Scaling Limits of Slim and Fat Trees

Vladislav Kargin[1]

Abstract

We consider Galton–Watson trees conditioned on both the total number of vertices n and the number of leaves k. The focus is on the case in which both k and n grow to infinity and $k = \alpha n + O(1)$, with $\alpha \in (0, 1)$. Assuming exponential decay of the offspring distribution, we show that the rescaled random tree converges in distribution to Aldous's Continuum Random Tree with respect to the Gromov–Hausdorff topology. The scaling depends on a parameter σ^* which we calculate explicitly. Additionally, we compute the limit for the degree sequences of these random trees.

1 Introduction

In this paper, we investigate the Galton–Watson ("GW") trees conditioned both on the total number of vertices and on the number of leaves, when the total number of vertices is large and the number of leaves is far from its expected value. For a situation with small number of leaves, we can expect long leafless branches, while in the situation when a tree has a large number of leaves, the branches are short and bushy, and this is why we call these trees "slim" and "fat," respectively.

The general setup is as follows. Let L be a non-negative, integer-valued random variable with the probability mass function $\mathbb{P}(L = i) = w_i$, $i = 0, 1, 2, \ldots$, which is assumed to satisfy the following conditions:

$$\mathbb{E}L = 1, \text{ and for some } a > 0, \mathbb{E}e^{aL} < \infty. \tag{1}$$

✉ Vladislav Kargin
vkargin@binghamton.edu

1   Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902, USA

Springer

We also assume that the distribution of L is not trivial: $w_1 \neq 1$.

A random tree T is the family tree of a Galton–Watson process with one individual in generation 0 and with offspring of each individual distributed independently as the random variable L. We use ordered trees so that the information on the birth order of family tree branches is known. For a formal description, see Sect. 3 in the paper of Le Gall and Le Jan [23].

The probability space Ω is the space of all ordered rooted trees with the probability distribution $\mathcal{P}$ induced by the distribution of random variable L. We write $\Omega_n \subset \Omega$ for trees with n vertices and $\Omega_{k,n} \subset \Omega_n$ for trees with k leaves and n vertices. We say that $(k, n) \in PP(w)$ for an offspring distribution $w = (w_0, w_1, \ldots)$ if the set of trees $\Omega_{k,n}$ has positive probability. In this case, both Ω_n and $\Omega_{k,n}$ are endowed with conditional probability laws $\mathcal{P}_n$ and $\mathcal{P}_{k,n}$ induced from law $\mathcal{P}$ on Ω.

Note that a probability distribution $(w_j)_{j=0}^{\infty}$ can be replaced by an equivalent probability distribution $(\tilde{w}_j)_{j=0}^{\infty}$ without changing the measure on Ω_n (see Section 4 in Janson [16]). In particular, by Remark 4.3 in [16], any equivalent probability distribution is determined by its expectation, and possible expectations are in $(0, \nu]$, where ν is a certain parameter. It is easy to check by using the definition of ν (see formulas (3.8) - (3.10) in [16]) that if $(w_j)_{j=0}^{\infty}$ has an exponentially declining tail, then $\nu \geq 1$. Therefore, for every such $(w_j)_{j=0}^{\infty}$ we can always find an equivalent probability distribution $(\tilde{w}_j)_{j=0}^{\infty}$ so that its expectation is 1. In particular, the condition $\mathbb{E}L = 1$ in (1) does not impose an additional restriction on measure of trees in Ω_n and $\Omega_{k,n}$ but is used only to fix a specific equivalent weight distribution.

We consider the case when the number of leaves k approximately proportional to the number of vertices n, more specifically, $k = \alpha n + O(1)$, where $\alpha \in (0, 1)$. Recall that if the offspring distribution $\{w_i\}_{i=0}^{\infty}$ has finite variance, then the expected number of leaves for a tree in Ω_n is $\sim w_0 n$ (see Theorem II.5.1 in [17]). We are mainly interested in the case when $\alpha \neq w_0$.

1.1 Main Results

In Theorem 1.2, we will show that as $n \to \infty$, a rescaled tree in $\Omega_{k,n}$ converges in the Gromov–Hausdorff topology to a random metric space called the Continuum Random Tree ("CRT"). This space was first introduced by Aldous in [1–3] and for details of the definition, we refer the reader to Section 2 in Le Gall's survey [22]. In order to formulate the result precisely, we introduce some additional notation in the next subsection.

1.1.1 α-Shifted Offspring Distribution

Let $\theta(t) := \sum_{i=0}^{\infty} w_i t^i$, where $\{w_i\}$ is the offspring distribution, and let ρ denote the radius of convergence for this series. By assumption (1), $\sum_{i=0}^{\infty} w_i = 1$ and $\sum_{i=1}^{\infty} i w_i = 1$. Therefore, $\theta(1) = 1$ and $\theta'(1) = 1$. In particular, $1 \leq \rho \leq \infty$. The assumption $w_1 \neq 1$ implies additionally that $w_0 \neq 0$.

Define

$$\widehat{\psi}(t) := \frac{t\theta'(t)}{\theta(t) - w_0}. \tag{2}$$

(Janson in [16] uses $\psi(t)$ to denote $\frac{t\theta'(t)}{\theta(t)}$.) In Lemma 2.1, we show that $\widehat{\psi}(t)$ is increasing on the interval $[0, \rho)$. Hence, we can define

$$\widehat{v} := \lim_{t\uparrow\rho} \widehat{\psi}(t) \in [1/(1 - w_0), \infty]. \tag{3}$$

($\widehat{v} \geq 1/(1 - w_0)$ because $\widehat{\psi}(1) = 1/(1 - w_0)$ and $\rho \geq 1$.)

Definition 1.1 Let $\alpha \in (0, 1)$, and let w_i^* be a probability mass distribution with unit mean, such that $w_0^* = \alpha$ and $w_j^* = Cw_j(t^*)^{j-1}$ for all $j > 0$ and for some $C > 0$ and $t^* > 0$. We call $w_0^*, w_1^*, \ldots$, the α-*shifted offspring distribution*. Let L^* denote a random variable that have the distribution $\mathbb{P}(L^* = j) = w_j^*$ for $j = 0, 1, 2, \ldots$.

In other words, in the α-shifted distribution, w_0^* is set to α and the remaining weights are exponentially re-weighted in such a way that the result is still a probability mass distribution with unit mean. It is easy to check that C and t^* can be determined from the following equations,

$$(1 - \alpha)t^*\theta'(t^*) = \theta(t^*) - w_0, \text{ and } C = \frac{1}{\theta'(t^*)}. \tag{4}$$

In Lemma 2.2, we show that these equations are solvable if $\alpha < 1 - 1/\widehat{v}$. (In particular, they are always solvable for $\alpha \leq w_0$. Intuitively, $\alpha > 1 - 1/\widehat{v}$ corresponds to the situation when the number of leaves is not sustainable by the structure of the offspring distribution.)

In Lemma 2.3, we show that the variance of the α-shifted distribution w_j^* is

$$(\sigma^*)^2 = \frac{t^*\theta''(t^*)}{\theta'(t^*)}. \tag{5}$$

1.1.2 Rescaled Convergence in the Gromov–Hausdorff Metric

Let us consider trees in $\Omega_{k,n}$ as metric spaces with a marked point ('root') by using the usual graph theoretic distance between vertices and extending it to the edges linearly. We will denote the distance between points u and v as $d(u, v)$.

Recall the definition of the Gromov–Hausdorff metric. If (E, δ) is a metric space, then the Hausdorff distance between compact subsets of E is defined as follows:

$$\delta_{\text{Haus}}(K, K') = \inf\{\alpha > 0 : K \subset U_\alpha(K') \text{ and } K' \subset U_\alpha(K)\},$$

where $U_\alpha(K) := \{x \in E : \delta(x, K) \le \alpha\}$. Then, if T and T' are two rooted compact metric spaces, with roots ρ and ρ', the Gromov–Hausdorff distance is defined as

$$d_{GH}(T, T') = \inf \left\{ \delta_{\text{Haus}}\big(\phi(T), \phi'(T')\big) \vee \delta\big(\phi(\rho), \phi'(\rho')\big) \right\},$$

where the infimum is over all choices of a metric space (E, δ) and all isometric embeddings $\phi : T \to E$ and $\phi' : T' \to E$. Two rooted compact metric spaces T_1, T_2 are equivalent if there is a root-preserving isometry between these two spaces and obviously $d_{GH}(T, T')$ depends only on the equivalence classes of T and T'. It is a fact that $d_{GH}(T, T')$ defines a metric on the set of equivalence classes of rooted compact metric spaces. (For the unrooted case, see Theorem 7.3.30 in Burago et al. [6]. The proof for the rooted case can be obtained after some adjustments.) Moreover, the set of isometry classes of real trees equipped with the Gromov–Hausdorff metric a complete and separable metric space. See Theorem 1 in a paper of Evans et al. [12].

Recall that we say that $(k, n) \in PP(w)$ for an offspring distribution $w = (w_0, w_1, \ldots)$ if the set of trees with n vertices and k leaves has a positive probability under the probability law $\mathcal{P}$ induced by this offspring distribution. We have the following convergence result.

Theorem 1.2 *Let the offspring distribution $w = (w_j)_{j=0,1,\ldots}$ satisfy Condition (1), and assume that $\big((k_i, n_i)\big)_{i=1}^{\infty}$ is a sequence in $PP(w)$ such that $n_i \to \infty$ and $k_i = \alpha n_i + O(1)$ where $\alpha \in (0, 1 - 1/\widehat{v})$. Let T_{n_i} be a tree chosen in Ω_{k_i,n_i} according to the probability law $\mathcal{P}_{k_i,n_i}$. Then,*

$$\frac{1}{\sqrt{n_i}} T_{n_i} \to \frac{2}{\sigma^*} \mathcal{T},$$

where $\mathcal{T}$ is the Aldous continuum random tree and the convergence is in distribution with respect to the Gromov–Hausdorff topology on the space of compact metric spaces.

1.1.3 Convergence of Contour Processes

Theorem 1.2 follows from a stronger result that asserts the convergence of the rescaled contour process associated with the tree $T_n \in \Omega_{k,n}$ to a Brownian excursion. For details of this implication, see the proof of Theorem 2.5 in Le Gall's survey [22]. The key fact is Lemma 2.4 in [22], which states that if T and T' are two real trees with contour functions $C(x)$ and $C'(x)$, then $d_{GH}(T, T') \le 2\|g - g'\|$, where $\|g - g'\|$ stands for the uniform norm of $g - g'$.

Before stating the stronger result, let us define the contour and related processes for an ordered rooted tree. While we need only the contour process to formulate the result, the other processes will be useful in the proof.

Let T be an ordered tree with n vertices.

The depth-first search (DFS): The sequence f_i, with $i = 0, \ldots, 2n-2$, is a sequence of vertices, defined as follows: (i) $f_0 = root$; (ii) given $f_i = v$, choose, if possible, the left-most child w of v that has not already been visited and set $f_{i+1} = w$. If this
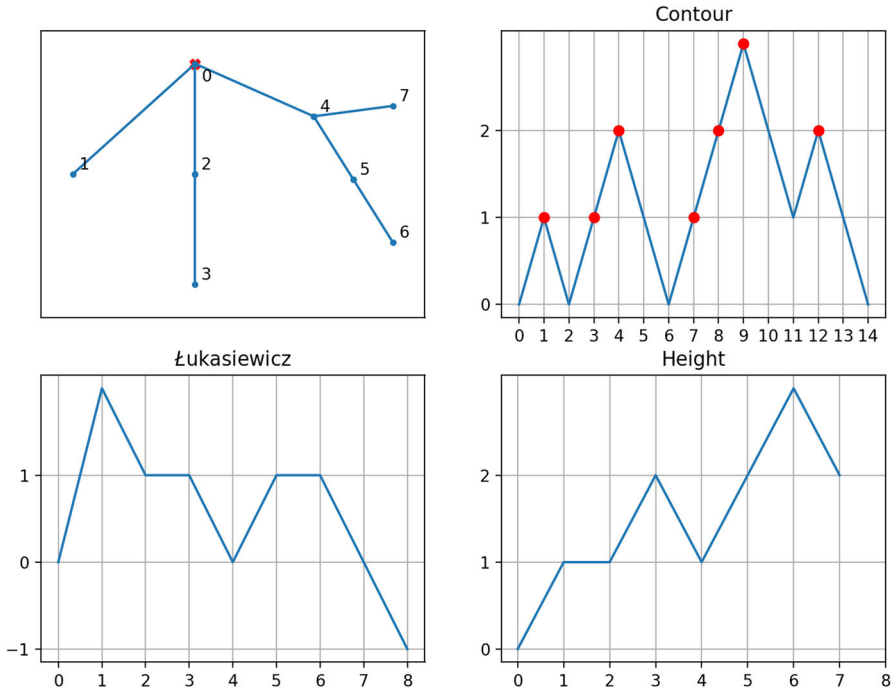
Fig. 1 The contour, Łukasiewicz, and height processes for a tree on 8 vertices. The red points $\big(m(l),\, H_n(l)\big)$ are an embedding of the height process in the contour process (see Definition 2.11)

is not possible, let f_{i+1} be the parent of v, (iii) the sequence ends when $f_{2n-2} = root$ and all vertices has been visited at least once.

The contour process $C_n(i)$, $i = 0, \ldots, 2n - 2$, is defined as follows:

$$C_n(i) = d(root,\, f_i) = \text{height of } f_i, \text{ the i-th vertex in the DFS.} \qquad (6)$$

Between the integer points, the path $C_n(t)$ is defined by linear interpolation. See Fig. 1 for an illustration.

We also define *the Łukasiewicz path process $S_n(i)$*, for $i = 0, \ldots, n$. (In the paper of Marckert and Mokkadem [24], this process is called the Depth First Queue Process (DFQP); we follow the terminology in the book of Flajolet and Sedgewick [14].) Assume that the vertices in a planar tree are denoted v_j, $j = 0, \ldots, n - 1$, in the order of their appearance in the Depth First Search (and without repetition, unlike f_i). Let ξ_i, $i = 1, \ldots, n$, be the number of children of vertex v_{i-1}. Then, $S_n(0) = 0$, and

$$S_n(j) = \sum_{i=1}^{j} (\xi_i - 1), \text{ for } 1 \leq j \leq n. \qquad (7)$$

Similar to the contour process, we extend the Łukasiewicz process to the intervals between integers by linearity. For illustration, see Fig. 1.

A Łukasiewicz path, by definition, is a piecewise continuous map from $[0, n] \to \mathbb{R}$ that has the following properties: its increments $S_n(j+1) - S_n(j) \in \{-1, 0, 1, \ldots\}$ for all integer $0 \leq j \leq n-1$; it is non-negative, $S_n(j) \geq 0$, for every $0 \leq j \leq n-1$; $S_n(0) = 0$, and $S_n(n) = -1$. The construction (7) maps an ordered rooted tree with n vertices to a Łukasiewicz path. In fact, every Łukasiewicz path can be obtained in this way, and therefore, this construction gives a bijective map from the set of ordered trees Ω_n to the set of Łukasiewicz paths on interval $[0, n]$.

Finally, we define the *height process* $H_n(i)$, $i = 0, \ldots, n-1$, as

$$H_n(i) = d(root, v_i) = \text{height of } v_i. \tag{8}$$

Now, let us formulate the key result. In the theorem below, the convergence of stochastic processes is the weak convergence with respect to the space of continuous functions on $[0, 1]$ with the topology induced by the uniform norm: $\|f\|_\infty = \sup_{t \in [0,1]} f(t)$.

Theorem 1.3 *Let the offspring distribution $w = (w_j)_{j=0,1,\ldots}$ satisfy Condition (1), and assume that $\big((k_i, n_i)\big)_{i=1}^\infty$ is a sequence in $PP(w)$ such that $n_i \to \infty$ and $k_i = \alpha n_i + O(1)$ where $\alpha \in (0, 1 - 1/\widehat{v})$. Let T_{n_i} be a tree chosen in Ω_{k_i, n_i} according to the probability law $\mathcal{P}_{k_i, n_i}$ and let C_{n_i} be the contour process for this tree. Then, as $n_i \to \infty$,*

$$\left(\frac{C_{n_i}(2n_i t)}{\sqrt{n_i}}\right)_{t \in (0,1)} \xrightarrow{weakly} \frac{2}{\sigma^*} e(t),$$

where $e(t)$ is a standard Brownian excursion, and σ^* is as defined in (5).

This theorem implies convergence results for continuous functionals on trees, in particular, the following result about the height of trees.

Corollary 1.4 *Let h_{n_i} be the height of a random tree $T_{n_i} \in \Omega_{k_i, n_i}$, and suppose that the assumptions of Theorem 1.3 are satisfied. Then, the distribution of $\sigma^* h_{n_i}/(2\sqrt{n_i})$ converges to the distribution of the maximum of a standard Brownian excursion.*

Proof The height of tree T_{n_i} corresponds to the maximum of the tree's contour process. The maximum is a continuous functional on functions in $C[0, 1]$ with the uniform metric. Therefore, by Theorem 1.3, the distribution of the maximum for the re-scaled contour process C_n converges to the distribution of the maximum of a standard Brownian excursion. □

1.1.4 Convergence of the Degree Sequence

Now, let $n_j(T)$ denote the number of vertices of out-degree j in tree T. (Recall that the out-degree of a vertex in a rooted tree is the number of edges that lead away from the root. For Galton–Watson trees, it is the number of children of an individual represented by the vertex.) We say that the sequence $(n_0(T), n_1(T), \ldots)$ is the *degree sequence* of tree T. A probability measure on trees $\Omega_{k,n}$ induces a probability measure on degree

sequences. By definition, $n_0(T) = k$ for a tree $T \in \Omega_{k,n}$. In addition, $\sum_{i=0}^{\infty} n_i(T) = n$, where n is the number of vertices in T, and therefore, $\{\frac{n_i(T)}{n}\}_{i=0}^{\infty}$ is a random probability measure.

Theorem 1.5 *Let the offspring distribution* $w = (w_j)_{j=0,1,\ldots}$ *satisfy Condition* (1), *and assume that* $(T_n \in \Omega_{k,n})$ *is a sequence of random trees with* $k \to \infty$, $n \to \infty$, *and* $(k, n) \in PP(w)$. *Suppose that* $k = \alpha n + O(1)$ *as* $n \to \infty$, *and* $\alpha \in (0, 1 - 1/\hat{v})$. *Then, almost surely,* $\frac{n_j(T_n)}{n} \to w_j^*$ *for* $j = 0, 1, 2, \ldots$, *where* $\{w_j^*\}$ *is the* α-*shifted offspring distribution.*

1.1.5 Examples

Here are some explicitly calculated α-shifted distributions together with the parameter $(\sigma^*)^2$.

Example 1.6 Consider GW trees with the geometric offspring distribution $w_j = 2^{-(j+1)}$, $j = 0, 1, \ldots$ conditioned to have n vertices. This distribution is an exponential rescaling of the weight sequence $w_j = 1$, where $j = 0, 1, \ldots$, and therefore, these trees have the uniform distribution on Ω_n. (This is Example 10.1 in Janson [16].) For this distribution, we calculate $\theta(t) = (2-t)^{-1}$ and $\hat{v} = +\infty$. Therefore, equations in (4) are solvable for every $\alpha \in (0, 1)$. The solution is $t^* = 2\alpha$ and $C = 4(1 - \alpha)^2$. This implies $w_0^* = \alpha$ and $w_j^* = \alpha^{j-1}(1 - \alpha)^2$ for $j = 1, 2, \ldots$. The variance is $(\sigma^*)^2 = 2\alpha/(1 - \alpha)$. Note that as α (i.e., the limiting ratio of the number of leaves to the number of vertices) increases, the variance parameter of the α-shifted offspring distribution, $(\sigma^*)^2$, also increases. In particular, we should expect that the height of the random tree in $\Omega_{k,n}$ decreases. (A tree with many leaves is "fat" and "short".)

Example 1.7 Consider unary-binary GW trees with weight sequence $w_0 = p$, $w_1 = 1 - 2p$, $w_2 = p$, and $w_i = 0$ for all other i, where $0 < p < 1/2$. In this case, $\hat{v} = 2$ and equations (4) are solvable for every $\alpha \in (0, \frac{1}{2})$. (It is not possible to solve these equations for $\alpha > 1/2$ because the number of leaves and the number of vertices are related by the inequality $n \geq 2k - 1$ in the unary-binary trees, with the equality reached for binary trees.) The setup of this example includes Examples 10.4 and 10.5 in Janson [16]. After α-shifting, we get $w_0^* = w_2^* = \alpha$, $w_1^* = 1 - 2\alpha$, with variance $(\sigma^*)^2 = 2\alpha$. We can see that, as α increases, the variance also increases, as in the previous example. However, in this case α is bounded from above; therefore, variance of the offspring distribution is bounded from above, and the tree cannot be made arbitrarily short.

1.1.6 Algorithm for Generating Trees in $\Omega_{k,n}$

The proof of Theorems 1.3 and 1.5 is based on a new algorithm that generates trees in $\Omega_{k,n}$ with weight sequence $w_1, w_2, \ldots$. In order to explain the algorithm, we introduce a useful map between ordered trees and balls-in-boxes allocations. The general idea of the algorithm is to generate a suitable random allocation and then, recover a random tree from this allocation.
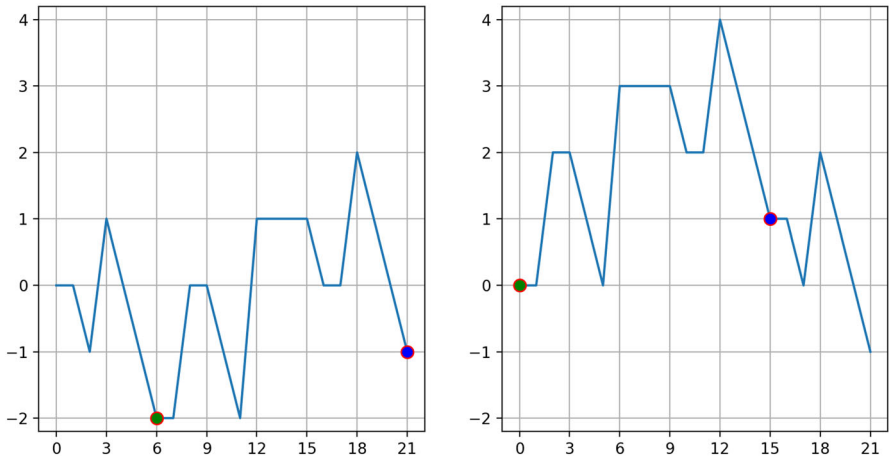
Fig. 2 An example of the cycle shift that sends a random walk bridge to a Łukasiewicz path

First, we define a map σ that sends an ordered tree on n vertices to the sequence of the vertex offspring sizes. In this sequence, the vertices are ordered by the DFS, but the repeats of the vertices are omitted:

$$\sigma : T \rightarrow \sigma(T) = (\xi_1, \ldots, \xi_n). \tag{9}$$

For example in Fig. 1, $\sigma(T) = (3, 0, 1, 0, 2, 1, 0, 0)$.

This sequence can be identified with an allocation of $n - 1$ balls (corresponding to edges) in n boxes (corresponding to vertices). The set of all possible allocations is denoted $B_{n-1,n}$. The map $\sigma : \Omega_n \rightarrow B_{n-1,n}$ is injective but not surjective.

By using the cycle lemma,[1] one can establish the following result.

Lemma 1.8 (Corollary 15.4 in [16]) *If* $(y_1, \ldots, y_n) \in B_{n-1,n}$, *then exactly one of the* n *cyclic shifts of* $(y_1, \ldots, y_n)$ *is the out-degree sequence* $\sigma(T)$ *of a tree* $T \in \Omega_n$.

The relevant cyclic shift can be identified as follows. First note that the sequence $y_1, \ldots, y_n$ corresponds to a path $W_0 = 0$, $W_j = \sum_{i=1}^{j}(y_i - 1)$, $j = 1, \ldots, n$, and that for this path $W_n = -1$. Let $m = \min_{0 \le j \le n} W_j$, and let $\widehat{j}$ be the first time when the path achieves this minimum. Then, the relevant cyclic shift of W_j is the path for the random walk with the steps given by $(y_{\widehat{j}+1} - 1, \ldots, y_n - 1, y_1 - 1, \ldots y_{\widehat{j}} - 1)$. It is easy to see that the resulting path is a Łukasiewicz path and, therefore, it corresponds to a tree $T \in \Omega_n$.

A cyclic shift is illustrated in Fig. 2.

---

[1] The cycle lemma appears first in a paper of Dvoretzky and Motzkin [11] and re-appears in different forms in many other papers. See Dershowitz and Zacks [7] and Note I.47 on p.75-76 of the book by Flajolet and Sedgewick [14] for an overview of the literature. The lemma is closely related to Vervaat's transform for the Brownian Bridge (see [31]) and is often called the discrete Vervaat transform (see Pitman [28], Exercise 1 in Section 5.1).

Now, we define an algorithm for the generation of trees in $\Omega_{k,n}$. Define the probability mass function $\widehat{w}_j^*$ as follows:

$$\widehat{w}_0^* := 0; \ \widehat{w}_j^* := \frac{w_j^*}{1 - \alpha}, \ \text{for } j = 1, 2, \ldots, \tag{10}$$

where w_j^* is the α-shifted distribution from Definition 1.1.[2]

Let the random sequence $\xi_1', \ldots, \xi_{n-k}'$ have the following distribution

$$\mathbb{P}(\xi_1' = j_1; \ldots; \xi_{n-k}' = j_{n-k}) = \begin{cases} \frac{1}{Z} \prod_{i=1}^{n-k} \widehat{w}_{j_i}^*, & \text{if } \sum_{i=1}^{n-k} j_i = n - 1, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

where Z is a normalizing constant. In other words, $\xi_1', \ldots, \xi_{n-k}'$ has the distribution of i.i.d. random variables with distribution $\widehat{w}_j^*$, conditioned to have the sum $n - 1$.

Algorithm A

(1) Generate a sample of $\xi_1', \ldots, \xi_{n-k}'$ with distribution as in (11).
(2) Add k zeros to the sequence ξ_i' and let $\xi'' = (\xi_i'')_{i \leq n}$ be a uniform permutation of the resulting sequence.
(3) Apply the (unique) cyclic shift to the resulting sequence $(\xi_i'')_{i \leq n}$ so that the result is in the image of the map σ defined in (9). Let the result be denoted $\xi_1, \ldots, \xi_n$.
(4) Obtain a tree $T \in \Omega_{k,n}$ as $\sigma^{-1}(\xi_1, \ldots, \xi_n)$.

It is assumed that the first step of this algorithm is implemented efficiently as described in Devroye [8].

Theorem 1.9 *Let the offspring distribution $w = (w_j)_{j=0,1,\ldots}$ satisfy Condition (1). Suppose that $\alpha \in (0, 1 - 1/\widehat{v})$ and $(k, n) \in PP(w)$. Then, algorithm A produces a tree $T \in \Omega_{k,n}$ with the probability distribution $\mathcal{P}_{k,n}$ corresponding to weights $w_0, w_1, \ldots$. If $k = \alpha n + O(1)$, then the expected running time of the algorithm is $O(n)$, as $n \to \infty$.*

Note that the essential corollary of this theorem is that for a random tree $T \in \Omega_{k,n}$ with the law $\mathcal{P}_{k,n}$, the distribution of $\sigma(T)$ equals the distribution of the random sequence $(\xi_1, \ldots, \xi_n)$ produced by the algorithm. This gives us the ability to study the Łukasiewicz and related processes for the random trees in $\Omega_{k,n}$.

We will first prove Theorem 1.9. After this theorem is proved, the general strategy is as follows. We use the fact that the sequence $(\xi_i'')_{i=1}^{n}$ from Algorithm A is exchangeable and apply tools from Billingsley [4] to show that the partial sums of this sequence converge (after rescaling) to the Brownian bridge. Then, properties of the discrete

---

[2] In the proofs of our theorems, the parameter α is given to us and satisfy the assumption that $k = \alpha n + O(1)$ as $n \to \infty$. If α is not known and the task is simply to generate the tree with k leaves and n vertices, then one can take $\alpha = k/n$.

Vervaat's transform ensure that the partial sums of $(\xi_i)_{i=1}^n$ converge after rescaling to a Brownian excursion. Then, we adapt tools from Marckert and Mokkadem [24] in order to show that this convergence implies the re-scaled convergence of the height and the contour processes to a Brownian excursion. The adaptation is necessary, because the sequence of i.i.d. random variables underlying $(\xi_i)_{i=1}^n$ is conditioned not only on the sum being equal to -1 but also on the number of increments -1 being equal to k. The changes to the argument are straightforward, but we give the argument in full detail for the convenience of interested reader.

Together, this gives a proof of Theorem 1.3 and, consequently, of Theorem 1.2.

For the proof of Theorem 1.5, we again start with results of Theorem 1.9. In this case, we note that it is enough to show that $n_{\xi'}(i)$, the number of occurrences of increment i in the sequence $(\xi_i')_{i=1}^{n-k}$, after normalization by $(n-k)^{-1}$ converges almost surely to $\hat{w}_i^*$. In order to show this, we use an idea from a paper of Broutin and Marckert [5]. Roughly speaking, the idea is that under certain conditions the events which are rare under a random walk with i.i.d. increments $\hat{\zeta}_j$ remain rare, even if a condition is imposed on the random walk (for example, if the variables $\hat{\zeta}_j$ are required to add up to a specific constant). By using this trick, we can extend the almost sure convergence for $n_{\hat{\zeta}}(i)$, where $(\hat{\zeta}_j)_{j=1}^{n-k}$ are i.i.d., to the almost sure convergence of $n_{\xi'}(i)$. This allows us to prove Theorem 1.5

The rest of the paper is organized as follows. In Sect. 1.2, we describe the relation of our results to previous work. In Sect. 2, we give detailed proofs. In Sect. 3, we discuss some remaining problems. Appendix is devoted to the proof of an auxiliary local limit theorem.

1.2 Relation to Previous Work

Excellent reviews of the vast literature about the asymptotic behavior of finite random trees can be found in a book [14] by Flajolet and Sedgewick, and in review articles [22] by Le Gall and [16] by Janson.

From early on, it was observed that different families of random trees often have some similar macroscopic properties. For example, the tree height distribution has the same functional form for many tree families, although the distribution parameters could depend on the specific choice of the family.

In a sequence of papers, [1–3], Aldous provided an explanation for this phenomenon by showing that certain sequences of finite random trees converge under rescaling in the Gromov–Hausdorff topology to universal objects that he called the *continuum random trees (CRTs)*. In particular, he showed that Galton–Watson trees with a finite-variance offspring distribution converge under rescaling to a special CRT, which he called the *Brownian CRT*. (We will sometime refer to this limit as the Aldous CRT.) The tree is characterized by the property that its contour process is the Brownian excursion. Aldous's proof of this result is based on graph theoretical arguments which showed that all such Galton–Watson trees converge under rescaling to the same limit independent of the specific offspring distribution. Then, proof was concluded by an investigation of a particular case of GW trees with geometric offspring distribution.

Marckert and Mokkadem in [24] gave a different proof of this result under the condition that the offspring distribution has exponentially declining tails. In their proof, they establish the convergence (under rescaling) of the Łukasiewicz process to the Brownian excursion and then, show that the paths of the height process must be uniformly close to the paths of the Łukasiewicz process. For this implication, they use a relation of Galton–Watson trees to queues and an explicit formula for the height process in terms of the Łukasiewicz process. Then, one can handle the height process by using methods from the classic theory about ladder heights (Feller [13], Chapter XII). Finally, the rescaled convergence of the height process implies the rescaled convergence of the contour process.

Duquesne in [9] extends Aldous's result to Galton–Watson trees that may have infinite variance. He relies heavily on the machinery developed in papers of Duquesne and Le Gall [10] and of Le Gall and Le Jan [23]. While their approach is similar to the approach of Marckert and Mokkadem in [24], however, in the case of infinite variance the Łukasiewicz process converges under rescaling to an excursion of a Lévy process, and it is not immediately clear what is the candidate for the rescaled limit of the associated height process. The authors of these papers defined a height process for the Lévy process and then, showed that the height process of the GW trees converges under rescaling to this continuous-time height process. In this way, they showed that Galton–Watson trees with infinite variance offspring distribution converge under rescaling to *Lévy CRTs*, encoded by these height processes. (There are several such CRTs, depending on the parameter of the Lévy process.)

An alternative proof of Duquesne's theorem was given by Kortchemski in [19]. The proof is still based on the results in Duquesne and Le Gall [10]. However, Kortchemski uses the fact that the probability measure of events conditioned on the size of a GW tree equal to n is absolutely continuous with respect to the probability measure of the events conditioned on the size of the tree being at least n, and this type of conditioning was already handled in Section 2.5 of [10]. A similar absolute continuity relation can be obtained for the continuous limit objects, and Kortchemski concludes the proof by showing the uniform convergence of the Radon–Nikodym densities.

By a similar method, in [18], Kortchemski proved that if one conditions a Galton–Watson tree on the number of leaves, instead of the total number of vertices, then one still has the convergence under rescaling to the Brownian CRT or a Lévy CRT. The key additional ingredient in the proof is the fact that a GW tree with k leaves has k/w_0 total vertices with high probability. He also proved an extension of this result to the case when one conditions the GW tree on the number of vertices that have degrees in a given non-empty set of integers.

The applicability of these universality results has been significantly enlarged in [15] by Haas and Miermont. They considered a big class of trees with Markov branching property. In random trees with this property, the subtrees above some given height are independent with a law that depends only on the tree size, and the size is measured either as the number of leaves or vertices. Their result is that these trees converge under rescaling to universal random objects, called self-similar fragmentation trees.

In particular, Haas and Miermont show in [15] that the convergence under rescaling holds in the case when the Galton–Watson trees are conditioned either on the total progeny, or on the number of leaves (but not both). In [29], Rizzolo used the method of

[15] to show the convergence under rescaling of the Galton–Watson trees conditioned on the number of vertices that have degrees in a specified set. The proof of these results is based on the theory of fragmentation processes, and this gives a completely different approach to results by Aldous, Duquesne, Le Gall, Le Jan, and Kortchemski.

In this paper, we use the approach based on random walks and their continuous-time limits and not on fragmentations processes.

The difference of this paper from earlier studies described above is that we focus on the situation when a Galton–Watson random tree is conditioned on both the number of total progeny and the number of leaves. In particular, we consider a situation in which these two requirements are at odds with each other. In a sense, the obtained results are about an atypical event since they are concerned with the structure of trees under an additional condition super-imposed on the usual conditional Galton–Watson tree.

A similar approach is taken in papers [20] by Labarbe and Marckert, and [5] by Broutin and Marckert. The paper [20] considers Bernoulli random walks (that is, walks with increments in $\{-1, +1\}$) and associated processes—bridges, meanders and excursions—conditioned on a given number of peaks. A well-known bijection maps the contour process of an ordered tree to a Bernoulli excursion, and the results in [20] imply the convergence of uniformly distributed ordered trees with a given number of leaves to the Aldous continuum random tree (see Section 1.1 in their paper).[3] This is a particular case of our Theorem 1.2 for the case when weights w_j are geometric, $w_j = 2^{-j-1}$.

For ordered trees with non-uniform distributions, the contour process is still an (interpolated) excursion with increments $\{-1, +1\}$, but its distribution is more complicated and its rescaled convergence is difficult to establish directly. So, instead of the contour process, we use the Łukasiewicz process, that has a more tractable distribution, in order to prove the rescaled convergence to the Brownian excursion. Then, we infer that this implies the convergence for the contour process as well, by applying the techniques developed by Marckert and Mokkadem in [24]. This allows us to derive the convergence under re-scaling for a broader class of random trees than that considered in the paper of Labarbe and Marckert [20].

From the technical viewpoint, the methods by which we obtain the re-scaled convergence of the Łukasiewicz process to the Brownian excursion, also differ from methods in [20]. In particular, in [20] the authors use combinatorics to explicitly evaluate the number of simple walk paths with a given number of peaks between two lattice points. Then, they use the Stirling formula and obtain the finite-dimensional distributions of the limit by calculating the asymptotics of these explicit expressions. The combinatorics at the heart of this approach becomes significantly harder for paths with a more general set of increments, and we choose to use a different approach.

For tightness, [20] uses a combinatorial trick that maps simple walks with increments ± 1 to walks with $\{0, 1\}$ increments in a non-trivial fashion: peaks and valleys of the original walk path correspond to increment 1, and the other points between steps of the original path to increment 0. This map is injective and converts conditioning on the number of peaks to conditioning on the final point of the walk, which is easier

---

[3] In addition, the paper [20] studies the counting process for leaves under DFS exploration of the tree. For extension of these results to GW trees, see a recent paper of Thévenin [30].

to handle. Then, it is noted that a formula for the inverse map involves summation of independent geometric random variables, and this fact is used in order to show that the tightness also holds for the original random walk. Additional conditioning on the final point of the original random walk (needed for bridges and excursions) is handled by using the explicit formulas for the number of paths with a given number of peaks.

This argument is hard to generalize to walks with more general increments (and so to Łukasiewicz walks) because of the absence of the required combinatorial trick.

Instead, we use the connection of discrete walks with the balls-in-boxes model, as described by Kolchin [17] and Janson [16], and with exchangeable random variables. This allows us to apply the results in Billingsley's book [4] that establish the re-scaled convergence for partial sums of a certain class of exchangeable random variables.

In a paper of Broutin and Marckert, [5], the entire degree sequence of a tree is specified.[4] More precisely, let $d_i(\tau)$ denote the number of vertices of out-degree i in the tree τ, and let $\vec{d}(\tau) = \{d_i(\tau)\}_{i=0}^{\infty}$ be the degree sequence of τ. Also, let $|\vec{d}| = \sum_{i=0}^{\infty} d_i$ and $\Delta(\vec{d}) = \max\{d_i\}$ be the size and the maximum degree of the tree. Assume that $\vec{d}^{(1)}, \vec{d}^{(2)}, \ldots$ be a sequence of degree sequences so that $|\vec{d}^{(j)}| \to \infty$, $\Delta(\vec{d}^{(j)}) = o(\sqrt{|\vec{d}^{(j)}|})$, and the measure $\vec{d}^{(j)}/|\vec{d}^{(j)}|$ converges to a probability measure $\mu_0, \mu_1, \ldots$ that has finite second moment.

Take the uniform measure on the set of trees with degree sequence $\vec{d}^{(j)}$ and let the correspondent random tree be denoted T_j. Broutin and Marckert showed that the sequence of trees T_j converges under rescaling to the Brownian CRT in the Gromov–Hausdorff topology. The proof is based on the approach of Marckert and Mokkadem [24], and it is similar to our approach. In particular, their proof of the convergence of Łukasiewicz process uses Billingsley's theorem 24.1 in [4]. In their main theorem, the conditions for applicability of Billingsley's theorem hold by assumption on $\Delta(\vec{d}^{(j)})$.

The results of [5] could be used to prove our main result, Theorem 1.2, provided one can show that as n grows, the degree sequences of the trees in $\Omega_{n,k}$ (after normalization) converge in probability to the α-shifted distribution and that they satisfy the condition that $\Delta\vec{d}^{(j)} = o\left(\sqrt{|\vec{d}^{(j)}|}\right)$.

In this paper, we choose a different, more direct approach to the proof of the main result. However, Theorem 1.5 contains the main ingredient of the different approach. Namely, in this theorem we show that as $n \to \infty$, the (normalized) degree sequence of a tree in $\Omega_{n,k}$ almost surely converges to the α-shifted distribution.

2 Proofs

2.1 Preliminary Results

We start with proving some results about the existence of the α-shifted distribution and about its variance.

Lemma 2.1 *The function $\widehat{\psi}(t)$ in (2) is an increasing continuous function on $(0, \rho)$.*

---

[4] See also an application of this approach to random forests in Lei [21], and an extension in a paper of Ojeda et al. [25] where different assumptions about the degree sequence are used.

Proof Let $\widehat{\theta}(t) = \theta(t) - w_0 = \sum_{i=1}^{\infty} w_i t^i$, and let

$$p_i(t) = \frac{w_i t^i}{\widehat{\theta}(t)} \text{ for } i = 1, 2, \ldots,$$

for $t \in (0, \rho)$, and zero otherwise. Then, for every $t \in (0, \rho)$, $p_i(t)$ is a probability mass function (pmf) of a discrete random variable X_t.

By definition, $\widehat{\psi}(t) = t\theta'(t)/\widehat{\theta}(t) = t\widehat{\theta}'(t)/\widehat{\theta}(t)$. Then, we calculate:

$$\begin{aligned}
t\widehat{\psi}'(t) &= \frac{t\widehat{\theta}'(t)}{\widehat{\theta}(t)} + \frac{t^2\widehat{\theta}''(t)}{\widehat{\theta}(t)} - \frac{t^2\widehat{\theta}'(t)^2}{\widehat{\theta}(t)^2} \\
&= \sum_{i=1}^{\infty} i p_i(t) + \sum_{i=1}^{\infty} i(i-1) p_i(t) - \left(\sum_{i=1}^{\infty} i p_i(t) \right)^2 \\
&= \mathbb{E}(X_t^2) - (\mathbb{E}X_t)^2 = \mathbb{V}\mathrm{ar}(X_t) > 0.
\end{aligned}$$

$\square$

Lemma 2.2 *Let $w_0, w_1, \ldots$ be a probability distribution with unit mean and finite variance. If $\alpha < 1 - 1/\widehat{v}$ with $\widehat{v}$ as defined in (3), then there exist $C > 0$ and $t^* > 0$ that solve problem (4).*

Proof We can write (4) as $C = 1/\theta'(t^*)$ and

$$\widehat{\psi}(t^*) = \frac{1}{1 - \alpha}. \tag{12}$$

By Lemma 2.1, the function $\widehat{\psi}(t)$ is continuous and increasing on the interval $(0, \rho)$, and we have: $\lim_{t \to 0} \widehat{\psi}(t) = 1$, $\lim_{t \to \rho} \widehat{\psi}(t) = \widehat{v}$. So if $(1 - \alpha)^{-1} < \widehat{v}$, then the equation (12) has a solution $t^* \in (0, \rho)$ by the intermediate value theorem, and we can set $C = 1/\theta'(t^*)$ since $\theta'(t^*)$ is well-defined and non-zero for $t \in (0, \rho)$. $\square$

Lemma 2.3 *The variance of the α-shifted distribution $\{w_j^*\}$, $j = 0, 1, \ldots$, is*

$$(\sigma^*)^2 = \frac{t^*\theta''(t^*)}{\theta'(t^*)}, \tag{13}$$

where t^* is the solution of equations (4).

Proof The generating function for the distribution $\{w_j^*\}$ is

$$\begin{aligned}
\theta^*(t) &:= \sum_{j=0}^{\infty} w_j^* t^j = \alpha + C w_1 t + C w_2 t^* t^2 + \ldots \\
&= \alpha + \frac{C}{t^*} \big(\theta(t^* t) - w_0 \big).
\end{aligned}$$

Then, the mean is $(\theta^*)'(1) = C\theta'(t^*) = 1$ by (4). For probability distributions with unit mean, the variance equals the second derivative of the generating function evaluated at $t = 1$. So we get that the variance is

$$(\theta^*)''(1) = Ct^*\theta''(t^*) = \frac{t^*\theta''(t^*)}{\theta'(t^*)}.$$

□

2.2 Proofs of Main Results

In this section, we prove Theorems 1.9, 1.3 and 1.5, about the generation algorithm, the convergence of the contour process, and the most likely type of the degree profile, respectively.

First, we establish some useful notation and give additional details about the relation of Galton–Watson trees and balls-in-boxes allocations.

Let w_k, $k = 0, 1, \ldots$ be a probability mass function with unit mean. It induces distributions on trees in Ω_n, as conditioned Galton–Watson trees, and on allocations of the balls-in-boxes model $B_{n-1,n}$. [The probability of an allocation $(y_1, \ldots y_n)$ is $\propto \prod_{i=1}^{n} w_{y_i}$.] For the injective map σ from (9), the push-forward probability distribution on $\sigma(\Omega_n)$ equals the distribution on $B_{n-1,n}$ conditioned to the image of the map σ.

The following result illustrates the relation between random trees and the balls-in-boxes model.

For a balls-in-boxes allocation $y = (y_1, \ldots, y_n)$, let $N_s(y)$ be the number of those y_i in allocation y that are equal to s. The *occupation profile* of an allocation y is the sequence $(N_s(y))_{s=0}^{\infty}$. In particular, $N_0(y)$ denotes the number of empty boxes in an allocation.

Note that a cyclic shift does not change the probability of an allocation and does not change numbers $N_s(y)$. This fact and the cyclic shift lemma (see Lemma 1.8) imply the following result. Recall that $n_s(T)$ denote the number of vertices of out-degree s in a random tree $T \in \Omega$. Then, the joint distribution of the degree sequence $(n_s(T))_{s=0}^{\infty}$ coincides with the joint distribution of the occupation profile $(N_s(y))_{s=0}^{\infty}$ for a random allocation y in $B_{n-1,n}$.

Lemma 2.4 (Lemma II.2.2 in [17]) *If the set Ω_n is not empty, then*

$$\mathbb{P}(n_s(T) = n_s; \ s = 0, \ldots, n-1) = \mathbb{P}(N_s(y) = n_s; \ s = 0, \ldots, n-1).$$

Now, we turn to the proofs of Theorems.

Proof of Theorem 1.9 Since σ is a one-to-one correspondence, for the first claim of the theorem it is enough to show that the distribution of the random sequence $(\xi_1, \ldots, \xi_n)$ produced in the third step of the algorithm equals to the distribution of $\sigma(T)$ for $T \in \Omega_{k,n}$ distributed according to $\mathcal{P}_{k,n}$.

Let v_i, $i = 0, \ldots, n - 1$, be the vertices of an ordered rooted tree $T \in \Omega_{k,n}$ listed in the DFS order. The probability of the tree is

$$p(T) \propto \prod_{i=0}^{n-1} w_{d(v_i)}.$$

The probability of an allocation $y = \sigma(T) = (y_1, \ldots, y_n) \in B(n - 1, n)$ is

$$p(y) \propto \prod_{j=1}^{n} w_{y_j}, \tag{14}$$

and the allocations are different for different T by the injectivity of map σ. If τ_i denotes a shift of allocation y by i positions then for all y in the image of σ, the allocations $\tau_i(y)$ are different for $i = 0, \ldots, n - 1$ by the Cyclic Lemma 1.8 and their union covers the set

$$B_k(n - 1, n) = \{y : y \in B(n - 1, n), \ N_0(y) = k\}.$$

Hence, the distribution (14) on the allocations in $\sigma(\Omega_{k,n})$ can be obtained provided that one can generate allocations in $B_k(n - 1, n)$ with distribution

$$p_k(y) \propto \prod_{i=1}^{n} w_{y_j} \propto w_0^k \prod_{j=1}^{\infty} w_j^{N_j(y)} \tag{15}$$

(The difference of (14) and (15) is that the latter is over all allocations in $B_k(n-1, n)$, while the former is only over the allocations in $\sigma(\Omega_{k,n}) \subset B_k(n - 1, n)$.)

The distribution (15) coincides with the distribution of random variables $\xi_1, \ldots, \xi_n$ which are i.i.d random variables $\zeta_1, \ldots, \zeta_n$ with distribution $w_0, w_1, \ldots$ conditional that $\sum_{i=1}^{n} \zeta_i = n - 1$ and that k of variables ζ_i take value 0.

By exchangeability of random variables $(\xi_1, \ldots, \xi_n)$, all locations for those random variables ξ_i that take value 0 are equally probable. Conditional on the choice of these locations, the remaining variables, which we denote $\xi' = (\xi'_1, \ldots, \xi'_{n-k})$, have the joint distribution

$$p(\xi') \propto \prod_{j=1}^{\infty} w_j^{N_j(\xi')} \text{ for } \xi' : \sum_{i=1}^{n-k} \xi'_i = n - 1.$$

Note that the distribution of $\xi'_1, \ldots, \xi'_{n-k}$ is the same as the probability distribution of an i.i.d sequence $\zeta'_1, \ldots, \zeta'_{n-k}$, with ζ'_1 having the distribution $\widehat{w}_0 = 0, \widehat{w}_i = w_i/(1 - w_0)$ conditional that $\sum_{i=1}^{n-k} \zeta'_i = n - 1$. By exponential re-weighting, which is possible by Condition (1), this distribution equals the probability distribution of an i.i.d sequence $\widehat{\zeta}_1, \ldots, \widehat{\zeta}_{n-k}$ with $\widehat{\zeta}_1$ having distribution $\widehat{w}_0^* = 0, \widehat{w}_i^* = w_i^*/(1 - \alpha)$ and conditioned to have the sum $\sum_{i=1}^{n-k} \widehat{\zeta}_i = n - 1$. In other words, the random variables $(\xi'_1, \ldots, \xi'_{n-k})$ have the distribution given in (11). This proves the first part

of the theorem and shows the validity of the algorithm: once we generated random variables $(\xi_1', \ldots, \xi_{n-k}')$, we reverse the steps of the above argument to generate the corresponding random tree.

In remains to estimate the speed of the algorithm. The main step is to generate $(\xi_1', \ldots, \xi_{n-k}')$, or, which is the same, a sequence $(\hat{\zeta}_1, \ldots, \hat{\zeta}_{n-k})$ Note that for $\hat{\zeta}_i$ we have $\mathbb{E}S_{\hat{\zeta}}(n-k) = (n-k)/(1-\alpha) \sim n$.)

A fast method for generating such sequences was developed by Devroye [8]. In application to our situation, the method is as follows.

(1) Generate a multinomial random vector $(N_1, N_2, \ldots,)$ with parameters $(n - k, \widehat{w}_1^*, \widehat{w}_2^*, \ldots)$. Repeat this step until a sample is obtained such that

$$\sum_{j=0}^{\infty} j N_j = n - 1.$$

(2) Form an array of length $n - k$ with N_j values j and randomly permute it. The result is the desired sequence $\xi_1', \ldots, \xi_{n-k}'$.

Let us analyze the running time of this algorithm.

By Theorem 1 in [8], there is a method to generate the required multinomial vector in expected time $\tau_n = o(n^{1/2})$ provided that the sequence $\widehat{w}_i^*$ has finite variance. (In fact, for a compactly supported distribution, this time is $O(1)$.)

The distribution of the sum $\sum_{j=0}^{\infty} j N_j$ is the same as the distribution of the sum of an i.i.d. sequence $\hat{\zeta}_1, \ldots, \hat{\zeta}_{n-k}$ where each r.v. has pmf $\{\widehat{w}_i^*\}$. The expectation of the pmf $\{\widehat{w}_i^*\}$ is $1/(1-\alpha)$ by its definition (10); hence, the expectation of the sum is $(n-k)/(1-\alpha) = n$. By using the local central limit theorem (valid under the assumption that the variance of pmf $\widehat{w}_i^*$ is finite), the probability that $\sum_{i=1}^{n-k} \hat{\zeta}_i = n-1$ is bounded from below by $\Omega(1/\sqrt{n})$. Therefore, the expected time that we draw a sample N_j such that $\sum_{j=0}^{\infty} j N_j = n - 1$ is $\tau_n \sqrt{n} = o(n)$.

The time for other operations (adding zeros and random shuffling) is linear in n.

Proof of Theorem 1.3 The proof consists of several steps. In the first step, we will show that the Łukasiewicz path of a random tree in $\Omega_{k,n}$ converges to a Brownian excursion. This will be accomplished in Lemmas 2.7 and 2.8 by analyzing random sequences produced by Algorithm A.

In the second step of the proof, we are going to show that the Łukasiewicz path of a random tree from $\Omega_{k,n}$ is close to the height process of the tree. This result is formulated in Lemma 2.10.

In the third step, we will show that with high probability the height process is close to the contour process. See Lemma 2.12.

The final step is to bring all these results together and show that the contour process for a tree in $\Omega_{k,n}$ converges to the Brownian excursion.

In all lemmas in this section (Lemmas 2.5 - 2.14), we assume that the assumptions of Theorem 1.3 are satisfied. That is, condition (1) holds, and the sequence of pairs (k, n) is such that $(k, n) \in PP(w)$ and, as $n \to \infty$, $k = \alpha n + O(1)$, where $\alpha \in (0, 1 - 1/\widehat{v})$.

Step 1: Łukasiewicz path converges to Brownian excursion.

Let $\xi_1'', \ldots, \xi_n''$ be the sequence of random variables produced by Step 2 of Algorithm A. Some properties of these random variables are proved in Lemmas 2.5 and 2.6. Then, we prove the convergence of Łukasiewicz path to Brownian excursion in Lemmas 2.7 and 2.8.

Lemma 2.5 *As $n \to \infty$, we have:*

(a) *For each integer $s \geq 0$, $\mathbb{E}((\xi_1'')^s) \to \mathbb{E}((L^*)^s)$, where L_* is a random variable with distribution w_j^*. In particular, $\mathrm{Var}(\xi_1'') \to (\sigma^*)^2$;*

(a) *For each integer pair $s, t \geq 0$, $\mathrm{Cov}\left((\xi_1'')^s, (\xi_2'')^t\right) \to 0$.*

(a) *For every $\beta > 0$, there exists a $\gamma > 0$ such that*

$$\mathbb{P}(\max_j |\xi_j''| > n^\beta) \leq e^{-\gamma n^\beta}.$$

Proof (a) First, consider variables ξ_i'. They have the distribution:

$$\mathbb{P}\left(\xi_1' = j_1, \ldots, \xi_{n-k}' = j_{n-k}\right) = \mathbb{P}\left(\hat{\zeta}_1 = j_1, \ldots, \hat{\zeta}_{n-k} = j_{n-k} \,\Big|\, S_{\hat\zeta}(n-k) = n-1\right),$$

where $\hat{\zeta}_1, \ldots, \hat{\zeta}_{n-k}$ are i.i.d. random variables with distribution $\mathbb{P}(\hat{\zeta}_1 = j) = \widehat{w}_j^* = w_j^*/(1-\alpha)$, if $j \geq 1$, and 0 if $j = 0$. Here, $S_{\hat\zeta}(p)$ denotes $\sum_{i=1}^p \hat{\zeta}_i$. Therefore,

$$\mathbb{P}(\xi_1' = j) = \mathbb{P}\left(\hat{\zeta}_1 = j \,\Big|\, S_{\hat\zeta}(n-k) = n-1\right) = \frac{\mathbb{P}\left(\hat{\zeta}_1 = j, \sum_{i=2}^{n-k} \hat{\zeta}_i = n-1-j\right)}{\mathbb{P}\left(S_{\hat\zeta}(n-k) = n-1\right)}$$

$$= \mathbb{P}(\hat{\zeta}_1 = j) \frac{\mathbb{P}\left(S_{\hat\zeta}(n-k-1) = n-1-j\right)}{\mathbb{P}\left(S_{\hat\zeta}(n-k) = n-1\right)}.$$

The distribution of the random variable ξ_1'' can be obtained as a mix distribution: $\xi_1'' = 0$ with probability (k/n) and $\xi_1'' = \xi'$ with probability $(1 - k/n)$, where ξ' has the same distribution as ξ_1'. Also, the distribution L^* is a mix of the trivial distribution with probability α and the distribution of $\hat\zeta$ with probability $(1-\alpha)$. So, for $j > 0$,

$$\mathbb{P}(\xi_1'' = j) = (1 - k/n)\mathbb{P}(\xi_1' = j),$$
$$\mathbb{P}(L^* = j) = (1 - \alpha)\mathbb{P}(\hat{\zeta}_1 = j),$$

and therefore,

$$\mathbb{P}(\xi_1'' = j) = \mathbb{P}(L^* = j) \frac{1 - k/n}{1 - \alpha} \frac{\mathbb{P}\left(S_{\hat\zeta}(n-k-1) = n-1-j\right)}{\mathbb{P}\left(S_{\hat\zeta}(n-k) = n-1\right)}.$$

Next, both k and n grow to infinity in such a way that $k/n \to \alpha \in (0, 1)$, hence $\mathbb{E}S_{\hat{\zeta}}(n - k) = (n - k)\mathbb{E}\hat{\zeta}_1 = n(1 - k/n)/(1 - \alpha) = n + o(1)$ and similarly $\mathbb{E}S_{\hat{\zeta}}(n - k - 1) = n + o(1)$. We apply the local limit law (Theorem I.4.2 in Kolchin [17]) and find that for any $\varepsilon > 0$ we can find $n(\varepsilon)$ such that for all $n > n(\varepsilon)$ and all $j < n^{1/4}$,

$$\left| \frac{1 - k/n}{1 - \alpha} \frac{\mathbb{P}\Big(S_{\hat{\zeta}}(n - k - 1) = n - 1 - j\Big)}{\mathbb{P}\Big(S_{\hat{\zeta}}(n - k) = n - 1\Big)} - 1 \right| < \varepsilon.$$

In addition, this ratio of probabilities is bounded uniformly in j and n.

Then, for arbitrary $\varepsilon > 0$ and $n > n(\varepsilon)$, and for some $c > 0$.

$$\left| \mathbb{E}\big[(\xi_1'')^s\big] - \mathbb{E}\big[(L^*)^s\big] \right| = \varepsilon \sum_{j=0}^{\lfloor n^{1/4} \rfloor} j^s \, \mathbb{P}(L^* = j) + c \sum_{j=\lfloor n^{1/4} \rfloor + 1}^{\infty} j^s \, \mathbb{P}(L^* = j).$$

By assumption, the probability distribution of L has exponentially declining tails, which implies that L^* has exponentially declining tails, so the last term is bounded by

$$c \sum_{j=\lfloor n^{1/4} \rfloor + 1}^{\infty} j^s \, e^{-aj} \to 0,$$

as $n \to \infty$, and we conclude that all moments of random variable ξ_1'' converge to the corresponding moments of L^*, as $n \to \infty$. In particular, $\mathbb{V}\mathrm{ar}(\xi_1'') \to (\sigma^*)^2$.

(b) By a similar argument, we can write for $j_1 > 0$, $j_2 > 0$,

$$\mathbb{P}(\xi_1'' = j_1, \xi_2'' = j_2) = \mathbb{P}(L_1^* = j_1, L_2^* = j_2)\left(\frac{1 - k/n}{1 - \alpha}\right)^2$$

$$\times \frac{\mathbb{P}\Big(S_{\hat{\zeta}}(n - k - 2) = n - 1 - j_1 - j_2\Big)}{\mathbb{P}\Big(S_{\hat{\zeta}}(n - k) = n - 1\Big)},$$

where L_1^*, L_2^* are two independent random variables that have the α-shifted distribution. By using the local limit law, we find that the ratio of probabilities approaches 1 as $n \to \infty$, uniformly for $j_1 < n^{1/4}$, $j_2 < n^{1/4}$. In addition, this ratio is bounded uniformly in n, j_1, j_2. Together with assumption (1), this ensures that $\mathbb{E}\big((\xi_1'')^s(\xi_2'')^t\big) \to \mathbb{E}\big((L_1^*)^s(L_2^*)^t\big)$ as $n \to \infty$, for all integer $s \geq 0, t \geq 0$. In particular, this implies that $\mathrm{Cov}\big((\xi_1'')^s, (\xi_2'')^t\big) \to 0$, as $n \to \infty$.

(c) We have

$$\mathbb{P}(\max_i \xi_i' \leq n^\beta) = \mathbb{P}\Big(\xi_1' \leq n^\beta, \ldots, \xi_{n-k}' \leq n^\beta\Big)$$

$$= \mathbb{P}\Big(\hat{\zeta}_1 \leq n^\beta, \ldots, \hat{\zeta}_{n-k} \leq n^\beta \,\big|\, S_{\hat{\zeta}}(n - k) = n - 1\Big)$$

$$= \left[\mathbb{P}(\hat{\zeta}_1 \leq n^\beta) \right]^{n-k} \frac{\mathbb{P}\left(S_{\hat{\zeta}}(n-k) = n-1 \middle| \hat{\zeta}_1 \leq n^\beta, \ldots, \hat{\zeta}_{n-k} \leq n^\beta \right)}{\mathbb{P}(S_{\hat{\zeta}}(n-k) = n-1)}.$$

Define truncated random variables $\hat{\zeta}_i^{(n)}$, $i = 1, \ldots, n-k$ as i.i.d random variables with distribution

$$\mathbb{P}(\hat{\zeta}_1^{(n)} = j) = \mathbb{P}\left(\hat{\zeta}_1 = j \middle| \hat{\zeta}_1 \leq n^\beta \right).$$

Then, we can write:

$$\mathbb{P}(\max_i \xi_i' \leq n^\beta) = \left[\mathbb{P}(\hat{\zeta}_1 \leq n^\beta) \right]^{n-k} \frac{\mathbb{P}\left(S_{\hat{\zeta}^{(n)}}(n-k) = n-1 \right)}{\mathbb{P}\left(S_{\hat{\zeta}}(n-k) = n-1 \right)}. \tag{16}$$

By the assumption on the offspring probabilities, $\widehat{\zeta}_1$ has exponentially declining tails and therefore, for some $a > 0$, we can write:

$$(n-k) \log \mathbb{P}(\hat{\zeta}_1 \leq n^\beta) \geq (n-k) \log(1 - e^{-an^\beta}) \geq -2n e^{-an^\beta},$$

where the last inequality follows because by a choice of a we can ensure $e^{-an^\beta} < 3/4$ and we have $\log(1-x) \geq -2x$ for $0 \leq x \leq 3/4$. Then,

$$\left[\mathbb{P}(\hat{\zeta}_1 \leq n^\beta) \right]^{n-k} \geq e^{-2n e^{-an^\beta}} \geq 1 - 2n e^{-an^\beta}.$$

This implies that for sufficiently small $\gamma > 0$, $\left[\mathbb{P}(\hat{\zeta}_1 \leq n^\beta) \right]^{n-k} \geq 1 - e^{-\gamma n^\beta}$.

To estimate the denominator in the ratio of probabilities in (16), we can use the local limit theorem (Theorem I.4.2 in Kolchin [17]). For the numerator, this theorem is not applicable directly, since the distribution of random variables $\hat{\zeta}_i^{(n)}$ is changing with n. However, it can be adapted to this more general setting along the lines in the proof of Theorem I.4.2 in Kolchin. (We provide the proof of the modified local limit theorem in "Appendix".)

Then, we find that, for $r > 0$,

$$\frac{\mathbb{P}\left(S_{\hat{\zeta}^{(n)}}(n-k) = n-1 \right)}{\mathbb{P}\left(S_{\hat{\zeta}}(n-k) = n-1 \right)} \to 1$$

as $n \to \infty$. It follows that for some smaller $\gamma > 0$, $\mathbb{P}(\max_i \xi_i' \leq n^\beta) \geq 1 - e^{-\gamma n^\beta}$. In a similar fashion, we can show that $\mathbb{P}(\min_i \xi_i' \geq -n^\beta) \geq 1 - e^{-\gamma n^\beta}$. The sequence $\xi_{(1)}'', \ldots, \xi_{(n)}''$ is different from the sequence $\xi_{(1)}', \ldots, \xi_{(n-k)}'$ only by addition of k zeros. It follows that some $\gamma > 0$, $\mathbb{P}(\max_i |\xi_i''| \leq n^\beta) \geq 1 - e^{-\gamma n^\beta}$, which implies the claim of the lemma. □

Lemma 2.6 *As $n \to \infty$,*

$$\frac{1}{n} \sum_{i=1}^{n} \left(\xi_i'' - 1 + \frac{1}{n} \right)^2 \xrightarrow{\mathbb{P}} (\sigma^*)^2.$$

Proof We have $\mathbb{E}\xi_i'' = 1 - 1/n$ and $\mathbb{E}\big[(\xi_i'' - 1 + \frac{1}{n})^2\big] \to (\sigma^*)^2$ by Lemma 2.5(a). Then,

$$\mathbb{Var}\left(\frac{1}{n} \sum_{i=1}^{n} \left(\xi_i'' - 1 + \frac{1}{n} \right)^2 \right) = \frac{1}{n^2} \Big[\sum_{i=1}^{n} \mathbb{Var}\left(\left(\xi_i'' - 1 + \frac{1}{n} \right)^2 \right)$$

$$+ 2 \sum_{i<j} \mathrm{Cov}\left(\left(\xi_i'' - 1 + \frac{1}{n} \right)^2, \left(\xi_j'' - 1 + \frac{1}{n} \right)^2 \right) \Big]$$

$$= \frac{1}{n^2} \Big[n \mathbb{Var}\left(\left(\xi_1'' - 1 + \frac{1}{n} \right)^2 \right) + n(n-1) \mathrm{Cov}\left(\left(\xi_1'' - 1 + \frac{1}{n} \right)^2, \left(\xi_2'' - 1 + \frac{1}{n} \right)^2 \right) \Big]$$

$$\to 0,$$

by Lemma 2.5(a) and (b). By Chebyshev's inequality, this implies the statement of the lemma. □

Let $S_n''(0) = 0$ and

$$S_n''(j) = \sum_{i=1}^{j} \xi_i'', \text{ where } j = 1, \dots, n,$$

and extend $S_n''(t)$ to all $t \in [0, n]$ by linear interpolation.

In the lemmas below, the convergence of stochastic processes is the weak convergence with respect to the space of continuous functions on [0, 1] with the topology induced by the uniform norm: $\|f\|_\infty = \sup_{t \in [0,1]} f(t)$.

Lemma 2.7 *As $n \to \infty$*

$$\left. \frac{S_n''(tn) - tn}{\sigma^* \sqrt{n}} \right|_{t \in [0,1]} \xrightarrow{weakly} b(t) \Big|_{t \in [0,1]},$$

where $b(t)$ is the standard Brownian bridge on the interval [0, 1], and σ^* is as defined in (5).

Proof Define $x_j = (\xi_j'' - 1 + 1/n)/(\sqrt{n}\sigma^*)$ and $X_n(t) = \sum_{j=1}^{[tn]} x_j$, with $X_n(t) = 0$ for $0 \le t < 1/n$. Then, $\sum_{j=1}^{n} x_j = 0$ and $\sum_{j=1}^{n} x_j^2 \xrightarrow{\mathbb{P}} 1$ by Lemma 2.6. In addition, by Lemma 2.5(c), $\max_j |x_j| \xrightarrow{\mathbb{P}} 0$. This means that the assumptions of Theorem 24.2 in Billingsley's book [4] are satisfied, and we have $X_n(t) \xrightarrow{weakly} b(t)$. This implies the claim of the lemma. □

Next, we consider the application of Step 3 in Algorithm A. By Theorem 1.9, this step produces the Łukasiewicz path of a random tree in $\Omega_{k,n}$.

Lemma 2.8 *As $n \to \infty$, the normalized Łukasiewicz path $S_n(tn)/(\sigma^*\sqrt{n})$ of a random tree from $\Omega_{k,n}$ weakly converges to a standard Brownian excursion $e(t)$.*

Proof Let $b_n''(t) = (S_n''(tn) - tn)/(\sigma^*\sqrt{n})$. By the previous lemma, $b_n''(t)$ converges to $b(t)$ in distribution. According to Vervaat's Theorem (Theorem 1 in [31]), the location of the absolute minimum of Brownian bridge $b(t)$ is unique with probability 1, and, if this time is denoted τ_m, the process $b(t + \tau_m) - b(\tau_m)$ is distributed as a Brownian excursion $e(t)$. (Here, the addition in the argument is modulo 1.)

Since $b_n''(t)$ converges to the Brownian bridge $b(t)$ in distribution, we can couple the random variables $b_n''(t)$ and $b(t)$, so that the time at which $b_n''(t)$ achieves its first minimum, τ_m'', converges in probability to τ_m. Then,

$$b_n''(t + \tau_m'') - b_n''(\tau_m'')\Big|_{t \in [0,1]} \xrightarrow{dist} b(t + \tau_m) - b(\tau_m)\Big|_{t \in [0,1]},$$

which is distributed as a Brownian excursion by Vervaat's theorem.

If we write $S_n(j) = \sum_{i=1}^{j} (\widehat{\xi}_i'' - 1)$, where $(\widehat{\xi}_i'')$ is the sequence (ξ_i'') after the cyclic shift, then the process $b_n''(t + \tau_m'') - b_n''(\tau_m'')$ is equal to $S_n(t)/(\sigma^*\sqrt{n})$. By Theorem 1.9, $S_n(t)$ is distributed as the Łukasiewicz path of a random tree from $\Omega_{k,n}$.

Hence, we have shown that the normalized Łukasiewicz path $S_n(tn)/(\sigma^*\sqrt{n})$ of a random tree from $\Omega_{k,n}$ converges to a standard Brownian excursion $e(t)$. $\qquad\square$

Step 2 Łukasiewicz path is close to the height process.

The height process for an ordered tree is defined in (8). In Lemma 1 of Marckert and Mokkadem [24], it was shown that if $S_n(j)$ is the Łukasiewicz path of an ordered tree, then the height process $H_n(l)$, $l = 0, \ldots, n-1$, can be written as follows:

$$H_n(l) = \#\left\{ i : 0 \le i \le l - 1, S_n(i) = \min_{i \le k \le l} \{S_n(k)\} \right\} \tag{17}$$

In words, $H_n(l)$ is the number of (weak) right minima of S_n on the interval $[0, l]$ (not counting the trivial right minimum at $i = l$).

For example, in Fig. 1, $H_8(3) = 2$, and this corresponds to the right minima at $i = 0$ and $i = 2$ of the Łukasiewicz path $S_8(i)$ on interval $[0, 3]$.

Now, let $W_n(j)$, $j = 0, \ldots, n$, be a path of a random walk with independent increments x_i that can take values $s = -1, 0, 1, \ldots$ with the following probabilities:

$$\mathbb{P}(x_i = s) = w_{s+1}^*, \tag{18}$$

where $\{w_k^*\}_{k \ge 0}$ is the α-shifted weight distribution from Definition 1.1. For a path of this random walk, we can also define the number of weak right minima on the interval $[0, l]$, $l = 0, \ldots, n-1$:

$$RM_n(l) = \#\left\{ i : 0 \le i \le l - 1, W_n(i) = \min_{i \le k \le l} \{W_n(k)\} \right\} \tag{19}$$

The key result about the process $RM_n(l)$ was proved in [24].

Lemma 2.9 (Corollary 5 in [24]) *Let $W_n(j)$, $j = 0, \ldots, n$, be a path of a random walk with independent increments that have the law as in (18), and let $RM_n(l)$ be as defined in (19). Then, for some $\gamma > 0$ and $v > 0$, we have*

$$\mathbb{P}\left(\sup_{0 \leq l < n} \left| W_n(l) - \min_{0 \leq i \leq l} W_n(i) - RM_n(l)\frac{(\sigma^*)^2}{2} \right| \geq n^{1/4+v} \right) \leq e^{-\gamma n^v}. \quad (20)$$

Note that if $W_n(j)$ happens to be an excursion, then $\min_{0 \leq i \leq l} W_n(i) = 0$ and in this case Lemma 2.9 says that $RM_n(l)$ after scaling by $(\sigma^*)^2/2$ is close to the excursion $W_n(l)$.

The proof of this result in [24] is based on a reduction in the problem to a problem about ladder heights of a random walk, for which appropriate tools are available in Feller's book ([13], Chapter XII).

By using this result, we can complete the second step of the proof.

Lemma 2.10 *Let $S_n(l)$ and $H_n(l)$ be the Łukasiewicz and the height processes for a random tree in $\Omega_{k,n}$. For any $v > 0$, there exists constants $\gamma > 0$ and $N > 0$ such that for all $n \geq N$ and $(n, k) \in PP(w)$,*

$$\mathbb{P}\left(\sup_{0 \leq l < n} \left| S_n(l) - \frac{(\sigma^*)^2}{2} H_n(l) \right| \geq n^{1/4+v} \right) \leq e^{-\gamma n^v}. \quad (21)$$

Proof From (17) and (19), it can be seen that the probability of the event in (21) equals the probability of the event in (20), denoted as $\mathfrak{A}_n$, conditional on the event $\mathfrak{B}_n$ that $W_n(i)$, $i = 0, \ldots, n$ is an excursion with exactly k steps equal to -1. We have

$$\mathbb{P}(\mathfrak{A}_n | \mathfrak{B}_n) \leq \frac{\mathbb{P}(\mathfrak{A}_n)}{\mathbb{P}(\mathfrak{B}_n)} \leq \frac{e^{-\gamma n^v}}{\mathbb{P}(\mathfrak{B}_n)}.$$

It follows that it is enough to show that $\mathbb{P}(\mathfrak{B}_n)$ is sufficiently large. Let us introduce two other events: $\mathfrak{C}_n$ is the event that $W_n(i)$ has exactly k steps equal to -1 and $\mathfrak{E}_n$ is the event that $W_n(i)$ is an excursion. Then, $\mathfrak{B}_n = \mathfrak{C}_n \cap \mathfrak{E}_n$ and

$$\mathbb{P}(\mathfrak{B}_n) = \mathbb{P}(\mathfrak{E}_n | \mathfrak{C}_n)\mathbb{P}(\mathfrak{C}_n).$$

The probability of the increment -1 in the random walk W_n equals to $w_0^* = \alpha$, so by the local central limit theorem the probability of the event $\mathfrak{C}_n$ is asymptotically equal to $a/\sqrt{n}$ with $a > 0$.

In order to calculate $\mathbb{P}(\mathfrak{E}_n | \mathfrak{C}_n)$, we note that the distribution of the increments conditional that they are non-negative is given by

$$\mathbb{P}(x_j = s | x_j \geq 0) = \frac{w_{s+1}^*}{1 - \alpha}, \text{ for } s = 0, 1, 2, \ldots$$

The conditional expectation of the increment equals

$$\mathbb{E}(x_j | x_j \geq 0) = \frac{1}{1-\alpha} \sum_{s=0}^{\infty} s w_{s+1}^* = \frac{1}{1-\alpha} \sum_{k=1}^{\infty} (k-1) w_k^*$$

$$= \frac{1}{1-\alpha} \left(\sum_{k=1}^{\infty} k w_k^* - \sum_{k=1}^{\infty} w_k^* \right) = \frac{1}{1-\alpha} \left(1 - (1-\alpha) \right)$$

$$= \frac{\alpha}{1-\alpha}.$$

Then, the expectation of the sum $W_n(n)$ conditional that exactly k of the increments are equal to -1 is

$$\mathbb{E}(W_n(n) | \mathfrak{C}_n) = -k + \mathbb{E}(x_j | x_j \geq 0)(n-k)$$

$$= -\alpha n + \frac{\alpha}{1-\alpha}(1-\alpha)n + O(1) = O(1).$$

Again, by using the local central limit theorem, we find that the probability that W_n is a random walk bridge, conditional on the event $\mathfrak{C}_n$, is asymptotically equal to $b/\sqrt{n}$ with $b > 0$. Then, all n cyclic shifts of this random walk bridge have the same probability by exchangeability of the increments and exactly one of these shifts is an excursion by the cyclic lemma. Hence, conditional on $\mathfrak{C}_n$, the random walk W_n is an excursion with probability $\sim b/n^{3/2}$. Altogether we conclude that

$$\mathbb{P}(\mathfrak{B}_n) \sim \frac{ab}{n^{5/2}},$$

with $ab > 0$. Hence,

$$\mathbb{P}(\mathfrak{A}_n | \mathfrak{B}_n) \leq \frac{1}{ab} n^{5/2} e^{-\gamma n^{\nu}} \leq e^{-\gamma' n^{\nu}},$$

for a suitably chosen γ' and this completes the proof of the lemma. □

Step 3 The height process is close to the contour process

Here, we introduce a couple of purely combinatorial results from Marckert and Mokkadem [24] that relate the height and contour processes.

Definition 2.11 Let $T \in \Omega_n$ be a rooted planar tree on n vertices. For every $l = 0, \ldots, n-1$, define $m(l)$ as the first time when the Depth First Search reaches the vertex v_l.

It is clear that the sequence $\left(m(l) \right)_{l=0}^{n-1}$ is strictly increasing.
In the example in Fig. 1, we have:

1	0	1	2	3	4	5	6	7
m(l)	0	1	3	4	7	8	9	12

By definition of the contour process $C_n(t)$, it is easy to see that for all l, the points $(m(l), H_n(l))$ are located on the graph of the contour process. Note also that for all $k \in [m(l), m(l+1))$,

$$H_n(l+1) - 1 \leq C_n(k) \leq H_n(l). \tag{22}$$

(see Lemma 3 in [24]). Intuitively, $(m(l), H_n(l))$ interpolate the contour process, and $l \mapsto m(l)$ can be thought as a time change that connects the contour process to the height process. See the illustration in Fig. 1.

Another important combinatorial result proved in [24] is that for every $l \in \{0, \ldots, n-1\}$,

$$m(l) = 2l - H_n(l). \tag{23}$$

Since for large l, we can expect that $H_n(l)$ has the order $\sqrt{l}$ with large probability, this identity suggests that the time change function $m(l) \sim 2l$ with large probability, and that the deviations of $m(l)$ from $2l$ are of order $\sqrt{l}$.

Now, let us define $\tau(nt)$ as the integer such that

$$\lfloor 2nt \rfloor \in \left[m(\tau(nt)), m(\tau(nt) + 1) \right). \tag{24}$$

(Intuitively, this function plays the role of the inverse of $m(l)$ scaled by 2.)

Then, we have the following result.

Lemma 2.12 *For all $a > 0$ and $v > 0$, there exists such $\gamma > 0$ that*

$$\mathbb{P}\left(\sup_{t \in [0,1]} \left| H_n(\tau(nt)) - C_n(2nt) \right| \geq a n^{1/4+v} \right) \leq e^{-\gamma n^v}$$

Proof By (22), we have

$$\mathbb{P}\left(\sup_{t \in [0,1]} \left| H_n(\tau(nt)) - C_n(2nt) \right| \geq a n^{1/4+v} \right)$$

$$\leq \mathbb{P}\left(\sup_{l \in \{0,\ldots,n-2\}} \left| H_n(l) - H_n(l+1) + 1 \right| \geq a n^{1/4+v} \right)$$

$$\leq \mathbb{P}\left(\sup_{l \in \{0,\ldots,n-2\}} \left| H_n(l) - \frac{2}{(\sigma^*)^2} S_n(l) \right| \geq \frac{a}{3} n^{1/4+v} \right)$$

$$+ \mathbb{P}\left(\sup_{l \in \{0,\ldots,n-2\}} \left| H_n(l+1) - 1 - \frac{2}{(\sigma^*)^2} S_n(l+1) \right| \geq \frac{a}{3} n^{1/4+v} \right),$$

$$+ \mathbb{P}\left(\sup_{l \in \{0,\ldots,n-2\}} \frac{2}{(\sigma^*)^2} \left| S_n(l+1) - S_n(l) \right| \geq \frac{a}{3} n^{1/4+v} \right),$$

and by Lemmas 2.5(c) and 2.10, the sum of these three probabilities can be bounded by $e^{-\gamma n^v}$ with a suitable $\gamma > 0$. $\qquad \square$

By using Lemmas 2.10 and 2.12, we can relate the Łukasiewicz process $S(\tau(nt))$ and the contour process $C_n(nt)$. To move further, we need to estimate the difference between $\tau(nt)$ and nt.

Lemma 2.13 *Let $\tau(nt)$ be as in* (24) *for the random time $m(l)$ defined for a random tree in $\Omega_{k,n}$. Then,*

$$\mathbb{P}\left(\sup_{t \in [0,1]} \left| \tau(nt) - nt \right| > n^{1/2+\nu} \right) \leq e^{-\gamma n^\nu}.$$

Proof By the triangle inequality, we have

$$\sup_{t \in [0,1]} \left| 2\tau(nt) - 2nt \right| \leq \sup_{t \in [0,1]} \left| 2\tau(nt) - m(\tau(nt)) \right| + \sup_{t \in [0,1]} \left| m(\tau(nt)) - 2nt \right|.$$

By applying Equation (23) with $l = \tau(nt)$, we find that the first term is bounded by $\sup_t H_n(\tau(nt))$. For the second term, we note that by definition of $\tau(nt)$, it is bounded by

$$\sup_{t \in [0,1]} \left| m(\tau(nt) + 1) - m(\tau(nt)) \right| = \sup_{l \in \{0,\ldots,n-1\}} \left| m(l+1) - m(l) \right|$$

$$\leq \sup_{l \in \{0,\ldots,n-1\}} H_n(l).$$

Altogether, we have

$$\sup_{t \in [0,1]} \left| 2\tau(nt) - 2nt \right| \leq 2 \sup_{l \in \{0,\ldots,n-1\}} H_n(l).$$

By (17), $H_n(l)$ is the number of right minima for S_n on the interval $[0, l]$. It was shown in Lemma 5 in [24] that the number of right minima for the random walk W_n satisfies the bound

$$\mathbb{P}\left(\sup_{l \in \{0,\ldots,n-1\}} RM_n(l) \geq n^{1/2+\nu} \right) \leq e^{-\gamma n^\nu}$$

for a $\gamma > 0$. An inequality of this form also holds for the number of right minima of S_n by a conditioning argument similar to the argument that we used in the proof of Lemma 2.10.

So we have, for some $\gamma > 0$,

$$\mathbb{P}\left(\sup_{t \in [0,1]} \left| \tau(nt) - nt \right| \leq n^{1/2+\nu} \right) \geq \mathbb{P}\left(2 \sup_{l \in \{0,\ldots,n-1\}} H_n(l) \leq n^{1/2+\nu} \right) \geq 1 - e^{-\gamma n^\nu}$$

and

$$\mathbb{P}\left(\sup_{t \in [0,1]} \left| \tau(nt) - nt \right| > n^{1/2+\nu} \right) \leq e^{-\gamma n^\nu}.$$

$\square$

Lemma 2.14 *Let $S_n(t)$ and $C_n(t)$ be the Łukasiewicz and the contour processes for a tree $T \in \Omega_{k,n}$. Then, for every $v > 0$, there exist two positive constants, $\gamma > 0$ and $N > 0$, such that for all $n \geq N$,*

$$\mathbb{P}\left[\sup_{t \in [0,1]} \left| S_n(nt) - \frac{\sigma^2}{2} C_n(2nt) \right| \geq n^{1/4+v} \right] \leq e^{-\gamma n^v}.$$

Proof We have

$$\mathbb{P}\left(\sup_{t \in [0,1]} \left| S_n(nt) - \frac{\sigma^2}{2} C_n(2nt) \right| \geq n^{1/4+v} \right) \leq A + B + C + D,$$

where

$$A = \mathbb{P}\left(\sup_{t \in [0,1]} \left| S_n(nt) - S_n\big(\tau(nt)\big) \right| \geq \frac{1}{3} n^{1/4+v}, \sup_t \left| \tau(nt) - nt \right| < n^{1/2+v} \right),$$

$$B = \mathbb{P}\left(\sup_{t \in [0,1]} \left| S_n\big(\tau(nt)\big) - \frac{\sigma^2}{2} H_n\big(\tau(nt)\big) \right| \geq \frac{1}{3} n^{1/4+v} \right),$$

$$C = \mathbb{P}\left(\frac{\sigma^2}{2} \sup_{t \in [0,1]} \left| H_n\big(\tau(nt)\big) - C_n(2nt) \right| \geq \frac{1}{3} n^{1/4+v} \right),$$

$$D = \mathbb{P}\left(\sup_{t \in [0,1]} \left| \tau(nt) - nt \right| > n^{1/2+v} \right).$$

The quantities B, C, and D can be estimated using Lemmas 2.10, 2.12, and 2.13, respectively. For quantity A, note that we can estimate a similar quantity for unconditional random walk W_n:

$$\mathbb{P}\left(\sup_{t \in [0,1]} \left| W_n(nt) - W_n\big(\tau(nt)\big) \right| \geq \frac{1}{3} n^{1/4+v}, \sup_t \left| \tau(nt) - nt \right| < n^{1/2+v} \right)$$

$$\leq \mathbb{P}\left(\sup_{k \leq n^{1/2+v}} \sup_{0 \leq j \leq n-k} \left| W_n(j) - W_n(j+k) \right| \geq \frac{1}{3} n^{1/4+v} \right)$$

$$\leq \sum_{j=1}^n \mathbb{P}\left(\sup_{k \leq n^{1/2+v}} \left| W_n(k) \right| \geq \frac{1}{3} n^{1/4+v} \right) \leq e^{-\gamma n^v}.$$

The last step holds by Theorem III.15 from Petrov's book [26] (see also Appendix in [24]), and here, we use the assumption that $\mathbb{E} e^{aL} < \infty$ for some $a > 0$.

This estimate remains valid for the process S_n instead of W_n, perhaps with a different $\gamma > 0$, by a conditioning argument similar to the argument that was used in the proof of Lemma 2.10. Therefore, we can find such $\gamma > 0$, that $A + B + C + D \leq e^{-\gamma n^v}$. □

Now, we can conclude the proof of Theorem 1.3. In Lemma 2.8, we proved that the normalized Łukasiewicz path $S_n(tn)/(\sigma^* \sqrt{n})$ of a random tree from $\Omega_{k,n}$ converges

to a standard Brownian excursion $e(t)$. Then, Lemma 2.14 implies that the process $\sigma^* C_n(2tn)/(2\sqrt{n})$ also converges to the excursion $e(t)$. This concludes the proof. $\square$

For the proof of Theorem 1.5, we need a lemma, which allows us to compare the probabilities of various events for unconditional random walks and corresponding random bridges. As in the proof of Theorem 1.9, let $\hat{\zeta}_1, \ldots, \hat{\zeta}_{n-k}$ be an i.i.d sequence with $\hat{\zeta}_1$ having distribution $\widehat{w}_0^* = 0$, $\hat{w}_i^* = w_i^*/(1-\alpha)$, $i = 1, \ldots$. Let

$$m := \lfloor (n-k)/2 \rfloor,$$

and let $\mathcal{F}_n^{(1/2)}$ denote the sigma-algebra generated by random variables $\hat{\zeta}_1, \ldots, \hat{\zeta}_m$.

Lemma 2.15 *Suppose $(n, k_n) \in PP(w)$, $k_n = \alpha n + O(1)$, and the distribution $\{w_i\}$ has finite variance. For every event $\mathfrak{B} \in \mathcal{F}_n^{(1/2)}$,*

$$\mathbb{P}(\mathfrak{B}|\hat{\zeta}_1 + \ldots + \hat{\zeta}_{n-k} = n - 1) \le c\mathbb{P}(\mathfrak{B}),$$

where c is a positive constant that depends on $\{w_i\}$ but not on n.

Proof For the joint probability, we have

$$\mathbb{P}(\mathfrak{B}, S_{\hat{\zeta}}(n-k) = n-1) = \sum_{x=0}^{\infty} \mathbb{P}\Big(\mathfrak{B}, S_{\hat{\zeta}}(m) = x\Big)$$
$$\times \mathbb{P}\Big(S_{\hat{\zeta}}(n-k-m) = n-1-x\Big)$$

Hence, for the conditional probability,

$$\mathbb{P}(\mathfrak{B}|S_{\hat{\zeta}}(n-k) = n-1) = \sum_{x=0}^{\infty} \mathbb{P}\Big(\mathfrak{B}, S_{\hat{\zeta}}(m) = x\Big)$$
$$\times \frac{\mathbb{P}\Big(S_{\hat{\zeta}}(n-k-m) = n-1-x\Big)}{\mathbb{P}(S_{\hat{\zeta}}(n-k) = n-1)}.$$

By using the local limit law (Theorem I.4.2 in Kolchin [17]), we have $\mathbb{P}(S_{\hat{\zeta}}(n-k) = n-1) \ge c_1/\sqrt{n}$ for some $c_1 > 0$. By Theorem 2.21 on p.76 in Petrov [27], $\sup_x \mathbb{P}\Big(S_{\hat{\zeta}}(n-k-m) = n-1-x\Big) \le c_2/\sqrt{n}$ for some $c_2 > 0$. Altogether, we get

$$\mathbb{P}(\mathfrak{B}|S_{\hat{\zeta}}(n-k) = n-1) \le \frac{c_2}{c_1} \sum_{x=0}^{\infty} \mathbb{P}\Big(\mathfrak{B}, S_{\hat{\zeta}}(m) = x\Big) = \frac{c_2}{c_1}\mathbb{P}(\mathfrak{B}).$$

$\square$

Proof of Theorem 1.5 Let

$$n_\xi(i) = \Big|\{j : \xi_j = i, 1 \le j \le n\}\Big|,$$

$$n_{\xi'}(i) = \Big|\{j : \xi'_j = i, 1 \le j \le n - k\}\Big|,$$

$$n_{\hat{\zeta}}(i) = \Big|\{j : \hat{\zeta}_j = i, 1 \le j \le n - k\}\Big|,$$

where sequences (ξ_j), (ξ'_j) and $(\hat{\zeta}_j)$ are as defined in the definition of Algorithm A and in the proof of Theorem 1.9. By Theorem 1.9, the random variable $n_i(T)$ has the same distribution as $n_\xi(i)$. For $i = 0$, we have $n_\xi(0) = k$ and $k/(n-1) \to \alpha$ by assumption, so $n_0(T)/(n-1) \to \alpha = w_0^*$. For $i > 0$, we can see from the algorithm that it is enough to show that $n_{\xi'}(i)/(n-k) \to \widehat{w}_i^* = w_i^*/(1-\alpha)$ almost surely.

Let us define additionally,

$$n_{\hat{\zeta}}^{(1/2)}(i) = \Big|\Big\{j : \hat{\zeta}_j = i, 1 \le j \le \frac{n-k}{2}\Big\}\Big|,$$

$$n_{\xi'}^{(1/2)}(i) = \Big|\Big\{j : \xi'_j = i, 1 \le j \le \frac{n-k}{2}\Big\}\Big|.$$

As before, for shortness of formulas, we define $m := \lfloor \frac{n-k}{2} \rfloor$.

Then,

$$\frac{n_{\hat{\zeta}}^{(1/2)}(i)}{m} = \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}[\hat{\zeta}_j = i],$$

which implies $\mathbb{E}\Big(\frac{n_{\hat{\zeta}}^{(1/2)}(i)}{m}\Big) = \widehat{w}_i^*$. Then, by using the independence of random variables $\hat{\zeta}_j$,

$$\mathbb{E}\Big|\frac{n_{\hat{\zeta}}^{(1/2)}(i)}{m} - \widehat{w}_i^*\Big|^4 = \frac{1}{m^4}\mathbb{E}\Big(\sum_{j=1}^{m}(\mathbb{1}[\hat{\zeta}_j = i] - \widehat{w}_j^*)\Big)^4 = O(m^{-2}).$$

Then, by Markov's inequality:

$$\mathbb{P}\Big[\Big|\frac{n_{\hat{\zeta}}^{(1/2)}(i)}{m} - \widehat{w}_i^*\Big| > \varepsilon\Big] = O\Big(\frac{1}{\varepsilon^4 m^2}\Big)$$

By using Lemma 2.15, we infer that this estimate holds also for the conditioned variables ξ'_j:

$$\mathbb{P}\Big[\Big|\frac{n_{\xi'}^{(1/2)}(i)}{m} - \widehat{w}_i^*\Big| > \varepsilon\Big] = O\Big(\frac{1}{\varepsilon^4 m^2}\Big). \tag{25}$$

Let

$$n_{\xi'}^{(2/2)}(i) = \Big|\Big\{j : \xi'_j = i, m < j \le n - k\Big\}\Big|.$$

Since $\{\xi'_j\}_{j=1}^n$ is an exchangeable sequence, we have

$$\mathbb{P}\left[\left|\frac{n_{\xi'}^{(2/2)}(i)}{m} - \widehat{w}_i^*\right| > \varepsilon\right] = O\left(\frac{1}{\varepsilon^4 m^2}\right). \tag{26}$$

Together, Eqs. (25) and (26) imply that

$$\mathbb{P}\left[\left|\frac{n_{\xi'}(i)}{n-k} - \widehat{w}_i^*\right| > \varepsilon\right] = O\left(\frac{1}{\varepsilon^4((n-k)^2)}\right). \tag{27}$$

So, by Borel–Cantelli lemma, as $n \to \infty$,

$$\frac{n_{\xi'}(i)}{n-k} \to \widehat{w}_i^*,$$

almost surely. This implies the statement of the theorem. □

3 Conclusion and Remaining Questions

In this paper, we investigated the scaling limit of Galton–Watson trees conditioned on a fixed number of vertices n and leaves k. We considered the case when both k and n grow to infinity and $k = \alpha n + O(1)$, with $\alpha \in (0, 1 - 1/\widehat{v})$. The offspring distribution is assumed to have the exponentially declining tails ($\mathbb{E}e^{aL} < \infty$ for some $a > 0$).

In this case, we found that the limit under rescaling is the Aldous continuum random tree and calculated the relevant parameter. We also showed that the height of a tree in $\Omega_{k,n}$ converges under rescaling to the distribution of the maximum of the Brownian excursion. Finally, we showed that the normalized degree sequence of the tree converges almost surely to the α-shifted distribution.

Some other questions remain open. First, what happens when the assumption of exponentially declining tails is relaxed? Similar to a comment in Marckert and Mokkadem [24], we can argue that some number of finite moments and the assumption $\mathbb{E}L = 1$ should be sufficient to ensure that our convergence results stay intact. However, the infinite variance of the offspring distribution or power tails with $\mathbb{E}L < 1$ will likely lead to considerably different results.

Another question is what happens in the intermediate regime when both k and n grow but $k/n \sim n^{-\beta}$ or $\left(k/n - (1 - \widehat{v})\right) \sim n^{-\beta}$ with $0 < \beta < 1$. The first case is when the tree has very small but still growing number of leaves. The second case is about the regime when the number of leaves in the tree is close to the maximal capacity that the offspring distribution can sustain. The results in the paper of Labarbe and Marckert [20] about Bernoulli random walks suggest that even in this case the limit is Aldous's Brownian continuum random tree, although the scaling factor may be different from $n^{1/2}$.

Besides the scaling limit in the Gromov–Hausdorff topology, it could also be interesting to study the characteristics of the conditioned trees in these situations, such as their limiting degree sequences.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

Appendix A. Proof of a Local Limit Theorem

First, recall some basic definitions. A random variable X has a *lattice distribution* if, with probability 1, it takes values in a set $\{a + kh\}$ where $k \in \mathbb{Z}$ and a and $h > 0$ are fixed numbers. Number h is called a *step* of the distribution. A step h is called the maximal step if there is no a_1 and $h_1 > h$ so that all values of X are in the set $\{a_1 + kh_1\}$.

For every integer random variable ξ, we use $\xi^{(A)}$ to denote a truncated version of ξ; that is, $\xi^{(A)}$ has the distribution

$$\mathbb{P}(\xi^{(A)} = j) = \mathbb{P}(\xi = j | \xi \leq A).$$

Theorem A.1 *Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d non-negative integer random variables. Assume that for some $\alpha > 0$, $\mathbb{E}e^{\alpha \xi_1} < \infty$, and let $\mathbb{E}\xi_1 = a$, $\mathrm{Var}(\xi_1) = \sigma^2$. Suppose that the maximal step of the distribution of ξ_1 equals 1. Let $A_1, A_2, \ldots$ be a sequence of positive numbers such that $A_N = \Omega(N^\varepsilon)$ for some $\varepsilon > 0$. Then, as $N \to \infty$ and uniformly over integer n,*

$$\sigma \sqrt{N} \mathbb{P}(\xi_1^{(A_N)} + \ldots + \xi_N^{(A_N)} = n) - \frac{1}{\sqrt{2\pi}} e^{-\frac{(n - aN)^2}{2\sigma^2 N}} \to 0.$$

Before starting the proof of the theorem, let us derive several useful lemmas.

Lemma A.2 *Suppose an integer-valued random variable $\xi \geq 0$ has maximal step 1 Then, there exists an A_0 such that for all $A > A_0$, $\xi^{(A)}$ has maximal step 1.*

Proof Let $x_0 < x_1 < \ldots$ be the values of ξ and let $Y \in \mathbb{Z}^+$ be the set of differences $x_i - x_{i-1}$ for all $i = 1, 2, \ldots$. Then, the maximal step $h = \gcd(Y)$. Indeed, if h is a step, then all elements in Y are divisible by h, so h is a divisor of $\gcd(Y)$. Conversely, every x_j can be written as $x_0 + \sum_{i=1}(x_i - x_{i-1})$ and therefore, belongs to the set $x_0 + k \gcd(Y)$, so $\gcd(Y)$ is a step.

By our assumption, $gcd(Y) = 1$. Order elements of Y as $y_1, y_2, \ldots$. Then, for some finite k, $\gcd(y_1, \ldots, y_k) = 1$. Take an m such that all y_i with $1 \leq i \leq k$ belong to the set of differences $x_j - x_{j-1}$ with $1 \leq j \leq m$. Take $A_0 = x_m$. Then, for every random variable $\xi^{(A)}$ with $A \geq A_0 = x_m$, the corresponding set of differences $Y^{(A)}$ contains

all numbers $y_1, \ldots, y_k$ and therefore $gcd(Y^{(A)}) = 1$. This implies that the maximal step of $\xi^{(A)}$ equals 1. $\qquad\square$

Now, let $\xi_i, A_N, \xi_i^{(A_N)}$ be as assumed in the statement of Theorem A.1. In addition, let $a_N := \mathbb{E}\xi_1^{(A_N)}$, $\sigma_N^2 := \mathrm{Var}(\xi_1^{(A_N)})$, $m_3 := \mathbb{E}(|\xi_1 - a|^3)$, and $m_{3,N} := \mathbb{E}(|\xi_1^{(A_N)} - a_N|^3)$.

Lemma A.3 *Under assumptions on ξ_i, A_N, as in Theorem A.1, for $N \to \infty$, we have:*

(a) $(a_N - a)N^r \to 0$, *for every $r > 0$,*
(b) $\sigma_N - \sigma \to 0$,
(c) $m_{3,N} - m_3 \to 0$.

Proof Recall that for $A_N \to \infty$, and $\alpha > 0$,

$$\int_{A_N}^{\infty} x e^{-\alpha x} \, \mathrm{d}x = O(A_N e^{-\alpha A_N}).$$

Then,

$$N^r \mathbb{E}|\xi_1 - \xi_1 \mathbf{1}(X \le A_N)| = N^r \sum_{j > A_N} j P(\xi_1 = j) = N^r O\left(\sum_{j > A_N} j e^{-\alpha j} \right),$$

and by comparison with the integral, we find that

$$N^r \mathbb{E}|\xi_1 - \xi_1 \mathbf{1}(X \le A_N)| = O(N^r A_N e^{-\alpha A_N}) \to 0, \tag{28}$$

as $N \to \infty$, under assumption that $A_N = \Omega(N^\varepsilon)$.

Then,

$$N^r \mathbb{E}|\xi_1 \mathbf{1}(X \le A_N) - \xi_1^{(A_N)}| = N^r \sum_{j \le A_N} j \mathbb{P}(\xi = j) \left| 1 - \frac{1}{\sum_{j \le A_N} \mathbb{P}(\xi = j)} \right|$$
$$= N^r O\left(\sum_{j > A_N} \mathbb{P}(\xi = j) \right) = N^r O(e^{-\alpha N}) \to 0, \tag{29}$$

as $N \to \infty$.

From (28) and (29) by the triangle inequality we find that

$$N^r \mathbb{E}|\xi_1 - \xi_1^{(A_N)}| \to 0,$$

which implies the first claim of the lemma. Clearly, the second and the third claims can be proved by similar arguments. $\qquad\square$

Finally, we need an estimate on characteristic functions. Let $f_N(t)$ be a characteristic function of $\xi_1^{(A_N)}$ and let

$$\tilde{f}_N(t) = f_N(t)e^{-ia_N t} \tag{30}$$

be the characteristic function of the centered r.v. $\xi_1^{(A_N)} - a_N$.

Lemma A.4 *Under assumptions on sequences $(\xi_i)_{i=1}^{\infty}$, $(A_N)_{N=1}^{\infty}$, as in Theorem A.1, we have:*

(a) $\tilde{f}_N(t) = 1 - \frac{\sigma_N^2 t^2}{2} + O(|t|^3)$, with the constant in O term not depending on N.
(b) there exists an $\varepsilon > 0$ and N_0, such that if $|t| < \varepsilon$ and $N > N_0$, then

$$|\tilde{f}_N(t)| \le e^{-\sigma^2 t^2/4}.$$

Proof By using Taylor's formula with remainder, we write:

$$\left| \tilde{f}_N(t) - \left(1 - \frac{\sigma_N^2 t^2}{2}\right) \right| \le \frac{m_{3,N}|t|^3}{6},$$

which implies the first claim of the lemma in view of Lemma A.3(c). In addition, we have

$$\left| \tilde{f}_N(t) - \left(1 - \frac{\sigma^2 t^2}{2}\right) \right| \le \frac{|\sigma_N^2 - \sigma^2|t^2}{2} + \frac{|m_{3,N} - m_3||t|^3}{6} + \frac{m_3|t|^3}{6}, \tag{31}$$

By using Lemma A.3(b) and (c), we can find N_0 and $\varepsilon > 0$ such that for $|t| < \varepsilon$ and $N > N_0$ the right-hand side of (31) is smaller than $\frac{\sigma^2 t^2}{4}$, and that $\frac{\sigma^2 t^2}{2} < 1$. Then,

$$|\tilde{f}_N(t)| = \left| \tilde{f}_N(t) - \left(1 - \frac{\sigma^2 t^2}{2}\right) + \left(1 - \frac{\sigma^2 t^2}{2}\right) \right|$$

$$\le \frac{\sigma^2 t^2}{4} + \left(1 - \frac{\sigma^2 t^2}{2}\right) \le 1 - \frac{\sigma^2 t^2}{4} \le e^{-\sigma^2 t^2/4}.$$

$\square$

Now, we begin the proof of Theorem A.1. Let

$$P_N(n) = \mathbb{P}(\xi_1^{(A_N)} + \ldots + \xi_N^{(A_N)} = n).$$

Since the characteristic function of $\xi_1^{(A_N)}$ is $f_N(t)$, the characteristic function of $S_N = \xi_1^{(A_N)} + \ldots + \xi_N^{(A_N)}$ equals

$$[f_N(t)]^N = \sum_{n=-\infty}^{\infty} P_N(n)e^{itn}.$$

By inversion formula,

$$P_N(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itn} [f_N(t)]^N \, dt.$$

By using formula (30), parameterization $n = aN + z\sigma\sqrt{N}$, and the change of variable $x = t\sigma\sqrt{N}$, we rewrite this expression as

$$P_N(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-it(a-a_N)N} e^{-itz\sigma\sqrt{N}} [f_N(t)]^N \, dt$$

$$= \frac{1}{2\pi} \int_{-\pi\sigma\sqrt{N}}^{\pi\sigma\sqrt{N}} e^{-ix\frac{(a-a_N)}{\sigma}\sqrt{N}} e^{-ixz} \left[\tilde{f}_N\left(\frac{x}{\sigma\sqrt{N}}\right) \right]^N \frac{dx}{\sigma\sqrt{N}}$$

Since, by inversion formula,

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixz - x^2/2} \, dx,$$

we can write the difference

$$R_N = 2\pi \left(\sigma\sqrt{N} P_N(n) - \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \right),$$

as a sum of four integrals:

$$I_1 = \int_{-A}^{A} e^{-ixz} \left(e^{-ix\frac{(a-a_N)}{\sigma}\sqrt{N}} \left[\tilde{f}_N\left(\frac{x}{\sigma\sqrt{N}}\right) \right]^N - e^{-x^2/2} \right) dx,$$

$$I_2 = -\int_{|x|\geq A} e^{-ixz - x^2/2} \, dx,$$

$$I_3 = \int_{A \leq |x| \leq \varepsilon\sigma\sqrt{N}} e^{-ixz} \left(e^{-ix\frac{(a-a_N)}{\sigma}\sqrt{N}} \left[\tilde{f}_N\left(\frac{x}{\sigma\sqrt{N}}\right) \right]^N \right) dx,$$

$$I_4 = \int_{\varepsilon\sigma\sqrt{N} \leq |x| \leq \pi\sigma\sqrt{N}} e^{-ixz} \left(e^{-ix\frac{(a-a_N)}{\sigma}\sqrt{N}} \left[\tilde{f}_N\left(\frac{x}{\sigma\sqrt{N}}\right) \right]^N \right) dx,$$

where A and ε will be chosen later.

In order to show that $R_N \to 0$ for $N \to \infty$, we take an arbitrary $\delta > 0$ and show that by a choice of sufficiently large N the difference can be made smaller than δ.

Since $|I_2| \leq \int_{|x|\geq A} e^{-x^2/2} \, dx$, it can be made arbitrarily small by a choice of a sufficiently large A.

Next, we can choose a sufficiently small ε such that for all $N \geq N_0$, Lemma A.4(b) applies and for $|x| \leq \varepsilon\sigma\sqrt{N}$,

$$\left| \tilde{f}_N\left(\frac{x}{\sigma\sqrt{N}}\right) \right| \leq \exp\left(-\frac{\sigma^2}{4}\left(\frac{x}{\sigma\sqrt{N}}\right)^2 \right) = e^{-x^2/(4N)}.$$

Then, we can estimate I_3:

$$|I_3| \leq \int_{A \leq |x| \leq \varepsilon \sigma \sqrt{N}} \left| \tilde{f}_N \left(\frac{x}{\sigma \sqrt{N}} \right) \right|^N dx \leq \int_{A \leq |x|} e^{-x^2/4} dx,$$

which can be made arbitrarily small by a choice of a sufficiently large A.

Fix A and ε so that $|I_2| < \delta/4$ and $|I_3| < \delta/4$ for $N > N_0$.

Next, we address I_1. Let

$$g_N(x) := \left[e^{-i \frac{a - a_N}{\sigma \sqrt{N}} x} \tilde{f}_N \left(\frac{x}{\sigma \sqrt{N}} \right) \right]^N,$$

and note that this is the characteristic function of $(S_N - aN)/(\sigma \sqrt{N})$. We claim that the distribution of this random variable converges to the normal distribution with parameters $(0, 1)$. Indeed, by using Lemma A.4(a) and Lemmas A.3(a) and (b), we find that for a fixed x and $N \to \infty$,

$$\log g_N(x) = N \left[-i \frac{a - a_N}{\sigma \sqrt{N}} x + \log \left(\tilde{f}_N \left(\frac{x}{\sigma \sqrt{N}} \right) \right) \right]$$

$$= -i \frac{a - a_N}{\sigma} \sqrt{N} x + N \log \left(1 - \frac{\sigma_N^2 x^2}{2\sigma^2 N} + o \left(\frac{1}{N} \right) \right)$$

$$= -\frac{x^2}{2} + o(1),$$

which implies that $g_N(x) \to e^{-x^2/2}$ for every fixed x, and therefore, the distribution of $(S_n - aN)/(\sigma \sqrt{N})$ converges weakly to the standard normal distribution. Now, we invoke the fact that the weak convergence of probability measures implies the uniform convergence of characteristic functions on every finite interval. This implies that for every A, integral I_1 converges to zero as $N \to \infty$.

It remains to estimate integral I_4. We have

$$|I_4| \leq \int_{\varepsilon \leq \frac{|x|}{\sigma \sqrt{N}} \leq \pi} \left| \tilde{f}_N \left(\frac{x}{\sigma \sqrt{N}} \right) \right|^N dx = \sigma \sqrt{N} \int_{\varepsilon \leq |t| \leq \pi} |f_N(t)|^N dt.$$

By Lemma A.2, the maximal step of $\xi_1^{(A_N)}$ equals 1 for all sufficiently large N, which implies that

$$\max_{\varepsilon \leq |t| \leq \pi} |f_N(t)| = q < 1.$$

Therefore, $|I_4| \leq \sigma \sqrt{N} 2\pi q^N$ and $I_4 \to 0$ as $N \to \infty$.

The estimates of integrals I_1 and I_4 show that there exists N_1 such that $|I_1| \leq \delta/4$ and $|I_4| \leq \delta/4$ for $N > N_1$. Hence, the difference R_N for $N \to \infty$ approaches 0 uniformly over integer n.

References

1. Aldous, D.: The continuum random tree I. Ann. Probab. **19**(1), 1–28 (1991)
2. Aldous, D.: The continuum random tree II: an overview. In: Barlow, M.T., Bingham, N.H. (eds.) Proceedings of the Durham Symposium on Stochastic Analysis, 1990, pp. 23 – 70. Cambridge University Press (1991)
3. Aldous, D.: The continuum random tree III. Ann. Probab. **21**(1), 248–289 (1993)
4. Billingsley, P.: Convergence of Probability Measures, 1st edn. Wiley (1968)
5. Broutin, N., Marckert, J.F.: Asymptotic of trees with a prescribed degree sequence. Random Struct. Algorithms **44**, 290–316 (2014)
6. Burago, D., Burago, Y., Ivanov, S.: A Course in Metric Geometry. Graduate Studies in Mathematics, vol. 33. American Mathematical Society (2001)
7. Dershowitz, N., Zacks, S.: The cycle lemma and some applications. Eur. J. Comb. **11**, 35–40 (1990)
8. Devroye, L.: Simulating size-constrained Galton–Watson trees. SIAM J. Comput. **41**(1), 1–11 (2012)
9. Duquesne, T.: A limit theorem for the contour process of conditioned Galton–Watson trees. Ann. Probab. **31**(2), 996–1027 (2003)
10. Duquesne, T., Le Gall, J.F.: Random trees, Lévy processes and spatial branching processes. Astérisque 281 (2002)
11. Dvoretzky, A., Motzkin, T.: A problem of arrangements. Duke Math. J. **14**, 305–313 (1947)
12. Evans, S.N., Pitman, J., Winter, A.: Rayleigh processes, real trees, and root growth with re-grafting. Probab. Theory Relat. Fields **134**, 81–126 (2006)
13. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 2, 2nd edn. Wiley (1971)
14. Flajolet, P., Sedgewick, R.: Analytic Combinatorics. Cambridge University Press (2009)
15. Haas, B., Miermont, G.: Scaling limits of Markov branching trees with applications to Galton–Watson and random unordered trees. Ann. Probab. **40**(6), 2589–2666 (2012)
16. Janson, S.: Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. Probab. Surv. **9**, 103–252 (2012)
17. Kolchin, V.F.: Random Mappings, Russian edition, Nauka. (English translation: Optimization Software, New York, 1986) (1984)
18. Kortchemski, I.: Invariance principles for Galton–Watson trees conditioned on the number of leaves. Stochast. Process. Appl. **122**, 3126–3172 (2012)
19. Kortchemski, I.: A simple proof of Duquesne's theorem on contour processes of conditioned Galton–Watson trees. In: Donati-Martin, C. (ed.) Seminaire de Probabilities XLV. Lecture Notes in Mathematics, vol. 2078, pp. 537–558. Springer (2013)
20. Labarbe, J.-M., Marckert, J.-F.: Asymptotics of Bernoulli random walks, bridges, excursions and meanders with a given number of peaks. Electron. J. Probab. **12**, 229–261 (2007)
21. Lei, T.: Scaling limit of random forests with prescribed degree sequences. Bernoulli **25**, 2409–2438 (2019)
22. Le Gall, J.-F.: Random trees and applications. Probab. Surv. **2**, 245–311 (2005)
23. Le Gall, J.-F., Le Jan, Y.: Branching processes in Lévy processes: the exploration process. Ann. Probab. **26**, 213–252 (1998)
24. Marckert, J.-F., Mokkadem, A.: The depth-first processes of Galton–Watson trees converge to the same Brownian excursion. Ann. Probab. **31**(3), 1655–1678 (2003)
25. Ojeda, G.B., Holmgren, C., Thévenin, P.: Convergence of trees with a given degree sequence and of their associated laminations. arxiv:2111.07748 (2021)
26. Petrov, V.V.: Sums of Independent Random Variables. Springer, Berlin (1975)
27. Petrov, V.V.: Limit Theorems of Probability Theory. Oxford Studies in Probability, vol. 4. The Clarendon Press, Oxford (1995)
28. Pitman, J.: Combinatorial Stochastic Processes. Lecture Notes in Mathematics, vol. 1875. Springer (2002)
29. Rizzolo, D.: Scaling limits of Markov branching trees and Galton-Watson trees conditioned on the number of vertices with out-degree in a given set. Annales de l'Institut Henri Poincaré Probabilités et Statistiques **51**, 512–532 (2015)
30. Thévenin, P.: Vertices with fixed outdegrees in large Galton–Watson trees. Electron. J. Probab. **25**, 1–25 (2020)
31. Vervaat, W.: A relation between Brownian bridge and Brownian excursion. Ann. Probab. **7**(1), 143–149 (1979)