - \mu)$ .

### 4.3 Complete condensation in the type III regime

If the weight-sequence  $\mathbf{w}$  has type III, then it holds that  $\mu = 0$  and almost surely  $\xi = 0$  and  $\hat{\xi} = \infty$ . Here the Galton–Watson tree  $\mathcal{T}$  is always equal to a single point. Hence the tree



$\mathcal{T}^*$  is obtained by letting  $u_1$  have infinitely many offspring to the left and right of  $u_0$ , all of which (including  $u_0$ ) are leaves.

► **Proposition 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.

Here we use that if  $\omega_k = k!^\alpha$  with  $\alpha > 0$  a constant, then it is known [9, page 226, Example 19.36], that the largest degree in  $\mathcal{T}_n$  has size  $n + o_p(n)$ . But there are also other examples that exhibit a more irregular behaviour. In [9, page 227, Example 19.38] a weight-sequence is constructed such that along a subsequence  $n = n_k$ , for each  $j \geq 1$  the  $j$ th largest degree  $Y_{(j)}$  in  $\mathcal{T}_{n_k}$  satisfies  $Y_{(j)} = 2^{-j}n$  with high probability. This may be seen as incomplete condensation. It is clear that in this case the limit of  $(\mathcal{T}_n, v_0)$ , if it exists at all, must have a different shape than  $\mathcal{T}^*$ .

#### 4.4 Large nodes and truncated limits

Suppose that the weight sequences  $\mathbf{w}$  has type II or type III. The limit theorems in Subsections 4.2 and 4.3 work in settings of complete condensation, where the maximum degree of the tree  $\mathcal{T}_n$  satisfies

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

If we content ourselves with the vicinity of the vertex  $v_0$  up to and including the first vertex having large degree, we may obtain a limit theorem in complete generality. We are also going to construct a coupling to demonstrate how the vertex with infinite degree in the limit corresponds to a vertex with large degree in the simply generated tree  $\mathcal{T}_n$ .

Janson [9, Lemma 19.32] showed that there is a deterministic sequence  $\Omega_n$  that tends to infinity sufficiently slowly, such that for any sequence  $K_n \rightarrow \infty$  with  $K_n \leq \Omega_n$  it holds that the numbers  $N_k$  of vertices with outdegree  $k$  in the tree  $\mathcal{T}_n$  satisfy

$$\sum_{k \leq K_n} kN_k = \mu n + o_p(n) \quad \text{and} \quad \sum_{k > K_n} kN_k = (1 - \mu)n + o_p(n) \tag{1}$$

The sequence  $\Omega_n$  may be replaced by any sequence that tends to infinity more slowly. Hence we may assume without loss of generality that  $\Omega_n$  additionally satisfies  $\Omega_n = o(n)$ . Let  $\tilde{D}_n$  denote a random positive integer, that is independent from all previously considered random variables, with distribution given by  $\tilde{D}_n \stackrel{d}{=} (d_{\mathcal{T}_n}^+(o) \mid d_{\mathcal{T}_n}^+(o) > \Omega_n)$ . Here we let  $o$  denote the root-vertex of  $\mathcal{T}_n$ . That is,  $\tilde{D}_n$  is distributed like the root-degree conditioned to be "large". We form the random tree  $\tilde{\mathcal{T}}_n^*$  in a similar manner as the random tree  $\mathcal{T}^*$ . The vertex  $u_0$  becomes the root of an independent copy of the Galton–Watson tree  $\mathcal{T}$ . For  $i = 1, 2, \dots$  the vertex  $u_i$  receives offspring according to independent copy  $\hat{\xi}_i$  of  $\hat{\xi}$ , where a randomly chosen son gets identified with  $u_{i-1}$  and the rest become roots of independent copies of  $\mathcal{T}$ . We proceed in this way for  $i = 1, 2, \dots$  until it occurs that  $\hat{\xi}_i = \infty$ . When  $\hat{\xi}_1, \dots, \hat{\xi}_{i-1} < \infty$  and  $\hat{\xi}_i = \infty$ , then  $u_i$  receives  $\tilde{D}_n$  offspring vertices, such that a uniformly at random chosen one gets identified with  $u_{i-1}$ , and the rest get identified with the roots of independent copies

of  $\mathcal{T}$ . Rather than continuing with the spine as in the construction of the tree  $\mathcal{T}^*$ , we stop at this point, so that  $u_i$  becomes the root of this tree.

Given a pointed tree  $T^\bullet = (T, v)$  and an ancestor  $a$  of  $v$  we may consider the fringe subtree of  $T$  at  $a$  as pointed at the vertex  $v$ . Let  $v_0$  denote a uniformly at random selected vertex of the simply generated tree  $\mathcal{T}_n$ . Let  $H(\mathcal{T}_n, v_0, \Omega_n)$  denote the pointed fringe subtree of  $(\mathcal{T}_n, v_0)$  at the youngest ancestor of  $v_0$  that has outdegree bigger than  $\Omega_n$ . If no such vertex exists (which is unlikely to happen, as we are going to verify), set  $H(\mathcal{T}_n, v_0, \Omega_n) = \diamond$  for some fixed placeholder value  $\diamond$ .

► **Theorem 6.** *Suppose that the weight sequence  $\mathbf{w}$  has type II or III. Let  $\bar{\mathcal{T}}^*$  denote the pointed fringe subtree of the tree  $\mathcal{T}^*$  at its unique vertex with infinite degree. Then it holds that*

$$H(\mathcal{T}_n, v_0, \Omega_n) \xrightarrow{d} \bar{\mathcal{T}}^*.$$

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

The strength of this theorem is its generality, as we make no additional assumptions on the weight-sequence at all. It is suitable for applications where it is not necessary to look behind the large vertex.

We may still improve upon this. For each  $n$ , let  $\bar{\mathcal{T}}_n^*$  be constructed from  $\bar{\mathcal{T}}^*$  by pruning at its root vertex such that its outdegree becomes  $\tilde{D}_n$ . Of course we have to select one of the  $\tilde{D}_n$  ways of how much we prune from the left and right so that the total outdegree becomes  $\tilde{D}_n$ , and we choose an option uniformly at random.

For each integer  $m \geq 0$  we let  $\bar{V}^{[m]} \subset \mathcal{V}_\infty^\bullet$  denote the vertex set of the tree obtained from  $\mathcal{U}_\infty^\bullet$  by deleting all vertices with distance larger than  $m$  from the center vertex  $u_0$  and pruning so that the vertices  $u_i$ ,  $1 \leq i \leq m$  have outdegree  $(m, m)$  and the remaining vertices all have outdegree equal to  $m$ . The topology on the subspace  $\mathfrak{T}_{\text{lf}}^\bullet \subset \mathfrak{T}^\bullet$  of locally finite trees is induced by the metric

$$d_{\mathfrak{T}_{\text{lf}}^\bullet}(T_1^\bullet, T_2^\bullet) = 1 / \sup\{m \geq 0 \mid d_{T_1^\bullet}^+(v) = d_{T_2^\bullet}^+(v) \text{ for all } v \in \bar{V}^{[m]}\}.$$

This can be verified using the fact that a sequence  $(T_n)_n$  in  $\mathfrak{T}^\bullet$  converges towards an element  $T \in \mathfrak{T}^\bullet$  if and only if  $d_{T_n}^+(v)$  converges towards  $d_T^+(v)$  for each  $v \in \mathcal{V}_\infty^\bullet$ .

► **Theorem 7.** *Suppose that the weight sequence  $\mathbf{w}$  has type II or III. For any finite set of vertices  $x_1, \dots, x_r \in \mathcal{V}_\infty^\bullet$  it holds that*

$$d_{\text{TV}}((d_{H(\mathcal{T}_n, v_0, \Omega_n)}^+(x_i))_{1 \leq i \leq r}, (d_{\bar{\mathcal{T}}_n^*}^+(x_i))_{1 \leq i \leq r}) \rightarrow 0.$$

Equivalently, there is a coupling of  $(\mathcal{T}_n, v_0)$  and  $\bar{\mathcal{T}}_n^*$  such that  $d_{\mathfrak{T}_{\text{lf}}^\bullet}(H(\mathcal{T}_n, v_0, \Omega_n), \bar{\mathcal{T}}_n^*) \xrightarrow{P} 0$ .

In Equation (20.4) and the subsequent paragraph of [9], Janson also argues that if

$$\Delta(\mathcal{T}_n) = (1 - \mu)n + o_p(n), \tag{2}$$

then  $d_{\text{TV}}(\Delta(\mathcal{T}_n), \tilde{D}_n) \rightarrow 0$ . Hence in the complete condensation regime where (2) is assumed to hold, we may choose  $\tilde{D}_n$  in the coupling of Theorem 7 such that  $\tilde{D}_n = \Delta(\mathcal{T}_n)$  with probability tending to 1 as  $n$  becomes large. This yields the asymptotic location of the vertex with maximum degree with respect to the random vertex  $v_0$ .

## References

- 1 R. Abraham, A. Bouaziz, and J.-F. Delmas. Local limits of Galton-Watson trees conditioned on the number of protected nodes. *ArXiv e-prints*, 2015. [arXiv:1509.02350](https://arxiv.org/abs/1509.02350).
- 2 R. Abraham, J.-F. Delmas, and H. Guo. Critical Multi-Type Galton-Watson Trees Conditioned to be Large. *ArXiv e-prints*, 2015. [arXiv:1511.01721](https://arxiv.org/abs/1511.01721).
- 3 Romain Abraham and Jean-François Delmas. Local limits of conditioned Galton-Watson trees: the condensation case. *Electron. J. Probab.*, 19:no. 56, 29, 2014.
- 4 Romain Abraham and Jean-François Delmas. Local limits of conditioned Galton-Watson trees: the infinite spine case. *Electron. J. Probab.*, 19:no. 2, 19, 2014.
- 5 David Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1(2):228–266, 1991. URL: [http://links.jstor.org/sici?sici=1050-5164\(199105\)1:2<228:AFDFGF>2.0.CO;2-8&origin=MSN](http://links.jstor.org/sici?sici=1050-5164(199105)1:2<228:AFDFGF>2.0.CO;2-8&origin=MSN).
- 6 Inés Armendáriz and Michail Loulakis. Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic Process. Appl.*, 121(5):1138–1147, 2011. doi:10.1016/j.spa.2011.01.011.
- 7 Michael Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability. doi:10.1007/978-3-211-75357-6.
- 8 C. Holmgren and S. Janson. Fringe trees, Crump-Mode-Jagers branching processes and  $m$ -ary search trees. *ArXiv e-prints*, 2016. [arXiv:1601.03691](https://arxiv.org/abs/1601.03691).
- 9 Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012. doi:10.1214/11-PS188.
- 10 Svante Janson, Thordur Jonsson, and Sigurdur Örn Stefánsson. Random trees with superexponential branching weights. *J. Phys. A*, 44(48):485002, 16, 2011. doi:10.1088/1751-8113/44/48/485002.
- 11 Thordur Jonsson and Sigurdur Örn Stefánsson. Condensation in nongeneric trees. *J. Stat. Phys.*, 142(2):277–313, 2011. doi:10.1007/s10955-010-0104-8.
- 12 Igor Kortchemski. Limit theorems for conditioned non-generic Galton-Watson trees. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):489–511, 2015. doi:10.1214/13-AIHP580.
- 13 Sophie Pénişon. Beyond the  $Q$ -process: various ways of conditioning the multitype Galton-Watson process. *ALEA Lat. Am. J. Probab. Math. Stat.*, 13(1):223–237, 2016.
- 14 Robin Stephenson. Local convergence of large critical multi-type galton-watson trees and applications to random maps. *Journal of Theoretical Probability*, pages 1–47, 2016. doi:10.1007/s10959-016-0707-3.
- 15 B. Stufler. Local limits of large Galton-Watson trees rerooted at a random vertex. To appear in *Annales de l’Institut Henri Poincaré Probabilités et Statistiques*.