
A BRANCHING PROCESS APPROACH TO LEVEL- K PHYLOGENETIC NETWORKS

by

Benedikt Stuffer

Abstract. — The mathematical analysis of random phylogenetic networks via analytic and algorithmic methods has received increasing attention in the past years. In the present work we introduce branching process methods to their study. This approach appears to be new in this context. Our main results focus on random level- k networks with n labelled leaves. Although the number of reticulation vertices in such networks is typically linear in n , we prove that their asymptotic global and local shape is tree-like in a well-defined sense. We show that the depth process of vertices in a large network converges towards a Brownian excursion after rescaling by $n^{-1/2}$. We also establish Benjamini–Schramm convergence of large random level- k networks towards a novel random infinite network.

1. Introduction

Phylogenetic networks may be used to model the evolutionary history of species that have undergone reticulate events [1], such as horizontal gene transfer (by which genes are transferred across species) or hybrid speciation (by which lineages recombine to create a new one) [2].

The application of phylogenetic networks in evolutionary biology motivates the mathematical study of their number and shape, which has received increasing attention in recent literature. See for example [3, 4, 5, 6, 7, 8, 9] and references given therein. In the present work, we introduce branching process methods to their study. This approach appears to be new in this context and makes a fine addition to the current toolbox of analytic and algorithmic methods.

1.1. Asymptotic enumeration of level k -networks. — A binary rooted phylogenetic network N on a finite non-empty set X of leaves may be defined as a simple rooted directed graph with no directed cycles that satisfies the following constraints:

1. It's unique *root* has indegree 0 and outdegree 2.
2. All non-root vertices are either *tree nodes* (indegree 1, outdegree 2), *reticulation nodes* (indegree 2, outdegree 1), or *leaves* (indegree 1, outdegree 0).

Key words and phrases. — Phylogenetic networks, random graphs, branching processes.

The leaves are bijectively labelled with elements of X . It will be notationally convenient to additionally admit the network consisting of a single labelled root vertex with no edges.

There is an infinite number of such networks on a given set X . For this reason, one restricts to subclasses for which this number is finite. We are going to focus on level- k networks, with k denoting a fixed positive integer. Recall that a *cutvertex* in a connected graph is a vertex whose removal disconnects the graph. Similarly, a *bridge* is an edge whose removal disconnects the graph. A *block* (or *2-connected component*) is a connected induced subgraph that is maximal with the property of having no cutvertices of its own. The reader may consult books on the foundation of graph theory for further details [10]. We say the binary rooted phylogenetic network N is a *level- k* network, if the following conditions are met:

1. Any block of N contains at most k reticulation vertices of N .
2. Any block of N with at least 3 vertices contains at least 2 vertices that are the source of bridges of N .

Here we view directed edges as joining a *source* vertex to a *destination* vertex according to its orientation. The second condition ensures that there are only finitely many such networks on a given set X . In fact, their number satisfies the following asymptotic expression:

LEMMA 1.1. — *The number $N(k, n)$ of level- k networks on an n -element set satisfies*

$$(1.1) \quad N(k, n) \sim a_k n^{-3/2} \rho_k^{-n} n! \quad \text{as} \quad n \rightarrow \infty$$

for constants $a_k, \rho_k > 0$ that only depend on k .

For $k = 1$ and $k = 2$ this was already shown by [5] via analytic methods, who additionally calculated the involved constants, gave exact enumerating formulas, studied unrooted networks, and proved limit theorems for the numbers of undirected cycles and inner edges in random networks.

1.2. Global shape and limits of extremal parameters. — Note that each vertex v in N may be reached from the root by following a directed path that only traverses edges according to their orientation. There may be multiple such paths, and we denote the length of a shortest path by $h(v)$. Often, $h(v)$ is called the depth of height of v . The maximal height of vertices in N is denoted by $H(N)$. We let $|N|$ denote the number of vertices of N .

Throughout this paper, we fix a positive integer k . For each integer $n \geq 2$ we let \mathbf{N}_n be drawn uniformly at random among all level- k networks on the set $[n] := \{1, \dots, n\}$. If we order its vertices $v_0, \dots, v_{|\mathbf{N}_n|-1}$ we may form the corresponding height process $(h_t : 0 \leq t \leq |\mathbf{N}_n|)$ by linearly interpolating the values $h_i := h(v_i)$ for $0 \leq i \leq |\mathbf{N}_n| - 1$ and $h_{|\mathbf{N}_n|} := 0$. Of course, the height process depends on the order we choose.

THEOREM 1.2. — *We may couple \mathbf{N}_n with an ordering of its vertices such that the associated height process $(h_t : 0 \leq t \leq |\mathbf{N}_n|)$ satisfies*

$$(1.2) \quad (b_k n^{-1/2} h_{s|\mathbf{N}_n|} : 0 \leq s \leq 1) \xrightarrow{d} (\mathbf{e}(s) : 0 \leq s \leq 1) \quad \text{as } n \rightarrow \infty$$

for a constant $b_k > 0$ that only depends on k . Here $(\mathbf{e}(s) : 0 \leq s \leq 1)$ denotes Brownian excursion normalized to have duration 1. The same limit as in (1.2) holds if we form the height process using leaves only.

Questions concerning heights in various classes of random phylogenetic networks were raised in [7, Sec. 7] and [3, Sec. 8].

We prove Theorem 1.2 by establishing a new relation between distances in the random phylogenetic network \mathbf{N}_n and distances in a Galton–Watson tree conditioned on having n leaves. This allows us to apply invariance principles for the latter given in [11, 12].

An immediate consequence of Theorem 1.2 is that

$$(1.3) \quad b_k n^{-1/2} \mathbf{H}(\mathbf{N}_n) \xrightarrow{d} \sup_{0 \leq s \leq 1} \mathbf{e}(s).$$

Likewise, a uniformly selected vertex (or a uniformly selected leaf) u of \mathbf{N}_n satisfies

$$(1.4) \quad b_k n^{-1/2} h(u) \xrightarrow{d} \mathbf{e}(r)$$

with $0 \leq r \leq 1$ denoting a random point of the unit interval $[0, 1]$, selected independently according to the uniform distribution. We may also obtain tail-bounds and convergence of moments:

THEOREM 1.3. — *There are constants $C, c > 0$ such that for all $x > 0$ and $n \geq 2$*

$$(1.5) \quad \mathbb{P}(\mathbf{H}(\mathbf{N}_n) > x) \leq C \exp(-cx^2/n).$$

Moreover, all higher moments in (1.3) and (1.4) converge.

Our main tool for proving Theorem 1.3 is a similar tail-bound by [13] for the height of Galton–Watson trees conditioned to be large. Theorem 1.3 entails that the moments converge in (1.3), yielding

$$(1.6) \quad b_k n^{-1/2} \mathbb{E}[\mathbf{H}(\mathbf{N}_n)] \rightarrow \mathbb{E} \left[\sup_{0 \leq s \leq 1} \mathbf{e}(s) \right] = \sqrt{\pi/2}$$

and

$$(1.7) \quad b_k^p n^{-p/2} \mathbb{E}[\mathbf{H}(\mathbf{N}_n)^p] \rightarrow 2^{-p/2} p(p-1) \Gamma(p/2) \zeta(p)$$

for each integer $p \geq 2$. Here Γ refers to Euler’s gamma-function, and ζ to Riemann’s zeta-function. Compare with [13, Eq. (1.6)].

Instead of using lengths of shortest directed paths, we may also study the structure of the undirected graph \mathbf{G}_n underlying the network \mathbf{N}_n . The graph distance from the root to a vertex in this graph needs not coincide with the length of a shortest *directed* path. It might be shorter. If we write $\mathbf{H}(\mathbf{G}_n)$ for the maximal height with

respect to the graph distance, then by construction $H(\mathbf{G}_n) \leq H(\mathbf{N}_n)$. Hence, the tail bound of Theorem 1.3 also applies to \mathbf{G}_n :

$$(1.8) \quad \mathbb{P}(H(\mathbf{G}_n) > x) \leq C \exp(-cx^2/n).$$

Moreover, if we let $h(v)$ denote the graph distance of a vertex in \mathbf{G}_n from the root and construct the height process h for \mathbf{G}_n accordingly, then it is clear from the proof of Theorem 1.2 that there exists a constant $b'_k \geq b_k$ such that

$$(1.9) \quad (b'_k n^{-1/2} h_{s|\mathbf{G}_n} : 0 \leq s \leq 1) \xrightarrow{d} (\mathbf{e}(s) : 0 \leq s \leq 1)$$

Likewise, the same convergence holds for the height process of the leaves. For $k = 1$ the constants $b_1 = b'_1$ agree, however for $k \geq 2$ the expressions obtained for b_k and b'_k differ.

Finally, we may also study the maximal length of a directed or undirected path starting from the root in \mathbf{N}_n . (As opposed to the maximal length of a shortest directed or undirected path from the root to any vertex in \mathbf{N}_n .) It is clear from the proofs of Theorems 1.2 and 1.3 that (1.3), (1.5), (1.6), and (1.7) hold analogously for these parameters, if we replace the constant b_k by some constant $b''_k > 0$ for the maximal length of a directed path, or by some constant $b'''_k > 0$ for the maximal length of an undirected path.

The following theorem states that although phylogenetic networks are not trees, their global shape is tree-like:

THEOREM 1.4. — *Let \mathbf{G}_n be the graph underlying the random level- k phylogenetic network \mathbf{N}_n , let $b'_k > 0$ be as in (1.9). Let μ_n denote either the uniform measure on the vertices or on the leaves of \mathbf{G}_n . Let $(\mathcal{T}_e, d_{\mathcal{T}_e}, \mu_{\mathcal{T}_e})$ denote the Brownian continuum random tree. Then*

$$(1.10) \quad (\mathbf{G}_n, b'_k n^{-1/2} d_{\mathbf{G}_n}, \mu_n) \xrightarrow{d} (\mathcal{T}_e, d_{\mathcal{T}_e}, \mu_{\mathcal{T}_e}) \quad \text{as } n \rightarrow \infty.$$

in the Gromov–Hausdorff–Prokhorov sense.

The intuition behind Theorem 1.4 is that although \mathbf{G}_n contains a linear number of cycles, the global shape is tree-like as the cycles are so short (at most $O(\log n)$ in circumference) that they contract to points when rescaling distances by $n^{-1/2}$. Care has to be taken that they nevertheless influence the global shape. A path between two typical points in \mathbf{G}_n traverses roughly \sqrt{n} cycles, and they distort the distance on average by a stretch factor.

The Brownian continuum random tree $(\mathcal{T}_e, d_{\mathcal{T}_e}, \mu_{\mathcal{T}_e})$ was introduced and studied in the series of pioneering papers [14, 15, 16]. In some sense, it's a random “continuum” tree with uncountably many points. Formally, it may be defined as the random metric space \mathcal{T}_e corresponding to the random semi-metric

$$(1.11) \quad d(x, y) = \mathbf{e}(x) + \mathbf{e}(y) - 2 \inf_{\min(x, y) \leq t \leq \max(x, y)} \mathbf{e}(t), \quad x, y \in [0, 1]$$

on the unit-interval. The measure $\mu_{\mathcal{T}_e}$ is the push-forward of the uniform measure on $[0, 1]$ under the canonical surjection $[0, 1] \rightarrow \mathcal{T}_e$. The Brownian continuum random tree is universal in the sense that it describes the asymptotic geometry of a variety of

different models of random graphs [17, 18, 19, 20], hence linking seemingly unrelated models of random structures.

The notion of Gromov–Hausdorff–Prokhorov convergence was introduced in [21]. It allows us to assign a distance to a pair (X, d_X, μ_X) and (Y, d_Y, μ_Y) of compact metric spaces equipped with Borel probability measures. We refer the reader to [21, Sec. 6] for a detailed introduction.

1.3. Local shape and additive parameters. — How does the vicinity of the root or a random vertex (or random leaf) in \mathbf{N}_n evolve as n becomes large? The frequencies with which we observe given shapes converge, allowing us to describe the asymptotic local shape of \mathbf{N}_n via random networks having a countably infinite number of vertices.

In order to make this precise, we require some notation. For any integer $\ell \geq 0$ we may consider the restriction $U_\ell(\cdot)$ that maps a pair of a phylogenetic network N and one of its vertices v to the ℓ -neighbourhood subnetwork $U_\ell(N, v)$ induced by all vertices that may be reached from v via an undirected path of length at most ℓ . That is, all vertices that may be reached from v by crossing at most ℓ edges, regardless whether we follow the direction of the edges or not. We consider the vertices of $U_\ell(N, v)$ as unlabelled, and the edges as directed. If v is equal to the root vertex of N we may simply write $U_\ell(N)$.

THEOREM 1.5. — *There is a random infinite level- k network $\hat{\mathbf{N}}$ such that*

$$(1.12) \quad \mathbf{N}_n \xrightarrow{d} \hat{\mathbf{N}} \quad \text{as} \quad n \rightarrow \infty$$

in the local topology. Even stronger, for any sequence ℓ_n of positive integers satisfying $\ell_n = o(\sqrt{n})$ it holds that

$$(1.13) \quad d_{\text{TV}}(U_{\ell_n}(\mathbf{N}_n), U_{\ell_n}(\hat{\mathbf{N}})) \rightarrow 0.$$

The limit (1.12) states that for each integer $\ell \geq 0$ and each finite directed vertex-marked graph G it holds that

$$(1.14) \quad \mathbb{P}(U_\ell(\mathbf{N}_n) = G) \rightarrow \mathbb{P}(U_\ell(\hat{\mathbf{N}}) = G).$$

Thus, (1.13) is a much stronger statement, as it allows us to describe the asymptotic shape of larger neighbourhoods of the root. The assumption $\ell_n = o(\sqrt{n})$ is as general as possible: Thm. 1.2 implies that for any constant but arbitrarily small $\epsilon > 0$ the convergence in (1.13) fails to hold for $\ell_n = \lfloor \epsilon\sqrt{n} \rfloor$.

It is natural to also study what happens in the vicinity of a uniformly selected vertex of \mathbf{N}_n .

THEOREM 1.6. — *Let u_n denote a uniformly selected vertex of \mathbf{N}_n . There is a random infinite directed vertex marked graph $\hat{\mathbf{N}}^*$ such that*

$$(1.15) \quad (\mathbf{N}_n, u_n) \xrightarrow{d} \hat{\mathbf{N}}^* \quad \text{as} \quad n \rightarrow \infty$$

in the local topology. Moreover, for any integer $\ell \geq 0$ and any finite directed unlabelled vertex-marked graph G the number $N_{\ell, G}$ of vertices u in \mathbf{N}_n with $U_\ell(\mathbf{N}_n, u) = G$

satisfies

$$(1.16) \quad \frac{N_{\ell,G}}{|\mathbf{N}_n|} \xrightarrow{p} \mathbb{P}(U_\ell(\hat{\mathbf{N}}^*) = G).$$

Borrowing terminology from statistical physics, (1.15) may be called annealed local convergence, and (1.16) quenched local convergence.

We might also wonder what happens in the vicinity of a uniform random leaf w of \mathbf{N}_n . Theorem 1.6 actually entails a local limit for the vicinity of w as well. Indeed, it follows directly from Theorem 1.6 that

$$(1.17) \quad \frac{|\mathbf{N}_n|}{n} \xrightarrow{p} \frac{1}{\mathbb{P}(\text{marked vertex of } \hat{\mathbf{N}}^* \text{ is a leaf})}$$

and

$$(1.18) \quad (\mathbf{N}_n, w) \xrightarrow{d} (\hat{\mathbf{N}}^* \mid \text{marked vertex of } \hat{\mathbf{N}}^* \text{ is a leaf}).$$

Likewise, it follows directly that the number $N_{\ell,G}^{\text{leaf}}$ of leaves whose ℓ -neighbourhood equals G satisfies

$$(1.19) \quad \frac{N_{\ell,G}^{\text{leaf}}}{n} \xrightarrow{p} \mathbb{P}(U_\ell(\hat{\mathbf{N}}^*) = G \mid \text{marked vertex of } \hat{\mathbf{N}}^* \text{ is a leaf}).$$

Using general principles for locally convergent random graphs [22], Theorem 1.6 entails laws of large numbers for subgraph counts. That is, the number of copies of some given finite graph in \mathbf{N}_n concentrates at a constant multiple of n , with the constant factor being given by the expected value of some functional of $\hat{\mathbf{N}}^*$. See [22, Lem. 4.3] for details.

Notation. — Unless otherwise stated, all unspecified limits are taken as $n \rightarrow \infty$. The arrows \xrightarrow{p} , \xrightarrow{d} , and $\xrightarrow{a.s.}$ denote convergence in probability, convergence in distribution, and almost sure convergence. Equality in distribution is denoted by $\stackrel{d}{=}$. The total variation distance between measures and random variables is denoted by d_{TV} . For a vertex v in a directed graph N its *indegree* refers to the number of edges in N whose *destination* is v . The *outdegree* of v is the number of edges in N whose *source* is v . We will refer to a vertex w as a *child* of v if there is a directed edge from v to w in N . In a rooted undirected tree we use the notions “outdegree” and “child” as if all edges were directed as pointing away from the root.

2. A bijective encoding in terms of decorated trees

Throughout the following, we fix an integer $k \geq 1$ and use the short term *k-network* in order to refer to binary rooted phylogenetic level- k networks. The collection of k -networks on a given set X will be denoted by $\mathcal{N}[X]$. Recall that we admit the k -network consisting of a single labelled root vertex with no edges.

In [23, 24] a decompositions of level- k networks into smaller simple networks was developed. In this section we explain how this leads to a blow-up procedure for the random generation of k -networks.

2.1. Level- k networks are blow-ups of decorated trees. — Let N be a k -network and let v be a vertex of N with outdegree 2. We say the vertex v *splits* (or *splitting*), if N contains no vertex that may be reached via a directed path from both children of v . That is, we cannot walk from each of its two children to the same vertex by only crossing edges in accordance with their direction.

This terminology allows us to differentiate three types of k -networks:

1. The trivial k -network consisting of a single root vertex and nothing else.
2. k -networks where the root has outdegree 2 and splits.
3. k -networks where the root has outdegree 2 and does not split.

It is easy to describe a k -network N on a set X where the root has outdegree 2 and splits. N is obtained in a unique way by taking an unordered pair of k -networks with disjoint leaf sets that partition X , and adding a root vertex that is the source of two directed edges that point to the respective roots of these two k -networks.

Now, suppose that N is a k -network where the root has outdegree 2 and does not split. Then it is contained in a unique block B with at least 4 vertices. Vertices with indegree 1 and outdegree 1 in B are tree nodes of N , and hence the source of a bridge of N whose destination lies outside of B . Likewise, vertices of B with indegree 2 and outdegree 0 are reticulation nodes of N , and the source of a bridge of N whose destination lies outside of B . Let S be the network obtained from B by additionally adding these bridges and their destinations, which then correspond precisely to the leaves of S . We say S is a *simple network*, and B is its *core*. Thus, the network N may be obtained from the k -network S by identifying the leaves of S with the roots of smaller k -networks. The leaf sets of the k -networks attached to S in this way partition the set X into non-empty disjoint subsets. We may view the leaf set of a k -network attached to a leaf of S as the label of the leaf. Thus, any k -network N on a set X whose root does not split may be obtained in a unique way by forming a partition M of X into non-empty subsets, choosing a simple k -network S on M , and for each partition class $C \in M$ we identify the leaf of S labelled by C with the root of a k -network on C .

Letting $N(k, n), B(k, n) \geq 0$ denote the number of k -networks and simple k -networks on a given n -element set, it follows by standard combinatorial tools [25, 26, 27] that the exponential generating series

$$\mathcal{N}(z) := \sum_{n \geq 1} N(k, n) \frac{z^n}{n!} \quad \text{and} \quad \mathcal{B}(z) := \sum_{n \geq 2} B(k, n) \frac{z^n}{n!}$$

satisfy the equation

$$(2.1) \quad \mathcal{N}(z) = z + \frac{\mathcal{N}(z)^2}{2} + \mathcal{B}(\mathcal{N}(z)).$$

Moreover, it also follows that to each k -network N on a set X of at least two leaves we may associate a k -network $\text{head}(N)$, which is either a simple network on some partition (with non-empty partition classes) of X , or it is the cherry network (consisting of a root with two children) on a 2-partition of X (with non-empty

partition classes). The network N is obtained by identifying the leaves of its head with the roots of the subnetworks on the corresponding partition classes.

Whenever one of these subnetworks is not the trivial network consisting of a single leaf, it has a head-structure of its own and is constructed from this head structure and even smaller subnetworks. We may proceed recursively in this way until only trivial subnetworks are left. In this way, we may form a pair $\Lambda(N) := (T, \delta)$ of a rooted unordered tree T whose leaves are labelled bijectively with the element of X , and a function δ that assigns to each inner vertex v of T a head structure $\delta(v)$ as its *decoration*. We say (T, δ) is a *decorated tree*. The formal definition is as follows:

1. If a network N is trivial, that is it consists of a single leaf and nothing else, we let T be given by a rooted tree consisting of a single vertex (labelled like the leaf of N). As T has no inner vertices, δ is the trivial function with an empty domain.
2. If a network N has at least two leaves, it consists of the network $\text{head}(N)$ with some number $\ell \geq 2$ of (possibly trivial) subnetworks N_1, \dots, N_ℓ attached to the leaves of $\text{head}(N)$. We let T be the tree whose root o has ℓ children, such that the i th child is the root of the tree corresponding (recursively) to N_i . We set $\delta(o) = \text{head}(N)$ and extend δ according to the decorations in $\Lambda(N_1), \dots, \Lambda(N_\ell)$.

Thus:

LEMMA 2.1. — *For each finite non-empty set X , the function Λ ($= \Lambda_X$) is a bijection between the collection $\mathcal{N}[X]$ of k -networks on X , and the collection $\mathcal{P}[X]$ of pairs (T, δ) , where T is a rooted unordered tree whose leaves are bijectively labelled with the elements of X , and δ is a function that assigns to each inner vertex v of T a head structure (that is, a cherry network or a simple network) on the children of v (or equivalently the corresponding partition of X).*

Lemma 2.1 identifies the combinatorial class of k -networks with a special case of a class of Schröder-enriched parenthesizations. In general, if we have a class of combinatorial structures where each structure has a “size” given by a positive integer, we may form the corresponding class of Schröder-enriched parenthesizations. It is the collection of all unordered rooted trees where each inner vertex has a decoration given by a structure whose size agrees with the outdegree of the vertex. See [28] for details. In the present case, the structures are the head networks, and the “size” of a head network is its number of leaves.

The inverse Λ^{-1} function of Λ may be described as a blow-up procedure, that maps a decorated tree (T, δ) to the network obtained by “blowing up” each inner vertex v of T by its decoration $\delta(v)$. That is, we delete the edges between v and its children, identify the root of $\delta(v)$ with v , and identify each child of v with the corresponding leaf in $\delta(v)$.

2.2. A sampling procedure that uses simply-generated trees. — Let P_n be uniformly selected from the collection $\mathcal{P}[X]$ for $X := \{1, \dots, n\}$, $n \geq 2$. Since,

$\Lambda^{-1} : \mathcal{P}[X] \rightarrow \mathcal{N}[X]$ is a bijection, it follows that the k -network $\Lambda^{-1}(\mathbb{P}_n)$ is distributed like the uniform k -network \mathbb{N}_n on X . Hence we may assume without loss of generality that

$$(2.2) \quad \mathbb{N}_n = \Lambda^{-1}(\mathbb{P}_n).$$

Letting $H(k, i) \geq 0$ denote the number of head-networks with i leaves for all $i \geq 0$, and set

$$(2.3) \quad \mathcal{H}(z) := \sum_{i \geq 2} \frac{H(k, i)}{i!} z^i = \mathcal{B}(z) + z^2/2.$$

We refer to $\mathbf{w} := (H(k, i)/i!)_{i \geq 0}$ as a *weight-sequence*. For each set Y we let $\mathcal{H}[Y]$ denote the collection of head-structures with leaves labelled bijectively by the elements of Y . For ease of notation, we set

$$(2.4) \quad \mathcal{H}[d] := \mathcal{H}[\{1, \dots, d\}]$$

for all integers $d \geq 0$.

A (*planted*) *plane tree* is a rooted ordered unlabelled tree. The children of any of its vertices are endowed with a linear order. We refer the reader to [29, Sec. 1.2.2] for a detailed introduction to this type of trees. The following procedure, formulated in [19, Lem. 6.7] for general Schröder enriched parenthesizations, allows us to generate \mathbb{P}_n from a random (weighted) plane tree with n leaves:

PROPOSITION 2.2 ([19, Lem. 6.7]). — *Set $p_0 = 1$ and $p_i = H(k, i)/i!$ for all $i \geq 1$. The outcome (τ_n, δ_n) of the following procedure is distributed like \mathbb{P}_n .*

1. *Generate a random plane tree τ_n with n leaves according to its distribution*

$$(2.5) \quad \mathbb{P}(\tau_n = T) = \left(\sum_P \prod_{v \in P} p_{d_P^+(v)} \right)^{-1} \prod_{v \in T} p_{d_T^+(v)},$$

with the sum-index P ranging over the finite collection of plane trees with n leaves, where each internal vertex has outdegree at least two.

2. *For each inner vertex v of τ_n sample a head-structure*

$$(2.6) \quad \delta_n(v) \in \mathcal{H}[d_{\tau_n}^+(v)]$$

uniformly at random.

3. *Choose a bijection σ between the set of leaves of τ_n and X uniformly at random, and distribute labels to the leaves of τ_n accordingly.*

Note that the decorated tree \mathbb{P}_n is unordered, whereas in (τ_n, δ_n) the children of any vertex are endowed with a linear order. If we forget about these linear orders, we obtain a decorated unordered tree that is distributed like \mathbb{P}_n . Note also that the distribution of τ_n in (2.5) depends on the weight sequence \mathbf{w} .

3. Asymptotic enumeration

In this section we will prove Lemma 1.1 and identify τ_n as a critical Galton–Watson tree conditioned on having n leaves.

The number of possible simple components in k -networks is infinite. However, [23] developed a decomposition of simple networks in terms of so-called *generators*. A generator is a directed multi-graph obtained from the core of a simple network by contracting each vertex of the core with in-degree 1 and outdegree 1. Conversely, simple networks are precisely the networks obtained by blowing up edges of generators into paths (that is, replacing the edge by a path with at least one edge), and afterwards adding outgoing edges to all vertices with indegree 1 and outdegree 1, and all vertices with indegree 2 and outdegree 0. However, care has to be taken when a generator contains multi-edges, that is, pairs of vertices joined by exactly two edges. In this case, for each such pair we have to blow-up at least one of the two edges by a path with at least two edges.

For all $i, j, \ell \geq 0$ let $G(k, i, j, \ell)$ denote the number of generators with the following properties:

1. The number of vertices with indegree 2 is at most k .
2. There are i vertices with outdegree 0, labelled from 1 to i .
3. There are j edges that are not multi-edges, labelled from 1 to j .
4. There are ℓ pairs of multi-edges joining the same two vertices, and each pair carries a unique label from 1 to ℓ .

It follows that the series $\mathcal{H}(z)$ defined in Equation (2.3) satisfies

$$(3.1) \quad \mathcal{H}(z) = \frac{z^2}{2} + \sum_{i,j,\ell \geq 0} \frac{G(k, i, j, \ell)}{i!j!\ell!} z^i \frac{1}{(1-z)^j} \left(\frac{z}{1-z} + \frac{z^2}{2(1-z)^2} \right)^\ell.$$

The total number of generators that contain at most k vertices with indegree 2 is finite, see [23, 24]. Consequently, it follows from (3.1) that there is a bivariate polynomial $f(z, w) \neq 0$ with $f(0, w) = 0 = f(z, 0)$ such that

$$(3.2) \quad \mathcal{H}(z) = \frac{z^2}{2} + f(z, (1-z)^{-1}).$$

This crucial equation entails that $\mathcal{H}(z)$ has radius of convergence 1 and

$$(3.3) \quad \lim_{t \nearrow 1} \mathcal{H}'(t) = \infty.$$

Let $0 < t_0 < 1$ be the unique point with $\mathcal{H}'(t_0) = 1$. Let ξ be a random non-negative integer with distribution given by

$$(3.4) \quad \mathbb{P}(\xi = \ell) = \begin{cases} p_\ell t_0^{\ell-1} & \ell \geq 1 \\ 1 - \sum_{i \geq 1} p_i t_0^{i-1} & \ell = 0 \end{cases}.$$

By the choice of t_0 , this is well-defined and

$$(3.5) \quad \mathbb{E}[\xi] = \mathcal{H}'(t_0) = 1.$$

As $t_0 < 1$, it follows that ξ has finite exponential moments. That is, there is an $\epsilon > 0$ such that

$$(3.6) \quad \mathbb{E}[(1 + \epsilon)^\xi] < \infty.$$

Recall that a ξ -Galton–Watson τ is a random plane tree that starts with a single root vertex and where each vertex receives offspring according to an independent copy of ξ . Thus, if T is a plane tree with n leaves,

$$(3.7) \quad \mathbb{P}(\tau = T) = \prod_{v \in T} \mathbb{P}(\xi = d_T^+(v)) = \mathbb{P}(\xi = 0)^n t_0^{n-1} \prod_{v \in T} p_{d_T^+(v)}.$$

Thus, $\mathbb{P}(\tau = T)$ is equal to the product of $\mathbb{P}(\tau_n = T)$ and a factor that only depends on n (and not on T). Hence, τ_n is distributed like the result of conditioning τ on the event that its number $L(\tau)$ of leaves is equal to n . That is,

$$(3.8) \quad \tau_n \stackrel{d}{=} (\tau \mid L(\tau) = n).$$

Equations (3.5) and (3.6) allow us to apply a general result [24, Thm. 3.1] for the number of leaves in a critical Galton–Watson tree, yielding

$$(3.9) \quad \mathbb{P}(L(\tau) = n) \sim \sqrt{\frac{\mathbb{P}(\xi = 0)}{2\pi\mathbb{V}[\xi]}} n^{-3/2}$$

The probability generating function $\mathcal{Z}(z) := \mathbb{E}[z^{L(\tau)}]$ for $L(\tau)$ satisfies the recursive equation

$$(3.10) \quad \mathcal{Z}(z) = z\mathbb{P}(\xi = 0) + \sum_{\ell \geq 1} \mathbb{P}(\xi = \ell) \mathcal{Z}(z)^\ell = z(1 - \mathcal{H}(t_0)/t_0) + \mathcal{H}(t_0 \mathcal{Z}(z))/t_0.$$

Equation (3.5) entails that

$$(3.11) \quad t_0 = \sum_{i \geq 1} p_i t_0^i > \mathcal{H}(t_0).$$

Hence it follows from Equation (3.10) that

$$(3.12) \quad t_0 \mathcal{Z}(z/(t_0 - \mathcal{H}(t_0))) = z + \mathcal{H}(t_0 \mathcal{Z}(z/(t_0 - \mathcal{H}(t_0)))).$$

Equation (2.1) entails that $\mathcal{N}(z) = z + \mathcal{H}(\mathcal{N}(z))$, hence

$$(3.13) \quad \mathcal{N}(z) = t_0 \mathcal{Z}(z/(t_0 - \mathcal{H}(t_0))).$$

In other words, it holds for all $n \geq 1$

$$(3.14) \quad N(k, n) = \mathbb{P}(L(\tau) = n) t_0 (t_0 - \mathcal{H}(t_0))^{-n}.$$

Using (3.9), it follows that

$$(3.15) \quad N(k, n) \sim \sqrt{\frac{\mathbb{P}(\xi = 0)}{2\pi\mathbb{V}[\xi]}} t_0 n^{-3/2} (t_0 - \mathcal{H}(t_0))^{-n}$$

This proves Lemma 1.1 for $a_k := \sqrt{\frac{\mathbb{P}(\xi = 0)}{2\pi\mathbb{V}[\xi]}} t_0$ and $\rho_k = (t_0 - \mathcal{H}(t_0))^{-n}$.

4. The asymptotic global shape

In the present section we prove Theorems 1.2, 1.3, and 1.4.

4.1. Preparations. — The following observation is easy, but we are going to use it often enough to state it explicitly:

PROPOSITION 4.1. — *Let $d \geq 2$ be an integer. Then the number of vertices $|H|$ in a head structure $H \in \mathcal{H}[d]$ is bounded by $2(d+k)$.*

Proof. — This follows from the fact that any simple network has at most k reticulation vertices, and all other vertices either have outdegree 2 or 0:

We may build any simple network from a cherry network (consisting of a root vertex with two children) step by step, where in each step we either add two children to a leaf, or fuse two leaves together and add a single child to the newly created reticulation vertex. (Please note that not every network created in this way is a simple network.) The first kind of step increases the total number of vertices by two and increases the number of leaves by 1. The second kind of step leaves the total number of vertices invariant and decreases the number of leaves by 1. Letting s_1 and s_2 denote the number of steps of kind 1 and 2, the resulting network has $V := 3 + 2s_1$ vertices and $L := 2 + s_1 - s_2$ leaves. Since we can have at most k reticulation vertices, $s_2 \leq k$ holds, and hence

$$2(L+k) \geq 2(2+s_1) \geq V.$$

Since a head structure from $\mathcal{H}[d]$ has d leaves, it follows that it has at most $2(d+k)$ vertices. \square

We are also going to need a bound for the maximal outdegree of τ_n :

PROPOSITION 4.2. — *There is a constant $C > 0$ such that the maximal outdegree $\Delta(\tau_n)$ of τ_n satisfies*

$$(4.1) \quad \Delta(\tau_n) \leq C \log n$$

with a probability that tends to 1 as $n \rightarrow \infty$.

Proof. — Since $\mathbb{P}(\xi = 1) = 0$, it follows that τ_n has (almost surely) no inner vertices with outdegree 1. Consequently, the number $|\tau_n|$ of vertices of τ_n is bounded by the number of vertices of a binary tree with n leaves, that is,

$$(4.2) \quad |\tau_n| \leq 2n - 1.$$

Using (3.9), it follows that for any $x > 0$

$$(4.3) \quad \begin{aligned} \mathbb{P}(\Delta(\tau_n) > x) &= \frac{\mathbb{P}(L(\tau) = n, \Delta(\tau) > x)}{\mathbb{P}(L(\tau) = n)} \\ &\leq \mathbb{P}(L(\tau) = n)^{-1} 2n \mathbb{P}(\xi > x) \\ &\leq O(n^{5/2}) \mathbb{P}(\xi > x). \end{aligned}$$

By (3.6) it follows that there are constants $C, c > 0$ such that

$$(4.4) \quad \mathbb{P}(\xi > x) \leq C \exp(-cx)$$

for all $x > 0$. Taking $x = C' \log n$ for a sufficiently large constant $C' > 0$, it follows that

$$(4.5) \quad \mathbb{P}(\Delta(\tau_n) > C' \log n) \rightarrow 0$$

as $n \rightarrow \infty$. \square

4.2. Size-biased trees. — Since $\mathbb{E}[\xi] = 1$, we may define the size-biased random positive integer $\hat{\xi}$ by

$$(4.6) \quad \mathbb{P}(\hat{\xi} = i) = i\mathbb{P}(\xi = i).$$

For each integer $\ell \geq 0$, we define the size-biased Galton–Watson tree $\hat{\tau}^{(\ell)}$ as a random finite plane tree with a marked vertex having height ℓ . For $\ell = 0$, we let $\hat{\tau}^{(0)}$ denote an independent copy of the Galton–Watson tree with a marked vertex that coincides with its root vertex. Inductively, we define $\hat{\tau}^{(\ell)}$ for $\ell \geq 1$ as follows: Start with a root vertex that receives offspring according to an independent copy of $\hat{\xi}$. A child of the root is selected uniformly at random and identified with an independent copy of $\hat{\tau}^{(\ell-1)}$. Each of the remaining children of the root gets identified with an independent copy of τ , that is, a fresh independent copy for each remaining child.

Thus, the vertices of $\hat{\tau}^{(\ell)}$ that receive offspring according to $\hat{\xi}$ (as opposed to ξ) together with the marked vertex of $\hat{\tau}^{(\ell)}$ form a path of length ℓ from the root to the marked vertex of $\hat{\tau}^{(\ell)}$.

It is elementary to verify that for each finite plane tree T and each vertex v of T with height ℓ it holds that

$$(4.7) \quad \mathbb{P}\left(\hat{\tau}^{(\ell)} = (T, v)\right) = \prod_{u \in T} \mathbb{P}(\xi = d_T^+(u)) = \mathbb{P}(\tau = T).$$

Any locally finite rooted tree T may be decorated in a *canonical* random way by choosing for each inner vertex v a decoration from $\mathcal{H}[d_T^+(v)]$ uniformly at random, independently from the remaining decorations. Thus we may form canonical random decorations (τ, δ) and $(\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)})$. It follows from (4.7) that for any deterministic decoration γ of T

$$(4.8) \quad \mathbb{P}\left((\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)}) = (T, \gamma)\right) = \mathbb{P}(\tau = T) \prod_{u \in T, d_T^+(u) > 0} \frac{1}{|\mathcal{H}[d_T^+(u)]|} \\ = \mathbb{P}((\tau, \delta) = (T, \gamma)).$$

4.3. Heights in random networks. — We let the height $h_N(v)$ of a vertex u in a network N be the length of a shortest directed path from the root of N to u . Using the decorated trees (τ, δ) and $(\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)})$, we are now ready to prove the following:

LEMMA 4.3. — *Let η denote the length of a shortest directed path from the root of a uniformly selected network from $\mathcal{H}[\hat{\xi}]$ to a leaf that is uniformly selected among its $\hat{\xi}$ leaves. With a probability that tends to 1 as $n \rightarrow \infty$, any vertex v of τ_n has the property, that the corresponding vertex u in $\mathbf{N}_n = \Lambda^{-1}(\mathbf{P}_n)$ satisfies*

$$(4.9) \quad |h_{\mathbf{N}_n}(u) - \mathbb{E}[\eta]h_{\tau_n}(v)| \leq n^{3/8}.$$

Proof. — Let us take a closer look at the network $\Lambda^{-1}(\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)})$. The path v_0, \dots, v_ℓ from the root v_0 of $\hat{\tau}^{(\ell)}$ to its marked vertex v_ℓ corresponds to vertices u_0, \dots, u_ℓ in $\Lambda^{-1}(\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)})$. In particular, u_0 coincides with the root of $\Lambda^{-1}(\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)})$. Every path in $\Lambda^{-1}(\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)})$ from u_0 to u_ℓ must pass through $u_1, \dots, u_{\ell-1}$ and is entirely contained in the subnetworks corresponding to $\hat{\delta}^{(\ell)}(v_0), \dots, \hat{\delta}^{(\ell)}(v_{\ell-1})$. The length of a shortest directed path from u_0 to u_ℓ is distributed like the sum $\eta_1 + \dots + \eta_\ell$ of ℓ independent copies η_1, \dots, η_ℓ of η .

It follows from Equation (3.6) and Proposition 4.1 that η has finite exponential moments, that is,

$$(4.10) \quad \mathbb{E}[(1 + \epsilon)^\eta] < \infty.$$

for some $\epsilon > 0$. Let $\tilde{\eta} := \eta - \mathbb{E}[\eta]$. It follows that there is a constant $c > 0$ such that for all sufficiently small $\lambda > 0$

$$(4.11) \quad \mathbb{E}[\exp(\lambda\tilde{\eta})] \leq 1 + c\lambda^2 \quad \text{and} \quad \mathbb{E}[\exp(-\lambda\tilde{\eta})] \leq 1 + c\lambda^2.$$

Using Markov's inequality, it follows that for all $x > 0$

$$(4.12) \quad \begin{aligned} \mathbb{P}(\eta_1 + \dots + \eta_\ell - \ell\mathbb{E}[\eta] > x) &\leq \mathbb{P}\left(\exp\left(\lambda \sum_{i=1}^{\ell} (\eta_i - \mathbb{E}[\eta])\right) > \exp(\lambda x)\right) \\ &\leq \frac{\mathbb{E}[\exp(\lambda\tilde{\eta})]^\ell}{\exp(\lambda x)} \\ &\leq \frac{(1 + c\lambda^2)^\ell}{\exp(\lambda x)}. \end{aligned}$$

Repeating the same argument for $-\tilde{\eta}$ instead of $\tilde{\eta}$, we arrive at

$$(4.13) \quad \mathbb{P}(|\eta_1 + \dots + \eta_\ell - \ell\mathbb{E}[\eta]| > x) \leq 2 \frac{(1 + c\lambda^2)^\ell}{\exp(\lambda x)}.$$

Taking $x = n^{3/8}$ and $\lambda = n^{-1/4}$, it follows that uniformly for all $1 \leq \ell \leq \sqrt{n} \log n$

$$(4.14) \quad \mathbb{P}(|\eta_1 + \dots + \eta_\ell - \ell\mathbb{E}[\eta]| > n^{3/8}) \leq \exp(-\Theta(n^{1/8}))$$

for all $\ell \geq 1$.

The main result of [11] entails that the height of the tree τ_n admits a distributional limit when rescaled by $n^{-1/2}$. In particular, it is smaller than $\sqrt{n} \log n$ with a probability that tends to 1 as $n \rightarrow \infty$. Thus, the probability that a “bad” vertex exists in (τ_n, δ_n) such that Inequality (4.9) fails is bounded by

$$(4.15) \quad o(1) + \sum_{\ell=1}^{\sqrt{n} \log n} \mathbb{P}((\tau_n, \delta_n) \text{ contains a “bad” vertex with height } \ell).$$

We know by (4.2) that τ_n has at most $2n$ vertices in total. Critically applying Equation (4.8) and using Inequality (4.14) and Equation (3.9), it follows that the

sum in (4.15) may be bounded by

$$\begin{aligned}
(4.16) \quad \mathbb{P}(L(\tau) = n) & \sum_{\ell=1}^{\sqrt{n} \log n} \mathbb{P}((\tau, \delta) \text{ contains a "bad" vertex with height } \ell) \\
& \leq O(n^{3/2}) \sum_{\ell=1}^{\sqrt{n} \log n} (2n) \mathbb{P}(\text{the marked vertex of } (\hat{\tau}^{(\ell)}, \hat{\delta}^{(\ell)}) \text{ is bad}) \\
& \leq O(n^{5/2}) \sum_{\ell=1}^{\sqrt{n} \log n} \mathbb{P}(|\eta_1 + \dots + \eta_\ell - \ell \mathbb{E}[\eta]| > n^{3/8}).
\end{aligned}$$

By Inequality (4.14), it follows that this upper bound may be further bounded by $O(n^3 \log n) \exp(-\Theta(n^{1/8}))$, which tends to zero as $n \rightarrow \infty$. This completes the proof. \square

For ease of notation, in everything that follows we will simply consider a vertex $v \in \tau_n$ of the tree τ_N also as a vertex $v \in \mathbf{N}_n$ of the network \mathbf{N}_n . This saves us from repeatedly writing “the vertex of \mathbf{N}_n that corresponds to $v \in \tau_n$ ”.

Note that \mathbf{N}_n is likely to have more vertices than τ_n . Each head structure $\delta_n(v)$ for $v \in \tau_n$ may contribute a surplus of vertices that do not correspond to any vertices of τ_n . These are precisely the non-root non-leaf vertices of $\delta_n(v)$. We will need fine-grained information on the growth of the surplus:

LEMMA 4.4. — *Let $v_1, \dots, v_{|\tau_n|}$ denote the depth-first-search ordered list of vertices of τ_n . For each head structure H let $S(H)$ denote the number of surplus vertices of H , that is, the number of non-root non-leaf vertices of H . Let κ denote the number of surplus vertices of a uniform random head structure from $\mathcal{H}[\xi]$. With a probability that tends to 1 as $n \rightarrow \infty$,*

$$(4.17) \quad \left| \sum_{i=1}^{\ell} S(\delta_n(v_i)) - \ell \mathbb{E}[\kappa] \right| \leq n^{3/4}$$

holds for all $1 \leq \ell \leq |\tau_n|$.

Proof. — Note that κ has finite exponential moments: ξ has finite exponential moments by (3.6) and any head structure from $\mathcal{H}[\xi]$ has at most $2(\xi + k)$ vertices by Proposition 4.1. Hence $\kappa \leq 2(\xi + k)$, entailing that

$$(4.18) \quad \mathbb{E}[(1 + \epsilon)^\kappa] < \infty$$

for some $\epsilon > 0$.

Let $\kappa_1, \kappa_2, \dots$ denote independent copies of κ . The finite exponential moments property (4.18) allows us to argue analogously as for Inequality (4.13), yielding that there is a constant $c > 0$ such that for all sufficiently small $\lambda > 0$ and all $x > 0$ and integers $\ell \geq 1$

$$(4.19) \quad \mathbb{P}(|\kappa_1 + \dots + \kappa_\ell - \ell \mathbb{E}[\kappa]| > x) \leq 2 \frac{(1 + c\lambda^2)^\ell}{\exp(\lambda x)}.$$

Taking $x = n^{3/4}$ and $\lambda = n^{-1/2}$, it follows that uniformly for all $1 \leq \ell \leq 2n$

$$(4.20) \quad \mathbb{P}(|\kappa_1 + \dots + \kappa_\ell - \ell \mathbb{E}[\kappa]| > n^{3/4}) \leq \exp(-\Theta(n^{1/4})).$$

Recall that τ_n has at most $2n$ vertices by Inequality (4.2). Hence the probability that there is a “bad” integer $1 \leq \ell \leq |\tau_n|$ for which (4.17) fails is bounded by

$$(4.21) \quad \mathbb{P}(L(\tau) = n)^{-1} \sum_{m=n}^{2n} \sum_{\ell=1}^m \mathbb{P}(|\tau| = m, \ell \text{ is “bad”})$$

Let ξ_1, ξ_2, \dots denote independent copies of ξ , and for each $i \geq 1$ let $H(\xi_i)$ be uniformly selected from $\mathcal{H}[\xi_i]$. Thus, $S(H(\xi_1)), S(H(\xi_2)), \dots$ are independent copies of κ .

The depth-first-search ordered outdegrees of τ may be described by (ξ_1, \dots, ξ_L) with L denoting the first integer for which $\sum_{i=1}^L (\xi_i - 1) = -1$. The decorations $(\delta(v))_{v \in \tau}$ may be described by H_1, H_2, \dots . Thus the event that $|\tau| = m$ and that ℓ is “bad” implies that

$$(4.22) \quad \left| \sum_{i=1}^{\ell} S(H_i) - \ell \mathbb{E}[\kappa] \right| > n^{3/4}$$

Using (4.20) and (3.9), it follows that the upper bound (4.21) may be bounded further by

$$(4.23) \quad \Theta(n^{3/2}) 2n^2 \exp(-\Theta(n^{1/4})).$$

This bound tends to zero as $n \rightarrow \infty$. Hence the proof is complete. \square

We are now ready to prove Theorems 1.2, 1.3, and 1.4.

Proof of Thm. 1.2. — Let $v_1, \dots, v_{|\tau_n|}$ denote the depth-first-search ordered list of vertices of τ_n , which we will also consider as vertices of $\mathbf{N}_n = \Lambda^{-1}(\mathbf{P}_n)$. For each $1 \leq i \leq |\tau_n|$ let s_i denote the size of the surplus $S(\delta_n(v_i))$ and let $v_{i,1}, \dots, v_{i,s_i}$ denote the vertices of the surplus. Thus,

$$(4.24) \quad (u_1, \dots, u_{|\mathbf{N}_n|}) := (v_1, v_{1,1}, \dots, v_{1,s_1}, \dots, v_{|\tau_n|}, v_{|\tau_n|,1}, \dots, v_{|\tau_n|,s_{|\tau_n|}})$$

is an ordering of the vertices of \mathbf{N}_n .

There is a constant $C > 0$ such that with a probability that tends to 1 as $n \rightarrow \infty$, it holds for all $1 \leq i \leq |\tau_n|$ and all $1 \leq j \leq s_i$ that

$$(4.25) \quad |h_{\mathbf{N}_n}(v_i) - h_{\mathbf{N}_n}(v_{i,j})| \leq C \log n.$$

To see this, note that the difference of distances in (4.25) is bounded by the number of vertices $|\delta_n(v_i)|$ of the head structure $\delta_n(v_i)$. By Proposition 4.1, this is at most $2(d_{\tau_n}^+(v_i) + k)$. by Proposition 4.2 this may be further bounded by

$$2(d_{\tau_n}^+(v_i) + k) \leq 2(\Delta(\tau_n) + k) = O(\log n).$$

Hence (4.25) holds with high probability for all i and j .

Note further that for each $1 \leq r \leq |\mathbf{N}_n|$ there is a unique index $1 \leq i(r) \leq |\tau_n|$ such that either $u_r = v_{i(r)}$ or $u_r = v_{i(r),j}$ for some $1 \leq j \leq s_{i(r)}$. By Inequality (4.25) it follows that

$$(4.26) \quad |h_{\mathbf{N}_n}(u_r) - h_{\mathbf{N}_n}(v_{i(r)})| \leq C \log n$$

holds for all $1 \leq r \leq |\tau_n|$ with a probability that tends to 1 as $n \rightarrow \infty$. By Lemma 4.3 and the triangle inequality, it follows that

$$(4.27) \quad |h_{\mathbf{N}_n}(u_r) - \mathbb{E}[\eta]h_{\tau_n}(v_{i(r)})| \leq O(n^{3/8}).$$

Furthermore, Lemma 4.4 (where $S(\delta_n(v_i))$ corresponds to the notation s_i here) implies that with probability tending to 1 as $n \rightarrow \infty$

$$(4.28) \quad ||\mathbf{N}_n| - (1 + \mathbb{E}[\kappa])|\tau_n|| \leq n^{3/4}$$

and for all $1 \leq r \leq |\mathbf{N}_n|$

$$(4.29) \quad |r - (1 + \mathbb{E}[\kappa])i(r)| \leq O(n^{3/4}).$$

By the main result of [11], there is a constant $b > 0$ such that the height process $(h_{\tau_n}(v_{t|\tau_n}) : 0 \leq t \leq 1)$ (with linear interpolation between values $h_{\tau_n}(v_i)$ for $1 \leq i \leq |\tau_n|$) satisfies

$$(4.30) \quad (bn^{-1/2}h_{\tau_n}(v_{t|\tau_n}) : 0 \leq t \leq 1) \xrightarrow{d} (\mathbf{e}(s) : 0 \leq s \leq 1)$$

as random elements of the space of continuous function $\mathcal{C}([0, 1], \mathbb{R})$.

Now it comes all together: By (4.27), (4.28), (4.29), and (4.30) (and the fact $n \leq |\tau_n| \leq 2n$ from (4.2)) it follows that

$$(4.31) \quad (b\mathbb{E}[\eta]^{-1})n^{-1/2}h_{\mathbf{N}_n}(v_{t|\mathbf{N}_n}) : 0 \leq t \leq 1) \xrightarrow{d} (\mathbf{e}(s) : 0 \leq s \leq 1).$$

Thus, Equation (1.2) holds with

$$(4.32) \quad b_k := b/\mathbb{E}[\eta].$$

Let w_1, \dots, w_n denote the depth-first-search ordered list of leaves of τ_n . Thus, w_1, \dots, w_n also correspond to the leaves of \mathbf{N}_n . The convergence

$$(4.33) \quad (b\mathbb{E}[\eta]^{-1})n^{-1/2}h_{\mathbf{N}_n}(w_{tn}) : 0 \leq t \leq 1) \xrightarrow{d} (\mathbf{e}(s) : 0 \leq s \leq 1).$$

follows by analogous arguments, since the number of vertices between consecutive leaves in the depth-first-search ordered list of vertices of τ_n may be controlled in an analogous fashion as the surplus of the head structures, ensuring that

$$(4.34) \quad (bn^{-1/2}h_{\tau_n}(w_{tn}) : 0 \leq t \leq 1) \xrightarrow{d} (\mathbf{e}(s) : 0 \leq s \leq 1).$$

This completes the proof of Theorem 1.2. \square

Proof of Theorem 1.3. — We want to show that there are constants $C, c > 0$ such that for all $x > 0$ and $n \geq 2$

$$(4.35) \quad \mathbb{P}(H(\mathbf{N}_n) > x) \leq C \exp(-cx^2/n).$$

By (4.2) we know that $|\tau_n| \leq 2n$. By Proposition 4.1 it follows that the total number $|\mathbf{N}_n|$ of vertices in \mathbf{N}_n satisfies

$$\begin{aligned}
 (4.36) \quad |\mathbf{N}_n| &= 1 + \sum_{v \in \tau_n} (|\delta_n(v)| - 1) \\
 &\leq 1 + 2 \sum_{v \in \tau_n} (d_{\tau_n}^+(v) + k) \\
 &= 1 + 2(|\tau_n| - 1) + 2|\tau_n|k \\
 &\leq 4n(k + 1).
 \end{aligned}$$

Consequently, $H(\mathbf{N}_n) \leq 4n(k + 1)$.

Hence it suffices to show (4.35) for $x \leq 4n(k + 1)$. Moreover, we may always take C large enough (depending on c) so that $C \exp(-c) > 1$, implying that (4.35) is automatically fulfilled for all $0 < x < \sqrt{n}$. Thus, it suffices to show the existence of $c, C > 0$ such that (4.35) holds for all

$$(4.37) \quad \sqrt{n} \leq x \leq 4n(k + 1).$$

The main theorem of [13] establishes tail-bounds for the height of critical Galton–Watson trees conditioned on having n vertices if the offspring distribution has finite variance. In [19, Lem. 6.61, Eq. (6.41)] this result and further observations from [13] were used to deduce the similar bounds for blow-ups of such trees. By [12], [19, Sec. 6.1.7] we can view a Galton–Watson tree conditioned on having n leaves as a blow-up of a different Galton–Watson tree conditioned on having n vertices, allowing us to apply [19, Lem. 6.61, Eq. (6.41)] to deduce that there are constants $c_1, C_1 > 0$ such that for all $x > 0$ and $n \geq 1$

$$(4.38) \quad \mathbb{P}(H(\tau_n) > x) \leq C_1 \exp(-c_1 x^2/n).$$

Let $\epsilon > 0$ be given. We will choose a suitable value for ϵ later on. Inequality (4.38) entails that

$$(4.39) \quad \mathbb{P}(H(\tau_n) > \epsilon x) \leq C_1 \exp(-c_1 \epsilon^2 x^2/n).$$

A vertex of \mathbf{N}_n with maximal height may either be a vertex that pertains to τ_n or a vertex that lies in the surplus of some head structure. (The later case is possible since we look at directed paths. Thus a vertex with maximal height in \mathbf{N}_n is *not* necessarily a leaf of \mathbf{N}_n .) If such a vertex u lies in the surplus of some head structure $\delta_n(v)$ for $v \in \tau_n$, then

$$h_{\mathbf{N}_n}(u) \leq h_{\mathbf{N}_n}(v) + |\delta_n(v)|.$$

Hence, if $h_{\mathbf{N}_n}(u) > x$, then $h_{\mathbf{N}_n}(v) > x/2$ or $|\delta_n(v)| > x/2$. It follows that

$$\begin{aligned}
 (4.40) \quad \mathbb{P}(H(\tau_n) \leq \epsilon x, H(\mathbf{N}_n) > x) \\
 \leq \mathbb{P}(\max_{v \in \tau_n} |\delta_n(v)| > x/2) + \mathbb{P}(H(\tau_n) \leq \epsilon x, \max_{v \in \tau_n} h_{\mathbf{N}_n}(v) > x/2).
 \end{aligned}$$

By Proposition (4.1) and Equation (4.3) it follows that

$$(4.41) \quad \mathbb{P}\left(\max_{v \in \tau_n} |\delta_n(v)| > x/2\right) \leq \mathbb{P}(\Delta(\tau_n) > x/4 - k) \\ \leq O(n^{5/2})\mathbb{P}(\xi > x/4 - k).$$

Recall that by (4.37) we assumed that $\sqrt{n} \leq x \leq 4n(k+1)$. Hence, using Inequality (4.4) it follows that

$$(4.42) \quad \mathbb{P}\left(\max_{v \in \tau_n} |\delta_n(v)| > x/2\right) \leq O(n^{5/2}) \exp(-\Theta(x)) \\ \leq \exp(-\Theta(x)) \\ \leq \exp(-\Theta(x^2/n)).$$

In order to bound the probability for the event that simultaneously $H(\tau_n) \leq \epsilon x$ and $\max_{v \in \tau_n} h_{\mathbf{N}_n}(v) > x/2$, we may argue identically as for (4.16) to obtain

$$(4.43) \quad \mathbb{P}(H(\tau_n) \leq \epsilon x, \max_{v \in \tau_n} h_{\mathbf{N}_n}(v) > x/2) \leq O(n^{5/2}) \sum_{\ell=1}^{\lfloor \epsilon x \rfloor} \mathbb{P}(\eta_1 + \dots + \eta_\ell > x/2).$$

Here η_1, η_2, \dots denote independent copies of η .

Setting $\epsilon := 1/(4\mathbb{E}[\eta])$, it follows that $x/2 - \mathbb{E}[\eta]\ell \geq x/4$ for all $1 \leq \ell \leq \epsilon x$. Hence, by (4.13), it follows that there is a constant $c_2 > 0$ such that for all sufficiently small λ (independent of ℓ and x) and all $1 \leq \ell \leq \epsilon x$

$$(4.44) \quad \mathbb{P}(\eta_1 + \dots + \eta_\ell > x/2) \leq 2 \frac{(1 + c_2 \lambda^2)^\ell}{\exp(\lambda x/4)} \leq 2 \frac{(1 + c_2 \lambda^2)^{x/(4\mathbb{E}[\eta])}}{\exp(\lambda x/4)}$$

Choosing λ small enough, it follows that there are constants $C_2, c_3 > 0$ such that

$$(4.45) \quad \mathbb{P}(\eta_1 + \dots + \eta_\ell > x) \leq C_2 \exp(-c_3 x).$$

Recall that we assumed $\sqrt{n} \leq x \leq 4n(k+1)$ by (4.37), hence it follows from (4.43) that

$$(4.46) \quad \mathbb{P}(H(\tau_n) \leq \epsilon x, \max_{v \in \tau_n} h_{\mathbf{N}_n}(v) > x/2) \leq O(n^{7/2}) \exp(-c_3 x) \\ \leq \exp(-\Theta(x)) \\ \leq \exp(-\Theta(x^2/n)).$$

Combining (4.39), (4.40), (4.42) and (4.46), it follows that

$$(4.47) \quad \mathbb{P}(H(\mathbf{N}_n) > x) \leq C_3 \exp(-c_4 x^2/n)$$

for some constants $C_3, c_4 > 0$ that do not depend on x or n . \square

Proof of Thm. 1.4. — We define η' analogously to η , as the length of a shortest *undirected* path from the root of a uniformly selected network from $\mathcal{H}[\hat{\xi}]$ to a leaf that is uniformly selected among its $\hat{\xi}$ leaves. That is, the path may cross edges in any direction, regardless of their orientation.

Lemma 4.3 holds analogously for undirected paths. That is, interpreting the vertices of τ_n as part of \mathbf{G}_n ,

$$(4.48) \quad \sup_{v \in \tau_n} |\mathbf{h}_{\mathbf{G}_n}(v) - \mathbb{E}[\eta'] \mathbf{h}_{\tau_n}(v)| \leq n^{3/8}$$

holds with probability tending to 1 as $n \rightarrow \infty$. Here $\mathbf{h}_{\mathbf{G}_n}(v)$ denotes the graph distance $d_{\mathbf{G}_n}(o, v)$ from the root vertex o of \mathbf{G}_n to the vertex v . Hence Thm. 1.2 and all intermediate observations in its proof hold analogously for undirected paths, that is,

$$(4.49) \quad (b'_k n^{-1/2} \mathbf{h}_{s|\mathbf{G}_n|} : 0 \leq s \leq 1) \xrightarrow{d} (\mathbf{e}(s) : 0 \leq s \leq 1)$$

for

$$(4.50) \quad b'_k := b/\mathbb{E}[\eta']$$

as $n \rightarrow \infty$.

Given two vertices $v, w \in \tau_n$ there is a unique path Q in τ_n that joins them in τ_n , but there may be several different paths P in \mathbf{G}_n that join v and w in \mathbf{G}_n . However, we know that the vertices and edges of P are always entirely contained in the head-structures $(\delta_n(u))_{u \in Q}$. Let $\text{lca}(u, v)$ denote the lowest common ancestor of v and w in the tree τ_n . The distance d_{τ_n} in the tree τ_n satisfies

$$(4.51) \quad d_{\tau_n}(v, w) = \mathbf{h}_{\tau_n}(v) + \mathbf{h}_{\tau_n}(w) - 2\mathbf{h}_{\tau_n}(\text{lca}(v, w)).$$

A similar statement holds for the distance $d_{\mathbf{G}_n}$ in \mathbf{G}_n : If P_v and P_w are *shortest* paths in \mathbf{G}_n from v to the root o and from w to o , we may construct a shortest path P from v to w by following P_v until we encounter for the first time a vertex v' from $\delta_n(\text{lca}(v, w))$, walking from v' to the analogously defined vertex w' from $\delta_n(\text{lca}(v, w))$ by a path that lies entirely in $\delta_n(\text{lca}(v, w))$, and then following P_w (in its reverse direction) from w' back to w . This entails that

$$(4.52) \quad d_{\mathbf{G}_n}(v, w) = \mathbf{h}_{\mathbf{G}_n}(v) + \mathbf{h}_{\mathbf{G}_n}(w) - 2\mathbf{h}_{\mathbf{G}_n}(\text{lca}(v, w)) + R(v, w)$$

for an error term $R(v, w)$ satisfying

$$(4.53) \quad |R(v, w)| \leq 3|\delta_n(\text{lca}(v, w))|.$$

Recall that in Equation (4.24) we constructed a special ordering

$$(4.54) \quad (u_1, \dots, u_{|\mathbf{N}_n|}) := (v_1, v_{1,1}, \dots, v_{1,s_1}, \dots, v_{|\tau_n|}, v_{|\tau_n|,1}, \dots, v_{|\tau_n|,s_{|\tau_n|}})$$

of the vertices of \mathbf{N}_n . Here $v_1, \dots, v_{|\tau_n|}$ is a depth-first-search ordering of the vertices of τ_n , where, say, we always proceed along the left-most unvisited vertex. This way, it holds for all $1 \leq i \leq j \leq |\mathbf{G}_n|$ that

$$(4.55) \quad \mathbf{h}_{\mathbf{G}_n}(\text{lca}(v_i, v_j)) = \min_{i \leq r \leq j} \mathbf{h}_{\mathbf{G}_n}(v_r).$$

By Proposition 4.1 and Proposition 4.2, it follows analogously as for (4.26) that there is a constant $C' > 0$ such that

$$(4.56) \quad \max_{v \in \tau_n} |\delta_n(v)| \leq 2(\Delta(\tau_n) + k) \leq C' \log n$$

with a probability that tends to 1 as $n \rightarrow \infty$. It follows from Equation (4.52) that hence there is a constant $C > 0$ such that with a probability that tends to 1 as $n \rightarrow \infty$

$$(4.57) \quad \left| d_{\mathbf{G}_n}(u_i, u_j) - \left(h_{\mathbf{G}_n}(u_i) + h_{\mathbf{G}_n}(u_j) - 2 \min_{i \leq r \leq j} h_{\mathbf{G}_n}(u_r) \right) \right| \leq C \log n$$

for all $1 \leq i \leq j \leq |\mathbf{N}_n|$. Multiplying both sides of (4.57) by $b'_k n^{-1/2}$, Thm. 1.4 now follows from (4.49) and the definition (1.11) of the metric of the Brownian continuum random tree. \square

5. The asymptotic local shape

Using the notation from Section 4.2, we may set

$$(5.1) \quad (\hat{\tau}, \hat{\delta}) := (\hat{\tau}^{(\infty)}, \hat{\delta}^{(\infty)})$$

and

$$(5.2) \quad \hat{\mathbf{N}} := \Lambda^{-1}(\hat{\tau}, \hat{\delta}).$$

Here, by a slight abuse of notation, we use Λ^{-1} to denote the canonical extension of the blow-up procedure Λ^{-1} (defined for finite decorated trees) to infinite locally finite decorated trees. The infinite network $\hat{\mathbf{N}}$ is the local weak limit of \mathbf{N}_n as $n \rightarrow \infty$:

Proof of Thm. 1.5. — General results for the local convergence of conditioned Galton–Watson trees [30] imply that

$$(5.3) \quad \tau_n \xrightarrow{d} \hat{\tau}$$

in the local topology. That is, for any fixed constant $\ell \geq 0$ the probability for the ℓ -neighbourhood of the root of τ_n to assume a given shape converges as $n \rightarrow \infty$ to the probability for the ℓ -neighbourhood of the root of $\hat{\tau}$ to assume that shape.

By Skorokhod’s representation theorem we may assume that $\hat{\tau}, \tau_2, \tau_3, \dots$ are coupled such that

$$(5.4) \quad \tau_n \xrightarrow{a.s.} \hat{\tau}.$$

This entails

$$(5.5) \quad (\tau_n, \delta_n) \xrightarrow{a.s.} (\hat{\tau}, \hat{\delta}).$$

Hence

$$(5.6) \quad \mathbf{N}_n = \Lambda^{-1}(\tau_n, \delta_n) \xrightarrow{a.s.} \Lambda^{-1}(\hat{\tau}, \hat{\delta}) = \hat{\mathbf{N}}.$$

This proves the local weak convergence in (1.6).

Kersting [31, Thm. 5] described the asymptotic shape of $o(\sqrt{n})$ -neighbourhoods of critical Galton–Watson trees conditioned on having n vertices (if the offspring distribution has finite variance). Using a transformation from [12, 28], it follows that this also holds for Galton–Watson trees conditioned on having n leaves, yielding

$$(5.7) \quad d_{\text{TV}}(U_{\ell_n}(\tau_n), U_{\ell_n}(\hat{\tau})) \rightarrow 0$$

for any sequence ℓ_n of positive integers satisfying $\ell_n = o(\sqrt{n})$. Consequently, we may assume that $\hat{\tau}, \tau_1, \tau_2, \dots$ are coupled such that

$$(5.8) \quad U_{\ell_n}(\tau_n) = U_{\ell_n}(\hat{\tau})$$

holds with a probability that tends to 1 as $n \rightarrow \infty$. Consequently, we may construct the decorations in such a way that

$$(5.9) \quad (U_{\ell_n}(\tau_n), (\delta_n(v))_{v \in U_{\ell_n}(\tau_n)}) = (U_{\ell_n}(\hat{\tau}), (\hat{\delta}(v))_{v \in U_{\ell_n}(\hat{\tau})})$$

holds with a probability tending to 1 as $n \rightarrow \infty$. This entails that

$$(5.10) \quad U_{\ell_n}(\mathbf{N}_n) = U_{\ell_n}(\Lambda^{-1}(\tau_n, \delta_n)) = U_{\ell_n}(\Lambda^{-1}(\hat{\tau}, \hat{\delta})) = U_{\ell_n}(\hat{\mathbf{N}})$$

again holds with probability tending to 1 as $n \rightarrow \infty$. This implies Equation (1.13) and completes the proof. \square

The surplus vertices of the head structures influence the location of a uniformly selected vertex. To take them into account, we form the tree \mathcal{T}_n by colouring the vertices of τ_n blue and adding to each vertex $v \in \tau_n$ additional $S(\delta_n(v))$ red children. We define \mathcal{T} in same way by adding the surplus vertices of (τ, δ) as red children at the appropriate vertices, making \mathcal{T} a two-type Galton–Watson tree with offspring distribution (ξ, κ) , with κ defined as in Lemma 4.4. (This makes ξ and κ dependent on each other.) A uniformly selected vertex u_n of \mathbf{N}_n hence corresponds to a uniformly selected vertex of \mathcal{T}_n .

By a general result [32, Thm. 3] for conditioned multi-type Galton–Watson trees, there is a random 2-type tree $\hat{\mathcal{T}}^*$ such that

$$(5.11) \quad (\mathcal{T}_n, u_n) \xrightarrow{d} \hat{\mathcal{T}}^*$$

in the local topology as $n \rightarrow \infty$. The tree $\hat{\mathcal{T}}^*$ has a marked vertex with a finite number of descendants, and an infinite number of ancestors, each having a random finite number of descendants in total. (This agrees with the intuition that most vertices of \mathcal{T}_n are far from the root and have few descendants.) We may assign random decorations $\hat{\delta}^*(v)$, $v \in \hat{\mathcal{T}}^*$ in the same way as before (for δ , δ_n , and $\hat{\delta}$), allowing us to define the infinite network

$$(5.12) \quad \hat{\mathbf{N}}^* := \Lambda^{-1}(\hat{\mathcal{T}}^*, \hat{\delta}^*).$$

The network $\hat{\mathbf{N}}^*$ describes the asymptotic shape of the vicinity a random vertex of \mathbf{N}_n , similarly as $\hat{\mathbf{N}}$ describes the vicinity of the fixed root vertex of \mathbf{N}_n :

Proof of Thm. 1.6. — Applying Skorokhod’s representation theorem to (5.11), we may assume that $\hat{\mathcal{T}}^*, (\mathcal{T}_2, u_2), (\mathcal{T}_3, u_3), \dots$ are coupled such that

$$(5.13) \quad (\mathcal{T}_n, u_n) \xrightarrow{a.s.} \hat{\mathcal{T}}^*.$$

This entails

$$(5.14) \quad ((\mathcal{T}_n, u_n), \delta_n) \xrightarrow{a.s.} (\hat{\mathcal{T}}^*, \hat{\delta}^*).$$

Hence

$$(5.15) \quad (\mathbf{N}_n, u_n) = (\Lambda^{-1}(\mathcal{T}_n, \delta_n), u_n) \xrightarrow{a.s.} \hat{\mathbf{N}}^*.$$

This proves the local weak convergence in (1.15).

In order to prove the quenched convergence in (1.16), note that stating (1.16) to hold for all $\ell \geq 1$ and all vertex-marked graphs G is equivalent to stating that the random probability measure $\mathfrak{L}((\mathbf{N}_n, u_n) \mid \mathbf{N}_n)$ given by the uniform measure on the $|\mathbf{N}_n|$ many vertex-marked versions of \mathbf{N}_n satisfies

$$(5.16) \quad \mathfrak{L}((\mathbf{N}_n, u_n) \mid \mathbf{N}_n) \xrightarrow{d} \mathfrak{L}(\hat{\mathbf{N}}^*),$$

with $\mathfrak{L}(\hat{\mathbf{N}}^*)$ denoting the deterministic law of $\hat{\mathbf{N}}^*$. Now, [32, Thm. 3] ensures such a quenched limit for the uniform measure $\mathfrak{L}((\mathcal{T}_n, u_n) \mid \mathcal{T}_n)$ on the $|\mathcal{T}_n|$ -many ($= |\mathbf{N}_n|$ many) vertex marked versions of \mathcal{T}_n , with the deterministic limit measure given by the law $\mathfrak{L}(\hat{\mathcal{T}}^*)$ of $\hat{\mathcal{T}}^*$. That is,

$$(5.17) \quad \mathfrak{L}((\mathcal{T}_n, u_n) \mid \mathcal{T}_n) \xrightarrow{d} \mathfrak{L}(\hat{\mathcal{T}}^*).$$

Using the Chernoff bounds, this implies such a limit for the decorated versions:

$$(5.18) \quad \mathfrak{L}(((\mathcal{T}_n, u_n), \delta_n) \mid \mathcal{T}_n) \xrightarrow{d} \mathfrak{L}(\hat{\mathcal{T}}^*, \hat{\delta}^*).$$

The limit (5.16) now follows from (5.18) by applying the blow-up procedure Λ^{-1} and the continuous mapping theorem. This completes the proof. \square

References

- [1] L. van Iersel and V. Moulton, “Trinets encode tree-child and level-2 phylogenetic networks,” *Journal of Mathematical Biology* **68** (2014) 1707–1729.
- [2] C. Linder, B. Moret, L. Nakhleh, and T. Warnow, “Network (reticulate) evolution: Biology, models, and algorithms,” *The Pacific Symposium on Biocomputing* (01, 2004)
- [3] M. Fuchs, G.-R. Yu, and L. Zhang, “Asymptotic enumeration and distributional properties of galled networks,” [arXiv:2010.13324](https://arxiv.org/abs/2010.13324) [math.CO].
- [4] F. Bienvenu, A. Lambert, and M. Steel, “Combinatorial and stochastic properties of ranked tree-child networks,” [arXiv:2007.09701](https://arxiv.org/abs/2007.09701) [math.PR].
- [5] M. Bouvel, P. Gambette, and M. Mansouri, “Counting phylogenetic networks of level 1 and 2,” *Journal of Mathematical Biology* **81** (2020) 1357–1395.
- [6] M. Fuchs, B. Gittenberger, and M. Mansouri, “Counting phylogenetic networks with few reticulation vertices: tree-child and normal networks,” *Australas. J. Combin.* **73** (2019) 385–423.
- [7] C. McDiarmid, C. Semple, and D. Welsh, “Counting phylogenetic networks,” *Ann. Comb.* **19** no. 1, (2015) 205–224. <https://doi.org/10.1007/s00026-015-0260-2>.
- [8] G. Cardona and L. Zhang, “Counting and enumerating tree-child networks and their subclasses,” *J. Comput. System Sci.* **114** (2020) 84–104. <https://doi.org/10.1016/j.jcss.2020.06.001>.
- [9] M. Fuchs, B. Gittenberger, and M. Mansouri, “Counting phylogenetic networks with few reticulation vertices: Exact enumeration and corrections,” [arXiv:2006.15784](https://arxiv.org/abs/2006.15784) [math.CO].
- [10] B. Bollobás, *Modern graph theory*, vol. 184 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998. <https://doi.org/10.1007/978-1-4612-0619-4>.
- [11] I. Kortchemski, “Invariance principles for Galton-Watson trees conditioned on the number of leaves,” *Stochastic Process. Appl.* **122** no. 9, (2012) 3126–3172. <http://dx.doi.org/10.1016/j.spa.2012.05.013>.

- [12] D. Rizzolo, “Scaling limits of Markov branching trees and Galton-Watson trees conditioned on the number of vertices with out-degree in a given set,” *Ann. Inst. Henri Poincaré Probab. Stat.* **51** no. 2, (2015) 512–532. <https://doi.org/10.1214/13-AIHP594>.
- [13] L. Addario-Berry, L. Devroye, and S. Janson, “Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees,” *Ann. Probab.* **41** no. 2, (2013) 1072–1087. <https://doi.org/10.1214/12-AOP758>.
- [14] D. Aldous, “The continuum random tree. I,” *Ann. Probab.* **19** no. 1, (1991) 1–28. [http://links.jstor.org/sici?sici=0091-1798\(199101\)19:1<1:TCRTI>2.0.CO;2-B&origin=MSN](http://links.jstor.org/sici?sici=0091-1798(199101)19:1<1:TCRTI>2.0.CO;2-B&origin=MSN).
- [15] D. Aldous, “The continuum random tree. II. An overview,” in *Stochastic analysis (Durham, 1990)*, vol. 167 of *London Math. Soc. Lecture Note Ser.*, pp. 23–70. Cambridge Univ. Press, Cambridge, 1991. <http://dx.doi.org/10.1017/CB09780511662980.003>.
- [16] D. Aldous, “The continuum random tree. III,” *Ann. Probab.* **21** no. 1, (1993) 248–289. [http://links.jstor.org/sici?sici=0091-1798\(199301\)21:1<248:TCRTI>2.0.CO;2-1&origin=MSN](http://links.jstor.org/sici?sici=0091-1798(199301)21:1<248:TCRTI>2.0.CO;2-1&origin=MSN).
- [17] N. Curien, B. Haas, and I. Kortchemski, “The CRT is the scaling limit of random dissections,” *Random Structures Algorithms* **47** no. 2, (2015) 304–327. <http://dx.doi.org/10.1002/rsa.20554>.
- [18] K. Panagiotou, B. Stuffer, and K. Weller, “Scaling limits of random graphs from subcritical classes,” *Ann. Probab.* **44** no. 5, (2016) 3291–3334.
- [19] B. Stuffer, “Limits of random tree-like discrete structures,” *Probab. Surv.* **17** (2020) 318–477. <https://doi.org/10.1214/19-PS338>.
- [20] B. Stuffer, “Random enriched trees with applications to random graphs,” *Electronic Journal of Combinatorics* **25** no. 3, (2018) .
- [21] G. Miermont, “Tessellations of random maps of arbitrary genus,” *Ann. Sci. Éc. Norm. Supér. (4)* **42** no. 5, (2009) 725–781. <http://dx.doi.org/10.24033/asens.2108>.
- [22] V. Kurauskas, “On local weak limit and subgraph counts for sparse random graphs,” *ArXiv e-prints* (Apr., 2015) , [arXiv:1504.08103](https://arxiv.org/abs/1504.08103) [math.PR].
- [23] L. van Iersel, J. Keijsper, S. Kelk, L. Stougie, F. Hagen, and T. Boekhout, “Constructing level-2 phylogenetic networks from triplets,” in *Research in computational molecular biology*, vol. 4955 of *Lecture Notes in Comput. Sci.*, pp. 450–462. Springer, Berlin, 2008. https://doi.org/10.1007/978-3-540-78839-3_40.
- [24] P. Gambette, V. Berry, and C. Paul, “The structure of level- k phylogenetic networks,” in *Combinatorial pattern matching*, vol. 5577 of *Lecture Notes in Comput. Sci.*, pp. 289–300. Springer, Berlin, 2009. https://doi.org/10.1007/978-3-642-02441-2_26.
- [25] A. Joyal, “Une théorie combinatoire des séries formelles,” *Adv. in Math.* **42** no. 1, (1981) 1–82. [http://dx.doi.org/10.1016/0001-8708\(81\)90052-9](http://dx.doi.org/10.1016/0001-8708(81)90052-9).
- [26] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, vol. 67 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1998. Translated from the 1994 French original by Margaret Readdy, With a foreword by Gian-Carlo Rota.
- [27] P. Flajolet and R. Sedgewick, *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009. <http://dx.doi.org/10.1017/CB09780511801655>.
- [28] R. Ehrenborg and M. Méndez, “Schröder parenthesizations and chordates,” *J. Combin. Theory Ser. A* **67** no. 2, (1994) 127–139. [http://dx.doi.org/10.1016/0097-3165\(94\)90008-6](http://dx.doi.org/10.1016/0097-3165(94)90008-6).
- [29] M. Drmota, *Random trees*. SpringerWienNewYork, Vienna, 2009. <http://dx.doi.org/10.1007/978-3-211-75357-6>. An interplay between combinatorics and probability.

- [30] R. Abraham and J.-F. Delmas, “Local limits of conditioned Galton-Watson trees: the infinite spine case,” *Electron. J. Probab.* **19** (2014) no. 2, 19.
- [31] G. Kersting, “On the height profile of a conditioned galton-watson tree,” [arXiv:1101.3656](https://arxiv.org/abs/1101.3656) [math.PR].
- [32] B. Stufler, “Rerooting multi-type branching trees: The infinite spine case,” *Journal of Theoretical Probability* (2021) .
<https://link.springer.com/article/10.1007/s10959-020-01069-y>.

BENEDIKT STUFLENER, Vienna University of Technology • *E-mail* : benedikt.stufler at
tuwien.ac.at