The continuum random tree is the scaling limit of unlabelled unrooted trees

Benedikt Stufler

The continuum random tree is the scaling limit of unlabelled unrooted trees

Benedikt Stufler*

Abstract

We show that the uniform unlabelled unrooted tree with n vertices and vertex degrees in a fixed set converges in the Gromov–Hausdorff sense after a suitable rescaling to the Brownian continuum random tree. This confirms a conjecture by Aldous (1991). We also establish Benjamini–Schramm convergence of this model of random trees and provide a general approximation result, that allows for a transfer of a wide range of asymptotic properties of extremal and additive graph parameters from Pólya trees to unrooted trees.

1 Introduction and main results

Combinatorial trees are classical mathematical objects and crop up in a variety of fields [29, 16, 17]. In the present work we take a probabilistic approach to study unordered trees without labels. Here one distinguishes between Pólya trees, which have a root, and unlabelled (unrooted) trees. It has been a long-standing conjecture by Aldous [4, p. 55] that the continuum random tree (CRT) arises as scaling limit of these models of random trees. Marckert and Miermont [28] treated the case of binary unordered rooted trees. The convergence of random (unrestricted) Pólya trees was confirmed by Haas and Miermont [24] using new methods, and an alternative proof has been given later by Panagiotou and Stufler [30]. As was also mentioned in [24, p. 18], this does not settle the question regarding the convergence of random unlabelled unrooted trees. The main challenge for these structures is the complexity of their symmetries. Rooted trees have a simpler structure, as any automorphism is required to fix the root vertex. Our first main result confirms the CRT as scaling limit of unlabelled unrooted trees as their number of vertices becomes large, confirming Aldous' conjecture for these structures. We take a unified approach to cover all (sensible) cases of vertex degree restrictions.

Throughout, we let Ω denote a fixed set of positive integers containing 1 and at least one integer equal or larger than 3, and set $\Omega^* = \Omega - 1$. Let T_n be drawn uniformly at random from the unlabelled trees with n vertices and vertex-degrees in Ω, and let A_{n-1} denote the random Pólya tree selected uniformly among all such trees with $n-1$ vertices and outdegrees in the shifted set Ω^*. See Figure 1 and 2 for illustrations of these structures.

Theorem 1.1. *There is a constant e_Ω such that*

$$(\mathsf{T}_n, e_\Omega n^{-1/2} d_{\mathsf{T}_n}) \xrightarrow{d} (\mathcal{T}_\mathsf{e}, d_{\mathcal{T}_\mathsf{e}}) \tag{1.1}$$

in the Gromov–Hausdorff sense, as $n \equiv 2 \mod \gcd(\Omega^*)$ becomes large. Moreover, there are constants $C, c > 0$ such that the diameter $\mathsf{D}(\mathsf{T}_n)$ satisfies the tail bound

$$\mathbb{P}(\mathsf{D}(\mathsf{T}_n) \ge x) \le C \exp(-cx^2/n) \tag{1.2}$$

for all n and $x \ge 0$.

*École Normale Supérieure de Lyon, E-mail: benedikt.stufler@ens-lyon.fr; The author is supported by the German Research Foundation DFG, STU 679/1-1

MSC2010 subject classifications. Primary 60C05; secondary 05C80.
Keywords and phrases. combinatorial trees, continuum random tree, scaling limits

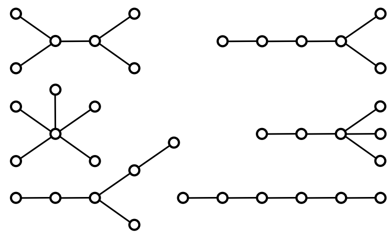

Figure 1: All unlabelled unrooted trees with 6 vertices.

The CRT $(\mathcal{T}_{\mathsf{e}}, d_{\mathcal{T}_{\mathsf{e}}})$ plays a central role in the study of the geometric shape of large discrete structures. It crops up as scaling limit for a variety of models [3, 10, 14, 15, 25, 31] and incited research in further directions [1, 2]. Although scaling limits describe asymptotic global properties, they do not contain information on local properties, such as the limiting degree distribution of a randomly chosen vertex in a graph. Such asymptotic local properties of random rooted structures are described by Benjamini–Schramm limits [5, 23, 8]. Our second main result establishes Benjamini–Schramm convergence for random unlabelled unrooted trees toward an infinite limit tree.

Theorem 1.2. *The random unrooted tree T_n converges in the Benjamini–Schramm sense toward an infinite rooted tree A_{Ω^*}, as $n \equiv 2 \mod \gcd(\Omega^*)$ becomes large. Even stronger, if v_n denotes a uniformly at random selected vertex of the tree T_n, then for each sequence $k_n = o(\sqrt{n})$ the radius k_n graph neighbourhood $V_{k_n}(\cdot)$ satisfies*

$$d_{\mathsf{TV}}(V_{k_n}(\mathsf{T}_n, v_n), V_{k_n}(\mathsf{A}_{\Omega^*})) \to 0. \tag{1.3}$$

Here d_{TV} denotes the total variation distance. Note that this form of convergence is best possible, as (1.3) fails if the order of k_n is comparable to $\sqrt{n}$. In the case $\Omega = \mathbb{N}$, Benjamini–Schramm convergence for T_n was independently obtained by Georgakopoulos and Wagner [22] using different techniques. Our methods for the proof of Theorems 1.1 and 1.2 are based on the cycle pointing decomposition established recently by Bodirsky, Fusy, Kang and Vigerske [11]. This novel and effective centering method differs fundamentally from classical approaches, such as the geometric center, and applies to arbitrary classes of combinatorial structures. We use it to approximate the random unlabelled unrooted tree T_n with n vertices and vertex outdegrees in a set Ω, by random Pólya trees with vertex outdegrees in the shifted set $\Omega^* = \Omega - 1$, whose random sizes concentrate around n. The approximation works not only for graph limits, but actually for a large range of additive and extremal graph parameters.

Theorem 1.3. *There are constants $C, c > 0$, a random number $K_n = n + O_p(1) \leq n$, and a coupling of the randomly sized Pólya tree A_{K_n} with a tree B_n having stochastically bounded size $n - K_n + 1$, such that the random tree $\bar{\mathsf{T}}_n$ obtained by identifying the root vertices of A_{K_n} and B_n satisfies*

$$d_{\mathsf{TV}}(\mathsf{T}_n, \bar{\mathsf{T}}_n) \leq C \exp(-cn)$$

for all n.

Theorem 1.3 establishes in full generality how a random unrooted tree may be approximated by a single large random rooted tree having the property, that when conditioned on having a fixed size, it is uniformly distributed among all Pólya trees with this size and the given vertex outdegree restrictions. This has far reaching consequences and underlines the advantages of this approach. It implies that for a very large set of graph theoretic properties (maximum degree, degree distribution, subtree counts, ...) everything known (present and future) about random Pólya trees also applies to random unlabelled unrooted trees, erasing the need to study uniform unrooted unordered trees directly. For example, Haas and Miermont [24, Thm. 9, Cor. 10] established Gromov–Hausdorff–Prokhorov scaling limits for uniform unordered rooted trees endowed with the uniform measure on their leaves or on all their vertices, if the

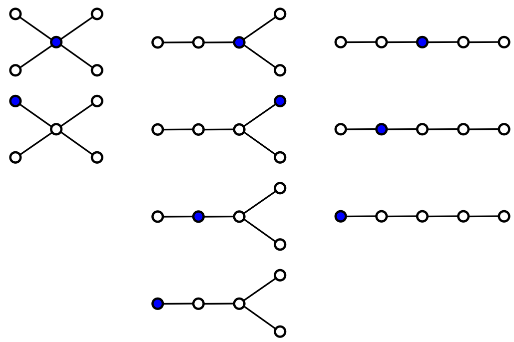

Figure 2: All Pólya trees with 5 vertices.

vertex out-degrees are restricted to a set of the form $\Omega^* = \mathbb{N}_0$, $\Omega^* = \{0, d\}$ or $\Omega^* = \{0, \dots, d\}$ for some $d \ge 2$. Using this result, it follows easily from Theorem 1.3 that the uniform vertex degree restricted unrooted tree T_n with vertex degrees in $\Omega = \Omega^* + 1$ also converges in the Gromov–Hausdorff–Prokhorov sense, thus strengthening the convergence of Theorem 1.1 for these cases. But again, it is not about for which cases of vertex-degree restrictions we may deduce convergence at the moment. The contribution of Theorem 1.3 is that "practically all" properties of random unordered rooted trees get transferred automatically to the unrooted case, regardless of the extend to which they are understood at present.

Thus, Theorem 1.3 provides a rigorous justification of the empirically backed and widely believed fact that rooted and unrooted trees behave asymptotically similarly. Note that this does *not* imply that almost all unrooted trees are asymmetric (meaning the absence of non-trivial symmetries) or possess as much possible root locations as vertices. Some discrete structures such as planar maps with half-edges as atoms have such properties, and hence a purely enumerative argument suffices to show that the asymptotic study of these objects is equivalent to the study of half-edge rooted planar maps. The case of unordered trees is different, as the probability for the random tree T_n to be asymmetric is bounded away from 1, as is the probability for the event that rooting it at each of its n vertices yields n distinct trees. Moreover, the approximation argument of Theorem 1.3 does not appear to work as well in the other direction. For example, the convergence of T_n (in the local sense, or in the sense of scaling limits) may be used to obtain convergence of a random Pólya-tree having a *random* number of vertices (depending on n), but, although this number concentrates, this is not sufficient to deduce convergence of a random Pólya tree with a deterministic size that becomes large. Hence the most economic approach is really to study Pólya trees and then transfer the results to random unlabelled unrooted trees. Furthermore, in [30] it was shown how asymptotic properties of conditioned Galton–Watson trees may be transferred to random Pólya trees, which by the results of the present work hence also apply to the unrooted model. As Galton–Watson trees are without doubt the best understood model of random trees in probability theory, it is natural to pave the way for building on this knowledge.

In [11] the cycle pointing method was developed for the enumeration and efficient sampling of discrete structures. The present work demonstrates for the important classical example of unlabelled trees how a combination with a probabilistic approach allows us to answer a large number of questions related to the study of asymptotic properties of random discrete structures. Due to the generality of the involved methods this will likely stimulate probabilistic applications to further classes of discrete structures, such as models of random unlabelled graphs.

1.1 Combinatorial applications of the scaling limit

A direct consequence of the scaling limit in Theorem 1.1 is that the rescaled diameter $e_\Omega n^{-1/2}\mathsf{D}(\mathsf{T}_n)$ converges weakly and in arbitrarily high moments toward the diameter $\mathsf{D}(\mathcal{T}_\mathsf{e})$ of the CRT. That is,

$$\mathbb{P}\left(n^{-1/2} e_\Omega \mathsf{D}(\mathsf{T}_n) > x\right) \to \mathbb{P}\left(\mathsf{D}(\mathcal{T}_\mathsf{e}) > x\right),$$

and
$$\mathbb{E}\left[\mathsf{D}(\mathsf{T}_n)^p\right] \sim e_\Omega^{-p} n^{p/2} \mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})^p\right].$$
The distribution of $\mathsf{D}(\mathcal{T}_\mathsf{e})$ is known and given by
$$\mathsf{D}(\mathcal{T}_\mathsf{e}) \stackrel{(d)}{=} \sup_{0\le t_1\le t_2\le 1}(\mathsf{e}(t_1)+\mathsf{e}(t_2)-2\inf_{t_1\le t\le t_2}\mathsf{e}(t)), \tag{1.4}$$
with $\mathsf{e} = (\mathsf{e}_t)_{0\le t\le 1}$ denoting Brownian excursion of length 1, and
$$\mathbb{P}\left(\mathsf{D}(\mathcal{T}_\mathsf{e}) > x\right) = \sum_{k=1}^{\infty}(k^2-1)\Big(\frac{2}{3}k^4x^4 - 4k^2x^2 + 2\Big)\exp(-k^2x^2/2). \tag{1.5}$$
Equations (1.4) and (1.5) have been established by Aldous [4, Ch. 3.4] using convergence of random discrete trees. Expression (1.5) was recently recovered directly in the continuous setting by Wang [34]. The moments of the diameter are given by:
$$\mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})\right] = \frac{4}{3}\sqrt{\pi/2}, \quad \mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})^2\right] = \frac{2}{3}\left(1+\frac{\pi^2}{3}\right), \quad \mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})^3\right] = 2\sqrt{2\pi}, \tag{1.6}$$
$$\mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})^k\right] = \frac{2^{k/2}}{3}k(k-1)(k-3)\Gamma(k/2)(\zeta(k-2)-\zeta(k)) \quad \text{for } k\ge 4. \tag{1.7}$$
The expression $\mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})\right] = \frac{4}{3}\sqrt{\pi/2}$ may be obtained as shown by Aldous [4, Sec. 3.4] using results of Szekeres [33], who proved the existence of a limit distribution for the diameter of rescaled random unordered labelled trees. The higher moments can be obtained in the same way by elaborated calculations, but it is more economic to deduce them by combining Theorem 1.1 with results by Broutin and Flajolet [12] as follows:

Consider the random tree τ_n drawn uniformly at random among all unlabelled trees with n leaves in which each inner vertex is required to have degree 3. In [12, Thm. 8] asymptotics of the form
$$\mathbb{E}\left[\mathsf{D}(\tau_n)^r\right] \sim c_r \lambda^{-r} n^{r/2}$$
were established by analytic methods, with λ an analytically given constant, and the constants c_r given by
$$c_1 = \frac{8}{3}\sqrt{\pi}, \quad c_2 = \frac{16}{3}(1+\frac{\pi^2}{3}), \quad c_3 = 64\sqrt{\pi},$$
$$c_r = \frac{4^r}{3}r(r-1)(r-3)\Gamma(r/2)(\zeta(r-2)-\zeta(r)) \quad \text{if } r\ge 4.$$
As τ_n has n leaves and hence $2n-1$ vertices in total, it follows by Theorem 1.1 that
$$(\tau_n, e_{\{0,2\}}(2n-1)^{-1/2}d_{\tau_n}) \xrightarrow{d} (\mathcal{T}_\mathsf{e}, d_{\mathcal{T}_\mathsf{e}})$$
and consequently, by the exponential tail-bounds for the diameter in Theorem 1.1, which imply arbitrarily high uniform integrability,
$$\mathbb{E}\left[\mathsf{D}(\tau_n)^r\right] \sim \mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})^r\right](e_{\{0,2\}}/\sqrt{2})^{-r}n^{r/2}.$$
It follows that
$$\mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})^r\right] = c_r(e_{\{0,2\}}/(\sqrt{2}\lambda))^r.$$
All that remains is to calculate the ratio $e_{\{0,2\}}/(\sqrt{2}\lambda)$, which is given by
$$e_{\{0,2\}}/(\sqrt{2}\lambda) = \mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})\right]/c_1 = 1/(2\sqrt{2}),$$
since $\mathbb{E}\left[\mathsf{D}(\mathcal{T}_\mathsf{e})\right] = 4/3\sqrt{\pi/2}$. This yields Equations (1.6) and (1.7).

Outline of the paper

In Section 2 we fix basic notions on graphs and discrete trees. Section 3 gives a brief account on Gromov–Hausdorff convergence and the continuum random tree. Section 4 recalls the notion of local weak convergence and results for random Pólya trees. Section 5 introduces the reader to the language of combinatorial species, and Section 6 to the technique of cycle pointing that is formulated using these notions. Section 7 recalls the concept of (Pólya-)Boltzmann samplers, which builds a bridge from combinatorial structures to random algorithms that sample these structures. Section 8 discusses extremal component sizes in random multisets. In Section 9 we present the proofs of our main results.

Notation

Throughout, we set

$$\mathbb{N} = \{1, 2, \ldots\}, \qquad \mathbb{N}_0 = \{0\} \cup \mathbb{N}, \qquad [n] = \{1, 2, \ldots, n\}, \qquad n \in \mathbb{N}_0.$$

we assume that all considered random variables are defined on a common probability space whose measure we denote by $\mathbb{P}$. All unspecified limits are taken as n becomes large, possibly along a shifted sublattice of the integers. We write $\xrightarrow{d}$ and $\xrightarrow{p}$ for convergence in distribution and probability, and $\overset{(d)}{=}$ for equality in distribution. An event holds with high probability, if its probability tends to 1 as $n \to \infty$. We let $O_p(1)$ denote an unspecified random variable X_n of a stochastically bounded sequence $(X_n)_n$. The total variation distance of measures and random variables is denoted by d_{TV}. For a sequence a_n that is eventually positive the notation $O(a_n)$ and $o(a_n)$ refer to unspecified deterministic sequences that are bounded by a multiple of a_n or whose order is negligible compared to a_n. Given a multi-variate power series $f(z_1, z_2, \ldots)$ we let $[z_1^{t_1} \cdots z_m^{t_m}]f(z_1, z_2, \ldots)$ denote the coefficient corresponding to the monomial $z_1^{t_1} \cdots z_m^{t_m}$.

2 Discrete trees

A *(labelled) graph* G consists of a non-empty set $V(G)$ of *vertices* (or *labels*) and a set $E(G)$ of *edges* that are two-element subsets of $V(G)$. The cardinality $|V(G)|$ of the vertex set is termed the *size* of G. Instead of $v \in V(G)$ we will often just write $v \in G$. Two vertices $v, w \in V(G)$ are said to be *adjacent* if $\{v, w\} \in E(G)$. An edge $e \in E(G)$ is adjacent to v if $v \in e$. The cardinality of the set of all edges adjacent to a vertex v is termed its *degree* and denoted by $d_G(v)$. We say the graph G is *connected* if any two vertices $u, v \in V(G)$ are connected by a path in G. The length of a shortest path connecting the vertices u and v is called the *graph distance* of u and v and it is denoted by $d_G(u, v)$. Clearly d_G is a metric on the vertex set $V(G)$. A graph G together with a distinguished vertex $v \in V(G)$ is called a *rooted* graph with root-vertex v. The *height* $\mathsf{h}(w)$ of a vertex $w \in V(G)$ is its distance from the root. The *height* $\mathsf{H}(G)$ of the entire graph is the supremum of the heights of the vertices in G. For any $k \geq 0$ we let $V_k(G, v)$ denote k*-neighbourhood* of the vertex v in G, that is, the subgraph induced by all vertices with distance at most k from v.

Two graphs G_1 and G_2 are termed *isomorphic*, if there is a bijection $\varphi : V(G_1) \to V(G_2)$ such that any two vertices $x, y \in V(G_1)$ are adjacent in G_1 if and only if $\phi(x)$ and $\phi(y)$ are adjacent in G_2. Any such bijection is termed an *isomorphism* between G_1 and G_2. Rooted graphs $G_1^\bullet = (G_1, o_1)$ and $G_2^\bullet = (G_2, o_2)$ are termed isomorphic, if there is a graph isomorphism ϕ from G_1 to G_2 that satisfies $\phi(o_1) = o_2$. An isomorphism class of (rooted) graphs is also called an *unlabelled (rooted) graph*. We will often not distinguish between such a class or any fixed representative of that class.

A *tree* T is a non-empty connected graph without cyclic subgraphs, that is, we cannot walk from one vertex to itself without crossing at least one edge twice. Any two vertices of a tree are connected by a unique path. Figure 1 depicts the list of all unlabelled trees with 6 vertices. If T is rooted, then the vertices $w' \in V(T)$ that are adjacent to a vertex w and have height $\mathsf{h}(w') = \mathsf{h}(w) + 1$ form the *offspring* *set* of the vertex w. Its cardinality is the *outdegree* $d_T^+(w)$ of the vertex w. Unlabelled rooted trees are also termed Pólya trees. Note that while any labelled tree with n vertices admits n different roots, this does not hold in the unlabelled setting. For example, as illustrated in Figure 2, there are 3 unlabelled trees with 5 vertices and each of them has a different number of rootings.

3 Scaling limits

We briefly recall several relevant results regarding the convergence of random rooted trees toward the continuum random tree.

3.1 Gromov–Hausdorff convergence

We introduce the required notions regarding the Gromov–Hausdorff convergence following Burago, Burago and Ivanov [13, Ch. 7] and Le Gall and Miermont [27]

3.1.1 The Hausdorff metric

Recall that given subsets A and B of a metric space (X, d), their *Hausdorff-distance* is given by

$$d_{\mathsf{H}}(A, B) = \inf\{\epsilon > 0 \mid A \subset U_\epsilon(B), B \subset U_\epsilon(A)\} \in [0, \infty],$$

where $U_\epsilon(A) = \{x \in X \mid d(x, A) \le \epsilon\}$ denotes the ϵ-*hull* of A. In general, the Hausdorff-distance does not define a metric on the set of all subsets of X, but it does on the set of all compact subsets of X ([13, Prop. 7.3.3]).

3.1.2 The Gromov–Hausdorff distance

The Gromov–Hausdorff distance allows us to compare arbitrary metric spaces, instead of only subsets of a common metric space. It is defined by the infimum of Hausdorff-distances of isometric copies in a common metric space. We are also going to consider a variation of the Gromov–Hausdorff distance given in [27] for *pointed* metric spaces, which are metric spaces together with a distinguished point.

Given metric spaces (X, d_X), and (Y, d_Y), and distinguished elements $x_0 \in X$ and $y_0 \in Y$, the Gromov–Hausdorff distances of X and Y and the pointed spaces $X^\bullet = (X, x_0)$ and $Y^\bullet = (Y, y_0)$ are defined by

$$\begin{aligned} d_{\mathsf{GH}}(X, Y) &= \inf_{\iota_X, \iota_Y} d_{\mathsf{H}}(\iota_X(X), \iota_Y(Y)) \in [0, \infty], \\ d_{\mathsf{GH}}(X^\bullet, Y^\bullet) &= \inf_{\iota_X, \iota_Y} \max\{d_{\mathsf{H}}(\iota_X(X), \iota_Y(Y)), d_E(\iota_X(x_0), \iota_Y(y_0))\} \in [0, \infty] \end{aligned}$$

where in both cases the infimum is taken over all isometric embeddings $\iota_X : X \to E$ and $\iota_Y : Y \to E$ into any common metric space (E, d_E), compare with Figure 3.

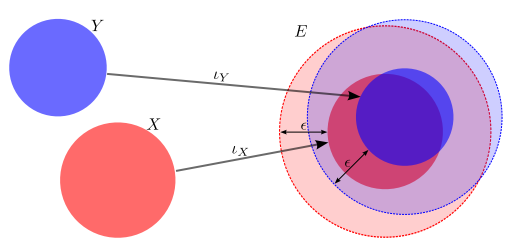

Figure 3: The Gromov–Hausdorff distance.

We will make use of the following characterisation of the Gromov–Hausdorff metric. Given two metric spaces (X, d_X) and (Y, d_Y) a *correspondence* between them is a relation $R \subset X \times Y$ such that any point $x \in X$ corresponds to at least one point $y \in Y$ and vice versa. If X and Y are pointed, we additionally require that the roots correspond to each other. The *distortion* of R is given by

$$\mathsf{dis}(R) = \sup\{|d_X(x_1, x_2) - d_Y(y_1, y_2)| \mid (x_1, y_1), (x_2, y_2) \in R\}.$$

Proposition 3.1 ([13, Thm. 7.3.25] and [27, Prop. 3.6])**.** *Given two metric spaces X, Y and pointed metric spaces $X^\bullet, Y^\bullet$ we have that*

$$d_{GH}(X, Y) = \frac{1}{2} \inf_R \mathsf{dis}(R), \quad \text{and} \quad d_{GH}(X^\bullet, Y^\bullet) = \frac{1}{2} \inf_R \mathsf{dis}(R),$$

where R ranges over all correspondences between X and Y (or $X^\bullet$ and $Y^\bullet$).

Using this reformulation of the Gromov–Hausdorff distance, one may check that it satisfies the following properties.

Lemma 3.2 ([13, Thm. 7.3.30] and [27, Thm. 3.5])**.** *Let X, Y, and Z be (pointed) metric spaces. Then the following assertions hold.*

i) $d_{GH}(X, Y) = 0$ *if and only if X and Y are isometric.*

ii) $d_{GH}(X, Z) \le d_{GH}(X, Y) + d_{GH}(Y, Z)$.

iii) If X and Y are bounded, then $d_{GH}(X, Y) < \infty$.

3.1.3 The space of isometry classes of compact metric spaces

In Section 3.1.1 we saw that the Hausdorff-distance defines a metric on the set of all compact subsets of a metric space. By Lemma 3.2 the Gromov–Hausdorff distance satisfies in a similar way the axioms of a (finite) pseudo-metric on the class of all compact metric spaces, and two metric spaces have Gromov–Hausdorff distance 0 if and only if they are isometric. Informally speaking, this yields a metric on the collection of all isometry classes of metric spaces, and in a similar way we may endow the collection of isometry classes of pointed metric spaces with a metric.

Note that from a formal viewpoint this construction is a bit problematic, since we are forming a collection of proper classes (as opposed to sets). A solution is presented as an exercise in [13, Rem. 7.2.5]:

Proposition 3.3. *Any set of pairwise non-isometric (pointed) metric spaces has cardinality at most $2^{\aleph_0}$, and there are specific examples of $2^{\aleph_0}$ many non-isometric (pointed) spaces.*

We may thus fix a representative of each isometry class of (pointed) metric spaces and let $\mathbb{K}$ (resp. $\mathbb{K}^\bullet$) denote the resulting sets of spaces. Lemma 3.2 now reads as follows.

Corollary 3.4 ([13, Thm. 7.3.30])**.** *The Gromov–Hausdorff distance defines a finite metric on the set $\mathbb{K}$ (resp. $\mathbb{K}^\bullet$) of representatives of isometry classes of (pointed) compact metric spaces.*

The metric spaces $\mathbb{K}$ and $\mathbb{K}^\bullet$ have nice properties, which make them suitable for studying random elements:

Proposition 3.5 ([27, Thm. 3.5] and [13, Thm. 7.4.15])**.** *The spaces $\mathbb{K}$ and $\mathbb{K}^\bullet$ are separable and complete, i.e. they are Polish spaces.*

3.2 The continuum random tree

An $\mathbb{R}$-*tree* is a metric space (X, d) such that for any two points $x, y \in X$ the following properties hold

1. There is a unique isometric map from the interval $\varphi_{x,y} : [0, d_f(x, y)] \to X$ satisfying $\varphi_{x,y}(0) = x$ and $\varphi_{x,y}(d_f(x, y)) = y$.
2. If $q : [0, d_f(x, y)] \to X$ is a continuous injective map, then
$$q([0, d_f(x, y)]) = \varphi_{x,y}([0, d_f(x, y)]).$$

$\mathbb{R}$-trees may be constructed as follows. Let $f : [0, 1] \to [0, \infty[$ be a continuous function satisfying $f(0) = f(1) = 0$. Consider the pseudo-metric d on the interval $[0, 1]$ given by

$$d(u, v) = f(u) + f(v) - 2 \inf_{u \le s \le v} f(s)$$

for $u \le v$. Let $(\mathcal{T}_f, d_{\mathcal{T}_f}) = ([0, 1]/\sim, \bar{d})$ denote the corresponding quotient space. We may consider this space as rooted at the equivalence class $\bar{0}$ of 0.

Proposition 3.6 ([27, Thm. 3.1])**.** *Given a continuous function $f : [0, 1] \to [0, \infty[$ satisfying $f(0) = f(1)$ the corresponding metric space $\mathcal{T}_f$ is a compact $\mathbb{R}$-tree.*

Hence, this construction defines a map from a set of continuous functions to the space $\mathbb{K}^\bullet$. It can be seen to be Lipschitz-continuous:

Proposition 3.7 ([27, Cor. 3.7])**.** *The map*

$$(\{f \in \mathcal{C}([0, 1], \mathbb{R}_{\ge 0}) \mid f(0) = f(1) = 0\}, \|\cdot\|_\infty) \to (\mathbb{K}^\bullet, d_{GH}), \qquad f \mapsto \mathcal{T}_f$$

is Lipschitz-continuous.

Hence we may define the continuum random tree as a random element of the polish space $\mathbb{K}^\bullet$.

Definition 3.8. *The random pointed metric space $(\mathcal{T}_{\mathsf{e}}, d_{\mathcal{T}_{\mathsf{e}}}, \bar{0})$ coded by the Brownian excursion of duration one $\mathsf{e} = (\mathsf{e}_t)_{0 \le t \le 1}$ is called the Brownian continuum random tree (CRT).*

Note that the Lipschitz-continuity (and hence measurability) of the above map ensures that the CRT is a random variable.

3.3 Scaling limits of random Pólya trees

It is known that for any subset $\Omega^* \subset \mathbb{N}_0$ containing zero and at least one integer $k \ge 2$, the Pólya tree A_n drawn uniformly at random from the set of all Pólya trees with n vertices and vertex outdegrees in the set Ω^* admits the CRT as scaling limit. That is, there is a constant c_{Ω^*} satisfying

$$(\mathsf{A}_n, c_{\Omega^*} n^{-1/2} d_{\mathsf{A}_n}) \xrightarrow{d} (\mathcal{T}_{\mathsf{e}}, d_{\mathcal{T}_{\mathsf{e}}}) \tag{3.1}$$

as random elements of the space $\mathbb{K}^\bullet$. See [28, 24, 30]. The diameter admits a tail-bound of the form

$$\mathbb{P}(\mathsf{D}(\mathsf{A}_n) \ge x) \le C \exp(-cx^2/n) \tag{3.2}$$

for all n and $x \ge 0$ [30, Thm. 1.2].

4 Local weak limits

We briefly recall relevant notions and concerning local weak convergence of random graphs.

4.1 The metric for local convergence

Given two rooted, locally finite (that is, the graph may have infinitely many vertices, but each vertex has only finitely many neighbours) connected graphs $G^\bullet = (G, o_G)$ and $H^\bullet = (H, o_H)$, we may consider the distance

$$d_{\mathsf{BS}}(G^\bullet, H^\bullet) = 2^{-\sup\{k \in \mathbb{N}_0 \,|\, V_k(G^\bullet) \simeq V_k(H^\bullet)\}}.$$

Here $V_k(G^\bullet) \simeq V_k(H^\bullet)$ denotes isomorphism of rooted graphs, that is, the existence of a graph isomorphism $\phi : V_k(G^\bullet) \to V_k(H^\bullet)$ satisfying $\phi(o_G) = o_H$. This defines a premetric on the collection of all rooted connected locally finite graphs.

If $\mathbb{B}$ denotes the collection of isomorphism classes of rooted locally finite connected graphs, then the (lift of) this distance defines a metric on $\mathbb{B}$ that is complete and separable. In other words, $(\mathbb{B}, d_{\mathsf{BS}})$ is a Polish space. Similarly as for the Gromov–Hausdorff metric, we may safely ignore the fact that $\mathbb{B}$ is a collection of proper classes (as opposed to sets). In order to be precise, we would only need to fix a representatives of each isomorphism class and work with the set of these representatives instead.

4.2 Benjamini–Schramm convergence of random Pólya trees

Let $\Omega \subset \mathbb{N}$ denote a subset containing 1 and at least one integer $k \geq 3$, and let $\Omega^* = \Omega - 1$ denote the corresponding shifted set. Let A_n denote the random tree drawn uniformly at random from the set of all Pólya trees with n vertices and vertex outdegdrees in Ω^*. Let u_n denote a uniformly at random drawn selected vertex of A_n. It was shown in [32, Thm. 6.22], that there is a random infinite rooted trees A_{Ω^*} such that for each sequence $k_n = o(\sqrt{n})$ the random vertex u_n has with high probability height strictly larger than k_n in the tree A_n and

$$d_{\mathsf{TV}}(V_{k_n}(\mathsf{A}_n, u_n), V_{k_n}(\mathsf{A}_{\Omega^*})) \to 0. \tag{4.1}$$

5 Combinatorial species of structures

The language of combinatorial species was developed by Joyal [26]. It is appropriate to use this framework in the context of combinatorial probability theory, as it allows for a systematic enumeration of a wide range of discrete structures. We recall the required theory and notations following [9, 26]. The language of *combinatorial classes* used by Flajolet and Sedgewick [21] is essentially equivalent in many aspects, although less emphasis is put on studying objects up to symmetry.

5.1 Combinatorial species of structures

Informally speaking, a *combinatorial species* is a collection of labelled discrete structures together with the information on when two such objects may be considered as structurally equivalent. The collection of labelled graphs together with the isomorphism relation is a natural example, that one may keep in mind to ease the understanding of the formal definition.

Formally, a species may be defined as a functor $\mathcal{F}$ that maps any finite set U of *labels* to a finite set $\mathcal{F}[U]$ of *$\mathcal{F}$-objects* and any bijection $\sigma : U \to V$ of finite sets to its (bijective) *transport function* $\mathcal{F}[\sigma] : \mathcal{F}[U] \to \mathcal{F}[V]$ *along* σ, such that composition of maps and the identity maps are preserved. Thus, for the species of graphs the labels correspond to the vertices and the transport functions to vertex relabellings.

An element $F_U \in \mathcal{F}[U]$ has size $|F_U| := |U|$ and two $\mathcal{F}$-objects F_U and F_V are termed *isomorphic* (or *structurally equivalent*) if there is a bijection $\sigma : U \to V$ such that $\mathcal{F}[\sigma](F_U) = F_V$. We will often just write $\sigma.F_U = F_V$ instead, if there is no risk of confusion. We say σ is an *isomorphism* from F_U to F_V. If $U = V$ and $F_U = F_V$ then σ is an *automorphism* of F_U. An isomorphism class of $\mathcal{F}$-structures is called an *unlabelled* $\mathcal{F}$-object. The isomorphism class of a given $\mathcal{F}$-object is also called its *isomorphism* *type*. By abuse of notation, we treat unlabelled objects as if they were regular objects. We will also write $F \in \mathcal{F}$ to state that F is an $\mathcal{F}$-object.

We say that a species $\mathcal{G}$ is a *subspecies* of $\mathcal{F}$, and write $\mathcal{G} \subset \mathcal{F}$, if $\mathcal{G}[U] \subset \mathcal{F}[U]$ for all finite sets U and $\mathcal{G}[\sigma] = \mathcal{F}[\sigma]|_U$ for all bijections $\sigma : U \to V$. For example, the species of connected graphs may be viewed as a subspecies of the species of graphs.

Given two species $\mathcal{F}$ and $\mathcal{G}$, an *isomorphism* $\alpha : \mathcal{F} \xrightarrow{\sim} \mathcal{G}$ from $\mathcal{F}$ to $\mathcal{G}$ is a family of bijections $\alpha = (\alpha_U : \mathcal{F}[U] \to \mathcal{G}[U])_U$ where U ranges over all finite sets, such that for all bijective maps $\sigma : U \to V$ the following diagram commutes.

$$\begin{tikzcd}
\mathcal{F}[U] \arrow[r, "\mathcal{F}[\sigma]"] \arrow[d, "\alpha_U"] & \mathcal{F}[V] \arrow[d, "\alpha_V"] \\
\mathcal{G}[U] \arrow[r, "\mathcal{G}[\sigma]"] & \mathcal{G}[V]
\end{tikzcd}$$

The species $\mathcal{F}$ and $\mathcal{G}$ are *isomorphic* if there exists an isomorphism from one to the other. This is denoted by $\mathcal{F} \simeq \mathcal{G}$. The idea behind this is that we consider two collections of discrete structures as equivalent, if there is a bijective correspondence between them that preserves the notion of structural equivalence. This is essential for the study of discrete structures considered up to symmetry. For example, two collections of n-sized graphs may admit a bijective correspondence, but if this bijection does not preserve graph isomorphisms, the corresponding classes of unlabelled graphs may have different cardinalities.

In the alternative language of combinatorial classes, two classes of combinatorial structures are already considered as equivalent, if there is a size-preserving bijection between them [21, Def. I.3]. For this reason the framework of combinatorial species is more convenient when studying structures up to symmetry, although it has a more formal character and is actually best understood in a category theoretic formulation.

We will mostly be interested in the species of trees, and make use of standard species such as the SET-species given by $\mathsf{SET}[U] = \{U\}$ for all U. Moreover, we let $\mathcal{X}$ the species with a single object of size 1.

5.2 Symmetries and generating power series

Letting $\tilde{f}_n$ denote the number of unlabelled $\mathcal{F}$-objects of size n, the *ordinary generating series* of $\mathcal{F}$ is defined by

$$\tilde{\mathcal{F}}(x) = \sum_{n=0}^{\infty} \tilde{f}_n x^n.$$

A pair (F, σ) of an $\mathcal{F}$-object together with an automorphism is called a *symmetry*. Its *weight monomial* is given by

$$w_{(F,\sigma)} = \frac{1}{n!} x_1^{\sigma_1} x_2^{\sigma_2} \cdots x_n^{\sigma_n} \in \mathbb{Q}[[x_1, x_2, \ldots]]$$

with n denoting the size of F and σ_i denoting the number of i-cycles of the permutation σ. (Recall that any permutation admits a unique decomposition into a product of disjoint cyclic permutations.) In particular σ_1 denotes the number of fixpoints. We may form the species $\mathsf{Sym}(\mathcal{F})$ of symmetries of $\mathcal{F}$. The *cycle index sum* of $\mathcal{F}$ is given by

$$Z_{\mathcal{F}} = \sum_{(F,\sigma)} w_{(F,\sigma)}$$

with the sum index (F, σ) ranging over the set $\bigcup_{n \in \mathbb{N}_0} \mathsf{Sym}(\mathcal{F})[n]$. The reason for studying cycle index sums is the following remarkable property.

Lemma 5.1 ([26, Sec. 3])**.** *Let U be a finite n-element set. For any unlabelled $\mathcal{F}$-object m of size n there are precisely $n!$ symmetries $(F, \sigma) \in \mathsf{Sym}(\mathcal{F})[U]$ having the property that F has isomorphism type m.*

From a probabilistic viewpoint, this observation guarantees that the isomorphism type of the first coordinate of a uniformly at random drawn element from $\mathrm{Sym}(\mathcal{F})[n]$ is uniformly distributed among all n-element unlabelled $\mathcal{F}$-objects. Lemma 5.1 also implies that the ordinary generating series and the cycle index sum are related by

$$\tilde{\mathcal{F}}(z) = Z_{\mathcal{F}}(z, z^2, z^3, \ldots). \tag{5.1}$$

See [26, Sec. 3, Prop. 9] for a more detailed justification.

For example, the cycle index sum Z_{SET} is easily calculated: For any integer $n \geq 0$ let $\mathfrak{S}_n$ denote the symmetric group of degree n. Then

$$Z_{\mathsf{SET}} = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} x_1^{\sigma_1} x_2^{\sigma_2} \cdots x_n^{\sigma_n}. \tag{5.2}$$

For any permutation σ let $(\sigma_1, \sigma_2, \ldots) \in \mathbb{N}_0^{(\mathbb{N})}$ denote its *cycle type*. Then to each element $m = (m_i)_i \in \mathbb{N}_0^{(\mathbb{N})}$ correspond only permutations of order $n := \sum_{i=1}^{\infty} i m_i$ and their number is given by $n!/\prod_{i=1}^{\infty}(m_i!\, i^{m_i})$. Hence we have

$$Z_{\mathsf{SET}} = \sum_{m \in \mathbb{N}_0^{(\mathbb{N})}} \prod_{i=1}^{\infty} \frac{x_i^{m_i}}{m_i!\, i^{m_i}} = \prod_{i=1}^{\infty} \sum_{m_i=0}^{\infty} \frac{x_i^{m_i}}{m_i!\, i^{m_i}} = \prod_{i=1}^{\infty} \exp\left(\frac{x_i}{i}\right) = \exp\left(\sum_{i=1}^{\infty} \frac{x_i}{i}\right).$$

If $(x_i)_i$ would denote a sequence of sufficiently fast decaying positive real-numbers, then this calculation could easily be justified. But they denote a countable set of formal variables, and hence one has every right to ask for a rigorous justification of this argument, in particular why the involved infinite products of formal variables vanish. A correct formalization is to define a topology on the set of formal power series and interpret these infinite products as actual limits with respect to this topology. We refer the inclined reader to [21, Appendix A.5] for an adequate discussion of these questions.

5.3 Operations on combinatorial species

The framework of combinatorial species offers a large variety of constructions that create new species from others. In the following let $\mathcal{F}$, $(\mathcal{F}_i)_{i \in I}$ and $\mathcal{G}$ denote species and U an arbitrary finite set. The *sum* $\sum_{i \in I} \mathcal{F}_i$ is defined by the disjoint union

$$(\sum_i \mathcal{F}_i)[U] = \bigsqcup_i \mathcal{F}_i[U]$$

if the right hand side is finite for all finite sets U. The *product* $\mathcal{F} \cdot \mathcal{G}$ is defined by the disjoint union

$$(\mathcal{F} \cdot \mathcal{G})[U] = \bigsqcup_{\substack{(U_1, U_2) \\ U_1 \cap U_2 = \emptyset, U_1 \cup U_2 = U}} \mathcal{F}[U_1] \times \mathcal{G}[U_2]$$

with componentwise transport. Thus, n-sized objects of the product are pairs of $\mathcal{F}$-objects and $\mathcal{G}$-objects whose sizes add up to n. Two such pairs are considered as structurally equivalent, if their components are. If the species $\mathcal{G}$ has no objects of size zero, we can form the *substitution* $\mathcal{F} \circ \mathcal{G}$ by

$$(\mathcal{F} \circ \mathcal{G})[U] = \bigsqcup_{\pi \text{ partition of } U} \mathcal{F}[\pi] \times \prod_{Q \in \pi} \mathcal{G}[Q].$$

An object of the substitution may be interpreted as an $\mathcal{F}$-object whose labels are substituted by $\mathcal{G}$-objects. The transport along a bijection σ is defined by applying the induced map

$$\overline{\sigma} : \pi \to \{\sigma(Q) \mid Q \in \pi\}, \quad Q \mapsto \sigma(Q)$$

of partitions to the $\mathcal{F}$-object, and the restricted maps $\sigma|_Q$ with $Q \in \pi$ to their corresponding $\mathcal{G}$-objects. We will often write $\mathcal{F}(\mathcal{G})$ instead of $\mathcal{F} \circ \mathcal{G}$. Explicit formulas for the generating series and cycle index sums of the discussed constructions are summarized in Table 1.

	OGF	Cycle index sum
$\sum_i \mathcal{F}_i$	$\sum_i \tilde{\mathcal{F}}_i(x)$	$\sum_i Z_{\mathcal{F}_i}(x_1, x_2, \ldots)$
$\mathcal{F} \cdot \mathcal{G}$	$\tilde{\mathcal{F}}(x)\tilde{\mathcal{G}}(x)$	$Z_{\mathcal{F}}(x_1, x_2, \ldots)Z_{\mathcal{G}}(x_1, x_2, \ldots)$
$\mathcal{F} \circ \mathcal{G}$	$Z_{\mathcal{F}}(\tilde{\mathcal{G}}(x), \tilde{\mathcal{G}}(x^2), \ldots)$	$Z_{\mathcal{F}}(Z_{\mathcal{G}}(x_1, x_2, \ldots), Z_{\mathcal{G}}(x_2, x_4, \ldots), \ldots)$

Table 1: Relation between combinatorial constructions and generating series.

5.4 Symmetries of composite structures

By Lemma 5.1, we know that random unlabelled structures may be studied using random symmetries. The coupled automorphism in the random symmetry gives us additional information on the shape of the object, and hence serves as a starting point for a probabilistic analysis. In this context, we will often encounter the following operation for composing cycles.

Definition 5.2. *Given a cyclic permutation $(a_1, \ldots, a_k)$ and a number $\ell \geq 1$, we may take ℓ "identical" copies $(a_1^i, \ldots, a_k^i)$, $1 \leq i \leq \ell$ and form a $k\ell$-cycle*

$$(a_1^1, \ldots, a_1^\ell, a_2^1, \ldots, a_2^\ell, \ldots, a_k^1, \ldots, a_k^\ell).$$

Any automorphism of an $\mathcal{F} \circ \mathcal{G}$ composite structure has only cycles of this form. The key idea is that such a composite symmetry is essentially composed out of an $\mathcal{F}$-symmetry where each atom is endowed with a $\mathcal{G}$-symmetry, such that symmetries belonging to a common cycle of the $\mathcal{F}$-symmetry are isomorphic copies of each other:

Let U be a finite set. Any element of $\mathrm{Sym}(\mathcal{F} \circ \mathcal{G})[U]$ consists of the following objects: a partition π of the set U, an $\mathcal{F}$-structure $F \in \mathcal{F}[\pi]$, a family of $\mathcal{G}$-structures $(G_Q)_{Q \in \pi}$ with $G_Q \in \mathcal{G}[Q]$, and a permutation $\sigma : U \to U$ that is an automorphism of the composite structure $(F, (G_Q)_{Q \in \pi}$. That is, we require the permutation σ to permute the partition classes and induce an automorphism

$$\bar{\sigma} : \pi \to \pi, \quad Q \mapsto \sigma(Q)$$

of the $\mathcal{F}$-object F. Moreover, for any partition class $Q \in \pi$ we require that the restriction $\sigma|_Q : Q \to \sigma(Q)$ is an isomorphism from G_Q to $G_{\sigma(Q)}$. For any cycle $\bar{\tau} = (Q_1, \ldots, Q_\ell)$ of the $\mathcal{F}$-automorphism $\bar{\sigma}$ it follows that for all $1 \leq i \leq \ell$ we have $\sigma^\ell(Q_i) = Q_i$ and hence the restriction $\sigma^\ell|_{Q_i} : Q_i \to Q_i$ is an automorphism of G_{Q_i}. Moreover, if we know $(G_{Q_1}, \sigma^\ell|_{Q_1})$ and the maps $\sigma|_{Q_i} = (\sigma|_{Q_1})^i$ for $1 \leq i \leq \ell - 1$, we can reconstruct the $\mathcal{G}$-objects $G_{Q_2}, \ldots, G_{Q_\ell}$ and the restriction $\sigma|_{Q_1 \cup \ldots \cup Q_\ell}$. Here any k-cycle $(a_1, \ldots, a_k)$ of the permutation $\sigma^\ell|_{Q_1}$ corresponds to the $k\ell$-cycle

$$(a_1, \sigma(a_1), \ldots, \sigma^{\ell-1}(a_1), a_2, \sigma(a_2), \ldots, \sigma^{\ell-1}(a_2), \ldots, a_k, \sigma(a_k), \ldots, \sigma^{\ell-1}(a_k))$$

of $\sigma|_{Q_1 \cup \ldots \cup Q_\ell}$. Thus any cycle ν of σ is of the form as in Definition 5.2 and corresponds to a cycle of the induced permutation $\bar{\sigma}$ whose length is a divisor of the length of ν.

Note that the maps $\sigma|_{Q_i}$ carry information about the labelling, but not really about the structure of the symmetry, as all $\mathcal{G}$-structure pertaining to a common cycle need to be isomorphic anyway. Up to relabelling, an $\mathcal{F} \circ \mathcal{G}$-symmetry is already fully described by its induced $\mathcal{F}$-symmetry and a family of $\mathcal{G}$-symmetries, one for each cycle of the $\mathcal{F}$-symmetry:

Proposition 5.3. *If we are given an $\mathcal{F}$-symmetry (m, σ_m) and for each of its cycles c a $\mathcal{G}$-symmetry (G_c, σ_c), then there is a canonical way to assemble an $\mathcal{F} \circ \mathcal{G}$ symmetry out of these objects. Here each atom of c receives an identical copy of G_c, and the cycles of σ_c are cloned and composed as in Definition 5.2. Up to relabelling, any $\mathcal{F} \circ \mathcal{G}$-symmetry may be constructed in this way.*

The details of the construction are as follows. For each cycle c of σ_m let Q_c denote the label set of the $\mathcal{G}$-object G_c. For every atom e of the cycle c set $Q_e := Q_c \times \{e\}$ and $(G_{Q_e}, \sigma_{Q_e}) := \mathrm{Sym}(\mathcal{G})[f_e](G_c, \sigma_c)$ with $f_e : Q_c \to Q_e$ the canonical bijection. For any label e of the $\mathcal{F}$-structure m set $f(e) := Q_e$ and let π denote the set of all sets Q_e. Thus $F := \mathcal{F}[f](m)$ is an $\mathcal{F}$-structure with label set π and $C := (\pi, F, (G_Q)_{Q\in\pi})$ is an $\mathcal{F} \circ \mathcal{G}$-structure. Let c be a cycle of σ_m and ν a cycle of σ_c. Fix an atom $b = b(c)$ of c and an atom $a = a(\nu)$ of ν. Let ℓ denote the length of c and k the length of ν. Form the composed cycle by

$$((a, b), \ldots, (a, c^{\ell-1}(b)), (\nu(a), b), \ldots, (\nu(a), c^{\ell-1}(b)), \ldots, (\nu^{k-1}(a), b), \ldots, (\nu^{k-1}(a), c^{\ell-1}(b))).$$

Then the product σ of all composed cycles (formed by all choices of c and ν) is an automorphism of the $\mathcal{F} \circ \mathcal{G}$-structure C. The composed cycles are pairwise disjoint, hence it does not matter in which order we take the product. Note that σ does not depend on the choice of the a's but different choices of the b's result in a different automorphism σ. More precisely, if for a given cycle c of σ_m we choose $c(b)$ instead of b, then the resulting automorphism is given by the conjugation $(\mathrm{id}, c)\sigma(\mathrm{id}, c)^{-1}$ instead of σ. But (id, c) is an automorphism of the $\mathcal{F} \circ \mathcal{G}$-structure C, hence the resulting symmetry $(C, (\mathrm{id}, c)\sigma(\mathrm{id}, c)^{-1})$ is isomorphic to (C, σ). This implies that the isomorphism type of (C, σ) does not depend on the choices of the a's and b's.

More details on this structural result are given in [11, Sec. 2.6.2], [26, Section 3] and [9, Section 4.3]. The main point for our purposes is to know that we may form an $\mathcal{F} \circ \mathcal{G}$-symmetry by taking an $\mathcal{F}$-symmetry and for each of its cycles c precisely $|c|$ identical copies of a $\mathcal{G}$-symmetry, which we compose into a permutation by composing the clones of each cycle of the $\mathcal{G}$-symmetry as in Definition 5.2.

6 Cycle pointing

When studying random graphs or other discrete structures, it is often convenient to select a root vertex. If the vertices are distinguishable by labels, this is easy and natural, as any labelled n-vertex object may be rooted in n different locations. However, if the structure has a non-trivial automorphism, then it corresponds to less than n pointed unlabelled structures, because pointing at two vertices in symmetric positions produces the same unlabelled structure. See for example Figure 2 for an illustration of the root locations in unlabelled trees.

In order to tackle this general enumerative problem, Bodirsky, Fusy, Kang and Vigerske [11] constructed an *unbiased* pointing operator, such that each unlabelled structure of size n gives rise to n pointed unlabelled structures.

6.1 The cycle pointing operator

The cycle pointing operator was constructed in [11] and maps a species $\mathcal{G}$ to the species $\mathcal{G}^\circ$ such that the $\mathcal{G}^\circ$-objects over a set U are pairs (G, τ) with $G \in \mathcal{G}[U]$ and τ a *marked* cycle of an arbitrary automorphism of G. Here we count fixpoints as 1-cycles. The transport is defined by $\sigma.(G, \tau) = (\sigma.G, \sigma\tau\sigma^{-1})$. Any subspecies $\mathcal{S} \subset \mathcal{G}^\circ$ is termed *cycle-pointed*. The *symmetric* cycle-pointed species $\mathcal{G}^\circledast \subset \mathcal{G}^\circ$ is defined by restricting to pairs (G, τ) with τ a cycle of length at least 2.

A *rooted symmetry* of the cycle-pointed species $\mathcal{S} \subset \mathcal{G}^\circ$ is a quadruple $((G, \tau), \sigma, v)$ such that (G, τ) is an $\mathcal{S}$-object, σ is an automorphism of G, τ is a cycle of σ and v is an atom of the cycle τ. Its *weight monomial* is given by

$$w_{((G,\tau),\sigma,v)} = \frac{t_\ell}{s_\ell} w_{(G,\sigma)}(s_1, s_2, \ldots)$$

with $w_{(G,\sigma)}$ denoting the weight of the symmetry (G, σ) and ℓ the length of the marked cycle τ. We may form the species $\mathrm{RSym}(\mathcal{S})$ of rooted symmetries of $\mathcal{S}$. The pointed cycle index sum of $\mathcal{S}$ is given by

$$\bar{Z}_{\mathcal{S}}(s_1, t_1; s_2, t_2; \ldots) = \sum_{(G,\tau,\sigma,v)} w_{(G,\tau,\sigma,v)} \in \mathbb{Q}[[s_1, t_1; s_2, t_2; \ldots]]$$

with the sum index ranging over the set $\bigcup_{n\in\mathbb{N}_0} \mathsf{RSym}(\mathcal{S})[n]$.

Let $\mathcal{G}^\circ_{(\ell)} \subset \mathcal{G}^\circ$ denote the subspecies given by all cycle pointed objects whose marked cycle has length ℓ. It follows from the definition of the pointed cycle index sum that

$$\bar{Z}_{\mathcal{G}^\circ_{(\ell)}} = \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathcal{G}}.$$

Since $\mathcal{G}^\circ = \sum_{\ell=1}^\infty \mathcal{G}^\circ_{(\ell)}$ it follows that

$$\bar{Z}_{\mathcal{G}^\circ} = \sum_{\ell=1}^{\infty} \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathcal{G}} \quad \text{and} \quad \bar{Z}_{\mathcal{G}^\circledast} = \sum_{\ell=2}^{\infty} \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathcal{G}}. \tag{6.1}$$

Lemma 6.1 ([11, Lem. 14])**.** *Let U be a finite set with n elements and fix an arbitrary linear order on U.*

1) *The following map is bijective:*

$$\begin{aligned} &\textit{RSym}(\mathcal{S})[U] \to \textit{Sym}(\mathcal{S})[U], \\ &M = ((G,\tau),\sigma,v) \mapsto ((\tau^{1-\ell(M)}.G,\tau),\sigma\tau^{\ell(M)-1}) \end{aligned}$$

with $\ell(M)$ defined as follows: let k denote the length of the cycle τ and u its smallest atom. Let $0 \le \ell(M) \le k-1$ be the unique integer satisfying $v = \tau^{\ell(M)}.u$.

2) *Any unlabelled cycle-pointed $\mathcal{S}$-object m of size n corresponds to precisely $n!$ rooted c-symmetries from RSym$(\mathcal{S})[U]$ having the property that the isomorphism type of the underlying $\mathcal{S}$-object equals m.*

In particular, the pointed cycle index sum relates to the ordinary generating series by

$$\tilde{\mathcal{S}}(x) = \bar{Z}_{\mathcal{S}}(x,x;x^2,x^2;\ldots). \tag{6.2}$$

Moreover, if we draw an element from $\mathsf{RSym}(\mathcal{S})[n]$ uniformly at random, then the isomorphism class of the corresponding cycle pointed structure is uniformly distributed among all unlabelled cycle-pointed $\mathcal{S}$-objects of size n. The main point of the cycle-pointing construction is evident from the following fact.

Lemma 6.2 ([11, Thm. 15])**.** *Any unlabelled $\mathcal{G}$-structure m of size n may be cycle-pointed in precisely n ways, that is, there exist precisely n unlabelled $\mathcal{G}^\circ$-structures with corresponding $\mathcal{G}$-structure m.*

Considered from a probabilistic viewpoint, this means that if we draw an unlabelled $\mathcal{G}^\circ$-structure of size n uniformly at random, then the underlying $\mathcal{G}$-object is also uniformly distributed. Moreover, Lemma 6.1 tells us that in order to sample the $\mathcal{G}^\circ$-object we may sample a rooted symmetry of this size uniformly at random.

Studying the random $\mathcal{G}^\circ$-object might be easier due to the additional information given by the marked cycle. Moreover, Lemma 6.2 implies that

$$\tilde{\mathcal{G}}^\circ(z) = z\frac{\mathrm{d}}{\mathrm{d}z}\tilde{\mathcal{G}}(z). \tag{6.3}$$

By Equations (6.1) and (5.2) the pointed cycle index sum of the species SET is given by

$$\bar{Z}_{\mathsf{SET}^\circ} = \sum_{\ell=1}^{\infty} \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathsf{SET}}(s_1,s_2,\ldots) = \exp\left(\sum_{i=1}^{\infty} s_i/i\right) \sum_{\ell=1}^{\infty} t_\ell. \tag{6.4}$$

As $\bar{Z}_{\mathsf{SET}^\circledast} = \bar{Z}_{\mathsf{SET}^\circ}(s_1,0;s_2,t_2;\ldots)$, the cycle index sum for $\bar{Z}_{\mathsf{SET}^\circledast}$ is given as in Equation (6.4), but with the sum index ℓ ranging from 2 to infinity.

6.2 Operations on cycle pointed species

Cycle pointed species come with a set of new operations introduced in [11]. If $\mathcal{S} \subset \mathcal{G}^\circ$ is a cycle-pointed species and $\mathcal{H}$ a species, then the *pointed product* $\mathcal{S} \star \mathcal{H}$ is the subspecies of $(\mathcal{G} \cdot \mathcal{H})^\circ$ given by all cycle-pointed objects such that the marked cycle consists of atoms of the $\mathcal{G}$-structure and the $\mathcal{G}$-structure together with this cycle belongs to $\mathcal{S}$. The corresponding pointed cycle index sum is given by

$$\bar{Z}_{\mathcal{S}\star\mathcal{H}} = \bar{Z}_{\mathcal{S}} Z_{\mathcal{H}}. \tag{6.5}$$

If $\mathcal{H}[\emptyset] = \emptyset$ we may form the *pointed substitution* $\mathcal{S} \circledcirc \mathcal{H} \subset (\mathcal{G} \circ \mathcal{H})^\circ$ as follows. Any $(\mathcal{G} \circ \mathcal{H})^\circ$-structure P has a marked cycle τ of some automorphism σ. By the discussion in Section 5.4, this cycle corresponds to a cycle on the $\mathcal{G}$-structure of P which does not depend on the choice of σ. Hence the $\mathcal{G}$-structure of P is cycle-pointed and we say P belongs to $\mathcal{S} \circledcirc \mathcal{H}$ if and only if this cycle pointed $\mathcal{G}$-structure belongs to $\mathcal{S}$. The corresponding pointed cycle index sum is given by

$$\bar{Z}_{\mathcal{S}\circledcirc\mathcal{H}} = \bar{Z}_{\mathcal{S}}(Z_{\mathcal{H}}(s_1, s_2, \ldots), \bar{Z}_{\mathcal{H}^\circ}(s_1, t_1; s_2, t_2; \ldots); Z_{\mathcal{H}}(s_2, s_4, \ldots), \bar{Z}_{\mathcal{H}^\circ}(s_2, t_2; s_4, t_4; \ldots); \ldots). \tag{6.6}$$

The sum $\sum_{i\in I} \mathcal{S}_i$ of a family of cycle pointed species $(S_i)_{i\in I}$ is defined as the sum of regular species in Section 5.3. Its pointed cycle index sum satisfies

$$\bar{Z}_{\sum_{i\in I} \mathcal{S}_i} = \sum_{i\in I} \bar{Z}_{\mathcal{S}_i}. \tag{6.7}$$

7 (Pólya-)Boltzmann samplers

Boltzmann samplers were introduced in [18, 19, 20] and generalized to Pólya–Boltzmann samplers in [11]. We briefly discuss the background to the extend required in our proofs.

7.1 Boltzmann models

The *Pólya–Boltzmann model* was introduced in [11]: Suppose that we are given a sequence of real numbers $s_1, s_2, \ldots \geq 0$ such that $0 < Z_{\mathcal{F}}(s_1, s_2, \ldots) < \infty$. Then we may consider the probability distribution on the set $\bigcup_{n=0}^\infty \mathrm{Sym}(\mathcal{F})[n]$ that assigns the probability weight

$$w_{(F,\sigma)} Z_{\mathcal{F}}(s_1, s_2, \ldots)^{-1} = \frac{s_1^{\sigma_1} s_2^{\sigma_2} \cdots}{n!} Z_{\mathcal{F}}(s_1, s_2, \ldots)^{-1}$$

for each n and symmetry $(F, \sigma) \in \mathrm{Sym}(\mathcal{F})[n]$. Here σ_i denotes the number of i-cycles of the permutation σ. The corresponding *Pólya–Boltzmann sampler* is denoted by $\Gamma Z_{\mathcal{F}}(s_1, s_2, \ldots)$, and simply refers to a random variable following this distribution, possibly with a description on how to sample it. When describing a sampling procedure the pseudo-code notation

$$(F, \sigma) \leftarrow \Gamma Z_{\mathcal{F}}(s_1, s_2, \ldots) \tag{7.1}$$

means that we let (F, σ) denote a random $\mathcal{F}$-symmetry that is independent from all previously considered random variables and sampled according to a Pólya–Boltzmann distribution for the species $\mathcal{F}$ with parameters $(s_i)_i$.

Remark 7.1. *In the special case $(s_i)_i = (x^i)_i$ for some $x > 0$, for each fixed n it holds that all outcomes with size n are equally likely. This means that $\Gamma Z_{\mathcal{F}}(x, x^2, \ldots)$ conditioned on having a given deterministic size n follows the uniform distribution. By Lemma 5.1 the n-sized symmetries from $\mathrm{Sym}(\mathcal{F})[n]$ are in a $n : 1$ relation to the unlabelled n-sized $\mathcal{F}$-objects. Thus, the $\mathcal{F}$-object corresponding to the conditioned Pólya Boltzmann sampler is uniformly distributed among all n-sized $\mathcal{F}$-objects.*

A Pólya–Boltzmann model for random cycle pointed species is given by a probability measure on random rooted symmetries: Let $\mathcal{S}$ be a cycle-pointed species. Given real non-negative numbers $(s_i, t_i)_{i \geq 1}$ such that $0 < \bar{Z}_{\mathcal{S}}(s_1, t_1; s_2, t_2; \ldots) < \infty$ we may consider the probability measure on the set $\bigcup_{n=0}^{\infty} \mathrm{RSym}(\mathcal{S})[n]$ that assigns probability weight

$$w_{((G,\tau),\sigma,v)} \bar{Z}_{\mathcal{S}}(s_1, t_1; s_2, t_2; \ldots)^{-1} = \frac{t_\ell s_1^{\sigma_1} \cdots s_{\ell-1}^{\sigma_{\ell-1}} s_\ell^{\sigma_\ell - 1} s_{\ell+1}^{\sigma_{\ell+1}} s_{\ell+2}^{\sigma_{\ell+2}} \cdots}{n! \bar{Z}_{\mathcal{S}}(s_1, t_1; s_2, t_2; \ldots)}$$

for each n to each rooted symmetry $((G,\tau),\sigma,v) \in \mathrm{RSym}(\mathcal{S})[n]$. Here ℓ denotes the lengths of the marked cycle τ. The corresponding Pólya–Boltzmann sampler of this model is denoted by $\Gamma\bar{Z}_{\mathcal{S}}(s_1, t_1; s_2, t_2; \ldots)$, and we use a similar notation as in (7.1) when describing sampling procedures.

Remark 7.2. *In the special case $(s_i, t_i)_i = (x^i, x^i)_i$ for some $x > 0$, for each fixed n we have that all outcomes with size n are equally likely. Hence conditioning $\Gamma\bar{Z}_{\mathcal{S}}(x, x; x^2, x^2; \ldots)$ on having size n yields the uniform distribution on RSym$(\mathcal{F})[n]$. By Lemma 6.1 we know that the rooted symmetries from RSym$(\mathcal{F})[n]$ are in an $n:1$ relation to the unlabelled n-sized cycle-pointed $\mathcal{S}$-objects. Thus, the $\mathcal{S}$-object corresponding to the conditioned Pólya–Boltzmann sampler follows the uniform distribution among all n-sized cycle pointed $\mathcal{S}$-objects.*

7.2 Rules for the construction of Boltzmann samplers

The sampling procedures described in the present exposition were established in [11, Prop. 38, Prop. 43].

7.2.1 Pólya–Boltzmann samplers

Let $\mathcal{F}$ denote a species and $(s_i)_{i\geq 1}$ non-negative real numbers such that

$$0 < Z_{\mathcal{F}}(x_1, x_2, \ldots) < \infty.$$

Products

Suppose that $\mathcal{F} = \mathcal{F}_1 \cdot \mathcal{F}_2$ is the product of two species $\mathcal{F}_1$ and $\mathcal{F}_2$. Then for any finite set U there is a bijection between the set Sym$(\mathcal{F})[U]$ and pairs (S_1, S_2) such that S_i is an $\mathcal{F}_i$-symmetry for all i and the label sets of the S_i partition the set U. This is due to the fact, that given an $\mathcal{F}$-symmetry $((F_1, F_2), \sigma) \in$ Sym$(\mathcal{F})[U]$ the permutation σ must leave the label set Q_i of the $\mathcal{F}_i$-object F_i invariant and satisfy $\sigma|_{Q_i}.F_i = F_i$, that is $(F_i, \sigma|_{Q_i}) \in$ Sym$(\mathcal{F}_i)[Q_i]$. The following procedure is a Pólya–Boltzmann sampler for the species $\mathcal{F}$.

1. For $i = 1, 2$ set
$$S_i \leftarrow \Gamma Z_{\mathcal{F}_i}(s_1, s_2, \ldots).$$
By the bijection for the symmetries of products, the pair (S_1, S_2) corresponds to an $\mathcal{F}$-symmetry (F, σ) over the (exterior) disjoint union U of the label-sets of the S_i.

2. Make a uniformly at random choice for a bijection ν from U to the set of integers $[n]$ with n denoting the size of U. Return the relabelled symmetry
$$\nu.(F, \sigma) = (\nu.F, \nu\sigma\nu^{-1}).$$

Substitution

Suppose that $\mathcal{F} = \mathcal{G} \circ \mathcal{H}$ with $\mathcal{H}[\emptyset] = \emptyset$ is the composition of a species $\mathcal{G}$ with another species $\mathcal{H}$. The symmetries of the substitution were discussed in detail in Section 5.4. The following procedure is a Pólya–Boltzmann sampler for $\mathcal{F}$.

1. Set
$$(G, \sigma) \leftarrow \Gamma Z_{\mathcal{G}}(Z_{\mathcal{H}}(s_1, s_2, \ldots), Z_{\mathcal{H}}(s_2, s_4, \ldots), \ldots).$$
That is, let (G, σ) denote a random $\mathcal{G}$-symmetry that follows a Pólya–Boltzmann distribution with parameters $Z_{\mathcal{H}}(s_1, s_2, \ldots), Z_{\mathcal{H}}(s_2, s_4, \ldots), \ldots$.

2. For each cycle τ of σ let $|\tau|$ denote its lengths and set
$$(H_\tau, \sigma_\tau) \leftarrow \Gamma Z_{\mathcal{H}}(s_{|\tau|}, s_{2|\tau|}, \ldots).$$
That is, the symmetries (H_τ, σ_τ), τ cycle of σ, are independent (conditional on σ) and follow Pólya–Boltzmann distributions.

3. For each cycle τ, make $|\tau|$ identical copies copies of (H_τ, σ_τ) and assemble an $\mathcal{F}$-symmetry (F, γ) out of (G, σ) and the copies of the (H_τ, σ_τ) as described in Proposition 5.3.

4. Choose bijection ν from the vertex set of (F, γ) to an appropriate sized set of integers $[n]$ and return the relabelled symmetry
$$\nu.(F, \gamma) = (\nu.F, \nu\gamma\nu^{-1}).$$

7.2.2 Pólya–Boltzmann samplers for cycle-pointed species

In the following, we suppose that $\mathcal{F}$ is a cycle pointed species and that $s_1, t_1, s_2, t_2, \ldots$ are non-negative real numbers such that
$$0 < \bar{Z}_{\mathcal{F}}(s_1, t_1; s_2, t_2; \ldots) < \infty.$$

Cycle pointed products

Suppose that $\mathcal{F} = \mathcal{G} \star \mathcal{H}$ with $\mathcal{G}$ a cycle-pointed species and $\mathcal{H}$ a species. Then for any finite set U there is a canonical choice for a bijection between the set $\mathrm{RSym}(\mathcal{F})[U]$ and tuples (S_1, S_2) with S_1 a rooted symmetry of $\mathcal{G}$, S_2 a symmetry of $\mathcal{G}$, such that the label sets of S_1 and S_2 form a partition of U. The following procedure is a Pólya–Boltzmann sampler for $\mathcal{F}$.

1. Set
$$S_1 \leftarrow \Gamma \bar{Z}_{\mathcal{G}}(s_1, t_1; s_2, t_2; \ldots).$$

2. Set
$$S_2 \leftarrow \Gamma Z_{\mathcal{H}}(s_1, s_2, \ldots).$$

3. Let U denote the exterior disjoint union of the label sets of S_1 and S_2. The tupel (S_1, S_2) corresponds to a rooted symmetry S over the set U.

4. Make a uniformly at random choice of a bijection ν from U to the set of integers $[n]$ with n denoting the size of U. Return the relabelled rooted symmetry $\nu.S$.

Cycle pointed substitution

Suppose that $\mathcal{F} = \mathcal{G} \circledcirc \mathcal{H}$ with $\mathcal{G}$ cycle-pointed and $\mathcal{H}[\emptyset] = \emptyset$. The symmetries of the substitution were discussed in detail in Section 5.4. The following procedure is a Pólya–Boltzmann sampler for $\mathcal{F}$.

1. Set
$$((G, \tau_0), \sigma, v_0) \leftarrow \Gamma \bar{Z}_{\mathcal{G}}(h_1, \bar{h}_1; h_2, \bar{h}_2; \ldots)$$
with parameters
$$h_i = Z_{\mathcal{H}}(s_i, s_{2i}, \ldots) \quad \text{and} \quad \bar{h}_i = \bar{Z}_{\mathcal{H}^\circ}(s_i, t_i; s_{2i}, t_{2i}; \ldots).$$

2. For each unmarked cycle τ of σ let $|\tau|$ denote its lengths and set
$$(H_\tau, \sigma_\tau) \leftarrow \Gamma Z_{\mathcal{H}}(s_{|\tau|}, s_{2|\tau|}, \ldots).$$

3. For the marked cycle τ_0 set
$$((H_{\tau_0}, c_{\tau_0}), \sigma_{\tau_0}, v_{\tau_0}) \leftarrow \Gamma Z_{\mathcal{H}^\circ}(s_{|\tau_0|}, t_{|\tau_0|}; s_{2|\tau_0|}, t_{2|\tau_0|}; \ldots).$$

4. Assemble an $\mathcal{F}$-symmetry (F, γ) out of the $\mathcal{G}$-symmetry (G, σ) and the $\mathcal{H}$-symmetries (H_τ, σ_τ) according to the construction of Proposition 5.3.

   Let c denote the cycle that gets composed out of the $|\tau_0|$ copies of the cycle c_{τ_0} in this construction. The marked vertex v_{τ_0} has $|\tau_0|$ copies (one for each atom of τ_0) and we let u denote the copy that corresponds to the marked atom v_0 of τ_0. Thus
$$((F, c), \gamma, u)$$
is a rooted symmetry of $\mathcal{F}$.

5. Choose a bijection ν from the vertex set of $((F, c), \gamma, u)$ to an appropriate sized set of integers $[n]$ and return the relabelled rooted symmetry
$$\nu.((F, c), \gamma, u) = ((\nu.F, \nu c \nu^{-1}), \nu \gamma \nu^{-1}, \nu.u).$$

8 Random multisets

If $\mathcal{F}$ is a species of structures with $\mathcal{F}[\emptyset] = \emptyset$, then unlabelled $\mathrm{SET} \circ \mathcal{F}$-objects are termed *multisets*. They consist of unordered collections of unlabelled $\mathcal{F}$-objects where each object is allowed to appear multiple times. The following preliminary observation is a consequence of a more general result established by Barbour and Granovsky [6, Thm. 2.2].

Lemma 8.1. *Suppose that*
$$[z^n]\tilde{\mathcal{F}}(z) = f(n) n^{-\beta} \rho^n$$
for some constants $\rho > 0$ and $\beta > 1$, and a function f that varies slowly at infinity. Then the largest component in a uniform n-sized multiset of unlabelled $\mathcal{F}$-structures has size $n + O_p(1)$.

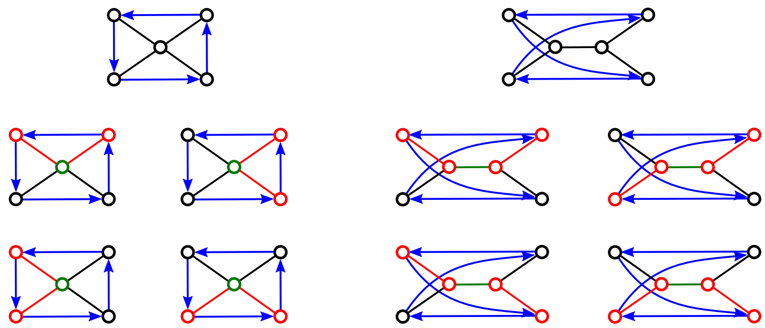

Figure 4: Two unlabelled cycle-pointed trees. The marked cycle is depicted in blue, connecting paths in red, and the cycle-pointing centers in green.

9 Proof of the main theorems

Throughout this section, let Ω be a set of positive integers containing the number 1 and at at least one integer equal or greater than 3. We let $\mathcal{F}$ denote the species of unrooted trees and $\mathcal{F}_\Omega$ its subspecies of trees with vertex degrees in the set Ω. Analogously, we let $\mathcal{A}$ denote the species of rooted trees and $\mathcal{A}_{\Omega^*}$ the subspecies of rooted trees with vertex outdegrees in the shifted set $\Omega^* = \Omega - 1$. In the following we will always assume that n denotes an integer satisfying $n \equiv 2 \mod \gcd(\Omega^*)$ and n large enough such that trees with n vertices and vertex degrees in the set Ω exist. Let ρ denote the radius of convergence of the generating series $\tilde{A}_{\Omega^*}(z)$.

We let (T_n, τ_n) denote a random cycle-pointed tree drawn uniformly from the unlabelled $\mathcal{F}_\Omega^\circ$-objects of size n. As discussed in Lemma 5.1, this implies that T_n is the uniform random unlabelled unrooted tree with n vertices and vertex degrees in the set Ω. Moreover, let A_{n-1} a random rooted tree drawn uniformly from the unlabelled $\mathcal{A}_{\Omega^*}$-objects of size $n-1$.

We let $c_{\Omega^*} > 0$ denote the constant from Equation (3.1) such that the uniformly drawn unlabelled rooted tree A_{n-1} satisfies

$$(\mathsf{A}_{n-1}, c_{\Omega^*} n^{-1/2} d_{\mathsf{A}_{n-1}}) \xrightarrow{d} (\mathcal{T}_{\mathsf{e}}, d_{\mathcal{T}_{\mathsf{e}}})$$

with respect to the Gromov–Hausdorff metric. Moreover, let $\hat{\mathsf{A}}_{\Omega^*}$ denote the infinite rooted tree from Equation (4.1) with

$$d_{\mathsf{TV}}(V_{k_n}(\mathsf{A}_{n-1}, u_{n-1}), V_{k_n}(\hat{\mathsf{A}}_{\Omega^*})) \to 0$$

for every sequence $k_n = o(\sqrt{n})$, with u_{n-1} denoting a uniformly at random selected vertex of A_{n-1}.

9.1 Decomposition of cycle-pointed trees

Given a cycle pointed tree (T, τ) such that the marked cycle τ has length at least 2 we may consider its *connecting paths*, i.e. the paths in T that join consecutive atoms of τ. Any such path has a middle, which is either a vertex if the path has odd length, or an edge if the path has even length. All connecting paths have the same lengths and by [11, Claim 22] they share the same middle, called the *center of symmetry*. See Figure 4 for an illustration.

The cycle pointing decomposition given in [11, Prop. 25] splits the species $\mathcal{F}_\Omega^\circ$ into three parts,

$$\mathcal{F}_\Omega^\circ \simeq \mathcal{X}^\circ \star (\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*}) + \mathsf{SET}_{\{2\}}^\circledast \circledcirc \mathcal{A}_{\Omega^*} + (\mathsf{SET}_\Omega^\circledast \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}. \tag{9.1}$$

Here

$$\mathcal{S} := \mathcal{X}^\circ \star (\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*})$$

corresponds to the trees with a marked fixpoint and the other summands to trees with a marked cycle of length at least two. More specifically,

$$\mathcal{E} := \mathsf{SET}_{\{2\}}^\circledast \circledcirc \mathcal{A}_{\Omega^*}$$

corresponds to the symmetric cycle pointed trees whose center of symmetry is an edge and

$$\mathcal{V} := (\mathsf{SET}_\Omega^{\circledast} \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}$$

to those whose center of symmetry is a vertex.

9.2 Enumerative properties

We start by collecting some basic enumerative facts. The following preliminary observation summarizes enumerative properties of Pólya trees with vertex degree restrictions.

Proposition 9.1 ([30, Prop. 4.1])**.** *The following statements hold.*

i) The radius of convergence ρ of the series $\tilde{\mathcal{A}}_{\Omega^*}(z)$ satisfies $0 < \rho < 1$ and $\tilde{\mathcal{A}}_{\Omega^*}(\rho) < \infty$.

ii) There is a positive constant d_{Ω^*} such that

$$[z^m]\tilde{\mathcal{A}}_{\Omega^*}(z) \sim d_{\Omega^*} m^{-3/2} \rho^{-m}$$

as the number $m \equiv 1 \mod \gcd(\Omega^*)$ tends to infinity.

iii) For any subset $\Lambda \subset \mathbb{N}$ the series

$$E^\Lambda(z, w) = z Z_{\mathsf{SET}_\Lambda}(w, \tilde{\mathcal{A}}_{\Omega^*}(z^2), \tilde{\mathcal{A}}_{\Omega^*}(z^3), \ldots)$$

satisfies

$$E^\Lambda(\rho + \epsilon, \tilde{\mathcal{A}}_{\Omega^*}(\rho) + \epsilon) < \infty$$

for some $\epsilon > 0$.

In [11, Prop. 24] the cycle-pointing decomposition was used in order to provide a new method for determining the asymptotic number of unlabelled unrooted trees. This may be extended to the case of vertex degree restrictions. A detailed justification is given in Section 9.5 below.

Proposition 9.2. *The series $\tilde{\mathcal{F}}_\Omega(z)$ and $\tilde{\mathcal{A}}_{\Omega^*}(z)$ both have the same radius of convergence ρ. Moreover, the following statements hold.*

i) There is a constant d'_{Ω^*} such that

$$[z^n]\tilde{\mathcal{F}}_\Omega(z) \sim d'_{\Omega^*} \rho^{-n} n^{-5/2}$$

as $n \equiv 2 \mod \gcd(\Omega^*)$ tends to infinity.

ii) For any set $\Lambda \subset \mathbb{N}$ the series

$$F^\Lambda(z, w) = \bar{Z}_{\mathsf{SET}_\Lambda^{\circledast}}(w, \tilde{\mathcal{A}}^\circ_{\Omega^*}(z); \tilde{\mathcal{A}}_{\Omega^*}(z^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}(z^2); \tilde{\mathcal{A}}_{\Omega^*}(z^3), \tilde{\mathcal{A}}^\circ_{\Omega^*}(z^3); \ldots)$$

satisfies $F^\Lambda(\rho + \epsilon, \tilde{\mathcal{A}}_{\Omega^*}(\rho) + \epsilon) < 0$ for some $\epsilon > 0$.

iii) The ordinary generating series

$$\widetilde{\mathsf{SET}_{\{2\}}^{\circledast} \circledcirc \mathcal{A}_{\Omega^*}}(z) = \tilde{\mathcal{A}}^\circ_{\Omega^*}(z^2)$$

has radius of convergence greater than ρ.

9.3 Approximation arguments

We are going to treat the classes $\mathcal{S}$, $\mathcal{E}$, and $\mathcal{V}$ separately.

9.3.1 The class $\mathcal{E}$ of symmetric cycle pointed trees whose center of symmetry is an edge

The event $(\mathsf{T}_n, \tau_n) \in \mathcal{E}$ is so unlikely, that we will be able to neglect this case:

Lemma 9.3. *There are constants $C, c > 0$, such that for all n*

$$\mathbb{P}((\mathsf{T}_n, \tau_n) \in \mathcal{E}) \le C \exp(-cn).$$

Geometrically speaking, this can be explained by the fact that any unlabelled cycle pointed tree from $\mathcal{E}$ corresponds bijectively to a cycle pointed Pólya tree from $\mathcal{A}^{\circ}_{\Omega^*}$ having precisely half of its size. Compare with Figure 5. The number of such objects is roughly given by $\rho^{n/2}$, while the number of all cycle pointed trees in $\mathcal{F}^{\circ}_{\Omega}$ is roughly given by ρ^n, which is exponentially larger.

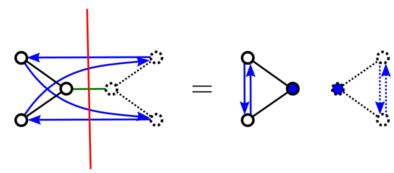

Figure 5: Any unlabelled $\mathcal{E} = \mathsf{SET}^{\circledast}_{\{2\}} \circledcirc \mathcal{A}_{\Omega^*}$ object corresponds to two identical copies of a cycle-pointed Pólya tree.

9.3.2 The class $\mathcal{S}$ of cycle pointed trees with a marked fixpoint

Lemma 9.4. *Let S_n be drawn uniformly at random from the unlabelled*

$$\mathcal{S} = \mathcal{X}^{\circ} \star (\mathsf{SET}_{\Omega} \circ \mathcal{A}_{\Omega^*})$$

objects of size n. Then the following properties hold.

a) *There are constants $C, c > 0$ such that for all n and $x \ge 0$ it holds that*

$$\mathbb{P}(\mathsf{D}(\mathsf{T}_n) \ge x) \le C \exp(-cx^2/n).$$

b) *There is a random number $K_n = n + O_p(1) \le n$ and a coupling of S_n with a partition into two rooted subtrees B_n, C_n that intersect only in their roots and satisfy $\mathsf{C}_n \overset{(d)}{=} \mathsf{A}_{K_n}$.*

The reason for this is, that each unlabelled $\mathcal{S} = \mathcal{X}^{\circ} \star (\mathsf{SET}_{\Omega} \circ \mathcal{A}_{\Omega^*})$ cycle pointed trees corresponds bijectively to a Pólya tree, in which each vertex degree must lie in Ω. That is, the outdegree of the root lies in Ω, and the outdegrees of all remaining vertices lie in Ω^*. Compare with Figure 6.

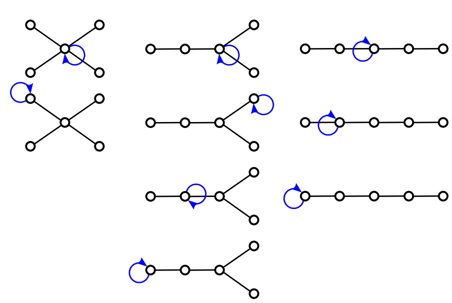

Figure 6: Unlabelled $\mathcal{S} = \mathcal{X}^{\circ} \star (\mathsf{SET}_{\Omega} \circ \mathcal{A}_{\Omega^*})$ cycle pointed trees correspond to Pólya trees, in which each vertex degree must lie in Ω.

9.3.3 The class $\mathcal{V}$ of symmetric cycle pointed trees whose center of symmetry is a vertex

Lemma 9.5. *Let* V_n *be drawn uniformly from the unlabelled*

$$\mathcal{V} = (\mathsf{SET}_\Omega^{\circledast} \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}$$

objects of size n. *Then the following statements hold.*

a) *There are constants* $C, c > 0$ *such that for all* $x \geq 0$ *and* n *we have the tail bound*

$$\mathbb{P}(\mathsf{D}(\mathsf{V}_n) \geq x) \leq C \exp(-cx^2/n).$$

b) *There is a random number* $K_n = n + O_p(1) \leq n$ *and a coupling of* V_n *with a partition into two rooted subtrees* B_n, C_n *that intersect only in their roots and satisfy* $\mathsf{C}_n \stackrel{(d)}{=} \mathsf{A}_{K_n}$.

The key point is that any unlabelled cycle pointed tree from $\mathcal{V}$ corresponds to a Pólya tree A from $\mathcal{A}_{\Omega^*}$ where each non-root vertex must have outdegrees in Ω^*, together with a number K of identical copies of a symmetrically cycle pointed Pólya tree A° from $A_{\Omega^*}^{\circledast}$, such that the sum of the root degrees of A and the K copies of A° lies in Ω. Compare with Figure 7.

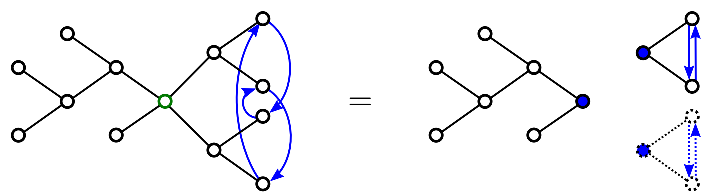

Figure 7: Decomposition of an unlabelled $\mathcal{V} = (\mathsf{SET}_\Omega^{\circledast} \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}$ object into a Pólya tree and a number of identical copies of a symmetrically cycle-pointed Pólya tree.

9.4 Proof of the main results: Theorems 1.1, 1.2, and 1.3

Having the approximation results from Section 9.3 at hand, we may verify our main results.

Proof of Theorem 1.3. Lemma 9.3 implies that the total variation distance between the unrooted tree T_n and a mixture of random $\mathcal{S}$-structures and $\mathcal{V}$-structures is exponentially small. Lemmas 9.4 and 9.5 imply that both S_n and V_n look like a large randomly sized Pólya tree with a stochastically bounded rest. Consequently their mixture looks like a large randomly sized Pólya tree with a small rest which is a mixture of the two stochastically bounded small trees corresponding to S_n and V_n. This completes the proof. □

Proof of Theorem 1.2. Theorem 1.3 reduces the problem to studying a certain random tree $\bar{\mathsf{T}}_n$, that consists of a $K_n = n + O_p(1)$-sized random Pólya tree A_{K_n} with a small tree B_n attached to its root. For the local limit, let u_n denote a uniformly at random drawn vertex of $\bar{\mathsf{T}}_n$, and let $k_n = o(\sqrt{n})$ denote a given sequence. It is clear that the random vertex u_n lies with high probability in the subtree A_{K_n}, and that, conditioned on this event, it is uniformly distributed among its vertices. Note that $K_n = n + O_p(1)$ implies that with high probability $K_n \geq n - \log n \to \infty$ and $k_n = o(\sqrt{n}) = o(\sqrt{K_n})$. By Equation (4.1) it follows that the radius k_n neighbourhood of a random vertex in A_{K_n} is close in total variation to the k_n neighbourhood of the infinite random tree $\hat{\mathsf{A}}_{\Omega^*}$, and that a random vertex in A_{K_n} has with high probability height strictly larger than k_n. In particular, with high probability the neighbourhood does not contain the root-vertex of A_{K_n} and is hence not influenced by the small tree B_n that gets attached to the root of A_{K_n} to form the tree $\bar{\mathsf{T}}_n$. This readily verifies that

$$d_{\mathsf{TV}}(V_{k_n}(\bar{\mathsf{T}}_n, u_n), V_{k_n}(\hat{\mathsf{A}}_{\Omega^*})) \to 0,$$

and hence completes the proof. □

Proof of Theorem 1.1. For the scaling limit, it suffices by Theorem 1.3 to consider the tree $\bar{\mathsf{T}}_n$, that consists of two rooted trees glued together at their root vertices, specifically a $K_n = n + O_p(1)$-sized random Pólya tree A_{K_n} and a small tree B_n. As $|\mathsf{B}_n| = O_p(1)$ it follows that with high probability it holds that, say, $|\mathsf{B}_n| \le n^{1/4}$. Hence

$$d_{\mathsf{GH}}(\bar{\mathsf{T}}_n/\sqrt{n}, \mathsf{A}_{K_n}/\sqrt{n}) \xrightarrow{p} 0. \tag{9.2}$$

Note that $\mathsf{K}_n \xrightarrow{d} \infty$ and the limit in (3.1) imply that

$$c_{\Omega^*}\mathsf{A}_{K_n}/\sqrt{K_n} \xrightarrow{d} \mathcal{T}_{\mathsf{e}}. \tag{9.3}$$

In particular, $\mathsf{D}(\mathsf{A}_{K_n}) = O_p(\sqrt{K_n})$ and hence

$$d_{\mathsf{GH}}(\mathsf{A}_{K_n}/\sqrt{n}, \mathsf{A}_{K_n}/\sqrt{K_n}) \le O_p(1)(1 - \sqrt{K_n/n}) \xrightarrow{p} 0.$$

Together with Equation (9.2) this implies that

$$d_{\mathsf{GH}}(\bar{\mathsf{T}}_n/\sqrt{n}, \mathsf{A}_{K_n}/\sqrt{K_n}) \xrightarrow{p} 0$$

and by the limit in (9.3) the scaling limit for $\bar{\mathsf{T}}_n$ follows.

The inclined reader may note that the arguments above work just as fine for the Gromov–Hausdorff–Prokhorov metric with respect to the uniform measure on the leaves or on all vertices, if we build on Gromov–Hausdorff–Prokhorov convergence for the random Pólya trees, which was established in [24] for the most important cases of vertex outdegree restrictions.

For the tail bound of the diameter, note that it suffices to show such a bound for $\mathbb{P}(\mathsf{D}(\mathsf{T}_n) \ge x)$ when $x \le n$. By Lemmas 9.3, 9.4 and 9.5 it follows that there are constants $C_i, c_i > 0$, for $i = 1, 2, 3$, such that

$$\begin{aligned}\mathbb{P}(\mathsf{D}(\mathsf{T}_n) \ge x) &\le \sum_{\mathcal{B} \in \{\mathcal{E}, \mathcal{S}, \mathcal{V}\}} \mathbb{P}(\mathsf{D}(\mathsf{T}_n) \ge x \mid (\mathsf{T}_n, \tau_n) \in \mathcal{B})\ \mathbb{P}((\mathsf{T}_n, \tau_n) \in \mathcal{B}) \\ &\le C_1 \exp(-c_1 n) + \sum_{i=2}^{3} C_i \exp(-c_i x^2/n).\end{aligned}$$

As we assumed that $x \le n$, it holds that

$$\exp(-c_1 n) \le \exp(-c_1 x^2/n).$$

Hence for a suitable choice of constants $C, c > 0$, it follows that

$$\mathbb{P}(\mathsf{D}(\mathsf{T}_n) \ge x) \le C \exp(-c x^2/n)$$

for all n and $x \ge 0$. □

9.5 Proof of the enumerative observation Proposition 9.2

Proof of Proposition 9.2. We start with Claim $iii)$. By the definition of the SET-species and the pointed cycle index sum it holds that $\bar{Z}_{\mathsf{SET}^{\circledast}_{\{2\}}} = t_2$. By Equations (6.6) and (6.2) it follows that

$$\widetilde{\mathsf{SET}^{\circledast}_{\{2\}} \circledcirc \mathcal{A}_{\Omega^*}}(z) = \bar{Z}_{\mathsf{SET}^{\circledast}_{\{2\}} \circledcirc \mathcal{A}_{\Omega^*}}(z, z; z^2, z^2; \ldots) = \tilde{\mathcal{A}}^{\circ}_{\Omega^*}(z^2).$$

By Proposition 9.1 we know that $0 < \rho < 1$. Hence the series

$$\tilde{\mathcal{A}}^{\circ}_{\Omega^*}(z) = z \frac{\mathsf{d}}{\mathsf{d}z} \tilde{\mathcal{A}}_{\Omega^*}(z)$$

has radius of convergence $\sqrt{\rho} > \rho$. This proves Claim iii).

We proceed with Claim ii). The series $\bar{Z}_{\mathsf{SET}^{\circledast}_{\Lambda}}$ is dominated coefficient-wise by the series $\bar{Z}_{\mathsf{SET}^{\circledast}}$, which by Equation (6.4) is given by

$$\bar{Z}_{\mathsf{SET}^{\circledast}}(s_1, t_1; s_2, t_2; \dots) = \exp\left(\sum_{k=1}^{\infty} s_k/k\right) \sum_{i=2}^{\infty} t_i.$$

It follows that $F^{\Lambda}(z, w)$ is dominated coefficient-wise by

$$\exp\left(w + \sum_{k=2}^{\infty} \tilde{A}_{\Omega^*}(z^k)/k\right) \sum_{i=2}^{\infty} \tilde{A}^{\circ}_{\Omega^*}(z^i).$$

Since $0 < \rho < 1$ and $\tilde{A}_{\Omega^*}(\rho) < \infty$ by Proposition 9.1, this series is finite for $z = \rho+\epsilon$ and $w = \tilde{A}_{\Omega^*}(\rho)+\epsilon$, if $\epsilon > 0$ is sufficiently small. This proves Claim ii).

In order to prove Claim i), note that by Equation (6.3) it holds that $\tilde{\mathcal{F}}^{\circ}_{\Omega}(z) = z\frac{\mathrm{d}}{\mathrm{d}z}\tilde{\mathcal{F}}_{\Omega}(z)$. Hence it suffices to study the ordinary generating series of the species $\mathcal{F}^{\circ}_{\Omega}$. By the cycle pointing decomposition in Equation (9.1) it holds that

$$\mathcal{F}^{\circ}_{\Omega} \simeq \mathcal{X}^{\circ} \star (\mathsf{SET}_{\Omega} \circ \mathcal{A}_{\Omega^*}) + \mathsf{SET}^{\circledast}_{\{2\}} \circledcirc \mathcal{A}_{\Omega^*} + (\mathsf{SET}^{\circledast}_{\Omega} \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}.$$

By the relation of pointed cycle index sums with ordinary generating series given in Equation (6.2) and the rules (6.7), (6.5), and (6.6) for cycle index sums of the the sum, product and substitution, it follows that

$$\tilde{\mathcal{F}}^{\circ}_{\Omega}(z) = zh(z, \tilde{A}_{\Omega^*}(z))$$

with

$$h(z, w) = E^{\Omega}(z, w) + F^{\Omega}(z, w) + \tilde{A}^{\circ}_{\Omega^*}(z^2)/z.$$

Here we let E^{Ω} be defined as in Proposition 9.1. The power series $h(z, w)$ has non-negative coefficients. By Claims i) and ii) and Proposition 9.1 it holds furthermore that

$$h(\tilde{A}_{\Omega^*}(\rho) + \epsilon, \rho + \epsilon) < \infty$$

for some $\epsilon > 0$. This means that the behaviour of the series $\tilde{A}_{\Omega^*}(z)$ at its singularities on the circle $|z| = \rho$ determines the asymptotic growth of the coefficients of the series $\tilde{\mathcal{F}}^{\circ}_{\Omega}(z)$.

Let us make this precise. A rooted tree with vertex outdegrees in the set Ω^* consists of a root-vertex together with an unordered list of fringe subtrees dangling from it, such that the total number of fringe subtrees lies in the set Ω^*, and each of these trees belongs to the species $\mathcal{A}_{\Omega^*}$. In the language of combinatorial species, this may concisely be expressed by

$$\mathcal{A}_{\Omega^*} \simeq \mathcal{X} \cdot \mathsf{SET}_{\Omega^*}(\mathcal{A}_{\Omega^*}).$$

By Equation (1) and the rules for the interplay of the cycle index sums with the operations on species summarized in Table 1 it follows that

$$\tilde{A}_{\Omega^*}(z) = E^{\Omega^*}(z, \tilde{A}_{\Omega^*}(z)).$$

By Proposition 9.1 it holds that

$$E^{\Omega^*}(\rho + \epsilon, \tilde{A}_{\Omega^*}(\rho) + \epsilon) < \infty$$

for some $\epsilon > 0$. By a general enumerative result by Bell, Burris and Yeats [7, Lem. 26, Cor. 12] and the rotational symmetry $\tilde{A}_{\Omega^*}(\zeta z) = \zeta \tilde{A}_{\Omega^*}(z)$ it follows that the series $\tilde{A}_{\Omega^*}(z)$ has dominant singularities (all of square-root type) precisely at the rotated points

$$\zeta^k, \qquad k = 0, \dots, d - 1$$

with

$$\zeta = e^{\frac{2\pi i}{d}} \qquad \text{and} \qquad d = \gcd(\Omega^*).$$

By the asymptotic expansion of the coefficients of $\tilde{\mathcal{A}}_{\Omega^*}(z)$ in Proposition 9.1 and a singularity analysis result for functions with multiple dominant singularities [21, Thm. VI.5] it follows that there is a constant $d' > 0$ with

$$[z^m]h(z, \tilde{\mathcal{A}}_{\Omega^*}(z)) \sim d' m^{-3/2} \rho^{-m} \tag{9.4}$$

as $m \equiv 1 \mod d$ becomes large. Consequently,

$$[z^n]\tilde{\mathcal{F}}_\Omega(z) \sim n^{-1}[z^n]\tilde{\mathcal{F}}_\Omega^\circ(z) \sim d'_{\Omega^*} n^{-5/2} \rho^{-n}$$

for some constant $d'_{\Omega^*} > 0$ as $n \equiv 2 \mod d$ becomes large. □

9.6 Proofs of the approximation arguments: Lemmas 9.3, 9.4, and 9.5

9.6.1 Cycle pointed trees whose cycle center is an edge

Proof of Lemma 9.3. The probability for this event is given by the ratio of unlabelled cycle pointed trees of $\mathcal{E}$ with n vertices, and the unlabelled cycle pointed trees in $\mathcal{F}_\Omega$ with n vertices. Hence

$$\mathbb{P}\left((\mathsf{T}_n, \tau_n) \in \mathcal{E}\right) = \frac{[z^n]\tilde{\mathcal{E}}(z)}{[z^n]\tilde{\mathcal{F}}^\circ(z)}.$$

By Proposition 9.2, $iii)$, the radius of convergence of the ordinary generating series $\tilde{\mathcal{E}}(z)$ is strictly larger than the radius of convergence ρ of $\tilde{\mathcal{F}}^\circ(z)$. This yields the claim. □

9.6.2 Cycle pointed trees whose cycle center is a fixpoint

It holds that

$$\mathcal{S} = \mathcal{X}^\circ \star (\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*}) \simeq \mathcal{X} \cdot (\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*}),$$

hence we do not require cycle pointing techniques in this case. Let (S_n, σ) be drawn uniformly at random from the set $\mathsf{Sym}(\mathcal{S})[n]$. Let π_n denote the corresponding partition. By the discussion in Section 5.4, σ induces an automorphism

$$\bar{\sigma} : \pi_n \to \pi_n$$

of the SET_Ω-object. Moreover, let $F_n \subset \pi_n$ denote the fixpoints of $\bar{\sigma}$, $f_n = |F_n|$ their number and for each fixpoint $Q \in F_n$ let (A_Q, σ_Q) denote the corresponding symmetry from $\mathsf{Sym}(\mathcal{A}_{\Omega^*})[Q]$. Let H_n denote the total size of the trees dangling from cycles with length at least 2. We are going to make the following observations.

Lemma 9.6. *The following statements hold.*

1) There are constants $C_1 > 0$ and $0 < \gamma < 1$ such that for all n and $x \geq 0$ we have that

$$\mathbb{P}\left(H_n \geq x\right) \leq C_1 n^{3/2} \gamma^x$$

and

$$\mathbb{P}\left(f_n \geq x\right) \leq C_1 n^{3/2} \gamma^x.$$

2) The maximum size of the individual trees corresponding to the fixpoints of $\bar{\sigma}$ satisfies

$$\max_{Q \in F_n} |\mathsf{A}_Q| = n + O_p(1).$$

3) There is a constant $C_2 > 0$ such that

$$\mathbb{E}\left[f_n\right] \leq C_2$$

for all n.

We are first going to argue how these claims suffice to prove Lemma 9.4:

Proof of Lemma 9.4. We start with Claim a), the tail bound for the diameter. First, it suffices to verify such a bound uniformly for all n and $\sqrt{n} \leq x \leq n$. If $\mathsf{D}(\mathsf{S}_n) \geq x$, then it must hold that $H_n \geq x/2$ or $\max_{Q \in F_n} \mathsf{H}(\mathsf{A}_Q) \geq x/2 - 1$. Hence

$$\mathbb{P}\left(\mathsf{D}(\mathsf{S}_n) \geq x\right) \leq \mathbb{P}\left(H_n \geq x/2\right) + \mathbb{P}\left(\max_{Q \in F_n} \mathsf{H}(\mathsf{A}_Q) \geq x/2 - 1\right). \tag{9.5}$$

By Claim 1) of Lemma 9.6, we know that there are constants $C_1 > 0$ and $0 < \gamma < 1$ with

$$\mathbb{P}\left(H_n \geq x/2\right) \leq C_1 n^{3/2} \gamma^{x/2}.$$

Hence there are constants $C_4, c_4 > 0$ such that for all n and $\sqrt{n} \leq x \leq n$ it holds that

$$\mathbb{P}\left(H_n \geq x/2\right) \leq C_4 \exp(-c_4 x^2/n). \tag{9.6}$$

It remains to bound the second summand in Equation (9.5). Let $F \subset [n]$ be a subset with $\mathbb{P}\left(F_n = F\right) > 0$. Given $F_n = F$, it follows by the discussion of the symmetries of composite structures in Section 5.4 that the symmetries $(\mathsf{A}_Q, \sigma_Q)_{Q \in F}$ are independent and for each $Q \in F$ we have that (A_Q, σ_Q) gets drawn uniformly at random from the set $\mathrm{Sym}(\mathcal{A}_{\Omega^*})[Q]$. That is, by Lemma 5.1, A_Q gets drawn uniformly at random from all $|Q|$-sized Pólya trees with outdegrees in the set Ω^*. By Inequality (3.2) it follows that there are positive constants C_5, c_5 such that uniformly for all n and x

$$\mathbb{P}\left(\max_{Q} \mathsf{H}(\mathsf{A}_Q) \geq x/2 - 1 \mid F_n = F\right) \leq C_5 \sum_{Q \in F} \exp(-c_4 x^2/|Q|) \leq |F| C_4 \exp(-c_5 x^2/n).$$

Hence

$$\mathbb{P}\left(\max_{Q} \mathsf{H}(\mathsf{A}_Q) \geq x/2 - 1\right) \leq C_5 \exp(-c_5 x^2/n) \sum_{F} \mathbb{P}\left(F_n = F\right)|F| \leq \mathbb{E}\left[f_n\right] C_5 \exp(-c_5 x^2/n). \tag{9.7}$$

By 3) it holds that

$$\mathbb{E}\left[f_n\right] \leq C_2$$

for all n. Thus, by Equations (9.5) and (9.6) it holds for some $C_6, c_6 > 0$ that

$$\mathbb{P}\left(\mathsf{D}(\mathsf{S}_n) \geq x\right) \leq C_4 \exp(-c_4 x^2/n) + C_2 C_5 \exp(-c_5 x^2/n) \leq C_6 \exp(-c_6 x^2/n)$$

uniformly for all n and $\sqrt{n} \leq x \leq n$. This verifies Claim a).

We continue with Claim b), the approximation argument. Make a canonical choice of a partition class of F_n with maximal size and let X_n denote the corresponding tree. Then, by Lemma 5.1, for all ℓ

$$(\mathsf{X}_n \mid |\mathsf{X}_n| = \ell) \overset{(d)}{=} \mathsf{A}_\ell. \tag{9.8}$$

Thus, setting $K_n = |X_n|$, it holds that $\mathsf{X}_n \overset{(d)}{=} \mathsf{A}_{K_n}$. By Claim 2) of Lemma 9.6 we have $|K_n| = n + O_p(1)$, hence the remainder that gets attached to the root of X_n to form the tree S_n is stochastically bounded. This completes the proof of Claim b). $\square$

It remains to verify Lemma 9.6.

Proof of Lemma 9.6. We start with the first claim. By the discussion of Boltzmann samplers in Section 7.2.1 regarding the product and substitution operation, the probability generating function of H_n is given by

$$\mathbb{E}\left[w^{H_n}\right] = \frac{[z^{n-1}]Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^2), \tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^3), \ldots)}{[z^{n-1}]Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \ldots)}. \tag{9.9}$$

Let us justify Equation (9.9) in more detail: By the product rule in Section 7.2.1 it suffices, to study $(n-1)$-sized symmetries of $\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*}$. The substitution rule in Section 7.2.1 tells us that a Boltzmann distributed symmetry of this composition with parameters $(\rho^i)_{i\geq 1}$ is obtained by first drawing a Pólya–Boltzmann distributed SET_Ω-symmetry with parameters $(\tilde{\mathcal{A}}_{\Omega^*}(\rho^i))_{i\geq 1}$, and then for each $j \geq 1$ and each j-cycle of the symmetry an unlabelled Boltzmann distributed symmetry of $\mathcal{A}_{\Omega^*}$ with parameters $(\rho^{ij})_{i\geq 1}$, of which i identical copies are attached to the SET_Ω-symmetry. Given a $k \in \Omega$ sized permutation ν, the probability for the SET_Ω-symmetry to assume this permutation is given by

$$\frac{\tilde{\mathcal{A}}_{\Omega^*}(\rho)^{\nu_1} \cdots \tilde{\mathcal{A}}_{\Omega^*}(\rho^k)^{\nu_k}}{k!\widetilde{\mathsf{SET}_\Omega \circ \mathcal{A}}_{\Omega^*}(\rho)} \tag{9.10}$$

Conditioned on this event, the probability generating function for the size of the resulting object is given by

$$\left(\frac{\tilde{\mathcal{A}}_{\Omega^*}(\rho z)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho)}\right)^{\nu_1} \left(\frac{\tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho^2)}\right)^{\nu_2} \cdots \left(\frac{\tilde{\mathcal{A}}_{\Omega^*}((\rho z)^k)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho^k)}\right)^{\nu_k}. \tag{9.11}$$

The exponents in the arguments are due to the fact that we attach i identical copies of each tree corresponding to an i-cycle. If we additionally want to keep track of the volume of the trees corresponding to cycles with length at least 2, we may form the corresponding bivariate probability generating function where w corresponds to this parameter and z to the total size by

$$\left(\frac{\tilde{\mathcal{A}}_{\Omega^*}(\rho z)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho)}\right)^{\nu_1} \left(\frac{\tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^2)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho^2)}\right)^{\nu_2} \cdots \left(\frac{\tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^k)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho^k)}\right)^{\nu_k}. \tag{9.12}$$

Multiplying (9.10) and (9.12) and summing over all outcomes that correspond to objects with size $n-1$ yields

$$\frac{[z^{n-1}]Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^2), \tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^3), \ldots)}{\widetilde{\mathsf{SET}_\Omega \circ \mathcal{A}}_{\Omega^*}(\rho)}. \tag{9.13}$$

Likewise multiplying (9.10) with (9.11) and summing up in the same way yields

$$\frac{[z^{n-1}]Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \ldots)}{\widetilde{\mathsf{SET}_\Omega \circ \mathcal{A}}_{\Omega^*}(\rho)}. \tag{9.14}$$

The quotient of (9.13) and (9.14) is the probability generating function for the random number H_n, and the expression obtained in this way agrees with Equation (9.9).

Having verified Equation (9.9) we proceed with the argument. Since $1 \in \Omega$ we may bound the denominator in (9.9) from below by $[z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(\rho z)$, and by Proposition 9.1 we have that

$$[z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(\rho z) \sim C n^{-3/2} \tag{9.15}$$

for some constant $C > 0$ as $n \equiv 2 \mod \gcd(\Omega^*)$ tends to infinity. Moreover, for all n the polynomial in the indeterminate w in the numerator is dominated coefficient wise by the series

$$Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}_{\Omega^*}((\rho w)^2), \ldots)$$

which by Proposition 9.1 has radius of convergence strictly greater than 1. In particular we have that

$$\sum_{k \geq x} [w^k] Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}_{\Omega^*}((\rho w)^2), \ldots) = O(\gamma^x)$$

for some constant $0 < \gamma < 1$. Hence there is a constant C' such that

$$\mathbb{P}(H_n \geq x) \leq C' n^{3/2} \gamma^x$$

for all n and x. By the discussion of Boltzmann samplers in Section 7.2.1 regarding the product and substitution operation, the probability generating function for the random number f_n is given by

$$\mathbb{E}\left[w^{f_n}\right] = \frac{[z^{n-1}] Z_{\mathsf{SET}_\Omega}(w\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \ldots)}{[z^{n-1}] Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \ldots)}. \tag{9.16}$$

The corresponding bound for the event $f_n \geq x$ follows by the same arguments as for the parameter H_n. This proves Claim 1).

We proceed with showing Claim 2). If $\Omega = \mathbb{N}$, then we may apply Lemma 8.1 to obtain that the largest component in a random $(n-1)$-sized multiset of unlabelled $\mathcal{A}_{\Omega^*}$-objects has size $n + O_p(1)$. By Claim 1) it follows that with high probability $H_n \leq \log^2 n$. Thus the largest component must correspond to a fixpoint, verifying Claim 2) for this special case. In order to treat the general case, it suffices by similar arguments to show that the largest component in a random $(n-1)$-sized unlabelled $\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*}$-object has size $n + O_p(1)$. However, we cannot apply Lemma 8.1 directly, and hence argue as follows.

We need to show that for any sequence $t_n \to \infty$ the probability for all components in the random $\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*}$-object to have size at most $n - t_n$ tends to zero. Using analogous arguments as in the justification of Equation (9.9), we may express this probability by the product of the normalizing factor

$$([z^{n-1}] Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \ldots))^{-1} \tag{9.17}$$

with the expression

$$\sum_{k \in \Omega} \frac{1}{k!} \sum_{\nu} \sum_{(a_{ij})_{i,j}} \left[z^{n-1} \prod_{i,j} x_{ij}^{a_{ij}} \right] \prod_{i,j} \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^i x_{ij}). \tag{9.18}$$

Here the sum index ν ranges over all permutations of the set $[k]$. The indices $(a_{ij})_{i,j}$ range over all families of numbers a_{ij} with $1 \leq i \leq n-1$, $1 \leq j \leq \nu_i$, and such that $a_{ij} \leq n - t_n$ for all i, j and

$$\sum_{\substack{1 \leq i \leq n-1 \\ 1 \leq j \leq \nu_i}} i a_{ij} = n - 1.$$

The indices i, j of the product range over all pairs of integers with $1 \leq i \leq n-1$ and $1 \leq j \leq \nu_i$.

Applying a standard result for the singularity analysis of functions with multiple dominant singularities [21, Thm. VI.5] we obtain analogously as in the justification of Equation (9.4) that the factor in Equation (9.17) is asymptotically equivalent to $n^{3/2}$ times a constant. Thus, showing that the largest component in a random unlabelled $n-1$-sized $\mathsf{SET}_\Omega \circ \mathcal{A}_{\Omega^*}$-object has size $n + O_p(1)$ is actually equivalent to showing that the expression in (9.18) multiplied by $n^{3/2}$ tends to zero as $n \equiv 2 \mod \gcd(\Omega^*)$ becomes large. Consider the species $\bar{\mathcal{A}}_{\Omega^*}$ where for each $k \in \mathbb{N}_0$ we set $\bar{\mathcal{A}}_{\Omega^*}[k] = \mathcal{A}_{\Omega^*}[k - \ell]$ for the smallest integer $\ell \geq 0$ satisfying $k - \ell \in \Omega^*$. Hence $\mathcal{A}_{\Omega^*}$ is a subspecies of $\bar{\mathcal{A}}_{\Omega^*}$, and $\tilde{\bar{\mathcal{A}}}_{\Omega^*}(z)$ has the same radius of convergence as $\tilde{\mathcal{A}}_{\Omega^*}(z)$.

We may apply Lemma 8.1 to the composition $\mathsf{SET} \circ \bar{\mathcal{A}}_{\Omega^*}$, yielding that the expression obtained from (9.18) by letting k range over $\mathbb{N}$ and replacing $\tilde{\mathcal{A}}_{\Omega^*}(\cdot)$ with $\tilde{\bar{\mathcal{A}}}_{\Omega^*}(\cdot)$ belongs to the class $o(n^{-3/2})$ of sequences that still tend to zero when multiplied by $n^{3/2}$. But this expression is clearly an upper bound to the expression in (9.18), yielding that (9.18) also belongs to $o(n^{-3/2})$. Hence the largest component in a random $(n-1)$-sized unlabelled $\mathsf{SET}_\Omega \circ \mathcal{A}_\Omega$ object has size $n + O_p(1)$. This verifies Claim 2).

It remains to prove Claim 3), i.e. we have to show that $\mathbb{E}[f_n] = O(1)$. If $\Omega \subset \mathbb{N}$ is bounded, then this is trivial. Otherwise it seems to require some work. By Equation (9.16) it follows that

$$\mathbb{E}[f_n] = \frac{[z^{n-1}] \left(s_1 \frac{\partial Z_{\mathsf{SET}_\Omega}}{\partial s_1} \right) (\tilde{\mathcal{A}}_{\Omega^*}(z), \tilde{\mathcal{A}}_{\Omega^*}(z^2), \ldots)}{[z^{n-1}] Z_{\mathsf{SET}_\Omega}(\tilde{\mathcal{A}}_{\Omega^*}(z), \tilde{\mathcal{A}}_{\Omega^*}(z^2), \ldots)}.$$

Since $1 \in \Omega$ the denominator is bounded from below by $[z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(z)$. By Proposition 9.1 it follows that

$$([z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(z))^{-1} = O(n^{3/2}\rho^n).$$

The power series in z in the numerator is bounded coefficient wise by

$$\left(s_1 \frac{\partial Z_{\mathsf{SET}}}{\partial s_1} \right) (\tilde{\mathcal{A}}_{\Omega^*}(z), \tilde{\mathcal{A}}_{\Omega^*}(z^2), \ldots) = \tilde{\mathcal{A}}_{\Omega^*}(z) \exp\left(\sum_{i=1}^{\infty} \tilde{\mathcal{A}}_{\Omega^*}(z^i)/i \right) = h(\tilde{\mathcal{A}}_{\Omega^*}(z))g(z)$$

with

$$h(w) = w \exp(w)$$

being analytic on $\mathbb{C}$ and

$$g(w) = \exp\left(\sum_{i \geq 2} \tilde{\mathcal{A}}_{\Omega^*}(z^i)/i \right)$$

having radius of convergence strictly larger than ρ since $\rho < 1$. By an identical argument as in the justification of Equation (9.4) we may apply the singularity analysis result [21, Thm. VI.5] to obtain

$$[z^{n-1}]h(\tilde{\mathcal{A}}_{\Omega^*}(z))g(z) = O(n^{-3/2}\rho^{-n}).$$

This concludes the proof. □

9.6.3 Symmetrically cycle pointed trees whose cycle center is a vertex

Recall that

$$\mathcal{V} = (\mathsf{SET}_\Omega^{\circledast} \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}.$$

Let $(\mathsf{V}_n, \tau_n, \sigma, v_n)$ be a rooted c-symmetry drawn uniformly at random from the set $\mathsf{RSym}(\mathcal{V})[n]$. In particular, V_n is distributed like the uniformly at random chosen unlabelled $\mathcal{V}$-object with size n. Let π_n denote the corresponding partition. By the discussion in Section 5.4, σ induces an automorphism

$$\bar{\sigma} : \pi_n \to \pi_n$$

of the SET_Ω-object. Moreover, let $F_n \subset \pi_n$ denote the fixpoints of $\bar{\sigma}$, $f_n = |F_n|$ their number and for each fixpoint $Q \in F_n$ let (A_Q, σ_Q) denote the corresponding symmetry from $\mathsf{Sym}(\mathcal{A}_{\Omega^*})[Q]$. Let H_n denote the total size of the trees dangling from cycles with length at least 2. We are going to make the following observations.

Lemma 9.7. *The following statements hold.*

1) *There are constants $C_1 > 0$ and $0 < \gamma < 1$ such that for all n and $x \ge 0$ we have that*
$$\mathbb{P}(H_n \ge x) \le C_1 n^{3/2}\gamma^x$$
and
$$\mathbb{P}(f_n \ge x) \le C_1 n^{3/2}\gamma^x.$$

2) *The maximum size of the trees corresponding to the fixpoints of $\bar{\sigma}$ satisfies*
$$\max_{Q \in F_n} |\mathsf{A}_Q| = n + O_p(1).$$

3) *There is a constant $C_2 > 0$ such that*
$$\mathbb{E}[f_n] \le C_2$$
for all n.

From these claims we may deduce Lemma 9.5 in an analogous manner as we deduced Lemma 9.4 from Lemma 9.6:

Proof of Lemma 9.5. As for Claim a), it suffices to show such a bound for all n and $\sqrt{n} \le x \le n$. Clearly it holds that
$$\mathbb{P}(\mathsf{D}(\mathsf{V}_n) \ge x) \le \mathbb{P}(H_n \ge x/2) + \mathbb{P}\left(\max_{Q \in F_n} \mathsf{H}(\mathsf{A}_Q) \ge x/2 - 1\right). \tag{9.19}$$
By Claim 1) of Lemma 9.7, we know that there are constants $C', C, c > 0$ and $0 < \gamma < 1$ with
$$\mathbb{P}(H_n \ge x/2) \le C' n^{3/2}\gamma^{x/2} \le C\exp(-cx^2/n). \tag{9.20}$$
As for the second summand in Equation (9.19), we may argue analogously as for Equation (9.7) that there constants $C^*, c^* > 0$ with
$$\mathbb{P}\left(\max_{Q} \mathsf{H}(\mathsf{A}_Q) \ge x/2 - 1\right) \le \mathbb{E}[f_n]\, C^* \exp(-c^* x^2/n).$$
Claim a) now follows from Equations (9.19), (9.20) and Claim 3) of Lemma 9.7.

It remains to verify Claim b). Let X_n denote the tree to a canonically chosen partition class of F_n with maximal size. Lemma 5.1 implies that for all ℓ
$$(\mathsf{X}_n \mid |\mathsf{X}_n| = \ell) \overset{(d)}{=} \mathsf{A}_\ell. \tag{9.21}$$
Hence $\mathsf{X}_n \overset{(d)}{=} \mathsf{A}_{K_n}$ for $K_n := |X_n|$. By Claim 2) of Lemma 9.7 we know that $|K_n| = n + O_p(1)$. This completes the proof of Claim b). □

It remains to verify Lemma 9.7.

Proof of Lemma 9.7. We start with Claim 1). Using the Boltzmann-sampling methods from Section 7.2.2, we obtain that the probability generating function of H_n is given by
$$\mathbb{E}\left[w^{H_n}\right] = \frac{[z^{n-1}]\bar{Z}_{\mathsf{SET}_\Omega^\circledast}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}^\circ_{\Omega^*}(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}((\rho w z)^2); \ldots)}{[z^{n-1}]\bar{Z}_{\mathsf{SET}_\Omega^\circledast}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}^\circ_{\Omega^*}(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}((\rho z)^2); \ldots)}. \tag{9.22}$$
A detailed justification of this fact goes as follows. By the product rule in Section 7.2.2 it suffices to consider $(n-1)$-sized rooted symmetries of $\mathsf{SET}_\Omega^\circledast \circ \mathcal{A}_{\Omega^*}$. The composition rule states that to sample such

a symmetry according according to the Boltzmann model with parameters $(\rho^i, \rho^i)_{i\ge 1}$, we may start with a Pólya–Boltzmann distributed rooted symmetry of $\mathsf{SET}_\Omega^{\circledcirc}$ with parameters $(\tilde{\mathcal{A}}_{\Omega^*}(\rho^i), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho^i))_{i\ge 1}$. Then, for each $j \ge 1$ and each unmarked j-cycle a symmetry of $\mathcal{A}_{\Omega^*}$ is sampled according to a Pólya–Boltzmann distribution with parameters $(\rho^{ij})_i$, and for the marked cycle we let s denote its length and draw a rooted symmetry of $\mathcal{A}_{\Omega^*}$ according to a Pólya–Boltzmann distribution with parameters $(\rho^{si}, \rho^{si})_{i\ge 1}$. Given a $k \in \Omega$ sized permutation ν with a marked cycle having length $\ell \ge 2$ and a distinguished atom of this cycle, the probability for the rooted symmetry of $\mathsf{SET}_\Omega^{\circledcirc}$ to assume this value is given by

$$\frac{\tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho^\ell)\tilde{\mathcal{A}}_{\Omega^*}(\rho^\ell)^{\nu_\ell - 1}}{k!\bar{Z}_{\mathsf{SET}_\Omega^{\circledcirc}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho); \tilde{\mathcal{A}}_{\Omega^*}(\rho^2), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho^2); \ldots)} \prod_{\substack{1\le i\le k \\ i\ne\ell}} \tilde{\mathcal{A}}_{\Omega^*}(\rho^i)^{\nu_i}. \tag{9.23}$$

Conditioned on this event, the probability generating function for the size of the resulting object is given by

$$\frac{\tilde{\mathcal{A}}_{\Omega^*}^\circ((\rho z)^\ell)}{\tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho^\ell)} \prod_{\substack{1\le i\le k \\ i\ne\ell}} \left(\frac{\tilde{\mathcal{A}}_{\Omega^*}((\rho z)^i)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho^i)}\right)^{\nu_i - \mathbb{1}_{i=\ell}}. \tag{9.24}$$

The exponents $(\rho z)^i$ are due to the fact that for each object corresponding to an i-cycle we attach i identical copies, and likewise for the marked cycle. In order to keep track of the volume of the trees corresponding to cycles with length at least 2 we may form the bivariate probability generating function where the variable w corresponds to this parameter and z to the total size, given by

$$\frac{\tilde{\mathcal{A}}_{\Omega^*}^\circ((\rho w z)^\ell)}{\tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho^\ell)} \left(\frac{\tilde{\mathcal{A}}_{\Omega^*}(\rho z)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho)}\right)^{\nu_1} \prod_{\substack{2\le i\le k \\ i\ne\ell}} \left(\frac{\tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^i)}{\tilde{\mathcal{A}}_{\Omega^*}(\rho^i)}\right)^{\nu_i - \mathbb{1}_{i=\ell}}. \tag{9.25}$$

Multiplying (9.23) with (9.25) and summing over all outcomes with total size $n-1$ yields

$$\frac{[z^{n-1}]\bar{Z}_{\mathsf{SET}_\Omega^{\circledast}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^2), \tilde{\mathcal{A}}_{\Omega^*}^\circ((\rho w z)^2); \ldots)}{\bar{Z}_{\mathsf{SET}_\Omega^{\circledcirc}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho); \tilde{\mathcal{A}}_{\Omega^*}(\rho^2), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho^2); \ldots)}. \tag{9.26}$$

Similarly, multiplying (9.23) with (9.24) and summing over all outcomes with total size $n-1$ yields

$$\frac{[z^{n-1}]\bar{Z}_{\mathsf{SET}_\Omega^{\circledast}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \tilde{\mathcal{A}}_{\Omega^*}^\circ((\rho z)^2); \ldots)}{\bar{Z}_{\mathsf{SET}_\Omega^{\circledcirc}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho); \tilde{\mathcal{A}}_{\Omega^*}(\rho^2), \tilde{\mathcal{A}}_{\Omega^*}^\circ(\rho^2); \ldots)}. \tag{9.27}$$

The quotient of (9.26) and (9.27) is the probability generating function for the parameter H_n, and the expression obtained in this way agrees with Equation (9.22).

Having verified Equation (9.22), we proceed with the argument. Since $1 \in \Omega$ and there is a number $k \ge 3$ with $k \in \Omega$ it follows that the denominator in (9.22) is bounded from below by

$$[z^{n-1}]z^{k-1}\tilde{\mathcal{A}}_{\Omega^*}(\rho z) = [z^{n-k}]\tilde{\mathcal{A}}_{\Omega^*}(\rho z).$$

Clearly it holds that

$$n - k \equiv 1 \mod \gcd(\Omega^*).$$

By Proposition 9.1 it follows that

$$[z^{n-k}]\tilde{\mathcal{A}}_{\Omega^*}(\rho z) \sim C n^{-3/2}$$

as $n \equiv 2 \mod \gcd(\Omega^*)$ tends to infinity. The polynomial in the numerator in (9.22) with indeterminate w is bounded coefficient wise by the series

$$\bar{Z}_{\mathsf{SET}_\Omega^\circledast}(\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}^\circ_{\Omega^*}(\rho); \tilde{\mathcal{A}}_{\Omega^*}((\rho w)^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}((\rho w)^2); \ldots)$$

which does not depend on n and, by Proposition 9.2, has radius of convergence strictly larger than 1. It follows that there is a constant C' such that

$$\mathbb{P}\left(H_n \geq x\right) \leq C' n^{3/2} \gamma^x$$

for all n and x. By a similar argument as for Equation (9.22) the probability generating function for the random number number f_n is given by

$$\mathbb{E}\left[w^{f_n}\right] = \frac{[z^{n-1}]\bar{Z}_{\mathsf{SET}_\Omega^\circledast}(w\tilde{\mathcal{A}}_{\Omega^*}(\rho z), w\tilde{\mathcal{A}}^\circ_{\Omega^*}(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}((\rho z)^2); \ldots)}{[z^{n-1}]\bar{Z}_{\mathsf{SET}_\Omega^\circledast}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}^\circ_{\Omega^*}(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}((\rho z)^2); \ldots)}. \tag{9.28}$$

The corresponding bound for the event $f_n \geq x$ follows by the same arguments as for H_n. This proves Claim 1).

In order to verify Claim 2), it suffices to show that for any sequence $t_n \to \infty$ the probability for all components in the random $\mathsf{SET}_\Omega^\circledast \circledcirc \mathcal{A}_{\Omega^*}$-object to have size at most $n - t_n$ tends to zero. It follows then by Claim 1) that the largest component corresponds with high probability to a class of F_n, hence yielding Claim 2).

By analogous arguments as in the justification of Equation (9.22), the probability that all components have size at most $n - t_n$ may be expressed by the product of the normalizing factor

$$([z^{n-1}]Z_{\mathsf{SET}_\Omega^\circledast}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}^\circ_{\Omega^*}(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}((\rho z)^2); \ldots))^{-1} \tag{9.29}$$

with the expression

$$\sum_{k \in \Omega} \frac{1}{k!} \sum_{\nu, \tau} \sum_{t, (a_{ij})_{i,j}} \left[z^{n-1} y^t \prod_{i,j} x_{ij}^{a_{ij}}\right] \tilde{\mathcal{A}}^\circ_{\Omega^*}(y(\rho z)^{|\tau|}) \prod_{i,j} \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^i x_{ij}). \tag{9.30}$$

Here the sum index ν ranges over all permutations of the set $[k]$, and τ ranges over all cycles of ν with length at least 2. The indices t and $(a_{ij})_{i,j}$ range over all pairs of an integer $t \geq 1$ with a family of numbers a_{ij} with $1 \leq i \leq n-1$, $1 \leq j \leq \nu_i$, such that $t \leq n - t_n$ and $a_{ij} \leq n - t_n$ for all i, j and

$$|\tau| t + \sum_{\substack{1 \leq i \leq n-1 \\ 1 \leq j \leq \nu_i}} i a_{ij} = n - 1.$$

The indices i, j of the product in (9.30) range over all pairs of integers with $1 \leq i \leq n-1$ and $1 \leq j \leq \nu_i$.

Applying a standard result for the singularity analysis of functions with multiple dominant singularities [21, Thm. VI.5] we obtain similarly as in the justification of Equation (9.4) that the factor in Equation (9.29) is asymptotically equivalent to $n^{3/2}$ times a constant. Thus, showing that the largest component in a random unlabelled $n-1$-sized $\mathsf{SET}_\Omega^\circledast \circledcirc \mathcal{A}_{\Omega^*}$-object has with high probability size at least $n - t_n$ is actually equivalent to showing that the expression in (9.30) multiplied by $n^{3/2}$ tends to zero as $n \equiv 2 \mod \gcd(\Omega^*)$ becomes large.

We consider the species $\bar{\mathcal{A}}_{\Omega^*}$ where for each $k \in \mathbb{N}_0$ we set $\bar{\mathcal{A}}_{\Omega^*}[k] = \mathcal{A}_{\Omega^*}[k - \ell]$ for the smallest integer $\ell \geq 0$ satisfying $k - \ell \in \Omega^*$. Hence $\mathcal{A}_{\Omega^*}$ is a subspecies of $\bar{\mathcal{A}}_{\Omega^*}$, and $\tilde{\bar{\mathcal{A}}}_{\Omega^*}(z)$ has the same radius of convergence as $\tilde{\mathcal{A}}_{\Omega^*}(z)$. The expression in (9.30) has an upper bound obtained by modifying (9.30) such that k ranges over all non-negative integers, and $\tilde{\mathcal{A}}_{\Omega^*}$ and $\tilde{\mathcal{A}}^\circ_{\Omega^*}$ get replaced by $\tilde{\bar{\mathcal{A}}}_{\Omega^*}$ and $\tilde{\bar{\mathcal{A}}}^\circ_{\Omega^*}$.

Showing that this upper bound belongs to $o(n^{-3/2})$ is equivalent to showing that the largest component in a uniform random n-sized unlabelled SET$^{\circledast} \circledcirc \bar{\mathcal{A}}_{\Omega^*}$-object has with high probability size at least $n - t_n$.

Combining Equations (6.4),(6.6), (6.2) and (5.1) yields that the ordinary generating function of the species SET$^{\circledast} \circledcirc \bar{\mathcal{A}}_{\Omega^*}$ is given by

$$\exp\left(\sum_{i\geq 1} \tilde{\bar{\mathcal{A}}}_{\Omega^*}(z^i)/i\right) \sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(z^j). \tag{9.31}$$

The intuitive explanation of this formula is that an unlabelled SET$^{\circledast} \circledcirc \bar{\mathcal{A}}_{\Omega^*}$-object is a multiset of unlabelled $\bar{\mathcal{A}}_{\Omega^*}$-objects, accounting for the factor $\exp(\sum_{i\geq 1} \tilde{\bar{\mathcal{A}}}_{\Omega^*}(z^i)/i)$, together with a number $j \geq 2$ of identical copies of an unlabelled $\tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}$-object, accounting for the second factor $\sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(z^j)$.

By Lemma 8.1 we know that a large random multiset of $\bar{\mathcal{A}}_{\Omega^*}$-objects consists of a giant component with a stochastically bounded rest. So in order to show that a large random unlabelled SET$^{\circledast} \circledcirc \bar{\mathcal{A}}_{\Omega^*}$-object also consists of a giant component with a stochastically bounded rest, it suffices to show that the total size of the copies of the $\bar{\mathcal{A}}^{\circ}_{\Omega^*}$-object is stochastically bounded.

By Equation (9.31) and the substitution sampler rule in Section 7.2.2, we obtain that the probability for the total size of the $\bar{\mathcal{A}}^{\circ}_{\Omega^*}$-objects in a uniform random n-sized unlabelled SET$^{\circledast} \circledcirc \bar{\mathcal{A}}_{\Omega^*}$-object to equal a fixed integer k is given by

$$\frac{\left([z^{n-k}] \exp\left(\sum_{i\geq 1} \tilde{\bar{\mathcal{A}}}_{\Omega^*}(z^i)/i\right)\right)\left([z^k] \sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(z^j)\right)}{[z^n] \exp\left(\sum_{i\geq 1} \tilde{\bar{\mathcal{A}}}_{\Omega^*}(z^i)/i\right) \sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(z^j)} \tag{9.32}$$

Analogously as in the justification of Equation (9.4) we obtain that there is a constant $c > 0$ such that

$$[z^n] \exp\left(\sum_{i\geq 1} \tilde{\bar{\mathcal{A}}}_{\Omega^*}(z^i)/i\right) \sim c\rho^{-n} n^{-3/2}.$$

As the series $\sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(z^j)$ has radius of convergence strictly larger than ρ, it follows by an elementary real analytic method [21, Thm. VI.12] that

$$[z^n] \exp\left(\sum_{i\geq 1} \tilde{\bar{\mathcal{A}}}_{\Omega^*}(z^i)/i\right) \sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(z^j) \sim c\rho^{-n} n^{-3/2} \sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(\rho^j). \tag{9.33}$$

Thus, the expression in (9.32) is asymptotically equivalent to

$$\rho^k \left([z^k] \sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(z^j)\right) / \sum_{j\geq 2} \tilde{\bar{\mathcal{A}}}^{\circ}_{\Omega^*}(\rho^j). \tag{9.34}$$

As the limit probabilities in (9.34) sum up to 1 when k ranges over all non-negative integers, it follows that the total size of cycle-pointed part of the random unlabelled SET$^{\circledast} \circledcirc \bar{\mathcal{A}}_{\Omega^*}$-object is stochastically bounded. We have thus verified Claim 2).

It remains to prove Claim 3). By Equation (9.28) it follows that

$$\mathbb{E}\left[w^{f_n}\right] = \frac{[z^{n-1}]\left(s_1 \frac{\partial \bar{Z}_{\mathsf{SET}^{\circledast}_{\Omega}}}{\partial s_1}\right)(\tilde{\mathcal{A}}_{\Omega^*}(z), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}(z); \tilde{\mathcal{A}}_{\Omega^*}(z^2), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}(z^2); \ldots)}{[z^{n-1}]\bar{Z}_{\mathsf{SET}^{\circledast}_{\Omega}}(\tilde{\mathcal{A}}_{\Omega^*}(z), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}(z); \tilde{\mathcal{A}}_{\Omega^*}(z^2), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}(z^2); \ldots)}. \tag{9.35}$$

As $1 \in \Omega$ the denominator is bounded from below by $[z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(z)$. By Proposition 9.1 it follows that

$$([z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(z))^{-1} = O(n^{3/2}\rho^n). \tag{9.36}$$

The power series in z in the numerator is bounded coefficient wise by

$$\left(s_1 \frac{\partial \bar{Z}_{\mathsf{SET}^\circledast}}{\partial s_1}\right)(\tilde{\mathcal{A}}_{\Omega^*}(z), \tilde{\mathcal{A}}^\circ_{\Omega^*}(z); \tilde{\mathcal{A}}_{\Omega^*}(z^2), \tilde{\mathcal{A}}^\circ_{\Omega^*}(z^2); \ldots) = \tilde{\mathcal{A}}_{\Omega^*}(z) \exp\left(\sum_{i\geq 1} \tilde{\mathcal{A}}_{\Omega^*}(z^i)/i\right) \sum_{j \geq 2} \tilde{\mathcal{A}}^\circ_{\Omega^*}(z^j).$$

By an elementary real analytic method [21, Thm. VI.12] and Equation (9.33) it follows that this sequence belongs to $O(\rho^{-n}n^{-3/2})$. Using Equation (9.36) this implies that the expression in (9.35) is bounded. This concludes the proof of Claim 3). $\square$

10 References

[1] C. Abraham and J.-F. Le Gall. Excursion theory for Brownian motion indexed by the Brownian tree. *ArXiv e-prints*, Sept. 2015.

[2] M. Albenque and C. Goldschmidt. The Brownian continuum random tree as the unique solution to a fixed point equation. *Electron. Commun. Probab.*, 20:no. 61, 14, 2015.

[3] M. Albenque and J.-F. Marckert. Some families of increasing planar maps. *Electron. J. Probab.*, 13:no. 56, 1624–1671, 2008.

[4] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

[5] O. Angel and O. Schramm. Uniform infinite planar triangulations. *Comm. Math. Phys.*, 241(2-3):191–213, 2003.

[6] A. D. Barbour and B. L. Granovsky. Random combinatorial structures: the convergent case. *J. Combin. Theory Ser. A*, 109(2):203–220, 2005.

[7] J. P. Bell, S. N. Burris, and K. A. Yeats. Counting rooted trees: the universal law $t(n) \sim C\rho^{-n}n^{-3/2}$. *Electron. J. Combin.*, 13(1):Research Paper 63, 64 pp. (electronic), 2006.

[8] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp. (electronic), 2001.

[9] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial species and tree-like structures*, volume 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.

[10] J. Bettinelli. Scaling limit of random planar quadrangulations with a boundary. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):432–477, 2015.

[11] M. Bodirsky, É. Fusy, M. Kang, and S. Vigerske. Boltzmann samplers, Pólya theory, and cycle pointing. *SIAM J. Comput.*, 40(3):721–769, 2011.

[12] N. Broutin and P. Flajolet. The distribution of height and diameter in random non-plane binary trees. *Random Structures Algorithms*, 41(2):215–252, 2012.

[13] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[14] A. Caraceni. The scaling limit of random outerplanar maps. *To appear in Annales de l'Institut Henri Poincaré Probabilités et Statistiques*.

[15] N. Curien, B. Haas, and I. Kortchemski. The CRT is the scaling limit of random dissections. *Random Structures Algorithms*, 47(2):304–327, 2015.

[16] M. Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.

[17] M. Drmota and B. Gittenberger. The shape of unlabeled rooted random trees. *European J. Combin.*, 31(8):2028–2063, 2010.

[18] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Random sampling from Boltzmann principles. In *Automata, languages and programming*, volume 2380 of *Lecture Notes in Comput. Sci.*, pages 501–513. Springer, Berlin, 2002.

[19] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combin. Probab. Comput.*, 13(4-5):577–625, 2004.

[20] P. Flajolet, É. Fusy, and C. Pivoteau. Boltzmann sampling of unlabelled structures. In *Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth Workshop on Analytic Algorithmics and Combinatorics*, pages 201–211. SIAM, Philadelphia, PA, 2007.

[21] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[22] A. Georgakopoulos and S. Wagner. Limits of subcritical random graphs and random graphs with excluded minors. *ArXiv e-prints*, Dec. 2015.

[23] O. Gurel-Gurevich and A. Nachmias. Recurrence of planar graph limits. *Ann. of Math. (2)*, 177(2):761–781, 2013.

[24] B. Haas and G. Miermont. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.*, 40(6):2589–2666, 2012.

[25] S. Janson and S. Ö. Stefánsson. Scaling limits of random planar maps with a unique large face. *Ann. Probab.*, 43(3):1045–1081, 2015.

[26] A. Joyal. Une théorie combinatoire des séries formelles. *Adv. in Math.*, 42(1):1–82, 1981.

[27] J.-F. Le Gall and G. Miermont. Scaling limits of random trees and planar maps. In *Probability and statistical physics in two and more dimensions*, volume 15 of *Clay Math. Proc.*, pages 155–211. Amer. Math. Soc., Providence, RI, 2012.

[28] J.-F. Marckert and G. Miermont. The CRT is the scaling limit of unordered binary trees. *Random Structures Algorithms*, 38(4):467–501, 2011.

[29] R. Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.

[30] K. Panagiotou and B. Stufler. Scaling limits of random Pólya trees. *ArXiv e-prints*, Feb. 2015.

[31] K. Panagiotou, B. Stufler, and K. Weller. Scaling limits of random graphs from subcritical classes. *Ann. Probab.*, 44(5):3291–3334, 2016.

[32] B. Stufler. Random enriched trees with applications to random graphs. *ArXiv e-prints*.

[33] G. Szekeres. Distribution of labelled trees by diameter. In *Combinatorial mathematics, X (Adelaide, 1982)*, volume 1036 of *Lecture Notes in Math.*, pages 392–397. Springer, Berlin, 1983.

[34] M. Wang. Height and diameter of Brownian tree. *Electron. Commun. Probab.*, 20:no. 88, 15, 2015.