Scaling limits of random trees and graphs

Benedikt Stufler

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität München

München, 30. Juni 2015

Scaling limits of random trees and graphs

Benedikt Stufler

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität München

vorgelegt von
Benedikt Stufler
aus München
am 30. Juni 2015

Erstgutachter:	Prof. Dr. Konstantinos Panagiotou
Zweitgutachter:	Prof. Dr. Jean-François Le Gall
Drittgutachter:	Prof. Dr. Michael Drmota
Tag der Disputation:	23. Oktober 2015

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Name, Vorname

Ort, Datum

Unterschrift Doktorand/in

Formular 3.2

In this thesis, we establish the scaling limit of several models of random trees and graphs, enlarging and completing the now long list of random structures that admit David Aldous' continuum random tree (CRT) as scaling limit. Our results answer important open questions, in particular the conjecture by Aldous [Ald91b, p. 55] for the scaling limit of random unlabelled unrooted trees. We also show that random graphs from subcritical graph classes admit the CRT as scaling limit, proving (in a strong from) a conjecture by Marc Noy and Michael Drmota [DN13, remark after Thm. 3.2], who conjectured a scaling limit for the diameter of these graphs. Furthermore, we provide a new proof for results by Bénédicte Haas and Grégory Miermont [HM12, Thm. 9] regarding the scaling limits of random Pólya trees, extending their result to random Pólya trees with arbitrary vertex-degree restrictions.

In dieser Arbeit ermitteln wir die Skalierungslimes mehrerer Modelle zufälliger Bäume und Graphen. Hierbei erweitern und vervollständigen wir die nun lange Liste zufälliger Strukturen, deren Skalierungslimes der Continuum Random Tree (CRT) von David Aldous ist. Unsere Resultate beantworten wichtige offene Fragen, insbesondere die Vermutung von David Aldous [Ald91b, p. 55] bezüglich des Skalierungslimes zufälliger Isomorphieklassen entwurzelter Bäume. Desweiteren beweisen wir, dass der CRT als Skalierungslimes zufälliger Graphen von subkritischen Klassen auftritt. Dies beweist (in einer starken Form) eine Vermutung von Marc Noy und Michael Drmota [DN13, Bemerkung nach Thm. 3.2], die einen Skalenlimes für den Durchmesser dieser Graphen vermuteten. Desweiteren geben wir einen neuen Beweis für Resultate von Bénédicte Haas und Grégory Miermont [HM12, Thm. 9] bezüglich des Skalierungslimes zufälliger Pólya Bäume. Hierbei erweitern wir dieses Resultat auf Pólya Bäume mit beliebigen Knotengrad Restriktionen.

I would like to thank Prof. Konstantinos Panagiotou, who supervised this thesis. He introduced me to the world of analytic combinatorics and random graphs, and his mathematical intuition impressed me time and again. He also enabled me to participate in many inspiring workshops and conferences, such as the fall school in Graz and several random geometry workshops in Cambridge.

I also thank all current and former members of the stochastic group in Munich for sharing so many enjoyable and motivating moments. I am particularly grateful to Prof. Vitali Wachtel, Prof. Markus Heydenreich and Prof. Merkl for interesting and helpful discussions, as I am to Jetlir Duraj, Felizitas Weidner, Eckhard Schlemm, Matthias Lienert, Jan Nagel, Johannes Kugler, Robert Graf, Simon Aumann, Steffen Seeliger, Andreas Groh, Martin Oelker, and Johannes Nissen-Meyer.

I would also like to thank Igor Kortchemski and Nicolas Curien for helpful suggestions and feedback on various topics.

I am deeply grateful to my parents Claudia and Vitus for their support and love. Finally, I deeply thank my beloved wife Milena for having faith in me, not only during the two years it took to write this thesis.

Table of Contents

Chapter 1

Introduction and main results

1.1 Preliminaries

The continuum random tree (CRT) was constructed by David Aldous [Ald91a, Ald91b, Ald93] and shown to be the scaling limit of critical Galton-Watson trees conditioned to be large, if the offspring distribution has finite (nonzero) variance. Since Aldous's pioneering work, the CRT has been identified as the limiting object of many different classes of discrete structures, in particular trees, see e.g. Marckert and Miermont [MM11], Haas and Miermont [HM12] and references therein, and planar maps, see e.g. Albenque and Marckert [AM08], Bettinelli [Bet15], Caraceni [Car], Curien, Haas and Kortchemski [CHK14] and Janson and Stefansson [JS15].

The preliminaries Chapter 2 intends to make our results accessible to a broad audience by recalling relevant notions and known results. More precisely, Section 2.1 gives a brief introduction to Aldous' scaling limit of conditioned Galton-Watson trees, recalling the notion of Gromov-Hausdorff convergence and the construction of the continuum random tree from Brownian excursion. Here we follow Le Gall and Miermont [LGM12], and the books by Burago, Burago and Ivanov [BBI01] and Diestel [Die10]. In Section 2.2 we give a concise introduction to the theory of combinatorial species, an algebraic framework for the systematic enumeration and decomposition of combinatorial objects. This section follows the original work by Joyal [Joy81] and the book [BLL98] by Bergeron, Labelle and Leroux. Section 2.3 discusses the cycle pointing operator, which is a valuable tool in the study of combinatorial structures up to symmetry. Here we follow the work by Bodirsky, Fusy, Kang and Vigerske [BFKV11]. In Section 2.4 we briefly set up the framework of Boltzmann samplers. It allows us to "mechanically" translate decompositions of combinatorial objects in the language of combinatorial species to random algorithms, that produce random objects following certain Boltzmann distributions. In the subsequent chapters we are going to make heavy use of this bridge from combinatorial species to random algorithms, in order study random trees and graphs. We emphasize the importance of Pólya-Boltzmann samplers introduced in [BFKV11], which generalize previous work by Duchon, Flajolet, Louchard and Schaeffer in [DFLS02] and [DFLS04], and the work [FFP07] by Flajolet, Fusy and Pivoteau. In Section 2.5 we close the preliminaries chapter by recalling a frequently used deviation inequality, found in almost any textbook on the subject. We advise that Subsection 2.2.4 and Section 2.3 are extended versions of some parts of the preliminaries section of the author's work [Stu14], and that Subsections 2.1.1, 2.1.2, 2.2.1, 2.2.3 and 2.2.5 follow closely certain parts of the preliminary section the author wrote for the work [PSW14] by Konstantinos Panagiotou, Kerstin Weller and the author.

1.2 Unlabelled (unrooted) trees

One of the main contributions of this thesis concerns random trees that are unordered and unlabelled. Here one distinguishes between Pólya trees, which have a root, and unlabelled (unrooted) trees, see Figure 1.1. It has been a long-standing conjecture

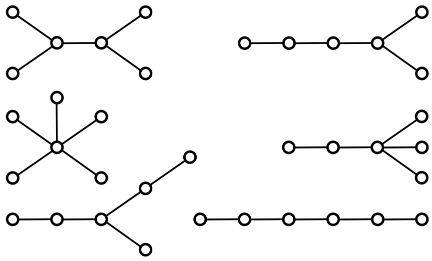

Figure 1.1: All unlabelled (unrooted) trees with 6 vertices.

by Aldous [Ald91b, p. 55] that the models "all Pólya trees with n-vertices equally likely" and "all unlabelled trees with n-vertices equally likely" admit the CRT as scaling limit. The convergence of binary Pólya trees, i.e. where the vertex outdegrees are restricted to the set $\{0, 2\}$, was shown by Marckert and Miermont [MM11] using an appropriate trimming procedure on trees. Later, Haas and Miermont [HM12] proved the conjecture for Pólya trees by establishing a general result on the scaling limits of random trees satisfying a certain Markov branching property and using these trees to approximate random Pólya trees. In this way, they established the convergence for Pólya trees without degree restrictions or with vertex outdegrees in a set of the form $\{0, 1, \dots d\}$ or $\{0, d\}$ for $d \geq 2$, remarking that the conjecture regarding unlabelled unrooted trees was still open. Chapter 3 settles this conjecture in the affirmative. It is an extended version of the the author's work [Stu14]. Our main result reads as follows.

Theorem 1.2.1. *Let T_n denote the uniform random unlabelled unrooted tree with n vertices. There is a constant $a > 0$ such that*

$$(\mathsf{T}_n, an^{-1/2}d_{\mathsf{T}_n}) \xrightarrow{(d)} (\mathcal{T}_\mathsf{e}, d_{\mathcal{T}_\mathsf{e}})$$

with respect to the Gromov-Hausdorff metric. Here we use the normalization by Le Gall and let $\mathcal{T}_\mathsf{e}$ denote the continuum random tree constructed from Brownian excursion $(\mathsf{e}(t))_{0 \leq t \leq 1}$.

The scaling constant a is precisely the same as for the case of Pólya-trees, i.e. it is given by $a = \sqrt{\pi/2}\kappa_\infty$ with κ_∞ denoting the constant such that the number of Pólya trees with n vertices is asymptotically given by $\kappa_\infty n^{-3/2}\rho^{-n}$ for some $\rho > 0$ [HM12]. The techniques of our proofs are based on the cycle-pointing decomposition developed by Bodirsky, Fusy, Kang and Vigerske [BFKV11]. A direct consequence is that the diameter $\mathrm{D}(\mathsf{T}_n)$ admits the scaling limit

$$\mathbb{P}(\mathrm{D}(\mathsf{T}_n) > a^{-1}xn^{1/2}) \to \mathbb{P}(\mathrm{D}(\mathcal{T}_\mathsf{e}) > x).$$

The distribution of the diameter is known and given by

$$\mathrm{D}(\mathcal{T}_\mathsf{e}) \overset{(d)}{=} \sup_{0 \leq t_1 \leq t_2 \leq 1} (\mathsf{e}(t_1) + \mathsf{e}(t_2) - 2 \inf_{t_1 \leq t \leq t_2} \mathsf{e}(t)) \tag{1.2.1}$$

and

$$\mathbb{P}(\mathrm{D}(\mathcal{T}_{\mathsf{e}}) > x) = \sum_{k=1}^{\infty}(k^2-1)\Big(\frac{2}{3}k^4x^4 - 4k^2x^2 + 2\Big)\exp(-k^2x^2/2). \tag{1.2.2}$$

Equations (1.2.1) and (1.2.2) the first moment $\mathbb{E}[\mathrm{D}(\mathcal{T}_{\mathsf{e}})] = 4/3\sqrt{\pi/2}$ have been known since the construction of the CRT by Aldous [Ald91b, Ch. 3.4], who used the convergence of random labelled trees to the CRT together with results by Szekeres [Sze83] regarding the diameter of these trees. Expression (1.2.2) was recently recovered directly in the continuous setting by Wang [Wan15]. We also provide exponential tailbounds for the diameter of the tree T_n:

Lemma 1.2.2. *Let T_n denote the uniform random unlabelled unrooted tree with n vertices. Then there are constants $C, c > 0$ such that for all n and $x \geq 0$ we have the following tail bound for the diameter:*

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

Given the limit distribution in (1.2.2) we may check that our tail-bound is essentially optimal. Lemma 1.2.2 implies that the rescaled diameter $a\mathrm{D}(\mathsf{T}_n)/\sqrt{n}$ is p-uniformly integrable for any $p \geq 1$. Hence it convergences towards the diameter $\mathrm{D}(\mathcal{T}_{\mathsf{e}})$ of the CRT not only in distribution, but also in arbitrarily high moments. Since $\mathbb{E}[\mathrm{D}(\mathcal{T}_{\mathsf{e}})] = 4/3\sqrt{\pi/2}$ it follows in particular that

$$\mathbb{E}[\mathrm{D}(\mathsf{T}_n)] \sim \frac{4}{3\kappa_\infty}n^{1/2}$$

asymptotically as n tends to infinity. If we consider trees with constraints on the vertex degrees we also have to deal with restrictions on the size of the tree:

Proposition 1.2.3. *Let Ω be a set of positive integers such that $1 \in \Omega$ and there is a $k \geq 3$ such that $k \in \Omega$. We let d denote the greatest common divisor of the nonzero elements of the shifted set $\Omega^* = \Omega - 1$. Then the following holds*

i) If there is a tree with n vertices and vertex degrees in Ω, then $n \equiv 2 \mod d$. Conversely, if $n \equiv 2 \mod d$ is large enough, then there always exists such a tree with n vertices.

ii) If there is a rooted tree with m vertices and vertex outdegrees in Ω^*, then $m \equiv 1 \mod d$. Conversely, if $m \equiv 1 \mod d$ is large enough, then there always exists such a tree with m vertices.

The proof of this well-known fact is by Schur's lemma, see for example Wilf [Wil06, Thm. 3.15.2]. For each subset $\Omega^* \subset \mathbb{N}_0$ containing 0 and at least one integer equal or larger than 2 there exists a constant $c_{\Omega*}$ such that the uniformly drawn random unlabelled rooted tree A_{n-1} with $n-1$ vertices and vertex outdegrees in Ω^* satisfies

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

as $n \equiv 2 \mod \gcd(\Omega^*)$ tends to infinity. For $\Omega^* = \{0, 2\}$ this was established by Marckert and Miermont [MM11]. Haas and Miermont [HM12] treated the cases $\Omega^* = \mathbb{N}_0$, $\Omega^* = \{0, b\}$ and $\Omega^* = \{0, 1, \dots, b\}$ for $b \geq 2$. The remaining cases are treated in Theorem 1.3.1 below. We provide the following extension of our main result:

Theorem 1.2.4. *Let Ω be a set of positive integers containing 1 and at least one integer equal or larger than 3, and set $\Omega^* = \Omega - 1$. Given an integer n with $n \equiv 2 \mod d$ we may consider the uniform random unlabelled unrooted tree T_n with n vertices and vertex-degrees in Ω. Then*

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

in the Gromov-Hausdorff sense as $n \equiv 2 \mod \gcd(\Omega^*)$ becomes large.

Let us fix the notation of Theorem 1.2.4, i.e. let Ω be a set of positive integers satisfying $1 \in \Omega$ and $k \in \Omega$ for at least one $k \geq 3$, and set $\Omega^* = \Omega - 1$. In order to ensure convergence of higher moments of extremal parameters, we show the following tail bound for the diameter.

Lemma 1.2.5. *Let T_n denote the uniform random unlabelled unrooted tree with n vertices and vertex-degrees in Ω. Then there are constants $C, c > 0$ such that for all $x \geq 0$ and n with $n \equiv 2 \mod \gcd(\Omega^*)$ we have that*

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

As an important ingredient in our proof we show a similar tail bound for the height of uniform random Pólya trees with arbitrary vertex-degree restrictions.

Lemma 1.2.6. *Let A_m denote the uniform random Pólya tree with m vertices and vertex out-degrees in the set Ω^*. Then there are constants $C, c > 0$ such that for all $x \geq 0$ and m with $m \equiv 1 \mod \gcd(\Omega^*)$ we have that*

$$\mathbb{P}(H(\mathsf{A}_m) \geq x) \leq C \exp(-cx^2/m).$$

The tail-bounds imply that the rescaled height $m^{-1/2}\mathrm{H}(\mathsf{A}_m)$ and diameters $n^{-1/2}\mathrm{D}(\mathsf{T}_n)$ and $m^{-1/2}\mathrm{D}(\mathsf{A}_m)$ of unlabelled trees are arbitrarily high uniformly integrable. Together with the convergence towards the CRT, this implies

$$\mathbb{E}[\mathrm{D}^p(\mathsf{T}_n)] \sim \mathbb{E}[\mathrm{D}^p(\mathcal{T}_{\mathsf{e}})] n^{p/2}/c_{\Omega^*}^p,$$
$$\mathbb{E}[\mathrm{D}^p(\mathsf{A}_{n-1})] \sim \mathbb{E}[\mathrm{D}^p(\mathcal{T}_{\mathsf{e}})] n^{p/2}/c_{\Omega^*}^p,$$
$$\mathbb{E}[\mathrm{H}^p(\mathsf{A}_{n-1})] \sim \mathbb{E}[\mathrm{H}^p(\mathcal{T}_{\mathsf{e}})] n^{p/2}/c_{\Omega^*}^p,$$

as $n \equiv 2 \mod d$ tends to infinity. Parts of this result have already been obtained using analytic methods: Broutin and Flajolet performed a precise study of the height of unlabelled rooted binary trees and diameter of unlabelled unrooted ternary trees (i.e. the case $\Omega^* = \{0,2\}$ and $\Omega = \{1,3\}$) in [BF08] and [BF12], showing among other results convergence of arbitrarily high moments with exact expressions for their limit. Drmota and Gittenberger [DG10, Thm. 2] obtained the limit behaviour of the height of unlabelled rooted trees with precise expressions for the limits of arbitrarily high moments.

The distribution of the height $\mathrm{H}(\mathcal{T}_{\mathsf{e}})$ is known and given by

$$\mathrm{H}(\mathcal{T}_{\mathsf{e}}) \stackrel{(d)}{=} \sup_{0\le t_1 \le t_2 \le 1} \mathsf{e}(t) \tag{1.2.3}$$

and

$$\mathbb{P}(\mathrm{H}(\mathcal{T}_{\mathsf{e}}) > x) = 2\sum_{k=1}^{\infty}(4k^2x^2 - 1)\exp(-2k^2x^2), \tag{1.2.4}$$

see [Ald91b, Ch. 3.1]. Its moments are also known are known and given by

$$\mathbb{E}[\mathrm{H}(\mathcal{T}_{\mathsf{e}})] = \sqrt{\pi/2}, \qquad \mathbb{E}[\mathrm{H}(\mathcal{T}_{\mathsf{e}})^k] = 2^{-k/2}k(k-1)\Gamma(k/2)\zeta(k) \quad \text{for } k \ge 2.$$

This holds by standard results for Brownian excursion by Chung [Chu76], and Biane, Pitman and Yor [BPY01] for a proof using Equation (1.2.3), or by results of Rényi and Szekeres [RS67, Eq. (4.5)]) who calculated the moments of the limit distribution of the height of a class of trees that converges towards the CRT (by [Ald91a]). The moments of the diameter are also known:

$$\mathbb{E}[\mathrm{D}(\mathcal{T}_{\mathsf{e}})] = \frac{4}{3}\sqrt{\pi/2}, \quad \mathbb{E}[\mathrm{D}(\mathcal{T}_{\mathsf{e}})^2] = \frac{2}{3}\left(1+\frac{\pi^2}{3}\right), \quad \mathbb{E}[\mathrm{D}(\mathcal{T}_{\mathsf{e}})^3] = 2\sqrt{2\pi}, \tag{1.2.5}$$

$$\mathbb{E}[\mathrm{D}(\mathcal{T}_{\mathsf{e}})^k] = \frac{2^{k/2}}{3}k(k-1)(k-3)\Gamma(k/2)(\zeta(k-2)-\zeta(k)) \quad \text{for } k \ge 4. \tag{1.2.6}$$

The expression $\mathbb{E}[\mathrm{D}(\mathcal{T}_{\mathsf{e}})] = \frac{4}{3}\sqrt{\pi/2}$ may be obtained as described in Aldous [Ald91b, Sec. 3.4] by results of Szekeres [Sze83], who proved the existence of a limit distribution for the diameter of rescaled random unordered labelled trees. The higher moments could be obtained in the same way by elaborated calculations, or, we can deduce them by building on results by Broutin and Flajolet, who studied in [BF12] the random tree T_n that is drawn uniformly at random among all unlabelled trees with n leaves in which each inner vertex is required to have degree 3. Using analytic methods [BF12, Thm. 8], they computed asymptotics of the form

$$\mathbb{E}[\mathrm{D}(\mathsf{T}_n)^r] \sim c_r \lambda^{-r} n^{r/2}$$

with λ an analytically given constant, and the constants c_r given by

$$c_1 = \frac{8}{3}\sqrt{\pi}, \quad c_2 = \frac{16}{3}(1+\frac{\pi^2}{3}), \quad c_3 = 64\sqrt{\pi},$$
$$c_r = \frac{4^r}{3}r(r-1)(r-3)\Gamma(r/2)(\zeta(r-2)-\zeta(r)) \quad \text{if } r \ge 4.$$

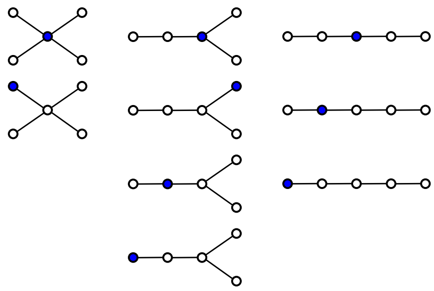

Figure 1.2: All Pólya trees with 5 vertices.

Since $\mathbb{E}[\mathrm{D}(\mathsf{T}_n)^r] \sim \mathbb{E}[\mathrm{D}(\mathcal{T}_\mathsf{e})^r] c_{\{0,2\}}^{-r} n^{r/2}$ it follows that $\mathbb{E}[\mathrm{D}(\mathcal{T}_\mathsf{e})^r] = c_r (c_{\{0,2\}}/\lambda)^r$. All that remains is to calculate the ratio $c_{\{0,2\}}/\lambda$, which is given by

$$c_{\{0,2\}}/\lambda = \mathbb{E}[\mathrm{D}(\mathcal{T}_\mathsf{e})]/c_1 = 1/(2\sqrt{2}),$$

since $\mathbb{E}[\mathrm{D}(\mathcal{T}_\mathsf{e})] = 4/3\sqrt{\pi/2}$. This yields Equations (1.2.5) and (1.2.6).

1.3 Pólya trees

Pólya trees are trees that are *rooted*, *unordered* and *unlabelled*, see Figure 1.2. They are named after George Pólya, who developed a framework based on generating functions in order to study their properties [Pól37]. The main difficulty of analysing these objects in a random setting is that they do not fit into well-studied models of random trees such as simply generated trees, a fact that was widely believed and which has been rigorously established by Drmota and Gittenberger [DG10, Thm. 1].

Marckert and Miermont [MM11] established the scaling limit of binary Pólya trees. Haas and Miermont [HM12] extended this result by using different methods, showing that the CRT is the scaling limit of random Pólya trees without degree restrictions or with vertex-outdegrees in a set of the form $\{0, d\}$ or $\{0, \dots, d\}$ for $d \geq 2$. However, the question about the convergence of Pólya trees with *arbitrary* degree restrictions has remained open since.

Chapter 4, which is an extended version of the work [PS15] by Konstantinos Panagiotou and the author of the present thesis, settles this question by presenting a proof of the fact that uniform random Pólya trees with arbitrary degree restrictions converge towards the CRT.

Theorem 1.3.1. *Let Ω^* be an arbitrary set of nonnegative integers containing zero and at least one integer greater than or equal to two. Let A_n denote the uniform random Pólya tree with n vertices and vertex outdegrees in Ω^*. Then there exists a constant $c_{\Omega^*} > 0$ such that*

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

with respect to the Gromov-Hausdorff topology, as $n \equiv 1 \mod \gcd(\Omega^*)$ *tends to infinity.*

Our proof is very short, almost elementary, and reveals a striking structural property. Roughly speaking, the idea is to draw random pairs (T_n, σ_n) of a *labelled* rooted tree T_n with n vertices and an automorphisms σ_n of T_n in such a way, that the isomorphism class corresponding to T_n is distributed like the uniform random Pólya tree. The fixpoints of σ then form a subtree $\mathcal{T}_n$ of T_n, which is distributed like a critical Galton-Watson tree conditioned on having a random size which concentrates around a constant multiple of n. In particular, the rescaled fixpoint tree $n^{-1/2}\mathcal{T}_n$ converges weakly towards a constant multiple of the CRT. Moreover, the non-fixpoints of σ_n form small subtrees (typically of order $O(\log n)$), that are attached to the fixpoints. Hence they do not contribute much to the geometric shape of T_n, yielding that the Gromov-Hausdorff distance between the rescaled trees $n^{-1/2}T_n$ and $n^{-1/2}\mathcal{T}_n$ converges in probability to zero, completing the proof.

1.4 Random graphs from subcritical classes

Chapter 5 (except for Section 5.7, which has not been published previously) is an extended version of the work [PSW14] by Konstantinos Panagiotou, Kerstin Weller and the author of the present thesis. Our motivation is that although the CRT was identified as the scaling limit in various settings, little is known about the limiting behaviour of random graphs from complex graph classes. In this chapter we study in a unified way the asymptotic distribution of distances in random graphs from so-called *subcritical classes*.

Informally speaking, a class $\mathcal{C}$ of labelled, connected (simple) graphs is called subcritical, if for a typical graph with n vertices the largest block (i.e. inclusion maximal 2-connected subgraph) has $O(\log n)$ vertices. See Section 5.1.4 for a formal definition. Prominent examples of classes that are subcritical are outerplanar and series-parallel graphs. Subcritical graph classes have been the object of intense research in the last years, especially because of their close connection to the class of planar graphs. See for example Drmota and Noy [DN13], Bernasconi, Panagiotou and Steger [BPS09], Drmota, Fusy, Kang, Kraus and Rué [DFK+11], and Panagiotou and Steger [PS10]. However, with the notable exception of [DN13], most research on such random graphs has focused on *additive* parameters, like the number of vertices of a given degree; the fine study of *global* properties, like the distribution of the distances, poses a significant challenge.

Let C_n denote a random graph drawn uniformly from the set of connected graphs with n vertices of an arbitrary but fixed subcritical class $\mathcal{C}$. In [DN13, Thm. 3.2], Michael Drmota and Marc Noy established the following bound for the diameter

$$c_1\sqrt{n} \le \mathbb{E}[\mathrm{D}(\mathsf{C}_n)] \le c_2\sqrt{n \log n},$$

and conjectured a universal limit law for $\mathrm{D}(\mathsf{C}_n)/\sqrt{n}$. We prove this conjecture in a strong sense by showing convergence towards the CRT $\mathcal{T}_\mathsf{e}$.

Graph Class $\mathcal{C}$	Numerical approximation of $c_{(\mathcal{C})}$
Trees = $\mathsf{Forb}(C_3)$	0.5
$\mathsf{Forb}(C_4)$	0.58778
$\mathsf{Forb}(C_5)$	0.66433
Cacti Graphs	0.62973
Outerplanar Graphs	0.96038

Table 1.1: The scaling constant of several examples of subcritical graph classes.

Theorem 1.4.1. *Let $\mathcal{C}$ be a subcritical class of connected graphs and let C_n denote the random graph drawn uniformly from the graphs in $\mathcal{C}$ with n vertices. Then there exists a constant $c_{(\mathcal{C})}$ such that*

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

with respect to the Gromov-Hausdorff metric, as n becomes large. Here we restrict ourselves to values of n for which the graphs with n vertices in the class $\mathcal{C}$ exist.

In order to ensure convergence of higher moments of the diameter $\mathrm{D}(\mathsf{C}_n)$ (or the height with respect to a uniformly at random chosen root), we also show exponential tail-bounds.

Theorem 1.4.2. *Let $\mathcal{C}$ be a block-stable class of connected graphs and let C_n denote the random graph drawn uniformly from graphs in $\mathcal{C}$ with n vertices. Then there are constants $C, c > 0$ such that for all n and $x \geq 0$*

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

We also give explicit analytic expressions and numeric approximations for the scaling constant $c_{(\mathcal{C})}$ for various examples of subcritical graph classes, including the class $\mathcal{O}$ of outerplanar graphs for which we obtain $c_{(\mathcal{O})} \approx$. See Table 1.1 for an overview. The scaling limit of outerplanar *maps*, i.e. embeddings of outerplanar graphs on the sphere considered up to orientation preserving homeomorphisms, was established by Caraceni [Car] using different methods. See also the author's recent work [Stu15b] (not included in this thesis) for an alternative proof and the scaling limit of bipartite outerplanar maps.

We extend our result for the convergence towards the CRT to random graphs with independent link weights. That is, we fix a random variable $\omega > 0$ having finite exponential moments and assign an independent copy of ω to each edge of the random graph C_n. The first-passage percolation distance $d_{\mathrm{FPP}}(x, y)$ of two points x and y is then given by the minimum of all sums of weights along paths joining x and y.

Proposition 1.4.3. *Let $\mathcal{C}$ be a subcritical class of connected graphs and let C_n denote the random graph drawn uniformly from the graphs in $\mathcal{C}$ with n vertices. Then there exists a constant $d_{(\mathcal{C})}$ such that*

$$(\mathsf{C}_n, d_{(\mathcal{C})} n^{-1/2} d_{FPP}) \xrightarrow{(d)} (\mathcal{T}_\mathsf{e}, d_{\mathcal{T}_\mathsf{e}})$$

with respect to the Gromov-Hausdorff metric, as n becomes large.

A further extension is to the largest component of a random graph that is not necessarily connected:

Proposition 1.4.4. *Let $\mathcal{C}$ be a subcritical class of connected graphs and let G_n denote the random graph drawn uniformly from all labelled graphs with n vertices whose connected components lie in $\mathcal{C}$. Let H_n denote the largest component of G_n. Moreover, for simplicity assume that $\mathcal{C}$ contains all trees. Then*

$$(\mathsf{H}_n, c_{(\mathcal{C})} n^{-1/2} d_{FPP}) \xrightarrow{(d)} (\mathcal{T}_\mathsf{e}, d_{\mathcal{T}_\mathsf{e}})$$

with respect to the Gromov-Hausdorff metric, as n becomes large. Here $c_{(\mathcal{C})}$ is the same scaling constant as in Theorem 1.4.1.

As a conclusion, we remark that it is natural to ask whether random graphs in the unlabelled setting also admit the CRT as scaling limit. Such graphs have been studied by Drmota, Fusy, Kang, Kraus and Rué [DFK+11] and Bodirsky, Fusy and Kang [BFKV07]. The scaling limit of rooted unlabelled graphs from subcritical classes was established by the author of the present thesis in the work [Stu15a] (which is not included in this thesis). The scaling limit of unlabelled unrooted graphs is current work in progress by the author. Moreover, besides the scaling limit it is also interesting to ask whether random graphs from subcritical classes admit a Benjamini-Schramm limit. A paper that addresses this question (and answers it in the affirmative) is also in preparation by the author.

Including a work with several coauthors in a thesis obliges the author to clarify that his contribution to the project was substantial. In the early stages during the preparation of [PSW14] on which Chapter 5 is based, the three authors of [PSW14] devised a proof sketch of the scaling limit Theorem 1.4.1 which the author extended to a complete proof that is included in this thesis in Section 5.7. Having finished that, the author independently discovered a new proof of Theorem 1.4.1 by constructing a size-biased $\mathcal{R}$-enriched tree. The new proof, described in Sections 5.2 and 5.3, fully replaces the old, as it is simpler, much shorter, and may easily be extended to the first-passage percolation setting (see Section 5.5.1). Furthermore, the author independently obtained the exponential tail bounds in Theorem 1.4.2. In Section 5.6, the author carried out most of the calculations and writing.

Chapter 2

Preliminaries

2.1 The continuum random tree

We briefly recall classical results regarding the convergence of random plane trees towards the continuum random tree.

2.1.1 Graph theoretic notions

All graphs considered in the present work are undirected and simple. That is, a graph G consists of a non-empty set $V(G)$ of vertices 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 is the *size* of G. Following the graph theory book by Diestel [Die10], we recall and fix basic definitions and notations. 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 $d_G(v)$. A *path* P is a graph such that

$$V(P) = \{v_0, \ldots, v_\ell\}, \qquad E(P) = \{v_0v_1, \ldots, v_{\ell-1}v_\ell\}$$

with the v_i being distinct. The number of edges of a path is its *length*. We say P *connects* or *joins* its endvertices v_0 and v_ℓ and we often write $P = v_0v_1 \ldots v_\ell$. If P has length at least two we call the graph $C_\ell = P + v_0v_\ell$ obtained by adding the edge v_0v_ℓ a *cycle*. The *complete graph* with n vertices in which each pair of distinct vertices is adjacent is denoted by K_n.

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* $\mathrm{h}(w)$ of a vertex $w \in V(G)$ is its distance from the root. The *height* $\mathrm{H}(G)$ of the entire graph is the supremum of the heights of the vertices in G. A *tree* T is a non-empty connected graph without cycles. Any two vertices of a tree are connected by a unique path. If T is rooted, then the vertices $w' \in V(T)$ that are adjacent to a vertex w and have height $\mathrm{h}(w') = \mathrm{h}(w) + 1$ form the *offspring set* of the vertex w. Its cardinality is the *outdegree* $d^+(w)$ of the vertex w.

2.1.2 Plane trees and contour functions

The *Ulam-Harris tree* is an infinite rooted tree with vertex set $\cup_{n \ge 0} \mathbb{N}^n$ consisting of finite sequences of natural numbers. The empty string $\emptyset$ is the root and the offspring of any vertex v is given by the concatenations $v1, v2, v3, \ldots$. In particular, the labelling of the vertices induces a linear order on each offspring set. A *plane tree* is defined to be a subtree of the Ulam-Harris tree that contains the root such that the offspring set of each vertex v is of the form $\{v1, v2, \ldots, vk\}$ for some integer $k \ge 0$ depending on v.

Given a plane tree T of size n we consider its canonical *depth-first search* walk $(v_i)_{0\le i\le 2(n-1)}$ that starts at the root and always traverses the leftmost unused edge first. That is, v_0 is the root of T and given $v_0,\dots,v_i$ walk if possible to the leftmost unvisited son of v_i. If v_i has no sons or all sons have already been visited, then try to walk to the parent of v_i. If this is not possible either, being only the case when v_i is the root of T and all other vertices have already been visited, then terminate the walk. The corresponding heights $c(i) := \mathrm{h}(v_i)$ define the search-depth function c of the tree T. The *contour function* $C : [0, 2(n-1)] \to \mathbb{R}_+$ is defined by $C(i) = c(i)$ for all integers $0 \le i \le 2(n-1)$ with linear interpolation between these values, see Figure 2.1 for an example.

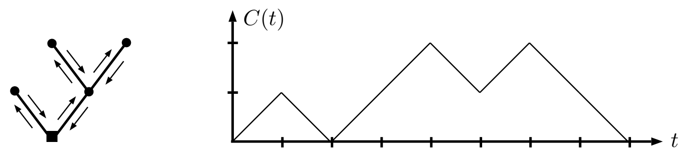

Figure 2.1: The contour function of a plane tree.

A typical model for random plane trees is that of Galton-Watson trees. The following result concerning the contour functions of conditioned Galton-Watson trees is due to Aldous [Ald93, Thm. 23], who stated it for aperiodic offspring distributions. See also Le Gall [LG10, Thm 6.1] (who stated it without aperiodicity requirements), as well as Duquesne [Duq03] and Kortchemski [Kor13] for further extensions.

Theorem 2.1.1. *Let $\mathcal{T}_n$ be a critical ξ-Galton-Watson tree conditioned on having n vertices, with the offspring distribution ξ having finite non-zero variance σ^2. Let C_n denote the contour function of $\mathcal{T}_n$. Then*

$$\left(\frac{\sigma}{2\sqrt{n}} C_n\big(t2(n-1)\big)\right)_{0\le t\le 1} \xrightarrow{(d)} \mathsf{e}$$

in $\mathcal{C}([0,1],\mathbb{R}_+)$, where $\mathsf{e} = (\mathsf{e}_t)_{0\le t\le 1}$ is a normalized Brownian excursion.

2.1.3 Gromov-Hausdorff convergence

Theorem 2.1.1 can be formulated as a convergence of random trees with respect to the Gromov-Hausdorff metric, which is a distance between compact metric spaces. We introduce the required notions following Burago, Burago and Ivanov [BBI01, Ch. 7] and Le Gall and Miermont [LGM12]

2.1.3.1 The Hausdorff metric

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

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

where $U_\epsilon(A) = \{x \in X \mid d(x, A) \leq \epsilon\}$ denotes the *ϵ-hull* of A. In general, the Hausdorff-distance does not define a metric on the set of all subsets of X, which is why we restrict ourselves to compact subsets.

Proposition 2.1.2 ([BBI01, Prop. 7.3.3])**.** *The Hausdorff distance d_H defines a metric on the set of compact subsets of X.*

Proof. The triangle inequality is easily seen to be satisfied for arbitrary subsets of X. If A and B are closed subsets of X, then $d_{\mathrm{H}}(A, B) = 0$ implies that $A = B$. Moreover, if A and B are bounded, then $d_{\mathrm{H}}(A, B) < \infty$. □

2.1.3.2 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 [LGM12] 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

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

where in both cases the infimum is taken over all isometric embeddings $\iota_X : X \to E$ and $\iota_Y : Y \to E$ into a common metric space (E, d_E), compare with Figure 2.2.

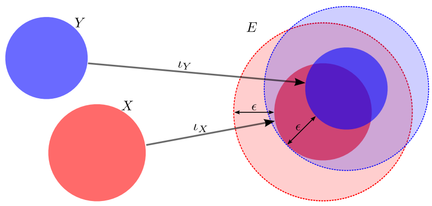

Figure 2.2: The Gromov-Hausdorff distance.

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 2.1.3 ([BBI01, Thm. 7.3.25] and [LGM12, 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 $Y^\bullet$).

Proof. We will only show the pointed case, as the (easier) regular case may be treated analogously.

In order to show "$\geq$", it suffices to show that $d_{\mathrm{GH}}(X,Y) < r$ implies $\mathsf{dis}(R) < 2r$ for some correspondence R. So, suppose that we are given $r > 0$ with $d_{\mathrm{GH}}(X,Y) < r$. Then there exists a metric space (E, d_E) and pointed subspaces $A^\bullet = (A, a_0)$ and $B^\bullet = (B, b_0)$ of E which are isometric copies of $X^\bullet$ and $Y^\bullet$, such that $d_{\mathrm{H}}(A,B) < r$ and $d_E(a_0, b_0) < r$. Let R be the correspondence given by $(a,b) \in R$ if and only if $d_E(a,b) < r$ for each $a \in A$ and $b \in B$. Note that the distinguished vertices a_0 and b_0 correspond to each other. Moreover, for each $(a,b), (a',b') \in R$ it holds that

$$d_E(a,a') \leq d_E(a,b) + d_E(b,b') + d_E(b',a') \leq 2r + d_E(b,b')$$

and similarly

$$d_E(b,b') \leq 2r + d_E(a,a').$$

Hence

$$\mathsf{dis}(R) \leq 2r.$$

In order to show "$\leq$", let R be an arbitrary correspondence between $X^\bullet$ and $Y^\bullet$ and set $r = \frac{1}{2}\mathsf{dis}(R)$. It suffices to show that there is a pseudo-metric d on the disjoint union $X \sqcup Y$ such that $d|_{X\times X} = d_X$, $d|_{Y\times Y} = d_Y$, $d_{\mathrm{H}}(X,Y) \leq r$ and $d(x_0, y_0) \leq r$. We define this by setting

$$d(x,y) = \inf\{d_X(x,x') + r + d_Y(y,y') \mid (x',y') \in \mathcal{R}\}$$

for each $x \in X$ and $y \in Y$. Note that this implies that $d(x,y) = r$ if $(x,y) \in R$. In particular, $d(x_0,y_0) = r$. Moreover, it follows that $Y \subset U_{r+\epsilon}(X)$ and $X \subset U_{r+\epsilon}(Y)$ for each $\epsilon > 0$. Thus $d_{\mathrm{H}}(X,Y) \leq r$ holds. It remains to check that the triangle inequality holds. To this end, suppose that $x_1, x_2 \in X$ and $y \in Y$. For each points x' and y' that correspond to each other, we have that

$$(d_X(x_1,x') + r + d_Y(y',y)) + (d_X(x_2,x') + r + d_Y(y',y)) \geq d_X(x_1,x_2) = d(x_1,x_2)$$

and consequently

$$d(x_1,y) + d(x_2,y) \geq d(x_1,x_2).$$

Similarly,

$$d(x_1,x_2) + (d_X(x_2,x') + r + d_Y(y',y)) \geq d_X(x_1,x') + r + d_Y(y',y) \geq d(x_1,y)$$

and hence

$$d(x_1,x_2) + d(x_2,y) \geq d(x_1,y).$$

The remaining cases are symmetric to the cases considered or trivial. $\square$

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

Lemma 2.1.4 ([BBI01, Thm. 7.3.30] and [LGM12, 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) $d_{GH}(X,Z) \leq d_{GH}(X,Y) + d_{GH}(Y,Z)$.

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

2.1.3.3 The space of isometry classes of compact metric spaces

In Section 2.1.3.1 we saw that the Hausdorff-distance defines a metric on the set of all compact subsets of a metric space. By Lemma 2.1.4 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. A solution is presented as an exercise in [BBI01, Rem. 7.2.5]:

Proposition 2.1.5. *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.*

Proof. The lower bound is easily checked, as the intervals $[0, \alpha]$, $\alpha > 0$ equipped with the restriction of the euclidian metric are pairwise non-isometric.

For the upper bound, note that any compact metric space has a countable basis and its isometry type is determined by the restriction of the metric to this basis. If the metric space is pointed, we may encode the distinguished root vertex either by distinguishing a vertex of the basis (if the root vertex happens to belong to the basis) or, if the root vertex does not belong to the basis, by an infinite subset of the basis whose unique accumulation point is the distinguished vertex. e Hence, the cardinality of any set M of pairwise non-isomorphic (pointed) metric spaces is bounded by the cardinality of $\mathbb{R}^{\mathbb{N}\times\mathbb{N}} \times 2^{\mathbb{N}}$ which equals $2^{\aleph_0}$. □

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 2.1.4 now reads as follows.

Corollary 2.1.6 ([BBI01, 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 very suitable for studying random elements:

Proposition 2.1.7 ([LGM12, Thm. 3.5] and [BBI01, Thm. 7.4.15])**.** *The spaces $\mathbb{K}$ and $\mathbb{K}^\bullet$ are separable and complete, i.e. they are Polish spaces.*

2.1.4 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

$$q([0, d_f(x, y)]) = \varphi_{x,y}([0, d_f(x, y)]).$$

We may construct $\mathbb{R}$-trees 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

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

for $u \leq v$. 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 2.1.8 ([LGM12, Thm. 3.1])**.** *Given a continous 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 2.1.9 ([LGM12, Cor. 3.7])**.** *The map*

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

is Lipschitz-continous.

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

Definition 2.1.10. *The random pointed metric space $(\mathcal{T}_{\mathsf{e}}, d_{\mathcal{T}_{\mathsf{e}}}, \bar{0})$ coded by the Brownian excursion of duration one $\mathsf{e} = (\mathsf{e}_t)_{0 \leq t \leq 1}$ 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.

Any plane tree is a pointed metric space with respect to the graph-metric and the root vertex $\emptyset$. Hence a random plane trees may be considered as random elements of the metric space $\mathbb{K}^\bullet$. The following invariance principle giving a scaling limit for certain random plane trees is due to Aldous [Ald93] and there exist various extensions. See for example Duquesne [Duq03], Duquesne and Le Gall [DLG05], Le Gall [LG10, p. 740], and Haas and Miermont [HM12].

Theorem 2.1.11. *Let $\mathcal{T}_n$ be a critical ξ-Galton-Watson tree conditioned on having n vertices, where ξ has finite non-zero variance σ^2. As n tends to infinity, $\mathcal{T}_n$ with edges rescaled to length $\frac{\sigma}{2\sqrt{n}}$ converges in distribution to the CRT, that is*

$$(\mathcal{T}_n, \frac{\sigma}{2\sqrt{n}} d_{\mathcal{T}_n}, \emptyset) \xrightarrow{(d)} (\mathcal{T}_{\mathsf{e}}, d_{\mathcal{T}_{\mathsf{e}}}, \bar{0})$$

in the metric space $(\mathbb{K}^\bullet, d_{GH})$.

Proof. Let $(C_n(t))_{0 \le t \le 2(n-1)}$ denote the contour function of the random plane tree $\mathcal{T}_n$. Then the random plane $\mathcal{T}_{f_n}$ with $f_n(t) = C_n(t2(n-1))$ may, informally described, be obtained from the tree $\mathcal{T}_n$ by replacing each discrete edge by a copy of the unit-interval $[0,1]$. In particular, in this coupling the Gromov-Hausdorff distance $d_{\mathrm{GH}}(\mathcal{T}_n, \mathcal{T}_{f_n})$ is bounded by a constant. Hence

$$d_{\mathrm{GH}}(n^{-1/2}\mathcal{T}_n, n^{-1/2}\mathcal{T}_{f_n}) = n^{-1/2} d_{\mathrm{GH}}(\mathcal{T}_n, \mathcal{T}_{f_n}) \xrightarrow{p} 0$$

Moreover, by Theorem 2.1.1 we have that $\frac{\sigma}{2\sqrt{n}} f_n(t)$ converges weakly to Brownian excursion e and hence $\frac{\sigma}{2\sqrt{n}}\mathcal{T}_{f_n}$ converges weakly to the CRT $\mathcal{T}_{\mathsf{e}}$. Thus

$$\frac{\sigma}{2\sqrt{n}}\mathcal{T}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}.$$

□

2.2 Combinatorial species

Combinatorial species were developed by Joyal [Joy81] and allow for a systematic study of a wide range of combinatorial objects. We are going to make heavy use of this framework and recall the required theory and notation following Bergeron, Labelle and Leroux [BLL98] and Joyal [Joy81]. The language of *combinatorial classes* used in the monumental book on analytic combinatorics by Flajolet and Sedgewick [FS09] is essentially equivalent in many aspects, although less emphasis is put on studying objects up to symmetry.

2.2.1 The category of combinatorial species

A *combinatorial 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. Formally, a species is a functor from the groupoid of finite sets and bijections to the category of finite sets and arbitrary maps. 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 $\sigma : U \to V$. 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.

$$\begin{array}{ccc} \mathcal{F}[U] & \xrightarrow{\mathcal{F}[\sigma]} & \mathcal{F}[V] \\ \downarrow{\scriptstyle \alpha_U} & & \downarrow{\scriptstyle \alpha_V} \\ \mathcal{G}[U] & \xrightarrow{\mathcal{G}[\sigma]} & \mathcal{G}[V] \end{array}$$

In other words, α is a natural isomorphism between these functors. The species $\mathcal{F}$ and $\mathcal{G}$ are *isomorphic* if there exists and isomorphism from one to the other. This is denoted by $\mathcal{F} \simeq \mathcal{G}$ or, by an abuse of notation committed frequently in the literature, just $\mathcal{F} = \mathcal{G}$. Formally, we may form the groupoid of combinatorial species with its objects given by species and its morphisms by natural isomorphisms.

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* 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 or an *isomorphism type*.

2.2.1.1 Examples

We will mostly be interested in subspecies of the species of finite simple graphs such as the species of trees. Moreover, we will make use of standard species such as the species of linear orders SEQ or the SET-species given by $\mathrm{SET}[U] = \{U\}$ for all U. Moreover let 0 denote the empty species, 1 the species with a single object of size 0 and $\mathcal{X}$ the species with a single object of size 1.

2.2.2 Symmetries and generating power series

The *exponential generating series* of a species $\mathcal{F}$ is defined as the formal power series

$$\mathcal{F}(x) = \sum_{n=0}^{\infty} f_n x^n \in \mathbb{Q}[[x]]$$

with f_n denoting the cardinality of the set of $\mathcal{F}$-objects $\mathcal{F}[n]$ with $[n] := \{1, \ldots, n\}$. 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. In particular σ_1 denotes the number of fixpoints. We may form the species $\mathrm{Sym}(\mathcal{F})$ 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 $\bigcup_{n \in \mathbb{N}_0} \mathrm{Sym}(\mathcal{F})[n]$. The reason for studying cycle index sums is the following remarkable property. Due to its importance, we provide a short proof.

Lemma 2.2.1 ([Joy81]). *Let U be a finite n-element set. For any unlabelled $\mathcal{F}$-object m of size n there are precisely $n!$ symmetries $(F, \sigma) \in Sym(\mathcal{F})[U]$ having the property that F has isomorphism type m.*

Proof. The symmetric group $G := \mathcal{S}(U)$ of the set U operates (from the left) via relabelling on the set $\mathcal{F}[U]$. The automorphisms of any object F are given by its stabilizer group G_F and its isomorphism class corresponds to its orbit $G.F$. Fix any $F \in \mathcal{F}[U]$ and let m denote its isomorphism type. By standard results on group actions, the map

$$G/G_F \to G.F, \qquad gG_F \to g.F$$

is well-defined and bijective. Let $T \subset G$ denote a (left) transversal of G_F in G, that is T contains precisely one element of each left coset with respect to G_F. Then the (distinct) $\mathcal{F}$-objects $t.F$, $t \in T$ are precisely the labelled $\mathcal{F}$-objects over U with isomorphism type m. Clearly, for any group element $g \in G$, the stabilizer of $g.F$ is given by its conjugated image

$$G_{g.F} = gG_F g^{-1}.$$

Hence the set of symmetries corresponding to the isomorphism type m is given by the distinct pairs $(t.F, \sigma)$, $t \in T$, $\sigma \in tG_F t^{-1}$. Hence the total number of symmetries is given by

$$|G.F \times G_F| = |G| = n!.$$

□

$\mathcal{F}$	$\mathcal{F}(x)$	$\tilde{\mathcal{F}}(x)$	$Z_{\mathcal{F}}(x_1, x_2, \ldots)$
SET	$\exp(x)$	$\exp(\sum_{i=1}^{\infty} x^i/i)$	$\exp(\sum_{i=1}^{\infty} x_i/i)$
SEQ	$1/(1-x)$	$1/(1-x)$	$1/(1-x_1)$
$\mathcal{X}$	x	x	x_1
0	0	0	0
1	1	1	1

Table 2.1: Generating series of some examples of combinatorial species.

From a probabilistic viewpoint, Lemma 2.2.1 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. This is crucial, as symmetries may be decomposed fairly systematically using the theory of species.

Moreover, it follows that the generating series and cycle index sum are related by

$$\mathcal{F}(z) = Z_{\mathcal{F}}(z, 0, 0, \ldots) \quad \text{and} \quad \tilde{\mathcal{F}}(z) = Z_{\mathcal{F}}(z, z^2, z^3, \ldots).$$

2.2.2.1 Examples

The generating series and cycle index sums of the examples of species mentioned so far are summarized in Table 2.1. The only non-trivial entry we need to check is the expression for the cycle index sum Z_{SET}, but this is easily established: For any integer $n \geq 0$ let $\mathcal{S}_n$ denote the symmetric group of order n. Then

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

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

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

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. We refer the inclined reader to [FS09, Appendix A.5] for an adequate discussion of these questions.

2.2.3 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 $\mathcal{F}$, $(\mathcal{F}_i)_{i\in\mathbb{N}}$ and $\mathcal{G}$ denote species and U an arbitrary finite set. The *sum* $\mathcal{F}+\mathcal{G}$ is defined by the disjoint union

$$(\mathcal{F}+\mathcal{G})[U] = \mathcal{F}[U] \sqcup \mathcal{G}[U].$$

More generally, the infinite sum $(\sum_i \mathcal{F}_i)$ may be defined by $(\sum_i \mathcal{F}_i)[U] = \bigsqcup_i \mathcal{F}_i[U]$ 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. 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 substition 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 $\overline{\sigma} : \pi \to \overline{\pi} = \{\sigma(Q) \mid Q \in \pi\}$ of partitions to the $\mathcal{F}$-object and the restriced maps $\sigma|_Q$ with $Q\in\pi$ to their corresponding $\mathcal{G}$-objects. We will often write $\mathcal{F}(\mathcal{G})$ instead of $\mathcal{F}\circ\mathcal{G}$. The *rooted* or *pointed* $\mathcal{F}$-species is given by

$$\mathcal{F}^\bullet[U] = \mathcal{F}[U]\times U$$

with componentwise transport. That is, a pointed object is formed by distinguishing a label, named the *root* of the object, and any transport function is required to preserve roots. The *derived* species $\mathcal{F}'$ is defined by

$$\mathcal{F}'[U] = \mathcal{F}[U\cup\{*_U\}]$$

with $*_U$ referring to an arbitrary fixed element not contained in the set U. (For example, we could take $*_U = U$.) The transport along a bijective map $\sigma : U \to V$ is done by applying the canonically extended bijection $\sigma' : U \sqcup \{*_U\} \to V \sqcup \{*_V\}$ with $\sigma'(*_U) = *_V$ to the object. Derivation and pointing are related by an isomorphism $\mathcal{F}^\bullet \simeq \mathcal{X}\cdot\mathcal{F}'$.

Note that $\mathcal{F}'^\bullet$ and $\mathcal{F}^{\bullet\prime}$ are in general different species. In $\mathcal{F}^{\bullet\prime}$ objects, the root and $*$-label may coincide, since

$$\mathcal{F}^{\bullet\prime}[U] = \mathcal{F}^\bullet[U\cup\{*_U\}]$$

implies that a $\mathcal{F}^{\bullet\prime}$-object over U is a $\mathcal{F}$-object over $U\cup\{*_U\}$ together with a distinguished element from $U\cup\{*_U\}$. On the other hand, $\mathcal{F}'^\bullet$-objects are always rooted at non-$*$-labels, since

$$\mathcal{F}'^\bullet[U] = \mathcal{F}'[U]\times U$$

	EGF	OGF	Cycle index sum
$\sum_i \mathcal{F}_i$	$\sum_i \mathcal{F}_i(x)$	$\sum_i \tilde{\mathcal{F}}_i(x)$	$\sum_i Z_{\mathcal{F}_i}(x_1, x_2, \ldots)$
$\mathcal{F} \cdot \mathcal{G}$	$\mathcal{F}(x)\mathcal{G}(x)$	$\tilde{\mathcal{F}}(x)\tilde{\mathcal{G}}(x)$	$Z_{\mathcal{F}}(x_1, x_2, \ldots)Z_{\mathcal{G}}(x_1, x_2, \ldots)$
$\mathcal{F} \circ \mathcal{G}$	$\mathcal{F}(\mathcal{G}(x))$	$Z_{\mathcal{F}}(\tilde{\mathcal{G}}(x), \tilde{\mathcal{G}}(x^2), \ldots)$	$Z_{\mathcal{F}}(Z_{\mathcal{G}}(x_1, x_2, \ldots), Z_{\mathcal{G}}(x_2, x_4, \ldots), \ldots)$
$\mathcal{F}^\bullet$	$x\frac{d}{dx}\mathcal{F}(x)$	$x(\frac{\partial}{\partial x_1} Z_{\mathcal{F}})(x, x^2, \ldots)$	$x_1 \frac{\partial}{\partial x_1} Z_{\mathcal{F}}(x_1, x_2, \ldots)$
$\mathcal{F}'$	$\frac{d}{dx}\mathcal{F}(x)$	$(\frac{\partial}{\partial x_1} Z_{\mathcal{F}})(x, x^2, \ldots)$	$\frac{\partial}{\partial x_1} Z_{\mathcal{F}}(x_1, x_2, \ldots)$

Table 2.2: Relation between combinatorial constructions and generating series.

implies that a $\mathcal{F}'^\bullet$-object over U is a $\mathcal{F}$-object over $U \cup \{*_U\}$ together with a distinguished element from U.

Explicit formulas for the generating series and cycle index sums of the discussed constructions are summarized in Table 2.2. The notation is quite suggestive: up to (canonical) isomorphism, each operation considered in this section is associative. Roughly described, this means that for each operation $\mu \in \{+, \cdot, \circ\}$ there is a "natural choice" for an isomorphism

$$(\mathcal{F}_1 \mu \mathcal{F}_2)\mu\mathcal{F}_3 \simeq \mathcal{F}_1 \mu(\mathcal{F}_2 \mu \mathcal{F}_3).$$

But this is only half of the story: for example, we may apply these isomorphisms in different orders in order to obtain an isomorphism from $((\mathcal{F}_1\mu\mathcal{F}_2)\mu\mathcal{F}_3)\mu\mathcal{F}_4$ to $\mathcal{F}_1\mu(\mathcal{F}_2\mu(\mathcal{F}_3\mu\mathcal{F}_4))$. But why should we end up with the same isomorphism, regardless of which order we choose? In order to answer this question adequately, the concept of monoidal categories is required, and we refer the inclined reader to [Joy81, Sec. 7] for a thorough discussion.

The sum and product are commutative operations (up to canonical isomorphisms) and satisfy the distributive law

$$\mathcal{F} \cdot (\mathcal{G}_1 + \mathcal{G}_2) \simeq \mathcal{F} \cdot \mathcal{G}_1 + \mathcal{F} \cdot \mathcal{G}_2. \tag{2.2.1}$$

for any two species $\mathcal{G}_1$ and $\mathcal{G}_2$. The operation of deriving a species is additive and satisfies a product rule and a chain rule, analogous to the derivative in calculus:

$$(\mathcal{F} \cdot \mathcal{G})' \simeq \mathcal{F}' \cdot \mathcal{G} + \mathcal{F} \cdot \mathcal{G}' \quad \text{and} \quad \mathcal{F}(\mathcal{G})' \simeq \mathcal{F}'(\mathcal{G}) \cdot \mathcal{G}'. \tag{2.2.2}$$

Recall that for the chain rule to apply we have to require $\mathcal{G}[\emptyset] = \emptyset$, since otherwise $\mathcal{F}(\mathcal{G})$ is not defined.

The species 0, 1, $\mathcal{X}$ are neutral elements in a certain sense, that is there are canonical isomorphisms

$$\mathcal{F} \simeq \mathcal{F} + 0 \simeq \mathcal{F} \cdot 1 \simeq \mathcal{F}(\mathcal{X}).$$

2.2.4 Decomposition of symmetries of the substitution operation

We are going to need detailed information on the structure of the symmetries of the composition $\mathcal{F} \circ \mathcal{G}$. The following is a standard decomposition given in [Joy81, BLL98, BFKV11]. 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, a $\mathcal{F}$-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 $\sigma : U \to U$. We require the permutation σ to permute the partition classes and induce an automorphism $\bar{\sigma} : \pi \to \pi$ 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 $\bar{\sigma}$ it follows that for all i we have $\sigma^\ell(Q_i) = Q_i$ and the restriction $\sigma^\ell|_{Q_i} : Q_i \to Q_i$ is an automorphism of G_{Q_i}. Conversely, if we know $(G_{Q_1}, \sigma^\ell|_{Q_1})$ and the maps $\sigma|_{Q_i} = (\sigma|_{Q_1})^i$ for $1 \le i \le \ell - 1$, 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 corresponds to a cycle of the induced permutation $\bar{\sigma}$ whose length is a divisor of the length of ν.

2.2.5 Combinatorial specifications

In this section we briefly recall Joyal's implicit species theorem that allows us to define combinatorial species up to unique isomorphism and construct recursive samplers that draw objects of a species randomly (see Section 2.4 below). In order to state the theorem we need to introduce the concept of *multisort species*. As it is sufficient for our applications, we restrict ourselves to the 2-sort case.

A 2-*sort species* $\mathcal{H}$ is a functor that maps any pair $U = (U_1, U_2)$ of finite sets to a finite set $\mathcal{H}[U] = \mathcal{H}[U_1, U_2]$ and any pair $\sigma = (\sigma_1, \sigma_2)$ of bijections $\sigma_i : U_i \to V_i$ to a bijection $\mathcal{H}[\sigma] : \mathcal{H}[U] \to \mathcal{H}[V]$ in such a way, that identity maps and composition of maps are preserved. The operations of sum, product and composition extend naturally to the multisort-context. Let $\mathcal{H}$ and $\mathcal{K}$ be 2-sort species and $U = (U_1, U_2)$ a pair of finite sets. The *sum* is defined by

$$(\mathcal{H} + \mathcal{K})[U] = \mathcal{H}[U] \sqcup \mathcal{K}[U].$$

We write $U = V + W$ if $U_i = V_i \cup W_i$ and $V_i \cap W_i = \emptyset$ for all i. The *product* is defined by

$$(\mathcal{H} \cdot \mathcal{K})[U] = \bigsqcup_{V+W=U} \mathcal{H}[V] \times \mathcal{K}[W].$$

The *partial derivatives* are given by

$$\partial_1 \mathcal{H}[U] = H[U_1 \cup \{*_{U_1}\}, U_2] \quad \text{and} \quad \partial_2 \mathcal{H}[U] = H[U_1, U_2 \cup \{*_{U_2}\}].$$

In order state Joyal's implicit species theorem we also require the substitution operation for multisort species; this will allow us to define species "recursively" up to (canonical) isomorphism. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be (1-sort) species and M a finite set. A structure of the *composition* $\mathcal{H}(\mathcal{F}_1, \mathcal{F}_2)$ over the set M is a quadrupel $(\pi, \chi, \alpha, \beta)$ such that:

1. π is partition of the set M.
2. $\chi : \pi \to \{1, 2\}$ is a function assigning to each class a sort.
3. α a function that assigns to each class $Q \in \pi$ a $\mathcal{F}_{\chi(Q)}$ object $\alpha(Q) \in \mathcal{F}_{\chi(Q)}[Q]$.
4. β a $\mathcal{H}$-structure over the pair $(\chi^{-1}(1), \chi^{-1}(2))$.

This construction is *functorial*: any pair of isomorphisms (or natural transformations) α_1, α_2 with $\alpha_i : \mathcal{F}_i \xrightarrow{\sim} \mathcal{G}_i$ *induces* an isomorphism (or natural transformation) $\mathcal{H}[\alpha_1, \alpha_2] : \mathcal{H}(\mathcal{F}_1, \mathcal{F}_2) \xrightarrow{\sim} \mathcal{H}(\mathcal{G}_1, \mathcal{G}_2)$.

Let $\mathcal{H}$ be a 2-sort species and recall that $\mathcal{X}$ denotes the species with a unique object of size one. A solution of the system $\mathcal{Y} = \mathcal{H}(\mathcal{X}, \mathcal{Y})$ is pair $(\mathcal{A}, \alpha)$ of a species $\mathcal{A}$ with $\mathcal{A}[\emptyset] = \emptyset$ and an isomorphism $\alpha : \mathcal{A} \xrightarrow{\sim} \mathcal{H}(\mathcal{X}, \mathcal{A})$. An isomorphism of two solutions $(\mathcal{A}, \alpha)$ and $(\mathcal{B}, \beta)$ is an isomorphism of species $u : \mathcal{A} \xrightarrow{\sim} \mathcal{B}$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\alpha} & \mathcal{H}(\mathcal{X}, \mathcal{A}) \\ \downarrow u & & \downarrow \mathcal{H}(\mathrm{id}, u) \\ \mathcal{B} & \xrightarrow{\beta} & \mathcal{H}(\mathcal{X}, \mathcal{B}) \end{array}$$

We may now state Joyal's implicit species theorem.

Theorem 2.2.2 ([Joy81], Théorème 6)**.** *Let $\mathcal{H}$ be a 2-sort species satisfying $\mathcal{H}(0, 0) = 0$. If $(\partial_2 \mathcal{H})(0, 0) = 0$, then the system $\mathcal{Y} = \mathcal{H}(\mathcal{X}, \mathcal{Y})$ has up to isomorphism only one solution. Moreover, between any two given solutions there is exactly one isomorphism.*

We say that an isomorphism $\mathcal{F} \simeq \mathcal{H}(\mathcal{X}, \mathcal{F})$ is a *combinatorial specification* for a species $\mathcal{F}$ with $\mathcal{F}[\emptyset] = \emptyset$, if the 2-sort species $\mathcal{H}$ satisfies the requirements of Theorem 2.2.2, i.e. if $\mathcal{H}(0, 0) = 0$.

Remark 2.2.3. *It is important to note how the solution is constructed in the proof [Joy81, Proof of théorème 6, p.52] of Theorem 2.2.2. Let $\mathcal{H}$ satisfy the requirements of Theorem 2.2.2. Define a sequence of (1-sort) species by*

$$\mathcal{A}_0 = 0 \quad \textit{and} \quad \mathcal{A}_{n+1} = \mathcal{H}(\mathcal{X}, \mathcal{A}_n).$$

We have a trivial "empty" natural transformation $\mathcal{A}_0 \xrightarrow{i_0} \mathcal{A}_1$ and may define recursively the natural transformation $i_n = \mathcal{H}(id, i_{n-1})$ from $\mathcal{H}(id, \mathcal{A}_{n-1}) = \mathcal{A}_n$ to

$\mathcal{H}(id, \mathcal{A}_n) = \mathcal{A}_{n+1}$. The solution in Theorem 2.2.2 is then obtained as the direct limit of the sequence

$$\mathcal{A}_0 \xrightarrow{i_0} \mathcal{A}_1 \xrightarrow{i_1} \mathcal{A}_2 \xrightarrow{i_2} \ldots .$$

This is possible, as Joyal argues in his proof, since for each integer $k \geq 0$ there is an N such that for all $n \geq N$ the natural transformation i_n induces an isomorphism from $\mathcal{A}_n^{[\leq k]}$ to $\mathcal{A}_{n+1}^{[\leq k]}$, the species obtained by restricting to objects of size equal or less than k.

2.3 Cycle pointing

Cycle pointing is a technique introduced by Bodirsky, Fusy, Kang and Vigerske [BFKV11] as means to study unlabelled graphs and trees. One of their main application is to the enumeration of unlabelled unrooted trees, providing a new proof for their asymptotic enumeration formula, that does not require the dissymmetry theorem.

2.3.1 The cycle pointing operator

Bodirsky, Fusy, Kang and Vigerske [BFKV11] introduced the cycle pointing operator which 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 c-symmetry* of the cycle-pointed species $\mathcal{S} \subset \mathcal{G}^\circ$ is a quadruple $((G, \tau), \sigma, v)$ such that (G, τ) is a $\mathcal{S}$-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 c-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 index ranging over the set $\bigcup_{n \in \mathbb{N}_0} \mathrm{RSym}(\mathcal{S})[n]$.

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

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

Lemma 2.3.1 ([BFKV11, 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:*

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

   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.*

Proof. 1) The inverse map is given as follows. Any symmetry $((G,\tau),\sigma) \in \mathrm{Sym}(\mathcal{S})[U]$ satisfies $\sigma\tau\sigma^{-1} = \tau$. Letting k denote the lengths of the marked cycle, this implies that there exists a unique integer $0 \le \ell \le k-1$ such that τ^ℓ is one of the disjoint cycles of σ. In order to see this, note first that if $\sigma\tau\sigma^{-1} = \tau$ then σ fixes the set of atoms V of the cycle τ, i.e. there exists a permutation ν of V which is a product of disjoint cycles of σ with $\nu\tau\nu^{-1} = \tau$. The symmetric group over V acts transitively on the $(k-1)!$-element set of k-cycles on V. Hence the stabilizer group of τ has k elements and must therefore agree with the powers $\mathrm{id}, \tau, \ldots, \tau^{k-1}$ of τ. Hence $\nu = \tau^\ell$ for some integer $0 \le \ell \le k-1$.

Let u denote the smallest atom of V. Then $((\tau^{\ell-1}.G,\tau), \tau^{1-\ell}\sigma, \tau^\ell.u)$ forms a rooted c-symmetry, i.e. an element of the set $\mathrm{RSym}(\mathcal{S})[U]$.

The two maps are inverse to each other and clearly preserve isomorphism types.

2) The bijection clearly preserves the isomorphism type of the $\mathcal{S}$-object corresponding to the symmetry. Hence the number of rooted c-symmetries corresponding to an unlabelled $\mathcal{S}$-object m of size n agrees with the numbers of symmetries corresponding to m, which by Lemma 2.2.1 equals $n!$. □

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

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

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

The main point of the cycle-pointing construction is evident from the following fact.

Lemma 2.3.2 ([BFKV11, Thm. 15]). *Any unlabelled $\mathcal{G}$-structure m of size n may be cycle-pointed in precisely n ways, i.e. there exist precisely n unlabelled $\mathcal{G}^\circ$-structures with corresponding $\mathcal{G}$-structure m.*

Proof. Any rooted c-symmetry over $[n]$ whose $\mathcal{G}$-object has type m may be obtained in a unique way by choosing a symmetry over $[n]$ whose $\mathcal{G}$-object has type m, selecting one of its atoms and marking the corresponding cycle. In particular, the numbers A and B of rooted c-symmetries and symmetries from $\mathrm{RSym}(\mathcal{G}^\circ)[n]$ and $\mathrm{Sym}(\mathcal{G})$ satisfy $A = nB$. By Lemma 2.2.1 we have that $B = n!$ and hence $A = n!n$. On the other hand, Lemma 2.3.1 implies that $A = Cn!$ with C the number of unlabelled cycle pointed structures corresponding m. Hence $C = n$. □

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. And studying the random $\mathcal{G}^\circ$-object might be easier due to the additional information given by the marked cycle. Moreover, Lemma 2.3.2 implies that

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

2.3.1.1 Example

The pointed cycle index sum of the species SET is given by

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

2.3.2 Operations on cycle pointed species

Cycle pointed species come with a set of new operations introduced in [BFKV11]. 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

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

The cycle-pointing operator obeys the following product rule

$$(\mathcal{G} \cdot \mathcal{H})^\circ \simeq \mathcal{G}^\circ \star \mathcal{H} + \mathcal{H}^\circ \star \mathcal{G}.$$

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 2.2.4, 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

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

2.4 (Pólya-)Boltzmann samplers

Boltzmann samplers were introduced in [DFLS02, DFLS04, FFP07] and generalized to Pólya-Boltzmann samplers in [BFKV11]. They form our main tool in the analysis of random discrete objects and we discuss the required notions and properties following these sources.

2.4.1 Boltzmann models

Given a species $\mathcal{F}$ and a real number $x \geq 0$ satisfying $0 < \mathcal{F}(x) < \infty$ we may consider the corresponding *Boltzmann model* for labelled objects. It is a probability measure on the set $\bigcup_{n=0}^{\infty} \mathcal{F}[n]$ that assigns the probability weight

$$\frac{x^n}{n!}\mathcal{F}(x)^{-1}$$

for each n to each $\mathcal{F}$-structure $F \in \mathcal{F}[n]$. Expressing $\mathcal{F}$ in terms of other species via the operations discussed aids in the construction of *Boltzmann samplers*, i.e. random generators that produce objects according to a Boltzmann model. We let $\Gamma\mathcal{F}(x)$ denote a Boltzmann sampler for labelled objects with parameter x. Note that $\Gamma\mathcal{F}(x)$ conditioned on having size n has the uniform distribution on $\mathcal{F}[n]$.

The Boltzmann model for unlabelled objects is defined similarly: For any integer n, let $\tilde{\mathcal{F}}[n]$ denote the set of unlabelled $\mathcal{F}$-objects with size n. Given a number $x \geq 0$ with $0 < \tilde{\mathcal{F}}(x) < \infty$, the Boltzmann distribution for unlabelled objects is a probability distribution on the set $\bigcup_{n=0}^{\infty} \tilde{\mathcal{F}}[n]$ that assigns the probability weight

$$x^n \tilde{\mathcal{F}}(x)^{-1}$$

for each n to each unlabelled $\mathcal{F}$-structure of size n. The corresponding Boltzmann sampler is denoted by $\Gamma\tilde{\mathcal{F}}(x)$.

The *Pólya-Boltzmann model* was introduced in [BFKV11]: 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 $\Gamma Z_{\mathcal{F}}(s_1, s_2, \ldots)$.

Lemma 2.2.1 directly implies the following crucial property, that shows how Boltzmann samplers for labelled and unlabelled objects are special cases of Pólya Boltzmann samplers.

Lemma 2.4.1 ([BFKV11, Lem. 36]). *Consider a species $\mathcal{F}$ having a Pólya-Boltzmann sampler $\Gamma Z_{\mathcal{F}}(s_1, s_2, \ldots)$. Then, for any parameter $x \geq 0$ with $0 < \tilde{\mathcal{F}}(x) < \infty$ a Boltzmann sampler $\Gamma\tilde{\mathcal{F}}(x)$ for unlabelled $\mathcal{F}$-objects is given by taking the isomorphism type of the $\mathcal{F}$-structure of the random symmetry*

$$\Gamma Z_{\mathcal{F}}(x, x^2, \ldots).$$

Given a parameter $y \geq 0$ with $0 < \mathcal{F}(y) < \infty$, a Boltzmann sampler $\Gamma\mathcal{F}(y)$ for labelled $\mathcal{F}$-objects is given by taking the $\mathcal{F}$-structure of the random symmetry

$$\Gamma Z_{\mathcal{F}}(y, 0, 0, \ldots).$$

A Pólya-Boltzmann model for random cycle pointed species is given by a probability measure on random rooted c-symmetries: Let $\mathcal{S}$ be a cycle-pointed species. Given real nonnegative 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 c-symmetry $((G, \tau), \sigma, v) \in \mathrm{RSym}[n]$. Here ℓ denotes the lengths of the marked cycle τ. The corresponding Pólya-Boltzmann sampler of this model is denoted by $\Gamma\bar{Z}_{\mathcal{S}}(s_1, t_1; s_2, t_2; \ldots)$.

2.4.2 Rules for the construction of Boltzmann samplers

Boltzmann samplers are traditionally denoted using some Pseudo-code notation. We are going to deviate from this tradition, by providing more detailed explanations of each step using words rather than improvised code. By this the author hopes to make the material more accessible to a larger audience. In the following we are always going to suppose that $\mathcal{F}$ is a species and $x, x_1, x_2, \ldots$ are nonnegative numbers such that sums $\mathcal{F}(x)$, and $Z_{\mathcal{F}}(x_1, x_2, \ldots)$ are positive and finite. If $\mathcal{F}$ is cycle-pointed, then we also assume that $s_1, t_1, s_2, t_2, \ldots$ are nonnegative numbers such that $\bar{Z}_{\mathcal{F}}(s_1, t_1; s_2, t_2; \ldots)$ is positive and finite.

Suppose that we are given a decomposition

$$\mathcal{F} = \mathcal{A}\mu\mathcal{B}$$

with $\mu \in \{+, \cdot, \circ\}$ one of the discussed operations of sum, product and substitution. In order to construct a (Pólya-)Boltzmann-sampler for the species $\mathcal{F}$, we may apply certain construction rules in order to obtain a sampler in terms of samplers for the species $\mathcal{A}$ and $\mathcal{B}$. In the following we summarize these construction rules for (Pólya-)Boltzmann samplers, following [DFLS04], [FFP07] and [BFKV11]. A treatment for (Pólya-)Boltzmann samplers in the more general context of weighted multisort species is currently in preparation by the author of this thesis.

2.4.2.1 The labelled case

Sums

Suppose that $\mathcal{F} = \sum_{i=1}^{\infty} \mathcal{F}_i$. Then the following procedure is Boltzmann sampler for $\mathcal{F}$.

1. Draw an integer $\ell \geq 1$ with probability
$$\mathbb{P}(\ell = i) = \mathcal{F}_i(x)/\mathcal{F}(x).$$
2. Return $\Gamma\mathcal{F}_i(x)$. That is, the result is a random Boltzmann-distributed $\mathcal{F}_i$-object.

Products

Suppose that $\mathcal{F} = \prod_{i=1}^{k} \mathcal{F}_i$. Then the following procedure is a Boltzmann sampler for $\mathcal{F}$.

1. For each $1 \leq i \leq k$ let
$$F_i \leftarrow \Gamma\mathcal{F}_i(x).$$
That is, let F_i, $1 \leq i \leq k$ be independent random variables such that F_i follows a Boltzmann-distribution for $\mathcal{F}_i$ with parameter x.
2. Let U denote the exterior disjoint union of the label-sets of the F_i. Hence, by a slight abuse of notation,
$$(F_1, \ldots, F_k) \in \mathcal{F}[U].$$
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 object $\nu.F$.

Substitution

Suppose that $\mathcal{F} = \mathcal{G} \circ \mathcal{H}$ with $\mathcal{H}[\emptyset] = \emptyset$. The following procedure is a Boltzmann sampler for $\mathcal{F}$.

1. Set
$$G \leftarrow \Gamma\mathcal{G}(y) \quad \text{with } y = \mathcal{H}(x).$$
That is, let G denote a random $\mathcal{G}$-object that follows a Boltzmann distribution with parameter $\mathcal{H}(x)$.

2. Let V denote the label set of G. For each atom $i \in V$ set
$$H_i \leftarrow \Gamma\mathcal{H}(x).$$
and let U_i denote the label set of H_i. Let U denote the exterior disjoint union of the label sets U_i. Hence, by a slight abuse of notation,
$$\pi := \{U_i \mid i \in V\}$$
is a partition of U. Setting
$$\sigma : V \to \pi,\ i \mapsto U_i$$
and $H_{U_i} := H_i$ for each $i \in V$ we have that
$$(\sigma.G, (H_Q)_{Q\in\pi}) \in (\mathcal{G} \circ \mathcal{H})[U].$$

3. 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 object
$$\nu.(\sigma.G, (H_Q)_{Q\in\pi}).$$

2.4.2.2 Pólya-Boltzmann samplers

Sums

Suppose that $\mathcal{F} = \sum_{i=1}^{\infty} \mathcal{F}_i$. Then the following procedure is a Pólya-Boltzmann sampler for $\mathcal{F}$.

1. Draw an integer $\ell \geq 1$ with probability
$$\mathbb{P}(\ell = i) = Z_{\mathcal{F}_i}(s_1, s_2, \ldots)/Z_{\mathcal{F}}(s_1, s_2, \ldots).$$

2. Return $\Gamma Z_{\mathcal{F}_\ell}(s_1, s_2, \ldots)$. That is, the result is a random $\mathcal{F}_\ell$-symmetry following a Pólya-Boltzmann distribution with the parameters $(s_1, s_2, \ldots)$.

Products

Suppose that $\mathcal{F} = \prod_{i=1}^{k} \mathcal{F}_i$. Then for any finite set U there is a bijection between the set $\mathrm{Sym}(\mathcal{F})[U]$ and tuples $(S_1, \ldots, S_k)$ such that S_i is a $\mathcal{F}_i$-symmetry for all i and the label sets of the S_i partition the set U. This is due to the fact, that given a $\mathcal{F}$-symmetry $((F_1, \ldots, F_k), \sigma) \in \mathrm{Sym}(\mathcal{F})[U]$ the permutation σ must leave the label set Q_i of the $\mathcal{F}_i$-object F_i invariant and satisfy $\sigma|_{Q_i}.F_i = F_i$, i.e. $(F_i, \sigma|_{Q_i}) \in \mathrm{Sym}(\mathcal{F}_i)[Q_i]$.

The following procedure is a Pólya-Boltzmann sampler for $\mathcal{F}$.

1. For each $1 \le i \le k$ set

$$(F_i, \sigma_i) \leftarrow \Gamma Z_{\mathcal{F}_i}(s_1, s_2, \ldots).$$

   That is, let $S_i := (F_i, \sigma_i)$, $1 \le i \le k$ be independent random variables such that S_i follows a Pólya-Boltzmann distribution for $\mathcal{F}_i$ with parameters $s_1, s_2, \ldots$.

2. By the bijection for the symmetries of products, the tupel $(S_1, \ldots, S_k)$ corresponds to an $\mathcal{F}$-symmetry (F, σ) over the (exterior) disjoint union U of the label-sets of the S_i. 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 symmetry

$$\nu.(F, \sigma) = (\nu.F, \nu\sigma\nu^{-1}).$$

Substitution

Suppose that $\mathcal{F} = \mathcal{G} \circ \mathcal{H}$ with $\mathcal{H}[\emptyset] = \emptyset$. The symmetries of the substitution were discussed in detail in Section 2.2.4. 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 a $\mathcal{F}$-symmetry (F, γ) out of (G, σ) and the copies of the (H_τ, σ_τ) as described in Section 2.2.4.

4. Choose bijection ν from the vertex set of (F, γ) to an appropriate sized set of integers $[n]$ and return the relabelled symmetry

$$\nu.(F, \gamma) = (\nu.F, \nu\gamma\nu^{-1}).$$

The Set construction

The following procedure is a Pólya-Boltzmann sampler for $\mathcal{F} = \mathrm{SET}$.

1. Let $(m_i)_{i \in \mathbb{N}}$ be an independent family of integers $m_i \geq 0$ such that m_i follows a Poisson-distribution with parameter s_i/i.

2. The sequence drawn in the previous step belongs almost surely to $\mathbb{N}_0^{\mathbb{N}}$. Let σ be a permutation with cycle type $(m_i)_i$.

3. Make a uniformly at random choice of a bijection ν from the label set of σ to an appropriate sized set of integers $[n]$ and return the SET-symmetry

$$(F, \nu\sigma\nu^{-1})$$

with $F = [n]$ the unique element from $\mathrm{SET}[n] = \{[n]\}$.

2.4.2.3 Pólya-Boltzmann samplers for cycle-pointed species

In the following, we suppose that $\mathcal{F}$ is a cycle pointed species.

Sums

Suppose that $\mathcal{F} = \sum_{i=1}^{\infty} \mathcal{F}_i$ with cycle-pointed species $\mathcal{F}_i$. Then the following procedure is a Pólya-Boltzmann sampler for $\mathcal{F}$.

1. Draw an integer $\ell \geq 1$ with probability

$$\mathbb{P}(\ell = i) = \bar{Z}_{\mathcal{F}_i}(s_1, t_1; s_2, t_2; \ldots)/Z_{\mathcal{F}}(s_1, t_1; s_2, t_2; \ldots).$$

2. Return $\Gamma\bar{Z}_{\mathcal{F}_\ell}(s_1, t_1; s_2, t_2; \ldots)$.

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 c-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 c-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 c-symmetry $\nu.S$.

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 2.2.4. The following procedure is a Pólya-Boltzmann sampler for $\mathcal{F}$.

1. Set
$$((G, \tau_0), \sigma, v) \leftarrow \Gamma \bar{Z}_{\mathcal{G}}(h_1, \bar{h}_1; h_2, \bar{h}_2; \ldots)$$
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 $\tau \neq \tau_0$ 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. For each cycle τ of σ (including the marked cycle τ_0), make $|\tau|$ identical copies copies of (H_τ, σ_τ), one for each atom of τ. Assemble a $\mathcal{F}$-symmetry (F, γ) out of (G, σ) and the copies of the (H_τ, σ_τ) as described in Section 2.2.4. Let c denote the cycle that gets composed out of the $|\tau_0|$ copies of the cycle c_{τ_0}. 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 c-symmetry of $\mathcal{F}$.

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

Cycle pointed Set constructions

The following procedure is a Pólya-Boltzmann sampler for $\mathcal{F} = \mathrm{SET}^\circ$.

1. Choose an integer $K \geq 1$ with distribution
$$\mathbb{P}(K = k) = t_k / \sum_{i=1}^{\infty} t_i.$$
2. Set
$$(G, \sigma) \leftarrow \Gamma Z_{\mathrm{SET}}(s_1, s_2, \ldots).$$
3. Add a disjoint cycle of length K to the permutation σ. Mark one of the atoms of this cycle uniformly at random.
4. Relabel the resulting rooted c-symmetry uniformly at random.

The sampler for the symmetrically cycle pointed species $\mathrm{SET}^{\circledast}$ is identical, only step 1. needs to be replaced with:

1'. Choose an integer $K \geq 2$ with distribution
$$\mathbb{P}(K = k) = t_k / \sum_{i=2}^{\infty} t_i.$$

2.4.3 Recursive Boltzmann samplers

The rules for the construction of (Pólya-)Boltzmann samplers may applied recursively in order to obtain a recursive procedure that is guaranteed to terminate almost surely. We are going to make this precise, following closely [BFKV11, Ch. 2.5].

Definition 2.4.2 ([BFKV11, Def. 7]). *A (standard) recursive specification with formal variables $x_1, \ldots, x_k$ over species $\mathcal{G}_1, \ldots, \mathcal{G}_\ell$ is a system ψ of equations*
$$x_1 = e_1, \ldots, x_k = e_k$$
where each e_i is of the form

- *$a + b$ or $a \cdot b$ with $a, b \in \{x_1, \ldots, x_k, \mathcal{G}_1, \ldots, \mathcal{G}_\ell\}$, or*
- *$a \circ b$ with $a \in \{\mathcal{G}_1, \ldots, \mathcal{G}_\ell\}$ and $b \in \{x_1, \ldots, x_k, \mathcal{G}_1, \ldots, \mathcal{G}_\ell\}$.*

We would like to recursively define a vector of species $(\mathcal{F}_1^{(n)}, \ldots, \mathcal{F}_k^{(n)})$ by starting with $\mathcal{F}_i^{(0)} = 0$ for each i and letting $\mathcal{F}_i^{(n+1)}$ be the result of replacing each occurence of a formal variable x_j, $1 \leq j \leq k$ in the expression e_i by $\mathcal{F}_j^{(n)}$. This is only possible if by doing so we never form a composition of species $\mathcal{A} \circ \mathcal{B}$ with $\mathcal{B}[\emptyset] \neq \emptyset$. If this never happens, then the specification is termed *well-founded.*

Given species $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and $\mathcal{D}$ with $\mathcal{A} \subset \mathcal{C}$ and $\mathcal{B} \subset \mathcal{D}$ we have that $\mathcal{A}\mu\mathcal{B} \subset \mathcal{C}\mu\mathcal{D}$ for any operator $\mu \in \{+, \cdot, \circ\}$. If ψ is well-founded, then it follows that for each finite set U and all $1 \leq i \leq k$

$$\mathcal{F}_i^{(0)}[U] \subset \mathcal{F}_i^{(1)}[U] \subset \mathcal{F}_i^{(2)}[U] \subset \dots .$$

If for each i and every finite set U this sequence stabilizes, i.e. if there is a number $N(U, i)$ such that $\mathcal{F}_i^{(n)}[U] = \mathcal{F}_i^{(N)}[U]$ for each $n \geq N(U, i)$, then we may define species $\mathcal{F}_1, \dots, \mathcal{F}_k$ by $\mathcal{F}_i[U] = \mathcal{F}_i^{(N(U,i))}[U]$ and $\mathcal{F}_i[\sigma] = \mathcal{F}_i^{(N(U,i))}[\sigma]$ for each bijection $\sigma : U \to V$. We say that each of the species $\mathcal{F}_1, \dots, \mathcal{F}_k$ is *decomposable* over the species $\mathcal{G}_1, \dots, \mathcal{G}_\ell$.

Theorem 2.4.3 ([BFKV11, Thm. 40])**.** *Suppose that the species $\mathcal{F}$ is decomposable over the species $\mathcal{G}_1, \dots, \mathcal{G}_\ell$. Then we may obtain a (recursive) Pólya-Boltzmann sampler $\Gamma Z_{\mathcal{F}}$ in terms of samplers $\Gamma Z_{\mathcal{G}_1}, \dots, \Gamma Z_{\mathcal{G}_\ell}$ from any corresponding well-founded system by applying the random generation rules from Section 2.4.2.*

Combining this with Lemma 2.4.1, we may thus build recursive Boltzmann samplers in the labelled and unlabelled setting by applying the corresponding construction rules to recursive specifications.

2.4.3.1 Examples

As an example, we are going to demonstrate this for the species $\mathcal{A}$ of rooted trees. It satisfies the combinatorial specification

$$\mathcal{A} \simeq \mathcal{X} \cdot \mathrm{SET}(\mathcal{A}). \qquad (*)$$

Setting $\mathcal{F}^{(0)} = 0$ and $\mathcal{F}^{(n+1)} = \mathcal{X} \cdot \mathrm{SET}(\mathcal{F}^{(n)})$, we have that for each finite set U the sequence

$$\mathcal{F}^{(0)}[U] \subset \mathcal{F}^{(1)}[U] \subset \mathcal{F}^{(2)}[U] \subset \dots .$$

stabilizes. This may be checked directly, but also follows from Remark 2.2.3, as $(*)$ satisfies the requirements of Theorem 2.2.2.

By Theorem 2.4.3 we may apply the rules for the construction of Pólya-Boltzmann samplers, obtaining the following Boltzmann sampler $\Gamma\tilde{\mathcal{A}}(x)$ for unlabelled rooted trees.

1. Start with a single root vertex v.
2. Let $(m_i)_{i \in \mathbb{N}}$ be an independent family of integers $m_i \geq 0$ such that m_i follows a Poisson distribution with parameter $\tilde{\mathcal{A}}(x^i)/i$.
3. For each $i \in \mathbb{N}$ and each $1 \leq j \leq m_i$ set

   $$A_{i,j} \leftarrow \Gamma\mathcal{A}(x^i).$$

   Make i identical copies $A_{i,j,k}$, $1 \leq k \leq i$ and attach them to the root-vertex v by adding an edge between v and the root of $A_{i,j,k}$ for each k.

A Boltzmann-sampler for $\Gamma\tilde{\mathcal{A}}(x)$ is also described in [BFKV11, Fig. 14, (1)]. However, the procedure given there seems to contain a typo, since it corresponds to attaching only copies of $A_{i,1}$ in step 3. (The paper is otherwise carefully written and contains amazing results, the author of this thesis would like to add.)

2.5 Deviation inequalities

We are going to make frequent use of the following deviation inequality for one-dimensional random walk, found in most textbooks on the subjects.

Lemma 2.5.1. *Let $(X_i)_{i\in\mathbb{N}}$ be an i.i.d. family of real-valued random variables with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[e^{\lambda X_1}] < \infty$ for all λ in some interval around zero. Then there are constants $\delta, c > 0$ such that for all $n \in \mathbb{N}$, $x \geq 0$ and $0 \leq \lambda \leq \delta$ it holds that*

$$\mathbb{P}(X_1 + \ldots + X_n \geq x) \leq \exp(cn\lambda^2 - \lambda x).$$

Proof. Let X denote a random variable that has the same distribution as all the $X_i's$. Since $\mathbb{E}[e^{\lambda X}] < \infty$ for $|\lambda| \leq \delta$, $\delta > 0$, we have that

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\mathbb{E}[|X|^k] = \mathbb{E}[e^{\lambda|X|}] \leq \mathbb{E}[e^{\lambda X}] + \mathbb{E}[e^{-\lambda X}] < \infty.$$

Since $\mathbb{E}[X] = 0$, it follows that there is a constant $c > 0$ such that

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}[e^{\lambda X} - \lambda X] \leq 1 + \sum_{k=2}^{\infty} \frac{|\lambda|^k}{k!}\mathbb{E}[|X|^k] \leq 1 + \lambda^2 c.$$

Applying Markov's inequality we obtain for $0 \leq \lambda \leq \delta$ and $x \geq 0$

$$\begin{aligned}\mathbb{P}(X_1 + \ldots + X_n \geq x) &\leq \mathbb{P}(\exp(\lambda(X_1 + \ldots + X_n)) \geq \exp(\lambda x)) \\ &\leq \mathbb{E}[e^{\lambda X}]^n \exp(-\lambda x) \\ &\leq (1 + c\lambda^2)^n \exp(-\lambda x)\end{aligned}$$

Using

$$\log(1 + c\lambda^2)^n = n\log(1 + c\lambda^2) \leq nc\lambda^2$$

for λ small enough (depending only on c), it follows that

$$\mathbb{P}(X_1 + \ldots + X_n \geq x) \leq \exp(nc\lambda^2 - \lambda x).$$

□

Chapter 3

The CRT is the scaling limit of unlabelled unrooted trees

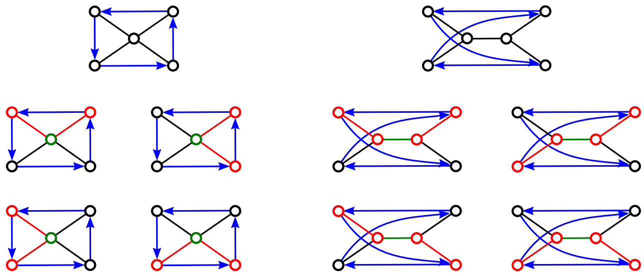

Figure 3.1: Two unlabelled cycle-pointed trees. The marked cycle is depicted in blue, connecting paths in red, and the cycle-pointing centers in green.

3.1 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, see Proposition 1.2.3. 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 the preliminaries section, 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 $n-1$.

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 [BFKV11, Claim 22] they share the same middle, called the *center of symmetry*. See Figure 3.1 for an illustration.

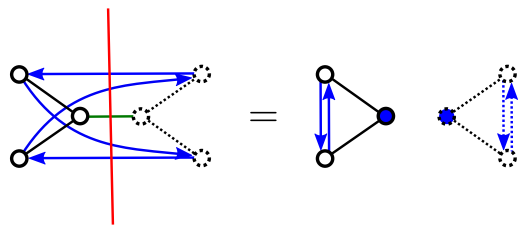

Figure 3.2: Any unlabelled $\mathcal{E} = \mathrm{SET}^{\circledast}_{\{2\}} \circledcirc \mathcal{A}_{\Omega^*}$ object corresponds to two identical copies of a cycle-pointed Pólya tree.

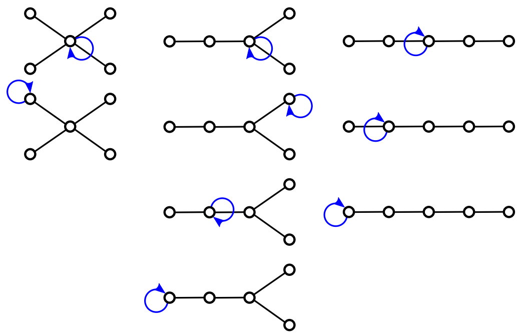

Figure 3.3: Unlabelled $\mathcal{S} = \mathcal{X}^\circ \star (\mathrm{SET}_\Omega \circ \mathcal{A}_{\Omega^*})$ objects correspond to Pólya trees.

The cycle pointing decomposition given in [BFKV11, Prop. 25] splits the species $\mathcal{F}_\Omega^\circ$ into three parts,

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

Here

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

corresponds to the trees with a marked fixpoint (compare with Figure 3.3) and the other summands to trees with a marked cycle of length at least two. More specifically,

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

corresponds to the symmetric cycle pointed trees whose center of symmetry is an edge (see Figure 3.2) and

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

to those whose center of symmetry is a vertex (compare with Figure 3.4). We are going to use this decomposition in order to show convergence of a rescaled uniform unlabelled $\mathcal{F}_\Omega$-object towards the continuum random tree.

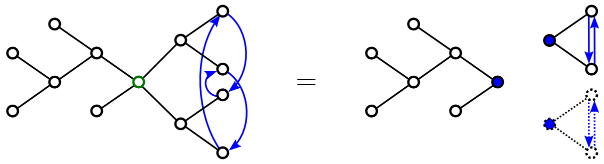

Figure 3.4: Decomposition of unlabelled $\mathcal{V} = (\mathrm{SET}_\Omega^\circledast \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}$ objects into a Pólya tree and a number of identical copies of a cycle-pointed Pólya tree.

3.1.1 A proof of Theorem 1.2.1 and Lemma 1.2.2

Of course, Theorem 1.2.1 and Lemma 1.2.2 are special cases of Theorem 1.2.4 and Lemma 1.2.5, respectively. Hence a separate treatment is not strictly necessary. However, we may take significant shortcuts in the unrestricted case $\Omega = \mathbb{N}$, which justify a redundant treatment.

Proof of Theorem 1.2.1 . Let $c_{\mathbb{N}_0}$ denote the scaling constant for the uniform unlabelled Pólya tree, i.e.

$$\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{A}_n \xrightarrow{(d)} \mathcal{T}_\mathsf{e}$$

with respect to the Gromov-Hausdorff metric. Let $f : \mathbb{K} \to \mathbb{R}$ be a bounded Lipschitz-continous function defined on the space of compact metric spaces equipped with the Gromov-Hausdorff metric. We are going to show the following three claims:

i) $\mathbb{P}((\mathsf{T}_n, \tau_n) \in \mathcal{E})$ converges to 0.

ii) $\mathbb{E}[f(\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{T}_n) \mid (\mathsf{T}_n, \tau_n) \in \mathcal{S}]$ converges to $\mathbb{E}[f(\mathcal{T}_\mathsf{e})]$.

iii) $\mathbb{E}[f(\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{T}_n) \mid (\mathsf{T}_n, \tau_n) \in \mathcal{V}]$ converges to $\mathbb{E}[f(\mathcal{T}_\mathsf{e})]$.

This implies that

$$\mathbb{E}[f(\frac{c_{\mathbb{N}}}{\sqrt{n}}\mathsf{T}_n)] \to \mathbb{E}[f(\mathcal{T}_\mathsf{e})]$$

and we are done. Claim i) follows from the fact that

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

and by Propositions 3.1.6 and 3.1.6 the radius of convergence of the series $\tilde{\mathcal{E}}(z)$ is strictly larger than the radius of convergence of $\tilde{\mathcal{F}}^\circ(z)$. Claim ii) follows directly from the convergence

$$\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{A}_n \xrightarrow{(d)} \mathcal{T}_\mathsf{e},$$

since $\Omega = \mathbb{N}$ implies that

$$\mathcal{S} = \mathcal{X}^\circ \star (\mathrm{SET} \circ \mathcal{A}) \simeq \mathcal{A}$$

and hence (T_n, τ_n) conditioned on belonging to $\mathcal{S}$ is distributed like the uniform random Pólya tree A_n. Claim iii) follows from Lemma 3.1.1 below. □

The proof for the tail bound of the diameter uses the same decomposition:

Proof of Lemma 1.2.2. We have to show that there are constants $C, c > 0$ such that for all n and $x \geq 0$ we have that

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

We may replace C by any larger constant and c by any smaller constant, hence it suffices to consider the case $\sqrt{n} \leq x \leq n$. Clearly we have that

$$\mathbb{P}(\mathrm{D}(\mathsf{T}_n) \geq x) \leq \sum_{\mathcal{B} \in \{\mathcal{E}, \mathcal{S}, \mathcal{V}\}} \mathbb{P}((\mathsf{T}_n, \tau_n) \in \mathcal{B}) \mathbb{P}(\mathrm{D}(\mathsf{T}_n) \geq x \mid (\mathsf{T}_n, \tau_n) \in \mathcal{B})$$

By Lemma 3.1.1 there are constants $C_1, c_1 > 0$ such that the summand for $\mathcal{B} = \mathcal{V}$ is bounded by $C_1 \exp(-c_1 x^2/n)$. The tree T_n conditioned on $(\mathsf{T}_n, \tau_n) \in \mathcal{S}$ is distributed like the uniform Pólya tree A_n. Hence by Lemma 1.2.6 there are constants $C_2, c_2 > 0$ such that the summand for $\mathcal{B} = \mathcal{S}$ is bounded by $C_2 \exp(-c_2 x^2/n)$. It follows from Propositions 3.1.6 and 3.1.6 and the expression

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

that there are constants $C_3 > 0$ and $0 < \gamma < 1$ with $\mathbb{P}((\mathsf{T}_n, \tau_n) \in \mathcal{E}) \leq C_3 \gamma^n$. Since $x \leq n$ we have that

$$\gamma^n \leq \exp(-c_3 x^2/n)$$

for some $c_3 > 0$. Hence the summand for $\mathcal{B} = \mathcal{E}$ is bounded by $C_3 \exp(-c_3 x^2/n)$. Thus

$$\mathbb{P}(\mathrm{D}(\mathsf{T}_n) \geq x) \leq \sum_{i=1}^{3} C_i \exp(-c_i x^2/n) \leq C \exp(-c x^2/n)$$

for some $C, c > 0$. □

It remains to show the following lemma which was used in both proofs.

Lemma 3.1.1. *Let V_n be a uniformly at random chosen unlabelled*

$$\mathcal{V} = (\mathrm{SET}^{\circledast} \circledcirc \mathcal{A}) \star \mathcal{X}$$

object with size n. Then

$$\frac{c_{\mathbb{N}_0}}{\sqrt{n}} \mathsf{V}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}$$

with respect to the Gromov-Hausdorff metric. Moreover, there are constants $C, c > 0$ such that for all n we have the following tail bound for the diameter

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

for all $x \geq 0$ and n.

Proof. We are first going to prove convergence towards the CRT. Let ρ denote the radius of convergence of $\tilde{\mathcal{F}}(z)$. By the rules for Pólya-Boltzmann samplers in Section refsec:pobosa, the following procedure draws a random Boltzmann distributed unlabelled $\mathcal{V}$-object with parameter ρ, i.e. each object with size k gets drawn with probability $\rho^k/\tilde{\mathcal{V}}(\rho)$. Compare with Figure 3.5.

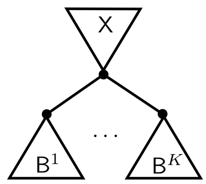

Figure 3.5: The Boltzmann distributed $(\mathrm{SET}^{\circledast} \circledcirc \mathcal{A}) \star \mathcal{X}$ object.

1. Draw a random unlabelled rooted tree X from $\mathcal{A}$ according to the Boltzmann distribution with parameter ρ.

2. Choose a random integer $K \geq 2$ with distribution given by

$$\mathbb{P}(K = k) = \tilde{\mathcal{A}}^{\circ}(\rho^k) / \sum_{i=2}^{\infty} \tilde{\mathcal{A}}^{\circ}(\rho^i).$$

3. Select a random Boltzmann distributed cycle-pointed rooted tree (B, ν) from the unlabelled $\mathcal{A}^{\circ}$-objects with parameter ρ^K.

4. Connect the root of X with the roots of K identical copies $(\mathsf{B}^1, \nu_1), \ldots, (\mathsf{B}^K, \nu_K)$ of (B, ν) by adding edges.

5. Compose the marked cycle τ out of atoms of the cycles $\nu_i = (a_i^1, \ldots, a_i^K)$ as follows (compare with Figure 3.4):

$$\tau = (a_1^1, \ldots, a_1^K, a_2^1, \ldots, a_2^K, \ldots, a_K^1, \ldots, a_K^K).$$

Let V denote the resulting cycle-pointed tree. By definition of the Boltzmann distribution we have that V conditioned on having size n is distributed like the uniform unlabelled $\mathcal{V}$-object V_n. The probability generating function of the total size of the K identical copies of B is given by

$$(\sum_{k \geq 2} \tilde{\mathcal{A}}^{\circ}((\rho z)^k)) / \sum_{i \geq 2} \tilde{\mathcal{A}}^{\circ}(\rho^i).$$

We have that $\rho < 1$ by Proposition 3.1.4, hence this series has radius of convergence strictly greater than 1. By Proposition 3.1.5 we know that $\mathbb{P}(|\mathsf{V}| = n) \sim d_{\Omega^*} n^{-3/2}$ for some constant $d_{\Omega^*} > 0$. Hence there is some constant $C > 0$ such that

$$\mathbb{P}(K|\mathsf{B}| \geq C \log(n) \mid |\mathsf{V}| = n) = O(n^{3/2}) \mathbb{P}(K|\mathsf{B}| \geq C \log(n)) = o(1).$$

Let X_n denote the random variable X conditioned on the event $|\mathsf{V}| = n$. Consider the correspondence $\mathcal{R}_n$ between the discrete metric spaces X_n and V_n given by

$$\mathcal{R}_n = \{(x, x) \mid x \in \mathsf{X}_n\} \cup (\{x_0\} \times (\mathsf{B}^1 \cup \ldots \cup \mathsf{B}^K)$$

with x_0 denoting the root of X_n. Then we have

$$\mathrm{dis}(\mathcal{R}_n) = O(\log(n))$$

with high probability. This implies that

$$d_{\mathrm{GH}}(\mathsf{X}_n/\sqrt{n}, \mathsf{V}_n/\sqrt{n}) \xrightarrow{p} 0.$$

Hence it suffices to show that

$$\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{X}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}.$$

For any positive integer ℓ we have that X_n conditioned on the event $|\mathsf{X}_n| = \ell$ is distributed like the uniform random unlabelled rooted tree A_ℓ with ℓ vertices. Hence for any bounded Lipschitz-continuous function $f : \mathbb{K} \to \mathbb{R}$ defined on the metric space $(\mathbb{K}, d_{\mathrm{GH}})$ of isometry classes of compact metric spaces we have that

$$\mathbb{E}[f(\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{X}_n)] = o(1) + \sum_{n - C\log(n) \le \ell \le n} \mathbb{E}[f(\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{A}_\ell)]\,\mathbb{P}(|\mathsf{X}_n| = \ell)$$

Moreover, the average value of the diameter $\mathrm{D}(\mathsf{A}_\ell)$ is known to satisfy

$$\mathbb{E}[\mathrm{D}(\mathsf{A}_\ell)] = O(\sqrt{\ell}),$$

see e.g. Lemma 1.2.6 below or [DG10, Thm. 2]. Hence

$$\mathbb{E}[d_{\mathrm{GH}}(\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{A}_\ell, \frac{c_{\mathbb{N}_0}}{\sqrt{\ell}}\mathsf{A}_\ell)] \le c_{\mathbb{N}_0}(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{\ell}})\mathbb{E}[\mathrm{D}(\mathsf{A}_\ell)] = o(1)$$

uniformly for all $n - C\log(n) \le \ell \le n$. Since

$$\mathbb{E}[f(\frac{c_{\mathbb{N}_0}}{\sqrt{\ell}}\mathsf{A}_\ell)] \to \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})]$$

it follows that

$$\mathbb{E}[f(\frac{c_{\mathbb{N}_0}}{\sqrt{n}}\mathsf{X}_n)] \to \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})].$$

This proves convergence towards the CRT.

It remains to show the tail bounds for the diameter of V_n. Let H denote the maximum length of a path in V that starts from the root of X and let H_n denote the corresponding random variable conditioned on the event $|\mathsf{V}| = n$. Since

$$\mathrm{D}(\mathsf{V}_n) \le 2\mathsf{H}_n$$

it suffices to show that there are constants $C, c > 0$ with

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

for all $x \ge 0$ and n. Since we may substitute C by any larger constants and c by any smaller constant it suffices to show this for the case $\sqrt{n} \le x \le n$. The event $\mathsf{H}_n \ge x$

implies that $\mathrm{H}(\mathsf{X}_n) \geq x$ or $|\mathsf{B}| \geq x$. Since X_n conditioned on the event $|\mathsf{X}_n| = \ell$ is distributed like the uniform Pólya tree A_ℓ, it follows by Lemma 1.2.6 below that there are constants $C_1, c_1 > 0$ such that for all $y \geq 0$ and n the probability $\mathbb{P}(\mathrm{H}(\mathsf{X}_n) \geq y)$ is bounded by

$$\begin{aligned}\sum_{\ell=1}^{n} \mathbb{P}(|\mathsf{X}_n| = \ell)\mathbb{P}(\mathrm{H}(\mathsf{A}_\ell) \geq y) &\leq \sum_{\ell=1}^{n} \mathbb{P}(|\mathsf{X}_n| = \ell) C_1 \exp(-c_1 y^2/\ell) \\ &\leq C_1 \exp(-c_1 y^2/n).\end{aligned}$$

Moreover, by Propositions 3.1.5 and 3.1.6 we know that there are constants $C_3 > 0$ and $0 < \gamma < 1$ such that for all $y \geq 0$ and n we have that

$$\mathbb{P}(|\mathsf{B}| \geq y \mid |\mathsf{V}| = n) \leq C_3 n^{3/2} \gamma^y.$$

It follows that there are constants $C_4, c_2 > 0$ such that we have uniformly for all $x \geq \sqrt{n}$

$$\mathbb{P}(\mathsf{H}_n \geq x) \leq C_1 \exp(-c_1 y^2/n) + C_3 n^{3/2} \gamma^x \leq C_4 \exp(-c_2 x^2/n).$$

This concludes the proof. □

3.1.2 A proof of Theorem 1.2.4 and Lemma 1.2.5

We start straight-away with the proof:

Proof of Theorem 1.2.4. Let $c_{\Omega^*} > 0$ denote the constant such that the uniformly drawn unlabelled rooted tree A_{n-1} satisfies

$$\frac{c_{\Omega^*}}{\sqrt{n-1}} \mathsf{A}_{n-1} \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}$$

with respect to the Hausdorff-Gromov metric.

The proof of Theorem 1.2.4 follows closely the proof of Theorem 1.2.1 in Section 3.1.2. The only difference lies in how we show convergence for the unlabelled $\mathcal{S} = \mathcal{X}^\circ \star (\mathrm{SET}_\Omega \circ \mathcal{A}_{\Omega^*})$ objects and the unlabelled $\mathcal{V} = (\mathrm{SET}_\Omega^\circledast \circledcirc \mathcal{A}_{\Omega^*}) \star \mathcal{X}$ objects. We treat these cases separately in Lemma 3.1.2 and Lemma 3.1.3 below. □

Proof of Lemma 1.2.5. The proof is analogous to the proof of Lemma 1.2.2. The only difference lies in how we show the tail bounds for the unlabelled $\mathcal{V}$-objects and unlabelled $\mathcal{S}$-objects. This is carried in out in Lemmas 3.1.2 and 3.1.3 below. □

Lemma 3.1.2. *Let S_n be drawn uniformly from the unlabelled*

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

objects of size n. Then we have

$$\frac{c_{\Omega^*}}{\sqrt{n}} \mathsf{S}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}.$$

with respect to the Gromov-Hausdorff metric. Moreover, there are constants $C, c > 0$ such that for all n and $x \ge 0$ we it holds that

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

Proof. We have that

$$\mathcal{S} \simeq \mathcal{X} \cdot (\mathrm{SET}_\Omega \circ \mathcal{A}_{\Omega^*})$$

, hence we do not require cycle pointing techniques in this case. Let (S_n, σ_n) be drawn uniformly at random from the set $\mathrm{Sym}(\mathcal{S})[n]$. Let π_n denote the corresponding partition. By the discussion in Section 2.2.4, σ_n induces an automorphism

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

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

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) For any $\delta > 0$ the maximum size $\max_{Q \in F_n} |\mathsf{A}_Q|$ of the trees corresponding to the fixpoints of $\bar{\sigma}_n$ satisfies

$$\mathbb{P}(\max_{Q \in F_n} |\mathsf{A}_Q| \le n - n^\delta) = o(1).$$

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

$$\mathbb{E}[f_n] \le C_2$$

for all n.

We may deduce the tail bound for the diameter as follows. First, it suffices to show such a bound for all $\sqrt{n} \le x \le n$. If $\mathrm{D}(\mathsf{S}_n) \ge x$, then we have $H_n \ge x/2$ or $\max_{Q \in F_n} \mathrm{H}(\mathsf{A}_Q) \ge x/2 - 1$. By 1), we have

$$\mathbb{P}(H_n \ge x/2) \le C_1 n^{3/2} \gamma^{x/2}$$

and there are constants $C_4, c_4 > 0$ such that

$$C_1 n^{3/2} \gamma^{x/2} \le C_4 \exp(-c_4 x^2/n)$$

for all n and $\sqrt{n} \le x \le n$. Let $\mathfrak{E}_n$ denote the event $\max_Q \mathsf{H}(\mathsf{A}_Q) \ge x/2 - 1$. It holds that

$$\mathbb{P}(\mathfrak{E}_n) \le \sum_F \mathbb{P}(F_n = F)\mathbb{P}(\mathfrak{E}_n \mid F_n = F).$$

with F ranging over all subsets of partitions of $[n]$ with $\mathbb{P}(F_n = F) > 0$. By the discussion of symmetries in Section 2.2.4 we have that given $F_n = F$, 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, A_Q gets drawn uniformly at random from all unlabelled Pólya trees with outdegrees in the set Ω^*. By Lemma 1.2.6 it follows that there are positive constants C_5, c_5 such that uniformly for all n and x

$$\mathbb{P}(\mathfrak{E}_n \mid F_n = F) \le C_5 \sum_{Q \in F} \exp(-c_4 x^2/|Q|) \le |F| C_4 \exp(-c_5 x^2/n).$$

It follows that

$$\mathbb{P}(\mathcal{E}_n) \le C_5 \exp(-c_5 x^2/n) \sum_F \mathbb{P}(F_n = F)|F| \le \mathbb{E}[f_n] C_5 \exp(-c_5 x^2/n).$$

By 3) we have that

$$\mathbb{E}[f_n] \le C_2$$

for all n. Thus, for some $C_6, c_6 > 0$, it holds that

$$\mathbb{P}(\mathrm{D}(\mathsf{S}_n) \ge x) \le C_4 \exp(-c_4 x^2/n) + C_2 C_5 \exp(-c_5 x^2/n) \le C_6 \exp(-c_6 x^2/n)$$

uniformly for all n and $\sqrt{n} \le x \le n$. Thus the claims 1) and 3) imply the tail bound for the diameter.

We may deduce the convergence towards the CRT as follows. Select one of the partition classes from F_n with maximal size uniformly at random and let X_n denote the corresponding tree. By claim 2) we have

$$\mathbb{P}(|\mathsf{X}_n| \le n - n^{1/4}) = o(1)$$

and thus

$$\mathbb{P}(d_{\mathrm{GH}}(\mathsf{X}_n, \mathsf{S}_n) \ge n^{1/4}) = o(1).$$

It follows that

$$d_{\mathrm{GH}}(c_{\Omega^*}\mathsf{S}_n/\sqrt{n}, c_{\Omega^*}\mathsf{X}_n/\sqrt{n}) \xrightarrow{p} 0.$$

Hence it suffices to show

$$c_{\Omega^*}\mathsf{X}_n/\sqrt{n} \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}.$$

Let $f : \mathbb{K} \to \mathbb{R}$ denote a bounded Lipschitz-continuos function defined on the space $(\mathbb{K}, d_{\mathrm{GH}})$ of isometry classes of compact metric spaces equipped with the Gromov-Hausdorff metric. By claim 2) it follows that

$$\mathbb{E}[f(\frac{c_{\Omega^*}}{\sqrt{n}}\mathsf{X}_n)] = o(1) + \sum_\ell \mathbb{P}(|\mathsf{X}_n| = \ell)\mathbb{E}[f(\frac{c_{\Omega^*}}{\sqrt{n}}\mathsf{X}_n) \mid |X_n| = \ell].$$

with the index of the sum ranging over all integers $n - n^{1/4} \leq \ell \leq n$ satisfying $\mathbb{P}(|\mathsf{X}_n| = \ell) > 0$, in particular $\ell \equiv 1 \mod \gcd(\Omega^*)$. Since $\ell > n/2$ we have by the discussion of the structure of symmetries in Section 2.2.4 that X_n conditioned $|\mathsf{X}_n| = \ell$ is distributed like a uniformly drawn Pólya tree A_ℓ of size ℓ with outdegrees in Ω^*. Hence

$$\mathbb{E}[f(\frac{c_{\Omega^*}}{\sqrt{n}}\mathsf{X}_n) \mid |X_n| = \ell] = \mathbb{E}[f(\frac{c_{\Omega^*}}{\sqrt{n}}\mathsf{A}_\ell)] = \mathbb{E}[f(\frac{c_{\Omega^*}}{\sqrt{\ell}}\mathsf{A}_\ell)] + R_\ell$$

with

$$|R_\ell| \leq C|\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{\ell}}|\mathbb{E}[\mathsf{D}(\mathsf{A}_\ell)]$$

for a fixed constant $C > 0$ that does not depend on ℓ. We have by Lemma 1.2.6 that

$$\mathbb{E}[\mathsf{D}(\mathsf{A}_\ell)] = O(\sqrt{\ell}),$$

hence

$$\sum_\ell R_\ell = o(1).$$

By assumption,

$$\mathbb{E}[f(\frac{c_{\Omega^*}}{\sqrt{\ell}}\mathsf{A}_\ell)] \to \mathbb{E}[f(\mathcal{T}_\mathsf{e})]$$

and hence it follows that

$$\mathbb{E}[c_{\Omega^*}\mathsf{X}_n/\sqrt{n}] \to \mathbb{E}[f(\mathcal{T}_\mathsf{e})].$$

Thus claim 2) implies that

$$c_{\Omega^*}\mathsf{S}_n/\sqrt{n} \xrightarrow{(d)} \mathcal{T}_\mathsf{e}.$$

It remains to verify claims 1) - 3). The probability generating function of H_n is given by

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

Since $1 \in \Omega$ we may bound the denominator from below by $[z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(\rho z)$ and by Proposition 3.1.5 we have that

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

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

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

which by Proposition 3.1.5 has radius of convergence strictly greater than 1. In particular we have that

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

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. The probability generating function for the random number f_n is given by

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

and the corresponding bound for the event $f_n \geq x$ follows by the same arguments. This proves claim 1).

We proceed with showing claim 2). Let x_n be a given sequence of positive numbers. The event

$$\max_{Q\in F_n} |\mathsf{A}_Q| \leq x_n$$

would imply that

$$n - 1 = H_n + \sum_{Q\in F_n} |A_Q| \leq H_n + x_n f_n.$$

In particular it holds that $H_n \geq (n-1)/2$ or $f_n \geq (n-1)/(2x_n)$. Thus, for

$$x_n = cn/\log(n)$$

with $c > 0$ a sufficiently small number, it follows by the tail bounds of claim 1) that

$$\mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| \leq x_n) = o(1).$$

Thus, setting

$$y_n = n - n^{2/3+\epsilon}$$

for any small $\epsilon > 0$, we have that

$$\mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| \leq y_n) = o(1) + \sum_{x_n\leq k\leq y_n} \mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| = k).$$

We can form any unlabelled $\mathcal{S}$-object by taking an ordered pair of unlabelled $\mathcal{A}_{\Omega^*}$-objects, connecting their roots by an edge, and declaring the root of the first object as the new root of the resulting tree. It follows that the number of unlabelled $\mathcal{S}$-objects with size n having the property that at least one of the subtrees dangling from the root has size k is bounded by $a_k a_{n-k}$ with $a_i = [z^i]\tilde{\mathcal{A}}_{\Omega^*}(z)$ for all i. Hence

$$\mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| = k) \leq a_k a_{n-k}/[z^n]\tilde{S}(z).$$

By Proposition 3.1.5 we know that $a_i \sim C i^{-3/2}\rho^{-i}$ as $i \equiv 1 \mod \gcd(\Omega^*)$ tends to infinity. Thus

$$\mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| \leq y_n) \leq o(1) + C' \sum_{x_n \leq k \leq y_n} (k(n-k)/n)^{-3/2}$$

for some $C' > 0$. Writing $k = n/2 + t$ we obtain $k(n-k)/n = ((n/2)^2 - t^2)/n$ and this quantity strictly decreases as $|t|$ grows. Hence we have $(k(n-k)/n)^{-3/2} \leq n^{2/3+\epsilon}(1 + o(1))$ uniformly for all $x_n \leq k \leq y_n$, and thus $\mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| \leq y_n) = o(1)$. Setting $z_n = n - n^{\frac{2}{3}(\frac{2}{3}+\epsilon)+\epsilon'}$ for a small $\epsilon' > 0$ we may repeat the same arguments to obtain

$$\begin{aligned}\mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| \leq z_n) &\leq o(1) + C' \sum_{y_n \leq k \leq z_n} (k(n-k)/n)^{-3/2} \\ &\leq o(1) + O(1)(z_n - y_n)(n^{\frac{2}{3}(\frac{2}{3}+\epsilon)+\epsilon'})^{-3/2}\end{aligned}$$

and this quantity tends to zero. We may repeat the same argument arbitrarily many times and hence obtain that for any $\delta > 0$ we have that

$$\mathbb{P}(\max_{Q\in F_n} |\mathsf{A}_Q| \leq n - n^{\delta}) = o(1).$$

This proves 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. We have that

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

Since $1 \in \Omega$ we have that the denominator is bounded from below by $[z^{n-1}]\tilde{\mathcal{A}}_{\Omega^*}(z)$. By Proposition 3.1.5 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

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

with

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

analytic on $\mathbb{C}$ and

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

having radius of convergence strictly larger than ρ since $\rho < 1$. By a singularity analysis using results from [BBY06] and [FS09, Thm. VI.5] it follows that

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

The detailed arguments are identical as in the proof of Proposition 3.1.6 below. This concludes the proof.

□

Lemma 3.1.3. *Let* V_n *be drawn uniformly from the unlabelled*

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

objects of size n*. Then we have*

$$\frac{c_{\Omega^*}}{\sqrt{n}} \mathsf{V}_n \xrightarrow{(d)} \mathcal{T}_\mathsf{e}.$$

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

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

Proof. The proof is analogous to the proof of Lemma 3.1.2, only with pointed cycle index sums replacing the role of cycle index sums. Let $(\mathsf{V}_n, \tau_n, \sigma_n, v_n)$ be a rooted c-symmetry drawn uniformly at random from the set $\mathrm{RSym}(\mathcal{S})[n]$. 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 2.2.4, σ_n induces an automorphism

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

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

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

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

and

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

2) For any $\delta > 0$ the maximum size $\max_{Q \in F_n} |\mathsf{A}_Q|$ of the trees corresponding to the fixpoints of $\bar{\sigma}_n$ satisfies

$$\mathbb{P}(\max_{Q \in F_n} |\mathsf{A}_Q| \leq n - n^\delta) = o(1).$$

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

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

for all n.

From these claims we may deduce the tail bounds for the diameter and the convergence towards the CRT in an identical manner as in the proof of Lemma 3.1.2. It remains to verify claims 1)-3). We start with claim 1). The probablity generating function of H_n is given by

$$\mathbb{E}[w^{H_n}] = \frac{[z^{n-1}]\bar{Z}_{\mathrm{SET}_{\Omega}^{\circledast}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho w z)^2), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}((\rho w z)^2); \ldots)}{[z^{n-1}]\bar{Z}_{\mathrm{SET}_{\Omega}^{\circledast}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho z), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}(\rho z); \tilde{\mathcal{A}}_{\Omega^*}((\rho z)^2), \tilde{\mathcal{A}}^{\circ}_{\Omega^*}((\rho z)^2); \ldots)}.$$

Since $1 \in \Omega$ and there is a number $k \geq 3$ with $k \in \Omega$ it follows that the denominator is bounded from below by

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

We have that

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

and thus, by Proposition 3.1.5, we have 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 with indeterminate w is bounded coefficient wise by the series

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

which does not depend on n and, by Proposition 3.1.6, has radius of convergence strictly larger than 1. It follows that 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. The probability generating function for the random number number f_n is given by

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

and the corresponding bound for the event $f_n \geq x$ follows by the same arguments. This proves claim 1).

We proceed with showing claim 2). Let x_n be a given sequence of positive numbers. The event

$$\max_{Q \in F_n} |\mathsf{A}_Q| \leq x_n$$

would imply that

$$n - 1 = H_n + \sum_{Q \in F_n} |A_Q| \leq H_n + x_n f_n.$$

In particular it holds that $H_n \geq (n-1)/2$ or $f_n \geq (n-1)/(2x_n)$. Thus, for

$$x_n = cn/\log(n)$$

with $c > 0$ a sufficiently small number, it follows by the tail bounds of claim 1) that

$$\mathbb{P}(\max_{Q \in F_n} |\mathsf{A}_Q| \leq x_n) = o(1).$$

Setting

$$y_n = n - n^{2/3+\epsilon}$$

for any small $\epsilon > 0$, we have that

$$\mathbb{P}(\max_{Q \in F_n} |\mathsf{A}_Q| \leq y_n) = o(1) + \sum_{x_n \leq k \leq y_n} \mathbb{P}(\max_{Q \in F_n} |\mathsf{A}_Q| = k).$$

Any unlabelled $\mathcal{V}$-object with a tree of size k dangling from the root that does not contain any vertex of the marked cycle can be formed by connecting the roots of an unlabelled $\mathcal{A}_{\Omega^*}$-object of size k and an unlabelled $\mathrm{SET}^{\circledast}_{\Omega^*} \circledcirc \mathcal{A}_{\Omega^*}$ object of size $n-k$. By a singularity analysis similiar to the proof of claim 3) in Lemma 3.1.2 we have that the number b_i of unlabelled $\mathrm{SET}^{\circledast}_{\Omega^*} \circledcirc \mathcal{A}_{\Omega^*}$-objects of size i is at most $O(i^{-3/2}\rho^{-i})$. It follows that

$$\mathbb{P}(\max_{Q \in F_n} |\mathsf{A}_Q| = k) \leq ([z^k]\tilde{\mathcal{A}}_{\Omega^*}(z))b_{n-k}/([z^n]\tilde{\mathcal{V}}(z)) = O((k(n-k)/n)^{-3/2})$$

uniformly for all $x_n \leq k \leq y_n$ and thus

$$\mathbb{P}(\max_{Q \in F_n} |\mathsf{A}_Q| \leq y_n) = o(1) + O(1) \sum_{x_n \leq k \leq y_n} (k(n-k)/n)^{-3/2}.$$

In order to finish the proof of claim 2) we may now follow precisely the same arguments as in the proof of claim 2) in Lemma 3.1.2.

Claim 3) follows by similar arguments as in the proof of claim 3) in Lemma 3.1.2. This completes the proof. □

3.1.3 A proof of Lemma 1.2.6

We have to show that there are constants $C, c > 0$ such that for all $x \geq 0$ and $m \geq 1$ with $m \equiv 1 \mod \gcd(\Omega^*)$ it holds that

$$\mathbb{P}(\mathrm{H}(\mathsf{A}_m) \geq x) \leq C \exp(-cx^2/m).$$

Proof of Lemma 1.2.6. Since we may replace C by any larger constant and c by any smaller constant, it suffices to pick a fixed constant M and show the claim for all $m \geq M$ and $\sqrt{m} \leq x \leq m$. By the rules governing Pólya-Boltzmann samplers in Sections 2.4.2.2 and 2.4.3 the following recursive procedure $\Gamma\tilde{\mathcal{A}}_{\Omega^*}(x)$ terminates almost surely and draws a random unlabelled $\mathcal{A}_{\Omega^*}$-object according to the Boltzmann distribution with parameters x for any $0 < x \leq \rho$.

1. Start with a root vertex v.

2. Draw a random permutation $\sigma(v)$ with size $|\sigma(v)|$ in the set Ω^* such that $\sigma(v)$ gets drawn with probability proportional to its weight

$$\frac{1}{|\sigma(v)|!}\tilde{\mathcal{A}}_{\Omega^*}(x)^{\sigma_1(v)}\tilde{\mathcal{A}}_{\Omega^*}(x^2)^{\sigma_2(v)}\cdots.$$

   Here $\sigma_i(v)$ denotes the number of i-cycles of the permutation $\sigma(v)$.

3. For each i draw $\sigma_i(v)$ independent copies $\mathsf{A}_1^i(v), \ldots, \mathsf{A}_{\sigma_i(v)}^i(v)$ of the recursively called sampler $\Gamma\tilde{\mathcal{A}}(x^i)$ and for each $1 \le j \le \sigma_i(v)$ attach the roots of i identical copies of $\mathsf{A}_j^i(v)$ to the root vertex v by adding edges.

Let A be a random tree drawn according to $\Gamma\tilde{\mathcal{A}}_{\Omega*}(\rho)$ and consider the subtree T given by the root-vertices of the trees generated by a call to the sampler with parameter ρ (as opposed to ρ^i for some $i \ge 2$). Then T is distributed like the result of drawing a Galton-Watson tree and discarding the orderings on the offspring sets, with the offspring distribution ξ given by the number of fixpoints of the random permutation drawn in step 2. The probability generating function of ξ is given by

$$\mathbb{E}[z^\xi] = Z_{\mathrm{SET}_{\Omega^*}}(z\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}_{\Omega^*}(\rho^2), \tilde{\mathcal{A}}_{\Omega^*}(\rho^3), \ldots)\rho/\tilde{\mathcal{A}}_{\Omega^*}(\rho).$$

Note that $\mathbb{E}[\xi] = 1$ and, by Proposition 3.1.5, $\mathbb{E}[z^\xi]$ has radius of convergence strictly larger than 1.

For any vertex v of T, the sum of vertices

$$S(v) := \sum_{i \ge 2}\sum_{j=1}^{\sigma_i(v)} i|\mathsf{A}_j^i(v)|$$

of the attached subtrees corresponding to cycles of lengths at least 2 has probability generating function

$$\mathbb{E}[z^{S(v)}] = Z_{\mathrm{SET}_{\Omega^*}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho), \tilde{\mathcal{A}}_{\Omega^*}((z\rho)^2), \tilde{\mathcal{A}}_{\Omega^*}((z\rho)^3), \ldots)\rho/\tilde{\mathcal{A}}_{\Omega^*}(\rho).$$

Again, by Proposition 3.1.5, this series has radius of convergence strictly larger than 1 and hence there is a constant $0 < \gamma < 1$ with

$$\mathbb{P}(S(v) \ge y) = O(\gamma^y)$$

uniformly for all $y \ge 0$.

Given $m \equiv 1 \mod \gcd(\Omega^*)$ let A_m, T_m and $(S_m(v))_{v \in \mathsf{T}_m}$ denote the random variables A, T and $(S(v))_{v \in \mathsf{T}}$ conditioned on the event $|\mathsf{A}| = m$. In particular, A_m is uniformly distributed among all Pólya trees of size m with outdegrees in the set Ω^*. If the height $\mathrm{H}(\mathsf{A}_m)$ of the tree A_m satisfies $\mathrm{H}(\mathsf{A}_m) \ge x$ then $\mathrm{H}(\mathsf{T}_m) \ge x/2$ or $S_m(v) \ge x/2$ for at least one vertex $v \in \mathsf{T}_m$. By the tail bounds for conditioned

Galton-Watson processes given in Addario-Berry, Devroye and Janson [ABDJ13] there exist constants $C_1, c_1 > 0$ such that for all ℓ and $y \geq 0$ we have that

$$\mathbb{P}(\mathrm{H}(\mathsf{T}) \geq y \mid |\mathsf{T}| = \ell) \leq C_1 \exp(-c_1 y^2/\ell).$$

Moreover, T_m conditioned on having size ℓ is distributed like T conditioned on having size ℓ. Thus the probability for the event $\mathrm{H}(\mathsf{T}_m) \geq x/2$ is bounded by

$$\sum_{\ell=1}^{m} \mathbb{P}(|\mathsf{T}_m| = \ell)\, \mathbb{P}(\mathrm{H}(\mathsf{T}) \geq x/2 \mid |\mathsf{T}| = \ell) \leq C_1 \exp(-\frac{c_1}{4} x^2/m).$$

By Proposition 3.1.5 and the definition of the Boltzmann-distribution, we have that asymptotically

$$\mathbb{P}(|\mathsf{A}| = m) \sim d_{\Omega^*} m^{-3/2}$$

for some constant d_{Ω^*}. In particular, there is a constant $C_2 > 0$ such that

$$\mathbb{P}(|\mathsf{A}| = m) \leq C_2 m^{-3/2}$$

for all m. Hence there is a constant $C_3 > 0$ such that for all x and m the probability for the event $S_m(v) \geq x/2$ for at least one vertex $v \in \mathsf{T}_m$ is bounded by

$$C_2 m^{3/2} \mathbb{P}(S(v) \geq x/2 \text{ for some } v \in \mathsf{T}, |\mathsf{A}| = m) \leq C_3 m^{5/2} \gamma^{x/2}.$$

We assumed that $\sqrt{m} \leq x \leq m$, hence

$$m^{5/2} \gamma^{x/2} \leq C_4 \exp(-c_2 x^2/m)$$

for some constants $C_4, c_2 > 0$. Thus there are constants $C_5, c_3 > 0$ such that

$$\mathbb{P}(\mathrm{H}(\mathsf{A}_m) \geq x) \leq C_1 \exp(-\frac{c_1}{4} x^2/m) + C_4 \exp(-c_2 x^2/m) \leq C_5 \exp(-c_3 x^2/m).$$

□

3.1.4 Enumerative properties

In this section we collect basic facts regarding the number of unordered unlabelled trees, which are frequently used in the proofs of the main theorems. Most of these are well-known (at least under less general assumptions), but we do provide proofs for the readers convenience.

Proposition 3.1.4. *The radius of convergence ρ of the series $\tilde{\mathcal{A}}_{\Omega^*}(z)$ satisfies $0 < \rho < 1$ and $\tilde{\mathcal{A}}_{\Omega^*}(\rho) < \infty$.*

Proof. The series $\tilde{\mathcal{A}}_{\Omega^*}(z)$ is dominated coefficentwise by the generating series $\tilde{\mathcal{A}}(z)$ of all rooted trees and it is known that $\tilde{\mathcal{A}}(z)$ is analytic at the origin (see e.g. Otter

[Ott48], Pólya [Pól37], Flajolet and Sedgewick [FS09]). Hence $\rho > 0$. As formal power series we have that

$$\tilde{\mathcal{A}}_{\Omega^*}(X) = XZ_{\mathrm{SET}_{\Omega^*}}(\tilde{\mathcal{A}}_{\Omega^*}(X), \tilde{\mathcal{A}}_{\Omega^*}(X^2), \ldots).$$

The coefficients of all involved series are nonnegative, hence we may lift this identity of formal power series to a identity of real numbers. By assumption, $0 \in \Omega^*$ and there is an integer $\ell \geq 2$ such that $\ell \in \Omega^*$. Thus, for all $0 < x < \rho$ it holds that

$$\tilde{\mathcal{A}}_{\Omega^*}(x) \geq x(1 + \frac{1}{\ell!} \sum_{\sigma \in S_\ell} \tilde{\mathcal{A}}_{\Omega^*}(x)^{\sigma_1} \tilde{\mathcal{A}}_{\Omega^*}(x^2)^{\sigma_2} \cdots \tilde{\mathcal{A}}_{\Omega^*}(x^\ell)^{\sigma_\ell}) \tag{$*$}$$

with S_ℓ denoting the symmetric group of degree ℓ and σ_i denoting the number of cycles of length i of the permuation σ. In particular, by considering the summand for $\sigma = \mathrm{id}$, we have that

$$\tilde{\mathcal{A}}_{\Omega^*}(x) \geq x(\tilde{\mathcal{A}}_{\Omega^*}(x))^\ell / \ell!.$$

Since $\ell \geq 2$ this implies that the limit $\lim_{x \uparrow \rho} \tilde{\mathcal{A}}(x)$ is finite and hence $\tilde{\mathcal{A}}_{\Omega^*}(\rho)$ is finite. Moreover, considering the summand in $(*)$ for σ a cycle of length ℓ yields that

$$\infty > \tilde{\mathcal{A}}_{\Omega^*}(\rho) \geq \rho(\tilde{\mathcal{A}}_{\Omega^*}(\rho^\ell))/\ell!.$$

This implies that $\rho \leq 1$ because otherwise $\tilde{\mathcal{A}}(\rho^\ell) = \infty$. If $\rho = 1$, then Inequality $(*)$ would imply that $\tilde{\mathcal{A}}_{\Omega^*}(1) \geq 1$. Applying $(*)$ again would then yield the clearly impossible inequality

$$\tilde{\mathcal{A}}_{\Omega^*}(1) \geq 1 + \tilde{\mathcal{A}}_{\Omega^*}(1).$$

Hence our premise cannot hold and thus $\rho < 1$. □

From this we obtain detailed information on the number of Pólya trees of a given size with outdegrees in Ω^*. This is a special case of [BBY06, Thm. 75]. See also [FS09, Thm. VII.4] for the aperiodic case.

Proposition 3.1.5. *The following two statements hold.*

i) 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.

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

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

satisfies

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

for some $\epsilon > 0$.

Proof. We have that

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

and for any Λ the series $E^\Lambda(z,w)$ is dominated coefficient-wise by

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

Since $\rho < 1$ it follows that there is an $\epsilon > 0$ such that

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

By a general enumeration result given in Bell, Burris and Yeats [BBY06, Thm. 28] it follows that

$$[z^m]\tilde{\mathcal{A}}_{\Omega^*}(z) \sim \gcd(\Omega^*)\sqrt{\frac{\rho E_z^{\Omega^*}(\rho, \tilde{\mathcal{A}}_{\Omega^*}(\rho))}{2\pi E_{ww}^{\Omega^*}(\rho, \tilde{\mathcal{A}}_{\Omega^*}(\rho))}}\rho^{-m}m^{-3/2}, \quad m \equiv 1 \mod \gcd(\Omega^*).$$

□

In [BFKV11, Prop. 24] the cycle-pointing decomposition was used in order to provide a new method for determining the asymptotic number of free trees. The argument used there can easily be extended to the case of vertex degree restrictions.

Proposition 3.1.6. *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

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

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

iii) The power series

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

has radius of convergence greater than ρ.

Proof. Let ρ denote the radius of convergence of $\tilde{\mathcal{A}}_{\Omega^*}(z)$. Claim iii) follows from the fact that $\rho < 1$ and the series

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

also has radius of convergence ρ. We proceed with claim ii). The series $\bar{Z}_{\mathrm{SET}^{\circledast}_{\Lambda}}$ is dominated coefficient-wise by the series

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

and hence $F^{\Lambda}(z, w)$ is dominated by

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

Since $\rho < 1$ this series is finite for $z = \rho + \epsilon$ and $w = \tilde{\mathcal{A}}_{\Omega^*}(\rho) + \epsilon$ if $\epsilon > 0$ is sufficiently small. In order prove claim i) we are going to perform a singularity analysis of the series $\tilde{\mathcal{F}}^{\circ}_{\Omega}(z)$. The cycle pointing decomposition

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

yields that the series $\tilde{\mathcal{F}}^{\circ}_{\Omega}(z) = z\frac{\mathrm{d}}{\mathrm{d}z}\tilde{\mathcal{F}}_{\Omega}(z)$ can be written in the form

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

with

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

Here we let E^{Ω} be defined as in Proposition 3.1.5. Set $d = \gcd(\Omega^*)$. We have that $\tilde{\mathcal{A}}_{\Omega^*}(z)$ satisfies the prerequisites of the type of power series studied in Jason, Stanley and Yeats [BBY06, Thm. 28]: Its dominant singularities (all of square-root type) are given by the rotated points

$$U = \{\omega^k \rho \mid k = 0, \ldots, d-1\}$$

with

$$\omega = e^{\frac{2\pi i}{d}}.$$

Moreover

$$\tilde{\mathcal{A}}_{\Omega^*}(\omega z) = \omega \tilde{\mathcal{A}}_{\Omega^*}(z)$$

for all z in a generalized Δ-region with wedges removed at the points of U. We have that $h(z, w)$ is a power series with nonnegative coefficients and by claim i) and ii) and Proposition 3.1.5 we have

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

for some $\epsilon > 0$. Hence the dominant singularities and their types are driven by the series $\tilde{\mathcal{A}}_{\Omega^*}(z)$. We may apply a standard result for the singularity analysis of functions with multiple dominant singularities [FS09, Thm. VI.5] and obtain that

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

for $m \equiv 1 \mod \gcd(\Omega^*)$ and $d'_{\Omega^*} > 0$ a constant. □

Chapter 4

Scaling limits of random Pólya trees

4.1 Proof of the main theorem

In the following Ω^* *will always denote a set of nonnegative integers containing zero and at least one integer greater than or equal to two. Moreover,* n *will always denote a natural number that satisfies* $n \equiv 1 \mod \gcd(\Omega^*)$ *and is large enough such that rooted trees with* n *vertices and outdegrees in* Ω^* *exist. We define the subspecies* $\mathrm{SET}_{\Omega^*} \subset \mathrm{SET}$ *by restricting to objects whose size lies in the set* Ω^**. We let* $\mathcal{A}_{\Omega^*}$ *denote the species of Pólya trees with vertex-outdegrees in the set* Ω^**. Clearly it satisfies an isomorphism of combinatorial species*

$$\mathcal{A}_{\Omega^*} \simeq \mathcal{X} \cdot \mathrm{SET}_{\Omega^*}(\mathcal{A}_{\Omega^*}) \tag{4.1.1}$$

Our starting point is constructing a Boltzmann-sampler for Pólya trees. We may apply the rules for the construction of Pólya-Boltzmann samplers in Sections 2.4.2.2, and 2.4.3, in order to obtain the following procedure.

Lemma 4.1.1. *The following recursive procedure* $\Gamma\tilde{\mathcal{A}}_{\Omega^*}(x)$ *terminates almost surely and draws a random Pólya tree with outdegrees in* Ω^* *according to the* Boltzmann distribution *with parameter* $0 < x \le \rho_{\Omega^*}$*, i.e. any tree with* n *vertices gets drawn with probability* $x^n/\tilde{\mathcal{A}}_{\Omega^*}(x)$*.*

1. *Start with a root vertex* v*.*

2. *Draw a random permutation* SET_{Ω^*}*-symmetry according to a (Pólya)-Boltzmann distribution with parameters* $(x^i)_{i\ge 1}$*. That is, let* $\sigma(v)$ *be a random permutation drawn from the union of permutation groups* $\bigcup_{k\in\Omega^*} \mathcal{S}_k$ *with distribution given by*

$$\mathbb{P}(\sigma(v) = \nu) = \frac{x}{\tilde{\mathcal{A}}_{\Omega^*}(x)} \frac{1}{k!} \tilde{\mathcal{A}}_{\Omega^*}(x)^{\nu_1} \tilde{\mathcal{A}}_{\Omega^*}(x^2)^{\nu_2} \cdots \tilde{\mathcal{A}}_{\Omega^*}(x^k)^{\nu_k}$$

   for each $k \in \Omega^*$ *and* $\nu \in \mathcal{S}_k$*. Here* ν_i *denotes the number of cycles of length* i *of the permutation* ν*. In particular,* ν_1 *is the number of fixpoints of* ν*.*

3. *If* $\sigma(v) \in \mathcal{S}_0$ *return the tree consisting of the root only and stop. Otherwise, for each cycle* τ *of* $\sigma(v)$ *let* $\ell_\tau \ge 1$ *denote its length and draw a Polya tree* A_τ *by an independent call to the sampler* $\Gamma\tilde{\mathcal{A}}_{\Omega^*}(x^{\ell_\tau})$*. Make* ℓ_τ *identical copies of the tree* A_τ *and connect their roots to the vertex* v *by adding edges. Return the resulting tree and stop.*

The Boltzmann distribution is a measure on Pólya trees with an arbitrary number of vertices. However, any tree with n vertices has the same probability, i.e., the distribution conditioned on the event that the generated tree has n vertices is *uniform.* This will allow us to reduce the study of properties of a random Pólya tree with exactly n vertices to the study of $\Gamma\tilde{\mathcal{A}}_{\Omega^*}$.

Proof of Theorem 1.3.1. We begin the proof with a couple of auxiliary observations about the sampler $\Gamma\tilde{\mathcal{A}}_{\Omega^*}(x)$ from Lemma 4.1.1. Let us fix $x = \rho_{\Omega^*}$ throughout. We may do so, since by Proposition 3.1.4 we have that $0 < \rho_{\Omega^*} < 1$ and $\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}) < \infty$.

Suppose that we modify Step 1 to "Start with a root vertex v. If the argument of the sampler is ρ_{Ω^*} (as opposed to $\rho_{\Omega^*}^i$ for some $i \geq 2$), then mark this vertex with the color blue.". Then the resulting tree is still Boltzmann-distributed, but comes with a colored subtree which we denote by $\mathcal{T}$. If we construct the sampler $\Gamma\tilde{\mathcal{A}}_{\Omega^*}(x)$ from a Pólya-Boltzmann sampler $\Gamma Z_{A_{\Omega^*}}(x, x^2, \ldots)$, then by the discussion in Section 2.2.4 the subtree $\mathcal{T}$ corresponds precisely to the fixpoints of the symmetry.

Note that $\mathcal{T}$ is distributed like a Galton-Watson tree without the ordering on the offspring sets. By construction, the offspring distribution ξ of $\mathcal{T}$ is given by the number of fixpoints of the random permutation drawn in Step 2. Thus, the probability generating function of ξ is

$$\mathbb{E}[z^\xi] = \frac{\rho_{\Omega^*}}{\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})} Z_{\mathrm{SET}_{\Omega^*}}(z\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}), \tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}^2), \tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}^3), \ldots). \tag{4.1.2}$$

Moreover, for any blue vertex v we may consider the forest $F(v)$ of the trees dangling from v that correspond to cycles of the permutation $\sigma(v)$ with length at least two. Let ζ denote a random variable that is distributed like the number of vertices $|F(v)|$ in $F(v)$. Then the probability generating function of ζ is

$$\mathbb{E}[z^\zeta] = \frac{\rho_{\Omega^*}}{\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})} Z_{\mathrm{SET}_{\Omega^*}}(\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}), \tilde{\mathcal{A}}_{\Omega^*}((z\rho_{\Omega^*})^2), \tilde{\mathcal{A}}_{\Omega^*}((z\rho_{\Omega^*})^3) \ldots). \tag{4.1.3}$$

Using Proposition 3.1.5 it follows from Equations (4.1.2) and (4.1.2) that the generating functions $\mathbb{E}[z^\xi]$ and $\mathbb{E}[z^\zeta]$ have radius of convergence strictly larger than one. Hence ξ and ζ have finite exponential moments. In particular, there are constants $c, c' > 0$ such that for any $s \geq 0$

$$\mathbb{P}(\xi \geq s), \mathbb{P}(\zeta \geq s) \leq ce^{-c's}. \tag{4.1.4}$$

Moreover, as we argue below, ξ has average value

$$\mathbb{E}[\xi] = \left(\frac{\partial}{\partial s_1} Z_{\mathrm{SET}_{\Omega^*}}\right)(\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}), \tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}^2), \ldots)\rho_{\Omega^*} = 1.$$

This can be shown as follows. Recall that the ordinary generating series satisfies the identity $\tilde{\mathcal{A}}_{\Omega^*}(z) = E(z, \tilde{\mathcal{A}}_{\Omega^*}(z))$ with the series $E(z, w)$ given by

$$E(z, w) = zZ_{\mathrm{SET}_{\Omega^*}}(w, \tilde{\mathcal{A}}_{\Omega^*}(z^2), \tilde{\mathcal{A}}_{\Omega^*}(z^3), \ldots).$$

In particular, we have that $F(z, \tilde{\mathcal{A}}_{\Omega^*}(z)) = 0$ with $F(z, w) = E(z, w) - w$. Suppose that $(\frac{\partial}{\partial w}F)(\rho, \tilde{\mathcal{A}}_{\Omega^*}(\rho)) \neq 0$. Then by the implicit function theorem the function $\tilde{\mathcal{A}}_{\Omega^*}(z)$ has an analytic continuation in a neighbourhood of ρ_{Ω^*}. But this contradicts Pringsheim's theorem [FS09, Thm. IV.6], which states that the series $\tilde{\mathcal{A}}_{\Omega^*}(z)$

must have a singularity at the point ρ_{Ω^*} since all its coefficients are nonnegative real numbers. Hence we have $(\frac{\partial}{\partial w}F)(\rho, \tilde{\mathcal{A}}_{\Omega^*}(\rho)) = 0$ which is equivalent to $\mathbb{E}[\xi] = 1$.

With all these facts at hand we proceed with the proof of the theorem. Slightly abusing notation, we let A_n denote the colored random tree drawn by conditioning the (modified) sampler $\Gamma\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})$ on having exactly n vertices. That is, if we ignore the colors, A_n is drawn uniformly among all Pólya trees of size n with outdegrees in Ω^*. Moreover, let $\mathcal{T}_n$ denote the colored subtree of A_n, and for any vertex v of $\mathcal{T}_n$ let $F_n(v)$ denote the corresponding forest that consists of non-blue vertices. We will argue that with high probability there is a constant $C > 0$ such that $|F_n(v)| \leq C \log n$ for all $v \in \mathcal{T}_n$. Indeed, note that by Proposition 3.1.5,

$$\mathbb{P}(|\Gamma\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})| = n) = \frac{\rho_{\Omega^*}^n}{\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})}[z^n]\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}) = \Theta(n^{-3/2}), \tag{4.1.5}$$

i.e. the probability is (only) polynomially small. Thus, for any $s \geq 0$, if we denote by $\zeta_1, \zeta_2, \ldots$ independent random variables that are distributed like ζ

$$\begin{aligned}\mathbb{P}(\exists v \in \mathcal{T}_n : |F_n(v)| \geq s) &= \mathbb{P}(\exists v \in \mathcal{T} : |F(v)| \geq s \mid |\Gamma\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})| = n)\\ &\leq O(n^{3/2})\mathbb{P}(\exists 1 \leq i \leq n : \zeta_i \geq s).\end{aligned}$$

Using (4.1.4) and setting $s = C \log n$ we get that $\mathbb{P}(\zeta_i \geq s) = o(n^{-5/2})$ for an appropriate choice of $C > 0$. Thus, by the union bound

$$\mathbb{P}(\forall v \in \mathcal{T}_n : |F_n(v)| \leq C \log n) = 1 - o(1). \tag{4.1.6}$$

The typical shape of A_n thus consists of a colored tree with small forests attached to each of its vertices, compare with Figure 4.1. In particular, we have that the Gromov-Hausdorff distance between the rescaled trees $\mathsf{A}_n/\sqrt{n}$ and $\mathcal{T}_n/\sqrt{n}$ converges in probability to zero. We are going to show that there is a constant $c_{\Omega^*} > 0$ such that $c_{\Omega^*}\mathcal{T}_n/\sqrt{n}$ converges weakly towards the Brownian continuum random tree $\mathcal{T}_\mathsf{e}$. This immediately implies that

$$c_{\Omega^*}\mathsf{A}_n/\sqrt{n} \xrightarrow{(d)} \mathcal{T}_\mathsf{e}$$

and we are done.

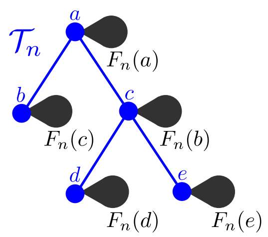

Figure 4.1: The typical shape of the random Pólya tree with n vertices.

We are going to argue that the number of vertices in $\mathcal{T}_n$ concentrates around a constant multiple of n. More precisely, we are going to show that for any exponent $0 < s < 1/2$ we have with high probability that

$$|\mathcal{T}_n| \in (1 \pm n^{-s})\frac{n}{1 + \mathbb{E}[\zeta]}. \tag{4.1.7}$$

To this end, consider the corresponding complementary event in the unconditioned setting

$$|\mathcal{T}| \notin (1 \pm n^{-s}) \frac{|\Gamma\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})|}{1+\mathbb{E}[\zeta]}.$$

If this occurs, then we clearly also have that

$$\sum_{v \in \mathcal{T}} (1 + |F(v)|) = |\Gamma\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})| \notin (1 \pm \Theta(n^{-s}))(1+\mathbb{E}[\zeta])|\mathcal{T}|.$$

Let $\mathcal{E}$ denote the corresponding event. From (4.1.6) we know that with high probability $|F_n(v)| = O(\log n)$ for all vertices v of $\mathcal{T}_n$. Hence, with high probability, say, $|\mathcal{T}_n| \geq n/\log^2 n$. Using again (4.1.5)

$$\mathbb{P}(\mathcal{E} \mid |\Gamma\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})| = n) = O(n^{3/2})\mathbb{P}(\frac{n}{\log^2 n} \leq |\mathcal{T}| \leq n, \mathcal{E}) + o(1).$$

By applying the union bound, the latter probability is at most

$$\sum_{n/\log^2 n \leq \ell \leq n} \mathbb{P}(\sum_{i=1}^{\ell} (1+\zeta_i) \notin (1 \pm \Theta(n^{-s}))(1+\mathbb{E}[\zeta])\ell).$$

Since the random variable ζ has finite exponential moments, we may apply the deviation inequality in Lemma 2.5.1 in order to bound this by

$$\sum_{n/\log^2 n \leq \ell \leq n} \exp(\ell(c\lambda^2 - \lambda\Theta(n^{-s})))$$

for all $0 \leq \lambda \leq \delta$ for some $\delta > 0$. Taking $\lambda = n^{-s/2}$, this may be bounded further by

$$n \exp(\frac{n^{1-s}}{\log^2 n}(c - \Theta(n^{s/2}))) = o(1).$$

Hence, (4.1.7) holds with probability tending to 1 as n becomes large. We are now going to prove that

$$\frac{\sqrt{(1+\mathbb{E}[\zeta])}\sigma}{2\sqrt{n}} \mathcal{T}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}} \tag{4.1.8}$$

with σ^2 denoting the variance of the random variable ξ. This implies that

$$c_{\Omega^*} \mathsf{A}_n/\sqrt{n} \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}} \quad \text{with} \quad c_{\Omega^*} = \frac{\sqrt{(1+\mathbb{E}[\zeta])}\sigma}{2} \tag{4.1.9}$$

and we are done. Note that σ and $\mathbb{E}[\zeta]$ may be computed explicitly from the expression of the probability generating functions in (4.1.2) and (4.1.3), in particular we obtain that σ^2 is given by

$$\begin{aligned}\sigma^2 &= \left(\frac{\partial^2}{\partial z^2}\mathbb{E}[z^\xi] + \frac{\partial}{\partial z}\mathbb{E}[z^\xi] - (\frac{\partial}{\partial z}\mathbb{E}[z^\xi])^2\right)(1) \\ &= \rho_{\Omega^*}\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})\frac{\partial^2 Z_{\mathrm{SET}_{\Omega^*}}}{\partial s_1^2}(\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}), \tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}^2), \dots) \\ &\quad + \rho_{\Omega^*}\frac{\partial Z_{\mathrm{SET}_{\Omega^*}}}{\partial s_1}(\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}), \tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}^2), \dots) \\ &\quad - \rho_{\Omega^*}^2\left(\frac{\partial Z_{\mathrm{SET}_{\Omega^*}}}{\partial s_1}(\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}), \tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}^2), \dots)\right)^2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\zeta] &= \left(\frac{\partial}{\partial z}\mathbb{E}[z^\zeta]\right)(1) \\ &= \frac{\rho_{\Omega*}}{\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*})}\sum_{i\ge 2}\left(\frac{\partial}{\partial s_i}Z_{\mathrm{SET}_{\Omega^*}}\right)(\tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}), \tilde{\mathcal{A}}_{\Omega^*}(\rho_{\Omega^*}^2), \dots)\, i\rho_{\Omega^*}^i\tilde{\mathcal{A}}'_{\Omega^*}(\rho_{\Omega^*}^i),\end{aligned}$$

where $\tilde{\mathcal{A}}'_{\Omega^*} = \frac{\partial}{\partial z}\tilde{\mathcal{A}}_{\Omega^*}$. Note that this expression is well-defined, since $0 < \rho_{\Omega^*} < 1$.

In order to show (4.1.9), let $f : \mathbb{K} \to \mathbb{R}$ denote a bounded, Lipschitz-continous function defined on the space $\mathbb{K}$ of isometry classes of compact metric spaces. Note that the tree $\mathcal{T}_n$ conditioned on having ℓ vertices is distributed like the tree $\mathcal{T}$ conditioned on having ℓ vertices. In particular, it is identically distributed to a ξ-Galton-Watson tree $\mathcal{T}^\xi$ conditioned on having ℓ vertices, which we denote by $\mathcal{T}_\ell^\xi$. Since (4.1.7) holds with high probability it follows that

$$\mathbb{E}[f(c_{\Omega^*}\mathcal{T}_n/\sqrt{n})] = o(1) + \sum_{\ell\in(1\pm n^{-s})\frac{n}{1+\mathbb{E}[\zeta]}}\mathbb{E}[f(c_{\Omega^*}\mathcal{T}_\ell^\xi/\sqrt{n})]\mathbb{P}(|\mathcal{T}| = \ell).$$

Let $\mathrm{D}(T)$ denote the diameter of T, i.e., the number of vertices on a longest path in T. Since f was assumed to be Lipschitz-continuous it follows that

$$\left|\mathbb{E}[f(c_{\Omega^*}\mathcal{T}_\ell^\xi/\sqrt{n})] - \mathbb{E}[f(\sigma\mathcal{T}_\ell^\xi/2\sqrt{\ell})]\right| \le a_{n,\ell}\mathbb{E}[\mathrm{D}(\mathcal{T}_\ell^\xi)/\sqrt{\ell}]$$

for a sequence $a_{n,\ell}$ with $\sup_\ell(a_{n,\ell}) \to 0$ as n becomes large. Moreover, the average rescaled diameter $\mathbb{E}[\mathrm{D}(\mathcal{T}_\ell^\xi)/\sqrt{\ell}]$ converges to a multiple of the expected diameter of the CRT $\mathcal{T}_{\mathsf{e}}$ as ℓ tends to infinity, see e.g. [ABDJ13]. In particular, it is a bounded sequence. Since

$$\mathbb{E}[f(\sigma\mathcal{T}_\ell^\xi/2\sqrt{\ell})] \to \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})]$$

as $\ell \to \infty$, it follows that

$$\mathbb{E}[f(c_{\Omega^*}\mathcal{T}_n/\sqrt{n})] \to \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})]$$

as n becomes large. This completes the proof. □

Chapter 5

Scaling limits of random graphs from subcritical classes

5.1 Preliminaries

5.1.1 Block-stable graph classes

Any graph may be decomposed into its *connected components*, i.e. its maximal connected subgraphs. These connected components allow a *block-decomposition* which we recall in the following. Let C be a connected graph. If removing a vertex v (and deleting all adjacent edges) disconnects the graph, we say that v is a *cutvertex* of C. The graph C is *2-connected*, if it has size at least three and no cutvertices.

A *block* of an arbitrary graph G is a maximal connected subgraph $B \subset G$ that does not have a cutvertex (of itself). It is well-known, see for example [Die10], that any block is either 2-connected or an edge or a single isolated point. Moreover, the intersection of two blocks is either empty or a cutvertex of a connected component of G. If G is connected, then the bipartite graph whose vertices are the blocks and the cutvertices of G and whose edges are pairs $\{v, B\}$ with $v \in B$ is a tree and called the *block-tree* of G.

Let $\mathcal{G}$ denote a subspecies of the species of graphs, $\mathcal{C} \subset \mathcal{G}$ the subspecies of connected graphs in $\mathcal{G}$ and $\mathcal{B} \subset \mathcal{C}$ the subspecies of all graphs in $\mathcal{C}$, that are 2-connected or consist of only two vertices joined by an edge. We say that $\mathcal{G}$ or $\mathcal{C}$ is a *block-stable* class of graphs, if $\mathcal{B} \neq 0$ and $G \in \mathcal{G}$ if and only if every block of G belongs to $\mathcal{B}$ or is a single isolated vertex. Block-stable classes satisfy the following combinatorial specifications that can be found for example in Joyal [Joy81], Bergeron, Labelle and Leroux [BLL98] and Harary and Palmer [HP73]:

$$\mathcal{G} \simeq \mathrm{SET} \circ \mathcal{C} \qquad \text{and} \qquad \mathcal{C}^\bullet \simeq \mathcal{X} \cdot (\mathrm{SET} \circ \mathcal{B}' \circ \mathcal{C}^\bullet). \tag{5.1.1}$$

The first correspondence expresses the fact that we may form any graph on a given vertex set U by partitioning U and constructing a connected graph on each partition class. The specification for rooted connected graphs, illustrated in Figure 5.1, is based on the construction of the block-tree. The idea is to interpret $\mathcal{B}' \circ \mathcal{C}^\bullet$-objects as graphs by connecting the roots of the $\mathcal{C}^\bullet$ objects on the partition classes and the $*$-vertex with edges according to the $\mathcal{B}'$-object on the partition. An object of $\mathrm{SET} \circ (B' \circ \mathcal{C}^\bullet)$ can then be interpreted as a graph by identifying the $*$-vertices of the $B' \circ \mathcal{C}^\bullet$ objects. This construction is compatible with graph isomorphisms, hence $\mathcal{C}' \simeq \mathrm{SET} \circ \mathcal{B}' \circ \mathcal{C}^\bullet$ and the second specification in (5.1.1) follows. By the rules for computing the generating series of species we obtain the equations

$$\mathcal{G}(x) = \exp(\mathcal{C}(x)) \qquad \text{and} \qquad \mathcal{C}^\bullet(x) = x \exp(\mathcal{B}'(\mathcal{C}^\bullet(x))). \tag{5.1.2}$$

The following lemma was given in Panagiotou and Steger [PS10] and Drmota et al. [DFK$^+$11] under some minor additional assumptions.

Lemma 5.1.1. *Let $\mathcal{C}$ be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. Then the exponential generating*

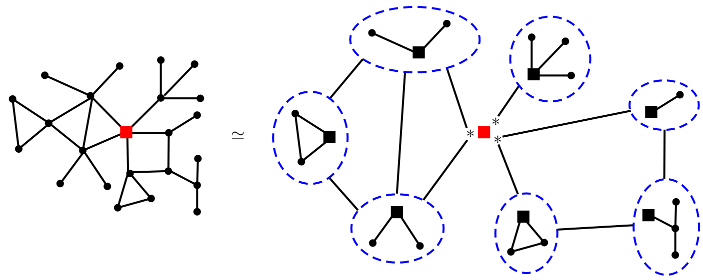

Figure 5.1: Decomposition of a rooted graph from $\mathcal{C}^\bullet$ into a $\mathcal{X} \cdot (\mathrm{SET} \circ \mathcal{B}' \circ \mathcal{C}^\bullet)$ structure. Labels are omitted and the roots are marked with squares.

series $\mathcal{C}(z)$ has radius of convergence $\rho < \infty$ and the sums $y := \mathcal{C}^\bullet(\rho)$ and $\lambda := \mathcal{B}'(y)$ are finite and satisfy

$$y = \rho \exp(\lambda). \tag{5.1.3}$$

Proof. It suffices to consider the case $\rho > 0$. By assumption we have $\mathcal{B} \neq 0$ and hence there is a $k \in \mathbb{N}$ such that $[z^k]B'(z) \neq 0$. Thus, by (5.1.2) we have, say, $\mathcal{C}^\bullet(z) = cz\mathcal{C}^\bullet(z)^{2k} + R(z)$ for some constant $c > 0$ and $R(z)$ a power series in z with nonnegative coefficients. This implies $\lim_{x \uparrow \rho} \mathcal{C}^\bullet(x) < \infty$ and thus ρ and $\mathcal{C}^\bullet(\rho)$ are both finite. The coefficients of all power series involved in (5.1.2) are nonnegative, and so it follows that $y = \rho \exp(\lambda)$ and thus $\lambda < \infty$. $\square$

We will only be interested in the case where $\mathcal{C}$ is analytic. The following observation (made for example also in [DN13]) shows that this is equivalent to requiring that $\mathcal{B}$ is analytic. We include a short proof for completeness.

Proposition 5.1.2. *Let $\mathcal{C}$ be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. Then $\mathcal{C}$ is analytic if and only if $\mathcal{B}$ is analytic.*

Proof. By nonnegativity of coefficients we see easily that $\rho > 0$ implies that $\mathcal{B}$ is analytic. Conversely, suppose that $\mathcal{B}(z)$ has positive radius of convergence $R > 0$. By the inverse function theorem, the block-stability equation $f(z) = z\exp(\mathcal{B}'(f(z)))$ has an analytic solution whose expansion at the point 0 agrees with the series $\mathcal{C}^\bullet(z)$ by Lagrange's inversion formula. Hence $\mathcal{C}$ is an analytic class. $\square$

5.1.2 $\mathcal{R}$-enriched trees

The class $\mathcal{T}^\bullet$ of rooted trees[1] is known to satisfy the decomposition

$$\mathcal{T}^\bullet \simeq \mathcal{X} \cdot \mathrm{SET}(\mathcal{T}^\bullet).$$

[1] *Arborescence* is the French word for rooted tree, hence the notation $\mathcal{A}$.

This is easy to see: in order to form a rooted tree on a given set of vertices, we choose a root vertex v, partition the remaining the vertices, endow each partition class with a structure of a rooted tree and connect the vertex v with their roots. More generally, given a species $\mathcal{R}$ the class $\mathcal{A}_{\mathcal{R}}$ of $\mathcal{R}$-enriched trees is defined by the combinatorial specification

$$\mathcal{A}_{\mathcal{R}} \simeq \mathcal{X} \cdot \mathcal{R}(\mathcal{A}_{\mathcal{R}}).$$

In other words, an $\mathcal{R}$-enriched tree is a rooted tree such that the offspring set of any vertex is endowed with an $\mathcal{R}$-structure. Natural examples are labeled ordered trees, which are SEQ-enriched trees, and plane trees, which are unlabeled ordered trees. Ordered and unordered tree families defined by restrictions on the allowed outdegree of internal vertices also fit in this framework. $\mathcal{R}$-enriched trees were introduced by Labelle [Lab81] in order to provide a combinatorial proof of Lagrange Inversion. They have applications in various fields of mathematics, see for example [ML14, DLL82, Lab10].

The combinatorial specification (5.1.1) together with Theorem 2.2.2 allows us to identify a block-stable graph class $\mathcal{C}^\bullet$ with the class $\mathcal{R}$-enriched trees where $\mathcal{R} = \mathrm{SET}(\mathcal{B}')$, that is, rooted trees from $\mathcal{T}^\bullet$ where the offspring set of each vertex is partitioned into nonempty sets and each of these sets carries a $\mathcal{B}'$-structure. Compare with Figure 5.2.

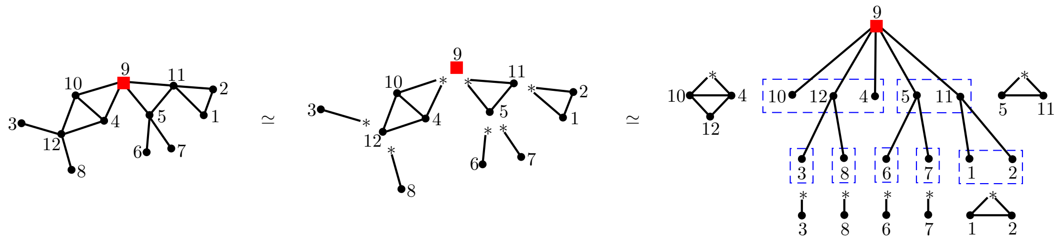

Figure 5.2: Correspondence of the classes $\mathcal{C}^\bullet$ and $\mathrm{SET}(\mathcal{B}')$-enriched trees.

Corollary 5.1.3. *Let $\mathcal{C}$ be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. Then there is a unique isomorphism between $\mathcal{C}^\bullet$ and the class $\mathcal{A}_{\mathrm{SET}\circ\mathcal{B}'}$ of pairs (T, α) with $T \in \mathcal{T}^\bullet$ and α a function that assigns to each $v \in V(T)$ a (possibly empty) set $\alpha(v) \in (\mathrm{SET} \circ \mathcal{B}')[M_v]$ of derived blocks whose vertex sets partition the offspring set M_v of v.*

Proof. By the isomorphism given in (5.1.1) the classes $\mathcal{A}_{\mathrm{SET}\circ\mathcal{B}'}$ and $\mathcal{C}^\bullet$ are both solutions of the system $\mathcal{Y} = \mathcal{H}(\mathcal{X}, \mathcal{Y})$ with $\mathcal{H}(\mathcal{X}, \mathcal{Y}) = \mathcal{X} \cdot \mathrm{SET} \circ \mathcal{B}' \circ \mathcal{Y}$. Joyal's Implicit Species Theorem 2.2.2 yields that there is a unique isomorphism between any two solutions. □

5.1.3 The classical Boltzmann sampler for block-stable classes

Let $\mathcal{C}$ be a block-stable class of connected graphs such that the radius of convergence ρ of the generating series $\mathcal{C}(z)$ is positive. The rooted class $\mathcal{C}^\bullet$ has a combinatorial specification given in (5.1.1) in terms of the subclass $\mathcal{B}$ of edges and 2-connected graphs. By Lemma 5.1.1, we know that $y = \mathcal{C}^\bullet(\rho)$ and $\lambda = \mathcal{B}'(y)$ are finite.

Since ρ is an admissible parameter for the Boltzmann-distribution of $\mathcal{C}^\bullet$, we may apply the rules for the construction of Boltzmann samplers given in Sections 2.4.2.2 and 2.4.3 in order to obtain an explicit sampler $\Gamma\mathcal{C}^\bullet(\rho)$. By the rule concerning products of species, we have to start with independent calls to the samplers $\Gamma\mathcal{X}(\rho)$ and $\Gamma(\mathrm{SET} \circ \mathcal{B}' \circ \mathcal{C}^\bullet)(\rho)$, and relabel uniformly at random afterwards. The sampler $\Gamma\mathcal{X}(\rho)$ generates (deterministically) a single root-vertex. The rule for compositions states that a Boltzmann sampler for $(\mathrm{SET} \circ \mathcal{B}') \circ \mathcal{C}^\bullet$ is obtained by starting with $\Gamma(\mathrm{SET} \circ \mathcal{B}')(y)$, and making independent calls to $\Gamma\mathcal{C}^\bullet(\rho)$ for each atom (i.e. non-$*$-vertex) of the result. Putting everything together, we obtain the following recursive procedure.

Corollary 5.1.4. *Let $\mathcal{C}$ be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. The following recursive procedure terminates almost surely and samples according to the Boltzmann distribution for $\mathcal{C}^\bullet$ with parameter ρ.*

$\Gamma\mathcal{C}^\bullet(\rho)$: $\gamma \leftarrow$ a single root vertex
  $M \leftarrow \Gamma(\mathrm{SET} \circ \mathcal{B}')(y)$
  for each derived block B in M
    merge the $*$-vertex of B with γ
    for each non $*$-vertex v of B
      $C \leftarrow \Gamma\mathcal{C}^\bullet(\rho)$
      merge v with the root of C
  return the resulting graph, relabeled uniformly at random

This procedure was used before in the study of certain block-stable graph classes, see for example [PS10]. Using the rules for the composition and the SET-species, we also obtain an explicit description of a Boltzmann sampler for the species $\mathrm{SET} \circ \mathcal{B}'$.

$\Gamma(\mathrm{SET} \circ \mathcal{B}')(y)$: $m \leftarrow \mathsf{Pois}(\lambda)$
  for $k = 1 \dots m$
    $B_k \leftarrow \Gamma\mathcal{B}'(y)$
  return $\{B_1, \dots, B_m\}$, relabeled uniformly at random

5.1.4 Subcritical graph classes

Let $\mathcal{C}$ be a block-stable class of connected graphs and $\mathcal{B}$ its subclass of all graphs that are 2-connected or a single edge. Assume that $\mathcal{B}$ is nonempty and analytic, hence $\mathcal{C}$ is analytic as well by Proposition 5.1.2. Denote by ρ and R the radii of convergence of the corresponding exponential generating series $\mathcal{C}(z)$ and $\mathcal{B}(z)$. By Lemma 5.1.1, we know that ρ, $y = \mathcal{C}^\bullet(\rho)$ and $\lambda = \mathcal{B}'(y)$ are finite quantities. The following proposition provides a coupling of a Boltzmann-distributed random graph drawn from the class $\mathcal{C}$ with a Galton-Watson tree. This will play a central role in the proof of the main theorem.

Proposition 5.1.5. *Let (T, α) denote the enriched tree corresponding to the Boltzmann Sampler $\Gamma\mathcal{C}^\bullet(\rho)$ given in Corollary 5.1.4. Then the rooted labeled unordered tree T is distributed like the outcome of the following process:*

1. *Draw a Galton-Watson tree with offspring distribution ξ given by the probability generating function $\varphi(z) = \exp(B'(yz) - \lambda)$.*
2. *Distribute labels uniformly at random.*
3. *Discard the ordering on the offspring sets.*

Proof. The sampler $\Gamma\mathcal{C}^\bullet(\rho)$ given in Corollary 5.1.4 starts with a single root-vertex and a set M of $\mathcal{B}'$-objects drawn according to $\Gamma(\mathrm{SET} \circ \mathcal{B}')(y)$. Each non-$*$-vertex of the blocks in M corresponds to an offspring vertex of the root in the tree T. Thus the root receives total offspring with size distributed according to $|\Gamma(\mathrm{SET} \circ \mathcal{B}')(y)|$, which by definition of the Boltzmann distribution has probability generating function $\exp(\mathcal{B}'(yz) - \lambda)$. For any offspring vertex, the sampler proceeds with a recursive call to $\Gamma\mathcal{C}^\bullet(\rho)$. After this recursive procedure terminates, the vertices of the resulting graph are relabeled uniformly at random. Thus T is distributed like a Galton-Watson tree with offspring distribution given by the pgf $\varphi(z)$, except that we neglect all orderings on the offspring sets and relabel the vertices uniformly at random after constructing the tree. □

Let ξ denote the offspring distribution given in Proposition 5.1.5. As discussed above, the rules governing Boltzmann samplers guarantee that the sampler $\Gamma\mathcal{C}^\bullet(\rho)$ terminates almost surely. Hence we have $1 \geq \mathbb{E}[\xi] = \varphi_\ell'(1) = y\mathcal{B}''(y) = \mathcal{B}'^\bullet(y)$ and in particular $y \leq R$. We define subcriticality depending on whether this inequality is strict.

Definition 5.1.6. *A block-stable class of connected graphs $\mathcal{C}$ is termed subcritical if $y < R$.*

Prominent examples of subcritical graph classes are trees, outerplanar graphs and series-parallel graphs; the class of planar graphs does not fall into this framework [DFK+11, BPS09], i.e. it satisfies $y = R$. The following lemma was proved in Panagiotou and Steger [PS10, Lem. 2.8] by analytic methods.

Lemma 5.1.7. *If $\mathcal{B}'^\bullet(R) \geq 1$, then $\mathcal{B}'^\bullet(y) = 1$. If $\mathcal{B}'^\bullet(R) \leq 1$, then $y = R$. In particular, $\mathcal{C}$ is subcritical if and only if $\mathcal{B}'^\bullet(R) > 1$.*

Thus, if $\mathcal{B}'^\bullet(R) \geq 1$, then the offspring distribution ξ has expected value 1 and variance

$$\sigma^2 = 1 + \mathcal{B}'''(y)y^2 = \mathbb{E}[|\Gamma\mathcal{B}'^\bullet(y)|]$$

with $\Gamma\mathcal{B}'^\bullet(y)$ denoting a Boltzmann sampler for the class $\mathcal{B}'^\bullet$ with parameter y. By Proposition 5.1.5 the size of the outcome of the sampler $\Gamma\mathcal{C}^\bullet(\rho)$ is distributed like the size of a ξ-Galton-Watson tree. Hence, by a standard asymptotic expression [Jan12, Thm. 18.11] , we obtain the following result, which was shown in [DFK+11] under stronger assumptions.

Corollary 5.1.8. *Let $\mathcal{C}$ be an analytic block-stable class of graphs, and let ξ be the distribution from Proposition 5.1.5. Suppose that $\mathcal{B}'^{\bullet}(R) \geq 1$ and $\mathcal{B}'''(y) < \infty$, i.e. ξ has finite variance. Let $d = \mathsf{span}(\xi)$. Then, as $n \equiv 1 \mod d$ tends to infinity,*

$$\mathbb{P}(|\Gamma\mathcal{C}^{\bullet}(\rho)| = n) \sim \frac{d}{\sqrt{2\pi\mathbb{E}[|\Gamma\mathcal{B}'^{\bullet}(y)|]}} n^{-3/2},$$

$$|\mathcal{C}_n| \sim \frac{yd}{\sqrt{2\pi\mathbb{E}[|\Gamma\mathcal{B}'^{\bullet}(y)|]}} n^{-5/2}\rho^{-n} n!.$$

5.2 A size-biased random $\mathcal{R}$-enriched tree

Let $\mathcal{C}$ be an analytic block-stable class of connected graphs and $\mathcal{B} \neq 0$ its subclass of graphs that are 2-connected or a single edge. As before we let ρ denote the radius of convergence of the exponential generating series $\mathcal{C}(z)$ and set $y = \mathcal{C}^{\bullet}(\rho)$. Recall that by Corollary 5.1.3 the class $\mathcal{C}^{\bullet}$ may be identified with the class of $\mathcal{R}$-enriched trees with $\mathcal{R} := \mathrm{SET} \circ \mathcal{B}'$, i.e. pairs (T, α) with $T \in \mathcal{T}^{\bullet}$ a rooted labeled unordered tree and α a function that assigns to each $v \in V(T)$ a (possibly empty) set $\alpha(v)$ of derived blocks whose vertex sets partition the offspring set of v.

An important ingredient in our forthcoming arguments will be an accurate description of the distribution of the blocks on sufficiently long paths in random graphs from $\mathcal{C}$. In order to study this distribution we will make use of a special case of a *size-biased* random $\mathcal{R}$-enriched tree. The use of size-biased structures to study distances for large random trees is a fruitful approach used in classic and recent literature (see e.g. Lyons, Pemantle and Peres [LPP95], and Addario-Berry, Devroye and Janson [ABDJ13]), and applying it to $\mathcal{R}$-enriched trees allows for a particular short and elegant proof of our main result.

Consider the species $\mathcal{A}_{\mathcal{R}}^{\bullet}$ of pointed enriched trees, that is of enriched trees $A = (T, \alpha)$ together with a distinguished vertex u of T. In order to avoid confusion, we call u the *outer root*, and the root of T the *inner root*. The directed path in T from the inner root to the outer root is termed the *spine* of the pointed enriched tree. The species $\mathcal{A}_{\mathcal{R}}^{\bullet}$ admits the following classical decomposition due to Labelle [Lab81, Thm. A]. First, we split the species into summands

$$\mathcal{A}_{\mathcal{R}}^{\bullet} \simeq \sum_{\ell=0}^{\infty} \mathcal{A}_{\mathcal{R}}^{(\ell)}$$

with $\mathcal{A}_{\mathcal{R}}^{(\ell)}$ denoting the subspecies of all pointed $\mathcal{R}$-enriched trees whose spine has length ℓ. Here the subspecies $\mathcal{A}_{\mathcal{R}}^{(0)}$ corresponds to the case in which the inner and outer root coincide, yielding $\mathcal{A}_{\mathcal{R}}^{(0)} \simeq \mathcal{A}_{\mathcal{R}}$. For $\ell \geq 1$ we are going to argue that there is an isomorphism

$$\mathcal{A}_{\mathcal{R}}^{(\ell)} \simeq \mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_{\mathcal{R}}) \cdot \mathcal{A}_{\mathcal{R}}^{(\ell-1)}, \tag{5.2.1}$$

as illustrated in Figure 5.3. Indeed, suppose that we are given an arbitrary $\mathcal{A}_{\mathcal{R}}^{(\ell)}$-object. The maximal (enriched) subtree rooted at the successor v of the inner root along the spine is an $\mathcal{A}_{\mathcal{R}}^{(\ell-1)}$-object, as the length of its spine is decreased by 1. If we cut this tree away and replace v with a $*$-vertex we are left with the inner root, accounting for the factor $\mathcal{X}$ in (5.2.1), together with an $\mathcal{R}'$-object whose non-$*$-labels are the roots of $\mathcal{A}_{\mathcal{R}}$-objects, accounting for the factor $\mathcal{R}'(\mathcal{A}_{\mathcal{R}})$.

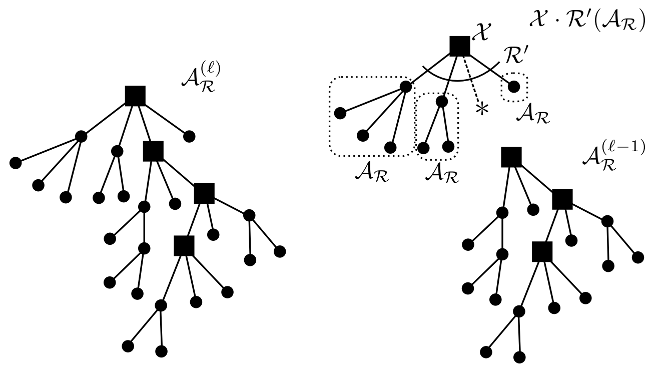

Figure 5.3: The decomposition of $\mathcal{A}_{\mathcal{R}}^{(\ell)}$, with the squares marking the vertices on the spine.

By iterating (5.2.1) we arrive at

$$\mathcal{A}_{\mathcal{R}}^{(\ell)} \simeq (\mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_{\mathcal{R}}))^{\ell} \mathcal{A}_{\mathcal{R}}. \tag{5.2.2}$$

Our size-biased $\mathcal{R}$-enriched tree will be given by a Boltzmann sampler $\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho)$ of the species $\mathcal{A}_{\mathcal{R}}^{(\ell)}$. Recall that $\rho > 0$ denotes the radius of convergence of the exponential generating series $\mathcal{C}^{\bullet}(z) = \mathcal{A}_{\mathcal{R}}(z)$ and that $y = \mathcal{A}_{\mathcal{R}}(\rho) < \infty$ by Lemma 5.1.1. Of course, we have to check whether ρ is an admissible parameter for the Boltzmann distribution, i.e. if $\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) < \infty$. This is easily confirmed, as the isomorphism in (5.2.2) yields

$$\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) = (\rho\mathcal{R}'(\mathcal{A}_{\mathcal{R}}(\rho)))^{\ell}\mathcal{A}_{\mathcal{R}}(\rho) = (\rho\mathcal{R}'(y))^{\ell} y = (\rho\mathcal{B}''(y)\mathcal{R}(y))^{\ell} y.$$

Using that $\mathcal{R}(y) = e^{\mathcal{B}'(y)}$ and applying Lemmas 5.1.1 and 5.1.7 we infer that the latter quantity is finite. Hence the Boltzmann distribution for $\mathcal{A}_{\mathcal{R}}^{(\ell)}$ with parameter ρ is well-defined and for any pointed enriched tree (A, u) from $\mathcal{A}_{\mathcal{R}}^{(\ell)}$ with k vertices it is given by

$$\mathbb{P}(\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) = (A, u)) = \rho^k / \mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) = \rho^k (\rho\mathcal{R}'(y))^{-\ell} / \mathcal{A}_{\mathcal{R}}(\rho).$$

Moreover, letting $\Gamma\mathcal{A}_{\mathcal{R}}(\rho)$ denote a Boltzmann sampler for $\mathcal{A}_{\mathcal{R}}$, we have that

$$\mathbb{P}(\Gamma\mathcal{A}_{\mathcal{R}}(\rho) = A) = \rho^k / \mathcal{A}_{\mathcal{R}}(\rho)$$

and hence

$$\mathbb{P}(\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) = (A,u)) = \big(\rho\mathcal{R}'(y)\big)^{-\ell}\,\mathbb{P}(\Gamma\mathcal{A}_{\mathcal{R}}(\rho) = A). \tag{5.2.3}$$

Equation (5.2.3) allows us to relate properties of the size-biased $\mathcal{R}$-enriched tree $\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho)$ to properties of a uniformly at random chosen enriched tree of a given size. We are going to apply the following general lemma in Section 5.3 in order to show that the blocks along sufficiently long paths in random graphs behave asymptotically like the spine of $\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho)$ for a corresponding integer ℓ.

Lemma 5.2.1. *Let $\mathcal{E}$ be a property of pointed $\mathcal{R}$-enriched trees (i.e. a subset of $\mathcal{A}_{\mathcal{R}}^\bullet$) and let $n \in \mathbb{N}$ be such that $\mathcal{A}_{\mathcal{R}}[n]$ is nonempty. Consider the function*

$$f : \mathcal{A}_{\mathcal{R}}[n] \to \mathbb{R},\ A \mapsto \sum_{v\in[n]} \mathbb{1}_{(A,v)\in\mathcal{E}}$$

counting the number of "admissible" outer roots with respect to $\mathcal{E}$. Let $\mathsf{A}_n \in \mathcal{A}_{\mathcal{R}}[n]$ be drawn uniformly at random. Then

$$\mathbb{E}[f(\mathsf{A}_n)] = \mathbb{P}(|\Gamma\mathcal{A}_{\mathcal{R}}(\rho)| = n)^{-1}\sum_{\ell=0}^{n-1}\big(\rho\mathcal{R}'(y)\big)^{\ell}\,\mathbb{P}(\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) \text{ has size } n \text{ and satisfies } \mathcal{E}).$$

Proof. First, observe that

$$\sum_{v=1}^{n}\mathbb{P}((\mathsf{A}_n,v)\in\mathcal{E}) = \sum_{\ell=0}^{n-1}\ \sum_{(A,u)\in\mathcal{E}\cap\mathcal{A}_{\mathcal{R}}^{(\ell)}[n]}\mathbb{P}(\mathsf{A}_n = A).$$

By (5.2.3) we have for all $(A,u)\in\mathcal{E}\cap\mathcal{A}_{\mathcal{R}}^{(\ell)}[n]$ that

$$\mathbb{P}(\Gamma\mathcal{A}_{\mathcal{R}}(\rho) = A \mid |\Gamma\mathcal{A}_{\mathcal{R}}(\rho)| = n) = \big(\rho R'(y)\big)^{\ell}\,\mathbb{P}(\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) = (A,u))\mathbb{P}(|\Gamma\mathcal{A}_{\mathcal{R}}(\rho)| = n)^{-1}.$$

This proves the claim. □

In order for Lemma 5.2.1 to be useful, we need information about the spine of the size-biased $\mathcal{R}$-enriched tree $\Gamma\mathcal{A}_{\mathcal{R}}(\rho)$. We are going to argue that the $\mathcal{R}$-structures along the spine are (up to relabelling of vertices) independent and follow a Boltzmann distribution for the pointed species $\mathcal{R}^\bullet$ with parameter y. We do this by constructing the Boltzmann sampler step by step following the corresponding rules.

We may apply the rules for products to the isomorphism in (5.2.1) in order to obtain the following sampling procedure for $\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho)$:

1. Draw a tuple of objects according to (independent) Boltzmann samplers $\Gamma\mathcal{X}(\rho)$, $\Gamma\mathcal{R}'(\mathcal{A}_{\mathcal{R}})(\rho)$ and $\Gamma\mathcal{A}_{\mathcal{R}}^{(\ell-1)}(\rho)$.

2. Relabel uniformly at random.

3. Apply the isomorphism in (5.2.1) in order to obtain an $\mathcal{A}^{(\ell)}$-structure from the resulting $\mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_\mathcal{R}) \cdot \mathcal{A}_\mathcal{R}^{(\ell-1)}$-structure.

The rule for the composition yields the following description for $\Gamma\mathcal{R}'(\mathcal{A}_\mathcal{R})(\rho)$.

1. Call $\Gamma\mathcal{R}'(y)$ and let R' denote the result.

2. For each non-$*$-label v of R' call $\Gamma\mathcal{A}_\mathcal{R}(\rho)$ and let A_v denote the result.

3. Relabel $(R', (R_v)_v)$ uniformly at random.

Note that since $\mathcal{R}^\bullet \simeq \mathcal{R} \cdot \mathcal{X}$ the sampler $\Gamma\mathcal{R}^\bullet(y)$ is given by sampling $\Gamma\mathcal{R}'(y)$ and relabelling all vertices including the $*$-vertex uniformly at random. Together with the sampler $\Gamma\mathcal{A}_\mathcal{R}(\rho)$ described in Section 5.1.3 we obtain the following procedure for the Boltzmann sampler $\Gamma\mathcal{A}_\mathcal{R}^{(\ell)}(\rho)$ which we call the *size-biased $\mathcal{R}$-enriched tree* (see also Figure 5.4):

Consider two kinds of vertices termed *normal* and *mutant*. We start with a single mutant root. Normal vertices have an independent copy of $\Gamma\mathcal{R}(y)$ as offspring. Mutant nodes have an independent copy of $\Gamma\mathcal{R}^\bullet(y)$ as offspring and the root in the $\mathcal{R}^\bullet$ object is declared mutant, unless its the ℓth copy of $\Gamma\mathcal{R}^\bullet(y)$. By the theory of recursive Boltzmann samplers obtained from combinatorial specifications this procedure terminates almost surely. The sampler $\Gamma\mathcal{A}_\mathcal{R}^{(\ell)}(\rho)$ is obtained by additionally distributing labels uniformly at random.

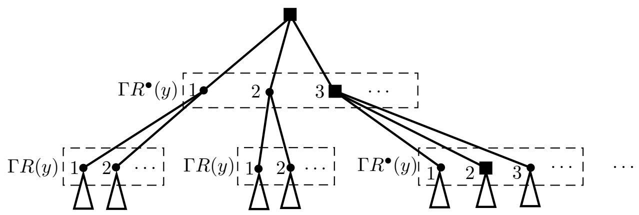

Figure 5.4: Illustration of the sampler for the size-biased $\mathcal{R}$-enriched tree.

Note that the $\mathcal{R}$-objects along the spine of the random enriched tree $\Gamma\mathcal{A}_\mathcal{R}^{(\ell)}(\rho)$ are drawn according to ℓ independent copies of $\Gamma\mathcal{R}^\bullet(y)$. In our setting we have that $\mathcal{R} = \text{SET} \circ \mathcal{B}'$, where $\mathcal{B} \neq 0$ denotes the subclass of blocks of the block-stable class $\mathcal{C}$. Using the chain rule for the derivative of species, we obtain

$$\mathcal{R}^\bullet \simeq (\text{SET} \circ \mathcal{B}') \cdot \mathcal{B}'^\bullet$$

and the sampler $\Gamma\mathcal{R}^\bullet(y)$ is given by independent calls of $\Gamma(\text{SET} \circ \mathcal{B}')(y)$ and $\Gamma\mathcal{B}'^\bullet(y)$. Hence, up to relabelling of vertices, the blocks along the spine are drawn according to ℓ independent copies of $\Gamma\mathcal{B}'^\bullet(y)$.

5.3 Convergence towards the CRT

Let $\mathcal{C}$ be an analytic subcritical class of connected graphs and $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. We let $\rho > 0$ denote the radius of convergence of the exponential generating series $\mathcal{C}(z)$ and set $y = \mathcal{C}^\bullet(\rho)$. As before we identify $\mathcal{C}^\bullet$ with the class $\mathcal{A}_\mathcal{R}$ of $\mathcal{R}$-enriched trees with $\mathcal{R} = \mathrm{SET} \circ \mathcal{B}'$. By Proposition 5.1.5 we know that if we draw an $\mathcal{R}$-enriched tree (T, α) according to the Boltzmann distribution with parameter ρ, then T is distributed like a ξ-Galton-Watson tree with $\xi := |\Gamma(\mathrm{SET} \circ \mathcal{B}')(y)|$, relabelling uniformly at random and discarding the ordering on the offspring sets.

Throughout this section let $n \equiv 1 \mod \mathsf{span}(\xi)$ denote a large enough integer such that the probability of a ξ-GWT having size n is positive. Let $\mathsf{C}_n \in \mathcal{C}_n$ be drawn uniformly at random and generate $\mathsf{C}_n^\bullet \in \mathcal{C}_n^\bullet$ by uniformly choosing a root from $[n]$. We let (T_n, α_n) be the corresponding enriched tree.

For any pointed derived block $B \in \mathcal{B}'^\bullet$ we let $\mathsf{d}(B) := d_B(*, \mathrm{root})$ denote the length of a shortest path connecting the $*$-vertex with the root. In this section we prove our main result, Theorem 1.4.1. More precisely, we are going to show that

$$\frac{\sigma}{2\kappa\sqrt{n}} \mathsf{C}_n^\bullet \xrightarrow{(d)} \mathcal{T}_\mathsf{e} \quad \text{and} \quad \frac{\sigma}{2\kappa\sqrt{n}} \mathsf{C}_n \xrightarrow{(d)} \mathcal{T}_\mathsf{e} \tag{5.3.1}$$

with respect to the (pointed) Gromov-Hausdorff metric. The constants are given by $\sigma^2 = \mathbb{E}[|\mathsf{B}|]$ and $\kappa = \mathbb{E}[|\mathsf{d}(\mathsf{B})|]$ with $\mathsf{B} \in \mathcal{B}'^\bullet$ a random block drawn according to the Boltzmann distribution with parameter $y = \mathcal{C}^\bullet(\rho)$, and in particular $\sigma^2 = 1 + \mathcal{B}'''(y)y^2$. To this end, we require a second metric on connected graphs:

Definition 5.3.1. *Let $C \in \mathcal{C}$. For any $x, y \in V(C)$ set $\bar{d}_C(x,y) := d_T(x,y)$ with (T, α) the enriched tree corresponding to (C, x), i.e. C rooted at the vertex x.*

Less formally speaking, $\bar{d}_C(x,y)$ denotes the minimum number of blocks required to cover the edges of a shortest path linking x and y. As the example illustrated in Figure 5.5 shows, the distance between x and y in the tree corresponding to a root $z \neq x, y$ might differ from $\bar{d}_C(x,y)$. The following lemma ensures that this difference is bounded.

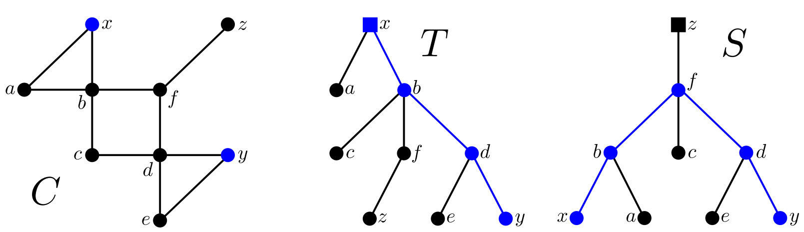

Figure 5.5: The trees T and S correspond to the rooted graph (C, x) and (C, z).

Lemma 5.3.2. *Let $C \in \mathcal{C}$ be a connected graph and x, y, z vertices of C. Let S be the tree corresponding to the graph C rooted at z. Then*

$$\bar{d}_C(x,y) \le d_S(x,y) \le \bar{d}_C(x,y) + 1.$$

Moreover, $\bar{d}_C$ is a metric on the vertex set $V(C)$.

Proof. Let T and S denote the trees corresponding to the graph C rooted at x and z. Consider the lowest common ancestor w of x and y in the tree S. We let $P = (p_1, \ldots, p_k)$ and $Q = (q_1, \ldots, q_\ell)$ denote the paths joining the vertices $x = p_1 = q_1$ and $y = p_k = q_\ell$ in T and S, respectively. If w lies on P, then $P = Q$ and consequently $d_S(x,y) = d_T(x,y)$. If w does not lie on P, then there is an index i with $Q = (p_1, \ldots, p_i, w, p_{i+1}, \ldots, p_k)$ and hence $d_S(x,y) = d_T(x,y) + 1$. Thus

$$d_S(x,y) = \bar{d}_C(x,y) + \mathbb{1}_{\{w \notin P\}}.$$

The case $z = y$ yields that $\bar{d}_C$ is symmetric. The triangle inequality follows from this fact and

$$\bar{d}_C(x,y) \le d_S(x,y) \le d_S(x,z) + d_S(z,y) = \bar{d}_C(z,x) + \bar{d}_C(z,y).$$

Clearly $\bar{d}_C$ is also reflexive and hence a metric. □

In the following lemma we apply the results on pointed enriched trees of Section 5.2.

Lemma 5.3.3. *Let $\mathcal{C}$ denote a subcritical class of connected graphs and set $\kappa = \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))]$. Then for all $s > 1$ and $0 < \epsilon < 1/2$ with $2\epsilon s > 1$ we have with a probability that tends to 1 as n becomes large that all $x, y \in V(\mathsf{C}_n)$ with $\bar{d}_{\mathsf{C}_n}(x,y) \ge \log^s(n)$ satisfy*

$$|d_{\mathsf{C}_n}(x,y) - \kappa \bar{d}_{\mathsf{C}_n}(x,y)| \le \bar{d}_{\mathsf{C}_n}(x,y)^{1/2+\epsilon}.$$

Proof. We denote $L_n = \log^s(n)$ and $t_\ell = \ell^{1/2+\epsilon}$. Let $\mathcal{E} \subset \mathcal{A}_{\mathcal{R}}^{\bullet} \simeq \mathcal{C}^{\bullet\bullet}$ with $\mathcal{R} = \mathrm{SET} \circ \mathcal{B}'$ denote the set of all bipointed graphs or pointed enriched trees $((C,x),y) \simeq ((T,\alpha),y)$, where we call x the *inner root* and y the *outer root*, such that

$$d_T(x,y) \ge L_{|T|} \quad \text{and} \quad |d_C(x,y) - \kappa d_T(x,y)| > t_{d_T(x,y)}.$$

We will bound the probability that there exist vertices x and y with $((\mathsf{C}_n,x),y) \in \mathcal{E}$. First observe that

$$\sum_{x,y \in [n]} \mathbb{P}(((\mathsf{C}_n,x),y) \in \mathcal{E}) = \sum_{((C,x),y) \in \mathcal{E}} \mathbb{P}(\mathsf{C}_n = C) = n \sum_{y=1}^{n} \mathbb{P}((\mathsf{C}_n^{\bullet}, y) \in \mathcal{E}).$$

By assumption we may apply Corollary 5.1.8 to obtain $\mathbb{P}(|\Gamma\mathcal{C}^{\bullet}(\rho)| = n) = \Theta(n^{-3/2})$. Moreover, Lemma 5.1.7 asserts that $\mathcal{B}'^{\bullet}(y) = 1$ and thus, with Lemma 5.1.1

$$\rho\mathcal{R}'(y) = \rho\mathcal{B}''(y)e^{\mathcal{B}'(y)} = y\mathcal{B}''(y) = 1.$$

Hence, by applying Lemma 5.2.1 we obtain that

$$\mathbb{P}(((\mathsf{C}_n, x), y) \in \mathcal{E} \text{ for some } x, y) \leq O(n^{5/2}) \sum_{\ell=L_n}^{n-1} \mathbb{P}(\Gamma \mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) \text{ has size } n \text{ and satisfies } \mathcal{E}).$$

The height of the outer root in the bipointed graph corresponding to $\Gamma A_R^{(\ell)}(\rho)$ is distributed like the sum of ℓ independent random variables, each distributed like the distance of the $*$-vertex and the root in the corresponding derived block of $\Gamma(\text{SET} \circ \mathcal{B}')^\bullet(y)$. Since $(\text{SET} \circ \mathcal{B}')^\bullet \simeq (\text{SET} \circ \mathcal{B}') \cdot \mathcal{B}'^\bullet$, these variables are actually $\mathsf{d}(\Gamma \mathcal{B}'^\bullet(y))$-distributed. Hence

$$\mathbb{P}(\Gamma \mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho) \in \mathcal{E}, |\Gamma \mathcal{A}_{\mathcal{R}}^{(\ell)}(\rho)| = n) \leq \mathbb{P}(|\eta_1 + \ldots + \eta_\ell - \ell \mathbb{E}[\eta_1]| > t_\ell)$$

with $(\eta_i)_i$ i.i.d. copies of $\eta := \mathsf{d}(\Gamma \mathcal{B}'^\bullet(y))$. Clearly we have that $\eta \leq |\Gamma \mathcal{B}'^\bullet(y)|$. Since $\mathcal{C}$ is subcritical it follows that there is a constant $\delta > 0$ such that $\mathbb{E}[e^{\eta\theta}] < \infty$ for all θ with $|\theta| \leq \delta$. Hence we may apply the standard moderate deviation inequality for one-dimensional random walk stated in Lemma 2.5.1 to obtain for some constant $c > 0$

$$\mathbb{P}(((\mathsf{C}_n, x), y) \in \mathcal{E} \text{ for some } x, y) \leq O(n^{7/2}) \exp(-c(\log n)^{2s\epsilon}) = o(1).$$

□

It remains to clarify what happens if $\bar{d}_{\mathsf{C}_n}$ is small. We prove the following statement for random graphs from block-stable classes that are not necessarily subcritical.

Proposition 5.3.4. *Let $\mathcal{C}$ be a block-stable class of connected graphs. Suppose that $\mathcal{B}'^\bullet(y) = 1$ and the offspring distribution ξ has finite second moment, i.e. $\mathcal{B}'''(y) < \infty$. Let $\mathsf{lb}(\mathsf{C}_n)$ denote the size of the largest block in C_n,*

1. *For any $x, y \in \mathsf{C}_n$ we have $d_{\mathsf{C}_n}(x, y) \leq \bar{d}_{\mathsf{C}_n}(x, y) \mathsf{lb}(\mathsf{C}_n)$.*
2. *If the offspring distribution ξ is bounded, then so is $\mathsf{lb}(\mathsf{C}_n)$. Otherwise, for any sequence K_n we have $\mathbb{P}(\mathsf{lb}(\mathsf{C}_n) \geq K_n) = O(n)\mathbb{P}(\xi \geq K_n)$.*

Proof. We have that $d_{\mathsf{C}_n} \leq \bar{d}_{\mathsf{C}_n}(\mathsf{lb}(\mathsf{C}_n) - 1)$ and $\mathsf{lb}(\mathsf{C}_n) = \mathsf{lb}(\mathsf{C}_n^\bullet) \leq \Delta(\mathsf{T}_n) + 1$ with $\Delta(\mathsf{T}_n)$ denoting the largest outdegree. Recall that $\Delta(\mathsf{T}_n)$ is distributed like the maximum degree of a ξ-Galton-Watson tree conditioned to have n vertices. By assumption, the offspring distribution ξ has expected value $\mathbb{E}[\xi] = \mathcal{B}'^\bullet(y) = 1$ and finite variance.

If ξ is bounded, then so is the largest outdegree of T_n. Otherwise, as argued in the proof of [Jan12, Eq. (19.20)], for any sequence K_n

$$\mathbb{P}(\Delta(\mathsf{T}_n) \geq K_n) \leq (1 + o(1))n\mathbb{P}(\xi \geq K_n). \tag{5.3.2}$$

Applying (5.3.2) yields $\mathbb{P}(\mathsf{lb}(\mathsf{C}_n) \geq K_n) \leq (1 + o(1))n\mathbb{P}(\xi \geq K_n)$ for any sequence K_n. □

Note that if $\mathcal{C}$ is subcritical then this implies that $\mathsf{lb}(\mathsf{C}_n) = O(\log n)$ with a probability that tends to 1: the definition of the Boltzmann model and the fact that y is smaller than the radius of convergence of $\mathcal{B}(z)$ guarantee that there is a constant $\beta < 1$ such that

$$\mathbb{P}(\xi = k) = \mathbb{P}(|\Gamma(\mathrm{SET} \circ \mathcal{B}')(y)| = k) = O(\beta^k).$$

Combined with the bounds of Lemma 5.3.3 this yields the following concentration result.

Corollary 5.3.5. *Let $\mathcal{C}$ be a subcritical class of connected graphs. Then for all $s > 1$ and $0 < \epsilon < 1/2$ with $2\epsilon s > 1$ we have with a probability that tends to 1 as n becomes large that for all vertices $x, y \in V(\mathsf{C}_n)$*

$$|d_{\mathsf{C}_n}(x,y) - \kappa \bar{d}_{\mathsf{C}_n}(x,y)| \le \bar{d}_{\mathsf{C}_n}(x,y)^{1/2+\epsilon} + O(\log^{s+1}(n)).$$

We may now prove the main theorem.

Proof of (5.3.1). Lemma 5.3.2 implies that $\bar{d}_{\mathsf{C}_n} \le d_{\mathsf{T}_n} \le \bar{d}_{\mathsf{C}_n} + 1$. By Corollary 5.3.5, and considering the distortion of the identity map as correspondence between the vertices of T_n and $\mathsf{C}_n^\bullet$, it follows that with a probability that tends to 1 as n becomes large

$$d_{\mathrm{GH}}(\mathsf{C}_n^\bullet/(\kappa\sqrt{n}), \mathsf{T}_n/\sqrt{n}) \le \mathrm{D}(\mathsf{T}_n)^{3/4}/\sqrt{n} + o(1).$$

Using the tail bounds given in [ABDJ13, Thm. 1.2] for the diameter $\mathrm{D}(\mathsf{T}_n)$ we obtain that $d_{\mathrm{GH}}(\mathsf{C}_n^\bullet/(\kappa\sqrt{n}), \mathsf{T}_n/\sqrt{n})$ converges in probability to zero. Recall that the variance of the offspring distribution ξ is given by $\sigma^2 = \mathbb{E}[|\Gamma\mathcal{B}'^\bullet(y)|]$. We have that $\frac{\sigma}{2\sqrt{n}}\mathsf{T}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}$ and thus $\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{C}_n^\bullet \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}}$. □

5.4 Exponential tail bounds for the height and diameter

In this section we provide a proof for Theorem 1.4.2. Our proof builds on results obtained in [ABDJ13]. Recall that (T_n, α_n) denotes the enriched tree corresponding to the graph $\mathsf{C}_n^\bullet$ and that T_n has a natural coupling with a ξ-Galton-Watson conditioned on having size n, see Proposition 5.1.5. With (slight) abuse of notation we also write T_n for the conditioned ξ-Galton-Watson tree within this section. We prove Theorem 1.4.2 by showing the following more general result for random graphs from block-stable classes that are not necessarily subcritical.

Theorem 5.4.1. *Let $\mathcal{C}$ be a block-stable class of connected graphs. Suppose that $\mathcal{C}$ satisfies $\mathcal{B}'^\bullet(y) = 1$ and the offspring distribution ξ has finite variance, i.e. $\mathcal{B}'''(y) < 1$. Then there are $C, c > 0$ such that for all $n, x \ge 0$*

$$\mathbb{P}(D(\mathsf{C}_n) \ge x) \le C\exp(-cx^2/n) \quad \textit{and} \quad \mathbb{P}(H(\mathsf{C}_n^\bullet) \ge x) \le C\exp(-cx^2/n).$$

As a main ingredient in our proof we consider the *lexicographic depth-first-search (DFS)* of the plane tree T_n by labeling the vertices in the usual way (as a subtree of the Ulam-Harris tree) by finite sequences of integers and listing them in lexicographic order $v_0, v_1, \ldots, v_{n-1}$. The search keeps a queue of Q_i^d nodes with $Q_0^d = 1$ and the recursion

$$Q_i^d = Q_{i-1}^d - 1 + d^+_{\mathsf{T}_n}(v_{i-1}).$$

The mirror-image of T_n is obtain by reversing the ordering on each offspring set and the *reverse* DFS Q_i^r is defined as the DFS of the mirror-image. Then $(Q_i^d)_{0 \le i \le n}$ and $(Q_i^r)_{0 \le i \le n}$ are identically distributed and satisfy the following bound given in [ABDJ13, Ineq. (4.4)]:

$$\mathbb{P}(\max_j Q_j^d \ge x) \le C \exp(-cx^2/n) \tag{5.4.1}$$

with $C, c > 0$ denoting some constants that do not depend on x or n.

Proof of Theorem 5.4.1. Since $\mathrm{D}(\mathsf{C}_n) \le 2\mathrm{H}(\mathsf{C}_n^\bullet)$ it suffices to show the bound for the height. Let $h \ge 0$. If $\mathrm{H}(\mathsf{C}_n^\bullet) \ge h$ then there exists a vertex whose height equals h. Consequently, we may estimate $\mathbb{P}(\mathrm{H}(\mathsf{C}_n^\bullet) \ge h) \le \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2)$ with $\mathcal{E}_1$ (resp. $\mathcal{E}_2$) denoting the event that there is a vertex v such that $h_{\mathsf{C}_n^\bullet}(v) = h$ and $h_{\mathsf{T}_n}(v) \ge h/2$ (resp. $h_{\mathsf{T}_n}(v) \le h/2$). By the tail bound [ABDJ13, Thm. 1.2] for the height of Galton-Watson trees we obtain

$$\mathbb{P}(\mathcal{E}_1) \le \mathbb{P}(\mathrm{H}(\mathsf{T}_n) \ge h/2) \le C_2 \exp(-c_2 h^2/(4n))$$

for some constants $C_2, c_2 > 0$. In order to bound $\mathbb{P}(\mathcal{E}_2)$ suppose that there is a vertex v with height $h_{\mathsf{C}_n^\bullet}(v) = h$ and $h_{\mathsf{T}_n}(v) \le h/2$. If a is a vertex of T_n and b one of its offspring, then $d_{\mathsf{C}_n^\bullet}(a,b) \le d^+_{\mathsf{T}_n}(a)$. Hence

$$\sum_{u \succ v} d^+_{\mathsf{T}_n}(u) \ge h_{\mathsf{C}_n^\bullet}(v) = h$$

with the sum index u ranging over all ancestors of v. Consider the lexicographic depth-first-search $(Q_i^d)_i$ and reverse depth-first-search $(Q_i^r)_i$ of T_n. Let j (resp. k) denote the index corresponding to the vertex v in the lexicographic (resp. reverse lexicographic) order. It follows from the definition of the queues that if $\mathcal{E}_2$ occurs

$$Q_j^d + Q_k^r = 2 + \sum_{u \succ v} d^+_{\mathsf{T}_n}(u) - h_{\mathsf{T}_n}(v) \ge h/2$$

and hence $\max(Q_j^d, Q_k^r) \ge h/4$. Since Q_j^d and Q_k^r are identically distributed it follows by (5.4.1) that

$$\begin{aligned}\mathbb{P}(\mathcal{E}_2) &\le \mathbb{P}(\max_i(Q_i^d) \ge h/4) + \mathbb{P}(\max_i(Q_i^r) \ge h/4) \\ &\le 2\mathbb{P}(\max_i(Q_i^d) \ge h/4) \\ &\le 2C \exp(-ch^2/(16n)).\end{aligned}$$

This concludes the proof. □

5.5 Extensions

In the following we use the notation from Section 5.3.

5.5.1 First passage percolation

Let $\omega > 0$ be a given random variable such that there is a $\delta > 0$ with $\mathbb{E}[e^{\theta\omega}] < \infty$ for all θ with $|\theta| \le \delta$. For any graph G we may consider the random graph $\hat{G}$ obtained by assigning to each edge $e \in E(G)$ a weight ω_e that is an independent copy of ω. The $d_{\hat{G}}$-distance of two vertices a and b is then given by

$$d_{\hat{G}}(a,b) = \inf\Big\{ \sum_{e\in E(P)} \omega_e \mid P \text{ a path connecting } a \text{ and } b \text{ in } G\Big\}.$$

We are going to prove Proposition 1.4.3 as follows. Let $\mathcal{C}$ be a subcritical class of connected graphs and $\mathcal{B}$ its subclass of graphs that are 2-connected or a single edge with its ends. Let $\mathsf{C}_n \in \mathcal{C}_n$ and $\mathsf{C}_n^\bullet \in \mathcal{C}_n^\bullet$ denote the uniform (rooted) random graphs. Form the link-weighted versions $\hat{\mathsf{C}}_n$ and $\hat{\mathsf{C}}_n^\bullet$ by assigning to each edge an independent copy of a random variable $\omega > 0$ having finite exponential moments. Then

$$\frac{\sigma}{2\hat{\kappa}\sqrt{n}}\hat{\mathsf{C}}_n^\bullet \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}} \quad \text{and} \quad \frac{\sigma}{2\hat{\kappa}\sqrt{n}}\hat{\mathsf{C}}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}} \tag{5.5.1}$$

with respect to the (pointed) Gromov-Hausdorff metric. The scaling constant $\hat{\kappa}$ is given by $\hat{\kappa} := \mathbb{E}[\mathsf{d}(\hat{\mathsf{B}})]$ with B drawn according to the Boltzmann sampler $\Gamma B'^\bullet(y)$ and $\mathsf{d}(\hat{\mathsf{B}})$ denoting the $d_{\hat{\mathsf{B}}}$-distance from the $*$-vertex to the root vertex.

Proof of (5.5.1). For any n let K_n denote the complete graph with n vertices. The idea is to generate $\hat{\mathsf{C}}_n$ by drawing C_n and $\hat{K}_n$ independently and assign the weights via the inclusion $E(\mathsf{C}_n) \subset E(K_n)$. By considering subsets $\mathcal{E} \subset \mathcal{C}^{\bullet\bullet} \times \mathbb{R}^{\cup_n E(K_n)}$ we may easily prove a weighted version of Lemma 5.2.1, i.e. the probability that the random pair $(\mathsf{C}_n^{\bullet\bullet}, \hat{K}_n)$ has some property $\mathcal{E}$ is bounded by

$$O(n^{5/2}) \sum_{\ell=0}^{n-1} \mathbb{P}(|\Gamma C^{\bullet(\ell)}(\rho)| = n, (\Gamma C^{\bullet(\ell)}(\rho), \hat{K}_n) \in \mathcal{E}).$$

This implies that the blocks along sufficiently long paths in the random graphs $\hat{\mathsf{C}}_n$ behave like independent copies of the weighted block $\hat{\mathsf{B}}$ with B drawn according to the Boltzmann sampler $\Gamma B'^\bullet(y)$. Hence, weighted versions of Lemma 5.3.3 and Proposition 5.3.4 may be deduced analogously to their original proofs with $\hat{\kappa}$ replacing κ and only minor modifications otherwise. Thus the scaling limit follows in the same fashion. $\square$

5.5.2 Random graphs given by their connected components

We study the case of an arbitrary graph consisting of a set of connected components. Let $\mathcal{G} \simeq \mathrm{SET} \circ \mathcal{C}$ denote a subcritical graph class given by its subclass $\mathcal{C}$ of connected graphs. For simplicity we are going to assume that all trees belong to the class $\mathcal{C}$.

Consider the uniform random graph $\mathsf{G}_n \in \mathcal{G}_n$. Of course we cannot expect G_n to converge to the Continuum Random Tree since it is disconnected with a probability that is bounded away from zero. Instead we study a uniformly chosen component H_n of maximal size. We are going to prove Proposition 1.4.4 by showing

$$\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{H}_n \xrightarrow{(d)} \mathcal{T}_{\mathsf{e}} \tag{5.5.2}$$

with respect to the Gromov-Hausdorff metric, where σ, κ are as in (5.3.1).

We are going to use the known fact that with a probability that tends to 1 as n becomes large the random graph G_n has a unique giant component with size $n + O_p(1)$. This follows for example from [MSW06, Thm. 6.4].

Lemma 5.5.1. *If $\mathcal{C}$ contains all trees, then the size of a largest component satisfies $|\mathsf{H}_n| = n + O_p(1)$.*

Proof of (5.5.2). Let $f : \mathbb{K} \to \mathbb{R}$ be a bounded Lipschitz-continuous function defined on the space of isometry classes of compact metric spaces. We will show that $\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{H}_n)] \to \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})]$ as n tends to infinity. Set $\Omega_n := \log n$. By Lemma 5.5.1 we know that $|\mathsf{H}_n| = n + O_p(1)$. Hence with a probability that tends to 1 as n becomes large we have that $n - |H_n| \le \Omega_n$ and thus

$$\mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{H}_n\right)\right] = o(1) + \sum_{0 \le k \le \Omega_n} \mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{H}_n\right) \mid |\mathsf{H}_n| = n-k\right] \mathbb{P}(|\mathsf{H}_n| = n-k).$$

The conditional distribution of G_n given the sizes $(s_i)_i$ of its components is given by choosing components $K_i \in \mathcal{C}[s_i]$ independently uniformly at random and distributing labels uniformly at random. In particular, as a metric space, H_n conditioned on $|\mathsf{H}_n| = n - k$ is distributed like the uniform random graph C_{n-k}. Thus, given $\epsilon > 0$ we have for n sufficiently large by Lipschitz-continuity

$$\mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{H}_n\right) \mid |\mathsf{H}_n| = n-k\right] = \mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{C}_{n-k}\right)\right] \in \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})] \pm \epsilon$$

for all $0 \le k \le \Omega_n$. Thus $|\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{H}_n)] - \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})]| \le \epsilon$ for sufficiently large n. Since $\epsilon > 0$ was chosen arbitrarily it follows that $\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{H}_n)] \to \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})]$ as n tends to infinity. □

5.6 The scaling factor of specific classes

In this section we apply our main results to several specific examples of subcritical graph classes. The notation that will be fixed throughout this section is as follows:

Graph Class	κ	H	c	ρ	y	λ	σ^2
Trees = $\mathsf{Forb}(C_3)$	1	2.50662	0.39894	0.36787	1	1	1
$\mathsf{Forb}(C_4)$	1	2.13226	0.20973	0.23618	0.27520	0.80901	1.38196
$\mathsf{Forb}(C_5)$	1.10355	1.88657	0.10987	0.06290	0.40384	1.85945	2.14989
Cacti Graphs	1.20297	1.99021	0.12014	0.23874	0.45631	0.64779	2.29559
Outerplanar Graphs	5.08418	1.30501	0.00697	0.13659	0.17076	0.22327	95.3658

Table 5.1: Numerical approximations of constants for examples of subcritical classes of connected graphs.

$\mathcal{C}$ denotes a subcritical class of connected graphs and $\mathcal{B}$ its subclass of 2-connected graphs and edges. The radius of convergence of $\mathcal{C}(z)$ is denoted by ρ. The constant $y = \mathcal{C}^\bullet(\rho)$ is the unique positive solution of the equation

$$y\mathcal{B}''(y) = 1.$$

By Lemma 5.1.1 this determines $\rho = y\exp(-\mathcal{B}'(y))$. Moreover, we set

$$\kappa = \mathbb{E}[\mathsf{d}(\Gamma\mathcal{B}'^\bullet(y))],$$

i.e. the expected distance from the $*$-vertex to the root in a random block chosen according to the Boltzmann distribution with parameter y. We call κ the *scaling factor* for $\mathcal{C}$. The offspring distribution ξ of the random tree corresponding to the sampler $\Gamma\mathcal{C}^\bullet(y)$ has probability generating function $\varphi(z) = \exp(\mathcal{B}'(yz) - \lambda)$ with $\lambda = \mathcal{B}'(y)$, see Proposition 5.1.5. Its variance is given by

$$\sigma^2 = 1 + \mathcal{B}'''(y)y^2 = \mathbb{E}[|\Gamma\mathcal{B}'^\bullet(y)|].$$

We let d denote the span of the offspring distribution. By applying our main results Theorems 1.4.1 and 1.4.1 we obtain

$$\mathbb{E}[\mathsf{H}(\mathsf{C}_n^\bullet)]/\sqrt{n} \to \kappa\sqrt{2\pi/\sigma^2} =: H \quad \text{as } n \to \infty \text{ with } n \equiv 1 \mod d$$

with $\mathsf{C}_n^\bullet \in \mathcal{C}_n^\bullet$ drawn uniformly at random. We call H the *expected rescaled height*. Moreover, Corollary 5.1.8 yields that

$$|\mathcal{C}_n| \sim c\, n^{-5/2}\rho_C^{-n} n! \quad \text{as } n \to \infty \text{ with } n \equiv 1 \mod d$$

with $c = yd/\sqrt{2\pi\sigma^2}$. In this section we derive analytical expressions for the relevant constants $\kappa, H, c, \rho, y, \lambda, \sigma^2$ for several graph classes; Table 5.1 provides numerical approximations. For a set of graphs M, we denote by $\mathsf{Forb}(M)$ the class of all connected graphs that contain none of the graphs in M as a topological minor; if M contains only 2-connected graphs, then it is easy to see that $\mathsf{Forb}(M)$ is block-stable, cf. [Die10]. For $n \geq 3$ we denote by C_n a graph that is isomorphic to a cycle with n vertices.

Remark 5.6.1. *The average blocksize* $\mathsf{b}(\mathsf{C}_n) \overset{(d)}{=} \mathsf{b}(\mathsf{C}_n^\bullet)$ *is concentrated around one plus the average size of* $\Gamma B'(y)$. *With high probability as* $n \equiv 1 \mod d$ *tends to infinity we have*

$$\mathsf{b}(\mathsf{C}_n) \in 1 + \mathbb{E}[|\Gamma\mathcal{B}'(y)|] \pm O(\log n/\sqrt{n}) \quad \textit{with} \quad \mathbb{E}[|\Gamma\mathcal{B}'(y)|] = 1/\lambda. \tag{5.6.1}$$

Proof of (5.6.1). The random graph $\mathsf{C}_n^\bullet$ is distributed like the Boltzmann-sampler $\Gamma\mathcal{C}^\bullet(\rho)$ conditioned on having size n. We may interpret the sampler $\Gamma\mathcal{C}^\bullet(\rho)$ as a deterministic procedure reading from an infinite i.i.d. list $(A_i)_{i\in\mathbb{N}}$ of random $\mathrm{SET}\circ\mathcal{B}'$-objects drawn according to the corresponding Boltzmann distribution with parameter y. The procedure starts by identifying the $*$-vertices of the blocks of the object A_1 with the root and marks the root as "touched". Then it recurses for every still untouched vertex, always using the leftmost unused $\mathrm{SET}\circ\mathcal{B}'$-object from the list. After $k \geq 1$ steps, the total size $1 + \sum_{i=1}^k |A_i|$ is greater or equal to k and the process stops if this sum is equal to n. Hence

$$\left\{|\Gamma\mathcal{C}^\bullet(\rho)| = n\right\} = \left\{\sum_{i=1}^n |A_i| = n-1, \sum_{i=1}^k A_i \geq k \text{ for all } k < n\right\}.$$

Each A_i is generated by drawing a $\mathsf{Pois}\,(\lambda_C)$-generated number m_i and sampling accordingly many i.i.d. blocks $B_1^{[i]}, \ldots, B_{m_i}^{[i]}$ using the sampler $\Gamma B'(y)$. If the procedure $\Gamma C^\bullet(\rho)$ terminates with an object of size n, then the total size of the derived blocks sum up to $n-1$. Hence the average block size b is given by

$$b = 1 + (n-1)/N$$

where $N = \sum_{i=1}^n m_i$ denotes the number of blocks. With foresight, let $\mathcal{E}$ denote the event $N \notin n\lambda_C(1 \pm a_n)$ with $a_n = \log n/\sqrt{n}$. Using Corollary 5.1.8 we get

$$\mathbb{P}(\mathcal{E} \mid |\Gamma\mathcal{C}^\bullet(\rho)| = n) = \Theta(n^{3/2})\,\mathbb{P}(\mathcal{E}, \sum_{i=1}^n |A_i| = n-1, \sum_{i=1}^k A_i \geq k \text{ for all } k < n).$$

The number of blocks does not depend on the order of the A_i's. We may apply the standard rotation argument to obtain

$$\mathbb{P}(\mathcal{E} \mid |\Gamma\mathcal{C}^\bullet(\rho)| = n) = \Theta(\sqrt{n})\,\mathbb{P}(\mathcal{E}, \sum_{i=1}^n |A_i| = n-1).$$

Applying the well-known Chernoff bounds yields that

$$\mathbb{P}(\sum_{i=1}^n m_i \notin \lambda_C(1 \pm a_n)) \leq 2\exp(-a_n^2 n\lambda_C/3) = o(n^{-1/2}).$$

By monotonicity it follows that $N \in n\lambda_C(1 \pm a_n)$ with high probability. Hence it holds that

$$\mathsf{b}(\mathsf{C}_n^\bullet) \in 1 + 1/\lambda \pm a_n$$

with high probability. Moreover $1/\lambda = \mathbb{E}[|\Gamma\mathcal{B}'(y)|]$ since $\mathcal{B}'^\bullet(y) = 1$, and the proof is complete. □

5.6.1 Trees

Let $\mathcal{C}$ be the class of trees, i.e. $\mathcal{B}$ consists only of the graph K_2. It is easy to see that the offspring distribution follows a Poisson distribution with parameter one. We immediately obtain:

Proposition 5.6.2. *For the class of trees we have $\kappa = 1$ and $\sigma^2 = 1$.*

5.6.2 $\mathsf{Forb}(C_4)$

Let $\mathcal{C}$ denote the connected graphs of the class $\mathsf{Forb}(C_4)$. Then each block is either isomorphic to K_2 or K_3. Hence $\mathcal{B}(z) = z^2/2 + z^3/6$. Moreover, for any $B \in \mathcal{B}$ and any two distinct vertices in B their distance is one. A simple computation then yields:

Proposition 5.6.3. *For the class $\mathsf{Forb}(C_4)$ we have $\kappa = 1$ and $\sigma^2 = (5 - \sqrt{5})/2$.*

5.6.3 $\mathsf{Forb}(C_5)$

Recall that the class $\mathsf{Forb}(C_5)$ consists of all graphs that do not contain a cycle with five vertices as a topological minor. Hence, a graph belongs to this class if and only if it contains no cycle of length at least five as subgraph.

Proposition 5.6.4. *For the class $\mathsf{Forb}(C_5)$ the constant y is the unique positive solution to $z\mathcal{B}''(z) = 1$, where $\mathcal{B}'(z)$ is given in (5.6.4). Moreover, we have*

$$\kappa = \left(2\,y^2 + 4\,y + 3\right) ye^y - \left(3\,y^2 + 12\,y + 4\right) y/2 \approx 1.10355.$$

and $\sigma^2 = 1 + B'''(y)y^2 \approx 2.14989$.

Before proving Proposition 5.6.4 we identify the unlabeled blocks of this class. This result (among extensions to $\mathsf{Forb}(C_6)$ and $\mathsf{Forb}(C_7)$) was given by Giménez, Mitsche and Noy [GMN13].

Proposition 5.6.5. *The unlabeled blocks of the class $\mathsf{Forb}(C_5)$ are given by*

$$K_2, K_4, (K_{2,m})_{m\geq 1}, (K_{2,m}^+)_{m\geq 2}. \tag{5.6.2}$$

Here K_n denotes the complete graph and $K_{m,n}$ the complete bipartite graph with bipartition $\{A_m, B_n\}$. The graph $K_{2,n}^+$ is obtained from $K_{2,n}$ by adding an additional edge between the two vertices from A_2.

Proof. We may verify (5.6.2) by considering the standard decomposition of 2-connected graphs: an arbitrary graph G is 2-connected if and only if it can be constructed from a cycle by adding H-paths to already constructed graphs H [Die10]. If $G \in \mathsf{Forb}(C_5)$, then so do all the graphs along its decomposition. In particular we must start with a triangle or a square. Since every edge of a 2-connected graph lies on a cycle, we may only add paths of length at most two in each step. In particular, for $m \geq 3$ a $K_{2,m}$ may only become a $K_{2,m}^+$ or $K_{2,m+1}$, and a $K_{2,m}^+$ may only become a $K_{2,m+1}^+$. Thus (5.6.2) follows by induction on the number of vertices. □

Proof of Proposition 5.6.4. With foresight, we use the decomposition

$$\mathcal{B} = \mathcal{S} + \mathcal{H} + \mathcal{P} \tag{5.6.3}$$

with the classes of labeled graphs $\mathcal{S}$, $\mathcal{H}$ and $\mathcal{P}$ defined by their sets of unlabeled graphs $\tilde{\mathcal{S}} = \{K_2, K_3, K_4, C_4\}$, $\tilde{\mathcal{H}} = \{K_{2,m} \mid m \geq 3\}$ and $\tilde{\mathcal{P}} = \{K^+_{2,m} \mid m \geq 2\}$. Any unlabeled graph from $\mathcal{H}$ or $\mathcal{P}$ with n vertices has exactly $\binom{n}{2}$ different labelings, since any labeling is determined by the choice of the two unique vertices with degree at least three. Hence

$$\mathcal{S}(x) = x^2/2 + x^3/6 + x^4/6, \quad \mathcal{H}(x) = \sum_{n\geq 5} \binom{n}{2} \frac{x^n}{n!} \quad \text{and} \quad \mathcal{P}(x) = \sum_{n\geq 4} \binom{n}{2} \frac{x^n}{n!}$$

and thus

$$\mathcal{B}'(x) = x(x+2)\mathrm{e}^x - x\left(15\,x + 2\,x^2 + 6\right)/6. \tag{5.6.4}$$

Solving the equation $\mathcal{B}'^{\bullet}(y) = 1$ yields

$$y \approx 0.40384.$$

First, let $\mathsf{H}_n \in \mathcal{H}'^{\bullet}_n$ with $n \geq 4$ be drawn uniformly at random. We say that a vertex lies on the left if it has degree at least three, otherwise we say it lies on the right. There are $n\binom{n+1}{2}$ graphs in the set $\mathcal{H}'^{\bullet}_n$ and precisely n^2 of those have the property that the $*$-vertex lies on the left. The distance of the root and the $*$-vertex equals two if they lie on the same side and one otherwise. Hence

$$\mathbb{E}[\mathsf{d}(\mathsf{H}_n)] = \frac{n}{\binom{n+1}{2}}\left(\frac{1}{n}\cdot 2 + \frac{n-1}{n}\cdot 1\right) + \left(1 - \frac{n}{\binom{n+1}{2}}\right)\left(\frac{2}{n}\cdot 1 + \frac{n-2}{n}\cdot 2\right).$$

For any n let $\mathsf{P}_n \in \mathcal{P}'^{\bullet}_n$ and $\mathsf{S}_n \in \mathcal{S}_n$ be drawn uniformly at random. Analogously to the above calculation we obtain

$$\mathbb{E}[\mathsf{d}(\mathsf{P}_n)] = \frac{n}{\binom{n+1}{2}} 1 + \left(1 - \frac{n}{\binom{n+1}{2}}\right)\left(\frac{2}{n}\cdot 1 + \frac{n-2}{n}\cdot 2\right)$$

and

$$\mathbb{E}[\mathsf{d}(\mathsf{S}_1)] = \mathbb{E}[\mathsf{d}(\mathsf{S}_2)] = 1, \qquad \mathbb{E}[\mathsf{d}(\mathsf{S}_3)] = \frac{1}{4}\cdot 1 + \frac{3}{4}\left(\frac{2}{3}\cdot 1 + \frac{1}{3}\cdot 2\right) = \frac{5}{4}.$$

Since $\mathcal{B}'^{\bullet}(y) = 1$ we have for any class $\mathcal{F} \in \{\mathcal{S}'^{\bullet}, \mathcal{H}'^{\bullet}, \mathcal{P}'^{\bullet}\}$ that

$$\mathbb{E}[\mathsf{d}(\Gamma\mathcal{B}'^{\bullet}(y)), \Gamma B'^{\bullet}(y) \in \mathcal{F}] = \sum_n \left([z^n]\mathcal{F}(yz)\right)\mathbb{E}[\mathsf{d}(\mathsf{F}_n)].$$

Summing up yields

$$\mathbb{E}[\mathsf{d}(\Gamma\mathcal{B}'^{\bullet}(y))] = \left(2\,y^2 + 4\,y + 3\right)ye^y - \left(3\,y^2 + 12\,y + 4\right)y/2 \approx 1.10355.$$

□

5.6.4 Cacti graphs

A cactus graph is a graph in which each edge is contained in at most one cycle. Equivalently, the class of cacti graphs is the block-stable class of graphs where every block is either an edge or a cycle. In the following $\mathcal{C}$ denotes the class of cacti graphs.

Proposition 5.6.6. *For the class of cacti graphs the constant y is the unique positive solution to $z\mathcal{B}''(z) = 1$, where $\mathcal{B}'(z)$ is given in* (5.6.5)*. Moreover, we have*

$$\kappa = \frac{y^4 - 2\,y^3 + 2\,y - 2}{(y^2 - 2\,y + 2)\,(1 + y)\,(y - 1)} \approx 1.20297.$$

and $\sigma^2 = 1 + \mathcal{B}'''(y)y^2 \approx 2.29559$.

Proof. By counting the number of labelings of a cycle, we obtain $|\mathcal{B}'_n| = n!/2$ for $n \geq 2$. It follows that

$$\mathcal{B}'(z) = z + \frac{z^2}{2(1-z)} \tag{5.6.5}$$

and hence $\mathcal{B}'^{\bullet}(z) = z + \frac{1}{2}\sum_{n\geq 2} nz^n = \frac{z^3-2z^2+2z}{2(z-1)^2}$. Solving the equation $\mathcal{B}'^{\bullet}(y) = 1$ yields

$$y = -\frac{1}{3}(17 + 3\,\sqrt{33})^{1/3} + \frac{2}{3}(17 + 3\,\sqrt{33})^{-1/3} + \frac{4}{3} \approx 0.45631.$$

Let $\Gamma\mathcal{B}'^{\bullet}(y)$ denote a Boltzmann-sampler for the class $\mathcal{B}'^{\bullet}$ with parameter y and for any $n \geq 1$ let $\mathsf{B}_n \in \mathcal{B}'^{\bullet}_n$ be drawn uniformly at random. Since $\mathcal{B}'^{\bullet}(y) = 1$, it follows that

$$\kappa = \mathbb{E}[\mathbb{E}[\mathsf{d}(\Gamma\mathcal{B}'^{\bullet}(y)) \mid |\Gamma\mathcal{B}'^{\bullet}(y)|]] = \sum_{n\geq 1} \mathsf{d}(\mathsf{B}_n)[z^n]\mathcal{B}'^{\bullet}(yz) = \mathsf{d}(B_1)y + \frac{1}{2}\sum_{n\geq 2} \mathsf{d}(\mathsf{B}_n)ny^n.$$

Clearly $\mathsf{d}(\mathsf{B}_1) = 1$ and for $n \geq 2$ we have that $\mathsf{d}(\mathsf{B}_n)$ is distributed like the distance from the $*$-vertex to a uniformly at random chosen root from $[n]$ in the cycle $(*, 1, 2, \ldots, n)$. Hence

$$\mathsf{d}(\mathsf{B}_n) = \begin{cases} \frac{2}{n}\sum_{i=1}^{n/2} i = \frac{n+2}{4}, & n \text{ is even} \\ \frac{n+1}{2n} + \frac{2}{n}\sum_{i=1}^{(n-1)/2} i = \frac{(n+1)^2}{4n}, & n \text{ is odd} \end{cases}.$$

Summing up over all possible values of n yields the claimed expression for κ. □

5.6.5 Outerplanar graphs

An outerplanar graph is a planar graph that can be embedded in the plane in such a way that every vertex lies on the boundary of the outer face. Any such embedding (considered up to continuous deformation) is termed an outerplanar map. The scaling limit of the model "all outerplanar maps with n vertices equally likely" was studied by Caraceni [Car], who established convergence to the CRT using a bijection by Bonichon, Gavoille and Hanusse [BGH05]. Our results allow us to study the model "all outerplanar graphs with n vertices equally likely", which is a different setting. Note also that the scaling factor obtained in the following differs from the one established for outerplanar maps.

Let $\mathcal{C}$ denote the class of connected outerplanar graphs and $\mathcal{B}$ the subclass consisting of single edges or 2-connected outerplanar graphs.

Proposition 5.6.7. *For the class of outerplanar graphs the constant y is the unique positive solution to $z\mathcal{B}''(z) = 1$, where $\mathcal{B}'(z) = (z+\mathcal{D}(z))/2$ and $\mathcal{D}$ is given in* (5.6.6). *Moreover,*

$$\kappa = \frac{y}{2} + \Big(1 - \frac{y}{2}\Big) \frac{8\,w^4 - 16\,w^3 + 4\,w - 1}{(4\,w^3 - 6\,w^2 - 2\,w + 1)\,(2\,w - 1)} \approx 5.0841 \textit{ with } w = \mathcal{D}(y)$$

and $\sigma^2 = 1 + \mathcal{B}'''(y)y^2 \approx 95.3658$.

Following [BPS10] we develop a specification of $\mathcal{B}'^{\bullet}$ that eventually will enable us to prove the above expressions of the relevant constants. Any 2-connected outerplanar graph has a unique Hamilton cycle, which corresponds to the boundary of the outer face in any drawing having the property that all vertices lie on the outer face. The edge set of a 2-connected outerplanar graph can thus be partitioned in two parts: the edges of the Hamilton cycle, and all other edges, which we refer to as the set of *chords*. Let $\mathcal{D}$ denote the class obtained from $\mathcal{B}'$ by orienting the Hamilton cycle of each object of size at least three in one of the two directions and marking the oriented edge whose tail is the $*$-vertex. The block consisting of a single edge is oriented in the unique way such that the $*$-vertex is the tail of the marked edge. We start with some observations.

Lemma 5.6.8. *We have that $\mathcal{B}'(z) = (z + \mathcal{D}(z))/2$ and*

$$\mathbb{E}[\mathsf{d}(\Gamma\mathcal{B}'^{\bullet}(x)] = \frac{x}{2\mathcal{B}'^{\bullet}(x)} + (1 - \frac{x}{2\mathcal{B}'^{\bullet}(x)})\mathbb{E}[\mathsf{d}(\Gamma\mathcal{D}^{\bullet}(x))].$$

Proof. We have an isomorphism $\mathcal{B}' + \mathcal{B}' =: 2\mathcal{B}' \simeq \mathcal{X} + \mathcal{D}$. Consequently, the classes $\mathcal{B}'^{\bullet}$ and $\mathcal{D}^{\bullet}$ obtained by additionally rooting at a non-$*$-vertex satisfy

$$2\mathcal{B}'^{\bullet} \simeq \mathcal{X} + \mathcal{D}^{\bullet}.$$

Hence the following procedure is a Boltzmann sampler for the class $\mathcal{B}'^{\bullet}$ with parameter x.

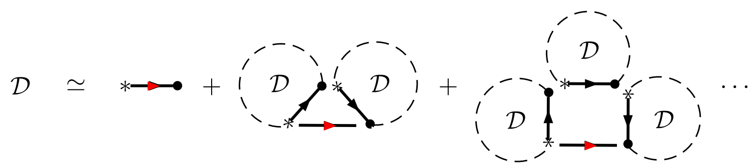

Figure 5.6: Recursive specification of the class $\mathcal{D}$.

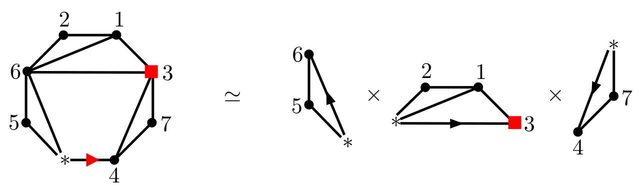

Figure 5.7: Decomposition of a $\mathcal{D}^\bullet$-object into a $\mathcal{D} \times \mathcal{D}^\bullet \times \mathcal{D}$-object. The root is marked with a square.

$\Gamma\mathcal{B}'^\bullet(x)$: $s \leftarrow \mathrm{Bern}(\frac{x}{2\mathcal{B}'^\bullet(x)})$
if $s = 1$ **then return** a single edge $\{*, 1\}$ rooted at 1
else return $\Gamma\mathcal{D}^\bullet(x)$ without the orientation

This concludes the proof. □

Hence it suffices to study the class $\mathcal{D}^\bullet$, see also Figures 5.6 and 5.7.

Lemma 5.6.9. *The classes $\mathcal{D}$ and $\mathcal{D}^\bullet$ satisfy*

$$\mathcal{D} = \mathcal{X} + \mathcal{D}\times\mathcal{D} + \mathcal{D}\times\mathcal{D}\times\mathcal{D} + \dots$$
$$\mathcal{D}^\bullet = \mathcal{X} + (\mathcal{D}^\bullet\times\mathcal{D} + \mathcal{D}\times\mathcal{D}^\bullet) + (\mathcal{D}^\bullet\times\mathcal{D}\times\mathcal{D} + \mathcal{D}\times\mathcal{D}^\bullet\times\mathcal{D} + \mathcal{D}\times\mathcal{D}\times\mathcal{D}^\bullet) + \dots$$

Their exponential generating functions are given by

$$\mathcal{D}^\bullet(z) = \frac{z(\mathcal{D}(z)-1)^2}{2(\mathcal{D}(z))^2 - 4\mathcal{D}(z) + 1} \quad \text{and} \quad \mathcal{D}(z) = \frac{1}{4}(1 + z - \sqrt{z^2 - 6z + 1}). \tag{5.6.6}$$

Proof. Let $B \in \mathcal{D}$ with $|B| \geq 2$ be a derived outerplanar block, rooted at an oriented edge $\overrightarrow{e}$ of its Hamilton cycle C such that the $*$-vertex is the tail of $\overrightarrow{e}$. Given a drawing of B such that C is the boundary of the outer face, the root face is defined to be the bounded face F whose border contains $\overrightarrow{e}$. Then B may be identified with the sequence of blocks along F, ordered in the reverse direction of the edge $\overrightarrow{e}$. This yields the decompositions illustrated in Figures 5.6 and 5.7. Solving the corresponding equations of generating functions yields (5.6.6). □

The equation determining $y = \mathcal{C}^\bullet(\rho)$ is $1 = \mathcal{B}'^\bullet(y) = \frac{1}{2}(y + \mathcal{D}^\bullet(y))$. We obtain that $y \approx 0.17076$ is the unique root of the polynomial $3\,z^4 - 28\,z^3 + 70\,z^2 - 58\,z + 8$ in the interval $[0, \frac{1}{2}]$ and hence $\sigma^2 = 1 + \mathcal{B}'''(y)y^2 \approx 95.3658$. It remains to compute κ.

Lemma 5.6.10. *We have that* $\mathbb{E}[\mathsf{d}(\Gamma\mathcal{D}^\bullet(y))] = \frac{8w^4 - 16w^3 + 4w - 1}{(4w^3 - 6w^2 - 2w + 1)(2w - 1)} \approx 5.46545$ *with* $w := \mathcal{D}(y) \approx 0.27578$.

Since $B'^\bullet(y) = 1$ this implies with Lemma 5.6.8 that

$$\kappa = \mathbb{E}[\mathsf{d}(\Gamma\mathcal{B}'^\bullet(y))] = \frac{y}{2} + \left(1 - \frac{y}{2}\right)\mathbb{E}[\mathsf{d}(\Gamma\mathcal{D}^\bullet(y))] \approx 5.08418,$$

and the this completes the proof of Proposition 5.6.7.

Proof of Lemma 5.6.10. The rules for Boltzmann samplers translate the specification of $\mathcal{D}^\bullet$ given in Lemma 5.6.9 into the following sampling algorithm.

$\Gamma\mathcal{D}^\bullet(x)$: $s \leftarrow$ drawn with $\mathbb{P}(s=i) = \begin{cases} \frac{x}{\mathcal{D}^\bullet(x)}, & i=2 \\ (i-1)(\mathcal{D}(x))^{i-2} & i \geq 3 \end{cases}$

if $s = 2$ **then**
  return a single directed edge $(*, 1)$
else
  $\gamma \leftarrow$ a cycle $\{v_1, v_2\}, \{v_2, v_3\}, \ldots, \{v_s, v_1\}$ with $v_1 = *$
  $t \leftarrow$ a number drawn uniformly at random from the set $[s-1]$
  $\gamma \leftarrow$ identify (v_t, v_{t+1}) with the root-edge of $\gamma_t \leftarrow \Gamma\mathcal{D}^\bullet(x)$
  for each $i \in [s-1] \setminus \{t\}$
    $\gamma \leftarrow$ identify (v_i, v_{i+1}) with the root-edge of $\gamma_i \leftarrow \Gamma\mathcal{D}(x)$
  end for
  root γ at the directed edge $(*, v_s)$
  return γ relabeled uniformly at random
endif

Given a graph H in $\mathcal{D}^\bullet$ let $S(H)$, $S'(H)$ denote the length of a shorted past in H from the root-vertex to the tail $v_1 = *$ or head v_s of the directed root-edge, respectively. Clearly, $S(H)$ and $S'(H)$ differ by at most one. It will be convenient to also consider their minimum $M(H)$. Let S, S' and M denote the corresponding random variables in the random graph D drawn according to the sampler $\Gamma\mathcal{D}^\bullet(x)$. For any integers $\ell, k \geq 0$ with $\ell + k \geq 1$ let $\mathsf{D}_{\ell,k}$ be the random graph D conditioned on the event that the graph is not a single edge and that in the root face $\{v_1, v_2\}, \{v_2, v_3\} \ldots, \{v_s, v_1\}$ the length of the path $v_1 v_2 \ldots v_t$ equals ℓ and the length of the path $v_{t+1} v_{t+2} \ldots v_s$ equals k. Note that the probability for this event equals

$$p_{\ell,k} = \mathbb{P}(s = \ell + k + 2)\mathbb{P}(t = \ell + 1 \mid s = \ell + k + 2) = (\mathcal{D}(x))^{k+\ell}.$$

We denote by $S_{\ell,k}$, $S'_{\ell,k}$ and $M_{\ell,k}$ the corresponding distances in the conditioned random graph $\mathsf{D}_{\ell,k}$. Summing over all possible values for k and ℓ we obtain

$$\mathbb{E}[S] = \frac{x}{\mathcal{D}^\bullet(x)} + \sum_{k+\ell \geq 1} \mathbb{E}[S_{\ell,k}] p_{\ell,k},$$
$$\mathbb{E}[S'] = \sum_{k+\ell \geq 1} \mathbb{E}[S'_{\ell,k}] p_{\ell,k},$$
$$\mathbb{E}[M] = \sum_{k+\ell \geq 1} \mathbb{E}[M_{\ell,k}] p_{\ell,k}.$$

Any shortest path from $*$ or v_s to the root-vertex of a $\mathcal{B}'^\bullet$-graph H ($\neq$ a single edge) must traverse the boundary of the root-face in one of the two directions until it reaches the root-edge of the attached $\mathcal{B}'^\bullet$-object H'. From there it follows a shortest path to the root in the graph H'. Hence for all $k, \ell \geq 0$ with $k + \ell \geq 1$ the following equations hold.

$$S_{\ell,k} \overset{(d)}{=} \min\{\ell + S, k + 1 + S'\},$$
$$S'_{\ell,k} \overset{(d)}{=} \min\{\ell + 1 + S, k + S'\},$$
$$M_{\ell,k} \overset{(d)}{=} \min\{\ell + S, k + S'\}.$$

Since S and S' differ by at most one, this may be simplified further depending on the parameters k and ℓ as follows:

$$S_{\ell,k} \overset{(d)}{=} \begin{cases} \ell + S, & \ell \leq k \\ \ell + M, & \ell = k+1 \\ k + 1 + S', & \ell \geq k+2 \end{cases},$$

$$S'_{\ell,k} \overset{(d)}{=} \begin{cases} k + S', & k \leq \ell \\ k + M, & k = \ell+1 \\ \ell + 1 + S, & k \geq \ell+2 \end{cases},$$

$$M_{\ell,k} \overset{(d)}{=} \begin{cases} \ell + S, & \ell \leq k-1 \\ \ell + M, & \ell = k \\ k + S', & \ell \geq k+1 \end{cases}.$$

Using this and (5.6.6), we arrive at the system of linear equations with parameter $w = \mathcal{D}(x)$ and variables $\mu_S = \mathbb{E}[S]$, $\mu_{S'} = \mathbb{E}[S']$ and $\mu_M = \mathbb{E}[M]$

$$\begin{aligned}
\mu_S &= \sum_{k\geq 1}\sum_{\ell=0}^{k}(\ell + \mu_S)w^{\ell+k} + \sum_{\ell\geq 1}(\ell + \mu_M)w^{2\ell-1} + \sum_{k\geq 0}\sum_{\ell\geq k+2}(k + 1 + \mu_{S'})w^{\ell+k} \\
&\quad + \frac{2w^2 - 4w + 1}{(w-1)^2}, \\
\mu_{S'} &= \sum_{\ell\geq 1}\sum_{k=0}^{\ell}(k + \mu_{S'})w^{\ell+k} + \sum_{k\geq 1}(k + \mu_M)w^{2k-1} + \sum_{\ell\geq 0}\sum_{k\geq \ell+2}(\ell + 1 + \mu_S)w^{\ell+k}, \\
\mu_M &= \sum_{k\geq 2}\sum_{\ell=0}^{k-1}(\ell + \mu_S)w^{\ell+k} + \sum_{\ell\geq 1}(\ell + \mu_M)w^{2\ell} + \sum_{k\geq 0}\sum_{\ell\geq k+1}(k + \mu_{S'})w^{\ell+k}.
\end{aligned}$$

Simplifying the equations yields the equivalent system $A \cdot (\mathbb{E}[S], \mathbb{E}[S'], \mathbb{E}[M])^\top = b$ with

$$A = \begin{pmatrix} 2w^4 - 4w^3 + 3w - 1 & -w^3 + w^2 & w^3 - 2w^2 + w \\ -w^3 + w^2 & 2w^4 - 4w^3 + 3w - 1 & w^3 - 2w^2 + w \\ -w^2 + w & -w^2 + w & 2w^4 - 4w^3 + w^2 + 2w - 1 \end{pmatrix}$$

and

$$b^\top = \left(2w^4 - 4w^3 - w^2 + 3w - 1 \quad -w \quad -w^2\right).$$

For $x = y \approx 0.17076$ we obtain $w \approx 0.27578$ and $\det(A) \approx -0.00799 \neq 0$. Solving the system of linear equations yields

$$\begin{aligned}
\mathbb{E}[S] &= \frac{8w^4 - 16w^3 + 4w - 1}{(4w^3 - 6w^2 - 2w + 1)(2w - 1)} \approx 5.46545, \\
\mathbb{E}[S'] &= \frac{\left(4w^3 - 8w^2 + 1\right) w}{(2w^2 - 3w + 1)(4w^3 - 6w^2 - 2w + 1)} \approx 5.31469 \\
\mathbb{E}[M] &= -\frac{w}{4w^4 - 10w^3 + 4w^2 + 3w - 1} \approx 5.01279.
\end{aligned}$$

and the proof is complete. □

5.7 An alternative proof of the main theorem

In this chapter we describe an alternative proof of the scaling limit in Theorem 1.4.1, without the size-biased random $\mathcal{R}$-enriched tree.

5.7.1 A size-biased random labelled Tree

In this section we derive a concentration inequality for the number of vertices of a given degree along a sufficiently long path in a random tree. We give an upper bound for the number of vertices whose degree exceed a certain limit. In the following we let ξ denote a random variable taking values in $\mathbb{N}_0$ with expected value $\mathbb{E}[\xi] = 1$ and finite nonzero variance $0 < \sigma^2 < \infty$. We let $\varphi(z) = \sum_{d \geq 0} \varphi_d z^d$ denote its probability generating function and $\mathsf{span}(\xi)$ its span. Moreover, n always denotes an integer with $n \equiv 1 \mod \mathsf{span}(\xi)$.

Lemma 5.7.1. *Let $\Gamma T^\bullet$ denote the procedure that draws from the class $\mathcal{T}^\bullet$ of labeled rooted trees as follows:*

$\Gamma T^\bullet$*:* $\gamma \leftarrow$ *a single "untouched" vertex (which will be the root)*
while *there are untouched vertices left*
$v \leftarrow$ *an arbitrary untouched vertex*
$m \leftarrow$ *a nonnegative integer drawn according to an i.i.d. copy of* ξ
attach m new untouched vertices to v
declare v "touched"
end while
return γ *relabeled uniformly at random*

Then each tree $T \in \mathcal{T}^\bullet$ with size n gets chosen with probability

$$\mathbb{P}(\Gamma T^\bullet = T) = \frac{1}{n!} \prod_{v \in V(T)} d^+(v)! \, \varphi_{d^+(v)}$$

We denote by $\Gamma T_n^\bullet$ the sampler conditioned on output size n.

Proof. Let $T \in \mathcal{T}^\bullet$ be a labeled rooted tree of size n and T be drawn according to $\Gamma T^\bullet$. We may endow the tree T with an ordering on each offspring set given by the order in which the vertices where declared "touched" by the procedure. Hence we may rephrase the sampler $\Gamma T^\bullet$ as follows:

1. Draw a Galton-Watson tree U_p with offspring distribution given by ξ.

2. Form the labeled plane tree T_p by distributing labels uniformly at random.

3. Forget about the orderings on the offspring sets.

Clearly we have $\mathsf{T} = T$ if and only if $\mathsf{T_p}$ is an embedding of the tree T. Thus, let T_p be one of the $\prod_{v \in V(T)} d^+(v)$ embedded versions of the tree T and U_p its corresponding unlabeled plane tree. Then

$$\mathbb{P}(\mathsf{T}_p = T_p) = \mathbb{P}(\mathsf{U}_p = U_p)\mathbb{P}(\mathsf{T}_p = T_p \mid \mathsf{U}_p = U_p).$$

The event $\mathsf{U}_p = U_p$ depends only on the inpendent choices of the outdegrees, hence

$$\mathbb{P}(\mathsf{U}_p = U_p) = \prod_{v \in V(T)} \varphi_{d^+(v)}.$$

Any labeling of the plane tree U_p corresponds to exactly one appropiate sized permutation, thus

$$\mathbb{P}(\mathsf{T}_p = T_p \mid \mathsf{U}_p = U_p) = \frac{1}{n!}.$$

Combining the above equations yields

$$\mathbb{P}(\mathsf{T} = T) = \frac{1}{n!} \prod_{v \in V(T)} d^+(v)!\, \varphi_{d^+(v)}.$$

□

We follow the notation from [Joy81, BLL98] and term a doubly rooted tree $V \in \mathcal{T}^{\bullet\bullet}$ a vertebrate. The directed path from the inner to the outer root is called the backbone. As illustrated in Figure 5.8, any vertebrate may be identified with the sequence of trees along its backbone, giving an isomorphism

$$\mathcal{T}^{\bullet\bullet} \simeq \sum_{k=1}^{\infty} (\mathcal{T}^{\bullet})^k \simeq \mathcal{T}^{\bullet} \cdot \mathrm{SEQ}(\mathcal{T}^{\bullet}).$$

See [Joy81] for details on the proof. We let $\mathcal{T}^{\bullet\bullet(\ell)}$ denote the class of vertebrates whose backbone has length ℓ.

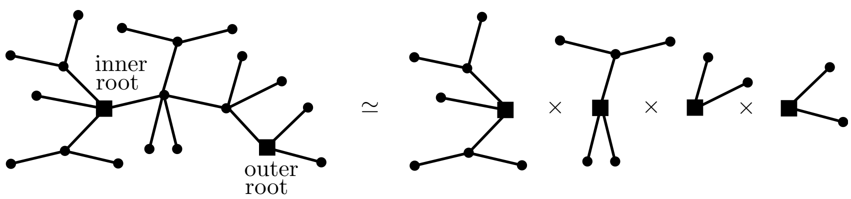

Figure 5.8: Correspondence of vertebrates and sequences of rooted trees. The roots are marked with squares.

The following lemma is a labeled nonplane version of a result given in [ABDJ13, Sec. 3].

Lemma 5.7.2. *Consider the following sampler that draws a random vertebrate from* $\mathcal{T}^{\bullet\bullet(\ell)}$.

$\Gamma T^{\bullet\bullet(\ell)}$: $\gamma \leftarrow$ *a directed path* $v_1, \ldots, v_{\ell+1}$ *of length* ℓ
(the first vertex will be the inner root, the last vertex the outer root)
$(d_i)_{1\leq i\leq \ell} \leftarrow$ *an i.i.d family of integers drawn according to the PGF* $\varphi'(z)$
$d_{\ell+1} \leftarrow$ *an integer drawn according to the PGF* $\varphi(z)$
$(T_{i,j})_{1\leq i\leq \ell+1, 1\leq j\leq d_i} \leftarrow$ *an i.i.d family of rooted trees drawn according to* $\Gamma T^{\bullet}$
for each $1 \leq i \leq \ell+1$ ***and*** $1 \leq j \leq d_i$
add an edge between the vertex v_i *and the root of the tree* $T_{i,j}$
return γ *relabeled uniformly at random*

Then for each vertebrate $(T, v, w) \in \mathcal{T}^{\bullet\bullet(\ell)}$ *we have that*

$$\mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} = (T, v, w)) = \mathbb{P}(\Gamma T^{\bullet} = (T, v)).$$

Proof. We may rephrase the procedure of the sampler as follows.

1. Draw a sequence $(\mathsf{U}_1^p, \ldots, \mathsf{U}_{\ell+1}^+)$ of independent unlabeled plane trees. Each tree U_i is generated like a Galton-Watson tree with offspring distribution given by $\varphi(z)$, except that the outdegree of the root is chosen according to $\varphi'(z)$ instead.

2. Let $(\mathsf{T}_1^p, \ldots, \mathsf{T}_{\ell+1}^p)$ denote the sequence of labeled plane trees obtained by distributing labels uniformly at random. Each labeling of the sequence corresponds to exactly one appropiate sized permutation.

3. Forget about the order on the offspring sets. The resulting sequence $(\mathsf{T}_1, \ldots, \mathsf{T}_{\ell+1})$ corresponds to a doubly rooted tree.

Let $V \in \mathcal{T}^{\bullet\bullet(\ell)}$ be a vertebrate whose backbone P has length ℓ. Then V corresponds to a sequence $(T_1, \ldots, T_{\ell+1})$ of subtrees such that the root of the tree T_i is the ith vertex of the backbone. For each i fix an arbitrary plane tree T_i^p corresponding to the tree T_i and let U_i^p denote the corresponding unlabeled plane tree. By considering the number of possible orderings of the offspring sets, we obtain

$$\mathbb{P}(\mathsf{T}_i = T_i \text{ for all } i) = \mathbb{P}(\mathsf{T}_i^p = T_i^p \text{ for all } i) \prod_{i=1}^{\ell+1} \prod_{v\in V(T_i)} d_{T_i}^+(v)!.$$

Since any labeling of a sequence of plane trees corresponds to exactly one appropiate sized permutation, we have that

$$\mathbb{P}(\mathsf{T}_i^p = T_i^p \text{ for all } i) = \frac{1}{n!}\mathbb{P}(\mathsf{U}_i^p = U_i^p \text{ for all } i).$$

The vertebrate V may be considered as a rooted tree in two ways and we let T denote the version whose root is the inner root of V. Then the multiset of outdegrees in $U_{\ell+1}^p$

and $V(T_{\ell+1}) \subset V(T)$ coincide. Since $\mathsf{U}^p_{\ell+1}$ is a Galton-Watson tree with offspring distribution given by the PGF $\varphi(z)$, it follows that

$$\mathbb{P}(\mathsf{U}^p_{\ell+1} = U^p_{\ell+1}) = \prod_{v \in V(T_{\ell+1})} \varphi_{d^+_T(v)}.$$

Let $1 \le i \le \ell$ and w_i denote the root of the tree T_i. We have $d^+_T(w_i) = d^+_{T_i}(w_i) + 1$ and $d^+_T(v) = d^+_{T_i}(v)$ for all vertices $v \in V(T_i)\backslash\{w_i\}$. The degree of the root of U^p_i was chosen according to $\varphi'(z)$, and the remaining outdegrees according to $\varphi(z)$. Since $[z^k]\varphi'(z) = (k+1)\varphi_{k+1}$, it follows that

$$\mathbb{P}(\mathsf{U}^p_i = U^p_i) = d^+_T(w_i) \prod_{v \in V(T_i)} \varphi_{d^+_T(v)}.$$

We obtain that

$$\mathbb{P}(\mathsf{U}^p_i = U^p_i \text{ for all } i) = \prod_{i=1}^{\ell+1} \mathbb{P}(\mathsf{U}^p_i = U^p_i) = \left(\prod_{i=1}^{\ell} d^+_T(w_i)\right)\left(\prod_{v \in V(T)} \varphi_{d^+_T(v)}\right).$$

Combining the above equations yields

$$\mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} = V) = \frac{1}{n!} \prod_{v \in V(T)} d^+_T(v)!\varphi_{d^+_T(v)} = \mathbb{P}(\Gamma T^\bullet = T).$$

□

Lemma 5.7.3. *Let $\mathcal{E}$ be a property of vertebrates (i.e. a subset of $\mathcal{T}^{\bullet\bullet}$) that is invariant under relabeling and consider the map*

$$B : \mathcal{T}^\bullet \to \mathbb{N}_0, \quad T \mapsto \sum_{v \in V(T)} \mathbb{1}_{[(T,v) \text{ satisfies } \mathcal{E}]}$$

Then we have the following bound for the expected value of $B(\Gamma T^\bullet_n)$

$$\mathbb{E}[B(\Gamma T^\bullet_n)] = \Theta(n^{3/2}) \sum_{l=0}^{n-1} \mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} \text{ has size } n \text{ and satisfies } \mathcal{E})$$

Proof. Let $\Gamma T^{\bullet\bullet}_n$ denote a random vertebrate obtained by sampling $\Gamma T^\bullet_n$ and choosing an outer root uniformly at random. Then

$$\mathbb{E}[B(\Gamma T^\bullet_n)] = \sum_{v=1}^{n} \mathbb{P}((\Gamma T^\bullet_n, v) \in \mathcal{E}) = n\mathbb{P}(\Gamma T^{\bullet\bullet}_n \in \mathcal{E})$$

Clearly, for any vertebrate $(T, v) \in \mathcal{E} \cap \mathcal{T}^{\bullet\bullet(\ell)}_n$ with $l \ge 0$ we have that

$$\mathbb{P}(\Gamma T^{\bullet\bullet}_n = (T,v)) = \frac{1}{n}\mathbb{P}(\Gamma T^\bullet_n = T) = \frac{1}{n}\mathbb{P}(\Gamma T^\bullet = T)\mathbb{P}(|\Gamma T^\bullet| = n)^{-1}$$

and Lemma 5.7.2 yields that

$$\mathbb{P}(\Gamma T^{\bullet} = T) = \mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} = (T, v))$$

Since $\mathbb{P}(|\Gamma T^{\bullet}| = n) = \theta(n^{-3/2})$, it follows that

$$\mathbb{E}[B(\Gamma T_n^{\bullet})] = \Theta(n^{3/2}) \sum_{\ell=0}^{n-1} \mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} \text{ has size } n \text{ and satisfies } \mathcal{E})$$

□

Corollary 5.7.4. *The following holds with high probability as n tends to infinity: all paths P in $\Gamma T_n^{\bullet}$ that start from the root satisfy*

$$|\{v \in V(P) \mid d^+(v) = d\}| \in \ell(P)d\varphi_d \pm \sqrt{\ell(P)d\varphi_d} \log n$$

for all integers $1 \leq d \leq n$ with $\ell(P)d\varphi_d \geq \log(n)^3$ or $\varphi_d = 0$. Here $\ell(P)$ denotes the length of the path P.

Proof. If $\varphi_d = 0$ then with probability 1 the random tree $\Gamma T_n^{\bullet}$ has no vertex of outdegree d. Hence it suffices to consider only integers d with $\varphi_d \neq 0$. For any such d let $\mathcal{E}_d \subset \mathcal{T}^{\bullet\bullet}$ denote the set of all vertebrates T whose backbone P satisfies $l(P)d\varphi_d \geq \log(m)^3$ with $m := |V(T)|$ and

$$|\{v \in V(P) \mid d^+(v) = d\}| \notin \ell(P)d\varphi_d \pm \sqrt{\ell(P)d\varphi_d} \log m.$$

Let $\mathcal{E} = \cup_d \mathcal{E}_d$ where the union is over all integers d with $\varphi_d \neq 0$. Consider the map

$$B : \mathcal{T}^{\bullet} \to \mathbb{N}_0, \quad T \mapsto \sum_{v \in V(T)} \mathbb{1}_{[(T,v) \text{ satisfies } \mathcal{E}]}.$$

By Markov's inequality it suffices to show that $\mathbb{E}[B(\Gamma T_n^{\bullet})]$ tends to zero. Lemma 5.7.3 yields

$$\mathbb{E}[B(\Gamma T_n^{\bullet})] = \Theta(n^{3/2}) \sum_{0 \leq \ell \leq n-1} \mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} \text{ has size } n \text{ and lies in } \mathcal{E}). \tag{$\star$}$$

Applying the union bound yields

$$\mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} \text{ has size } n \text{ and lies in } \mathcal{E}) \leq \sum_{\substack{1 \leq d \leq n-1 \\ \ell d\varphi_d \geq \log(n)^3}} \mathbb{P}(\Gamma T^{\bullet\bullet(\ell)} \text{ has size } n \text{ and lies in } \mathcal{E}_d). \tag{$\star\star$}$$

Recall that the sampler $\Gamma T^{\bullet\bullet(\ell)}$ starts by drawing integers $d_1, \ldots, d_\ell$ independently according to the PGF $\varphi'(z)$ and $d_{\ell+1}$ according to $\varphi(z)$. The outdegrees of the vertices of the backbone $v_1, \ldots, v_{\ell+1}$ are then given by $d^+(v_i) = d_i + 1$ for $1 \leq i \leq \ell$

and $d^+(v_{\ell+1}) = d_{\ell+1}$. In particular $\mathbb{P}(d^+(v_i) = d) = d\varphi_d$ for $1 \leq i \leq \ell$ and $\mathbb{P}(d^+(v_{\ell+1}) = d) = \varphi_d$. The event that $\Gamma T^{\bullet\bullet(\ell)}$ has size n and lies in $\mathcal{E}_d$ implies that

$$\sum_{i=1}^{\ell+1} \mathbb{1}_{d^+(v_i)=d} \notin \ell d\varphi_d \pm \sqrt{\ell d\varphi_d}\log n$$

The expected value of the sum is given by $\mu = \ell d\varphi_d + \varphi_d$. Since $\ell d\varphi_d \geq \log(n)^3$ a short calculation shows that for large n

$$\mu(1 \pm \delta) \subset \ell d\varphi_d \pm \sqrt{\ell d\varphi_d}\log n \quad \text{with } \delta = \frac{\log n}{2\sqrt{\mu}}.$$

Using monotonicity and Chernoff's bounds we obtain that the probabilty for the event $(\star)$ is bounded by

$$\mathbb{P}(\sum_{i=1}^{\ell+1} \mathbb{1}_{d^+(v_i)=d} \notin \mu(1 \pm \delta)) \leq 2\exp(-\delta^2\mu/3) = 2n^{-\log(n)/4}$$

Hence Equations $(\star)$ and $(\star\star)$ imply that

$$\mathbb{E}[B(\Gamma T_n^\bullet)] \leq \theta(n^{3/2})n^2 2n^{-\log(n)/4} = n^{-\theta(\log n)}$$

This completes the proof. $\square$

In the following we assume additionally that the PGF $\varphi(z)$ has radius of convergence $r > 1$. Note that $\limsup_{n\to\infty} \sqrt[n]{\varphi_n} = \frac{1}{r}$ implies

$$\varphi_n \leq (\frac{1}{r} + o(1))^n.$$

Clearly the same inequality also holds for the coefficients of arbitrary derivatives $\varphi^{(\ell)}(z)$.

Corollary 5.7.5. *Suppose that the PGF $\varphi(z)$ has radius of convergence strictly greater than one. Given $\epsilon > 0$ we may choose $D \geq 1$ large enough such that with high probability all paths P in $\Gamma T_n^\bullet$ that start from the root and have length $\ell(P) \geq \log(n)^2$ satisfy*

$$\frac{1}{\ell(P)} \sum_{v \in V(P)} d^+(v)\mathbb{1}_{[d^+(v)\geq D]} \leq \epsilon$$

Proof. Given $0 < \epsilon < 1$, let $\mathcal{E} \subset \mathcal{T}^{\bullet\bullet}$ denote the set of all vertebrates T whose backbone P has length at least $\log(m)^2$ with $m := |V(T)|$ and satisfies the inequality

$$\sum_{v \in V(P)} d^+(v)\mathbb{1}_{[d^+(v)\geq D]} > \ell(P)\epsilon \qquad (\star)$$

Consider the map

$$B : \mathcal{T}^\bullet \to \mathbb{N}_0, \quad T \mapsto \sum_{v \in V(T)} \mathbb{1}_{[(T,v) \text{ satisfies } \mathcal{E}]}.$$

By Markov's inequality it suffices to show that $\mathbb{E}[B(\Gamma T_n^\bullet)]$ tends to zero. Lemma 5.7.3 yields

$$\mathbb{E}[B(\Gamma T_n^\bullet)] \leq \Theta(n^{3/2}) \sum_{\ell \geq \log(n)^2} \mathbb{P}(\text{the backbone } P \text{ of } \Gamma T^{\bullet\bullet(\ell)} \text{ satisfies inequality } (\star)). \tag{$\star\star$}$$

Recall that the sampler $\Gamma T^{\bullet\bullet(\ell)}$ starts by drawing integers $d_1, \ldots, d_\ell$ independently according to the PGF $\varphi'(z)$ and $d_{\ell+1}$ according to $\varphi(z)$. The outdegrees of the vertices of the backbone $v_1, \ldots, v_{\ell+1}$ are then given by $d^+(v_i) = d_i + 1$ for $1 \leq i \leq \ell$ and $d^+(v_{\ell+1}) = d_{\ell+1}$. In particular $\mathbb{P}(d^+(v_i) = d) = d\varphi_d$ for $1 \leq i \leq \ell$ and $\mathbb{P}(d^+(v_{\ell+1}) = d) = \varphi_d$. By assumption on the offspring distribution we may choose the constant D large enough such that there is a constant $0 < \nu < 1$ with $d\varphi_d < \nu^d$ for all $d \geq D$. Let $\lambda > 0$ be a constant that is small enough such that $e^\lambda \nu < 1$. The probability that the backbone P of the vertebrate $\Gamma T^{\bullet\bullet(\ell)}$ satisfies inequality $(\star)$ can be bounded using Markov's inequality by

$$\frac{\mathbb{E}[\exp\left(\lambda \sum_i d^+(v_i) \mathbb{1}_{[d^+(v_i) \geq D]}\right)]}{e^{\epsilon \lambda \ell}} = \left(\frac{\mathbb{E}[\exp(\lambda d^+(v_1))]}{e^{\epsilon\lambda}} \right)^\ell \mathbb{E}[\exp(\lambda d^+(v_{\ell+1}))]$$

Since $e^\lambda \nu < 1$ by our choice of the constant λ, we get

$$\mathbb{E}[\exp(\lambda d^+(v_1))] = \sum_{d \geq D} e^{\lambda d} d\varphi_d + \mathbb{P}(d^+(v_1) < D) \leq \frac{(e^\lambda \nu)^D}{1 - e^\lambda \nu} + 1.$$

A similar calculation shows that $\mathbb{E}[\exp(\lambda d^+(v_{\ell+1}))]$ obeys the same upper bound. Since $e^{\epsilon\lambda} > 1$ we may choose D large enough such that

$$\frac{\mathbb{E}[\exp(\lambda d^+(v_1))]}{e^{\epsilon\lambda}} < \delta \quad \text{and} \quad \mathbb{E}[\exp(\lambda d^+(v_{\ell+1}))] \leq 2$$

for some constant $\delta < 1$. Applying this to Equation $(\star\star)$ yields

$$\mathbb{E}[B(\Gamma T_n^\bullet)] \leq \Theta(n^{3/2}) \sum_{\ell \geq \log(n)^2} \delta^\ell = \theta(n^{3/2}) \delta^{\log(n)^2} = n^{-\theta(\log n)}$$

This completes the proof. □

5.7.2 Alternative Boltzmann sampler

Let $\mathcal{C}$ be a nontrivial block stable class of connected graphs such that its exponential generating function $C(z)$ has positive radius of convergence ρ. Let $\mathcal{B}$ its subclass of all graphs that are biconnected or a single edge with its ends. Recall that we always let $y = C^\bullet(\rho)$ and $\lambda_C = B'(y)$.

In the following we will construct a nonrecursive Boltzmann Sampler for the class $\mathcal{C}^\bullet$ which allows us to apply the results on random trees of the previous section. Recall that the class $\mathcal{C}^\bullet$ may be identified with the class $\mathcal{A}$ of SET $\circ\, \mathcal{B}'$-enriched trees, i.e. pairs (T, α) with $T \in \mathcal{T}^\bullet$ and α a function that assigns to each vertex $v \in V(T)$ a (possibly empty) set $\alpha(v)$ of derived blocks whose vertex sets partition the offspring set of the vertex v.

Lemma 5.7.6. *Let* C *be drawn according to the Boltzmann sampler* $\Gamma C^\bullet(\rho)$ *and let* (T, α) *denote the corresponding enriched tree. Then the random tree* T *is distributed according to the sampler* $\Gamma T^\bullet$ *with offspring distribution given by the probability generating function*

$$\varphi(z) = \exp(B'(yz) - \lambda_C) =: \sum_{d \geq 0} \varphi_d z^d.$$

Let D *denote the root degree of the tree* T. *Consider the sequence* $\nu_1, \nu_2, \ldots, \nu_D$ *where* ν_i *counts the number of blocks of size* $i+1$ *in* C *that contain the root. Then for any* $d \geq 1$ *with* $\varphi_d \neq 0$, *and nonnegative integers* $n_1, \ldots, n_d$ *with* $\sum_i i n_i = d$, *we have*

$$\mathbb{P}(\nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_d = n_d \mid D = d) = \frac{1}{\exp(\lambda_C)\varphi_d} \prod_{i=1}^{d} \frac{([z^i]B'(yz))^{n_i}}{n_i!}.$$

We will denote this probability distribution on $\mathbb{N}_0^d$ *by PSeq(d).*

Proof. Recall that the sampler $\Gamma\mathcal{C}^\bullet(\rho)$ starts by drawing the number of blocks attached to the root according to the Poisson distribution with parameter λ_C, and proceeds by sampling m derived blocks $B_1, \ldots, B_m$ according to $\Gamma\mathcal{B}'(y)$. The degree D of the root in T is the sum of the sizes of these blocks. The PGF of the size of $\Gamma\mathcal{B}'(y)$ is given by $\mathcal{B}'(yz)/\lambda_C$. Hence the probability generating function for D is given $\exp(\mathcal{B}'(yz) - \lambda_C) = \varphi(z)$. After drawing the blocks the sampler marks the root as touched and repeats the steps for all untouched vertices. In other words the tree T is drawn by generating a unlabeled nonplane tree with offspring distribution corresponding to the PGF $\varphi(z)$ and distributing labeles uniformly at random afterwards. Hence T is distributed according to $\Gamma T^\bullet$.

Now, consider the sequence $\nu_1, \nu_2, \ldots, \nu_D$ where ν_i counts the number of blocks of size $i+1$ in C that contain the root. This means ν_i is the number of indices $1 \leq j \leq m$ such that the derived block B_j has size i and $D = \sum_i i\nu_i$. Hence for any $d \geq 1$ and nonnegative integers $n_1, \ldots, n_d$ with $\sum_i i n_i = d$ we have

$$\mathbb{P}(\nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_d = n_d \mid D = d) = \\ \varphi_d^{-1}\mathbb{P}(\nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_d = n_d, \nu_i = 0 \text{ for } i > d).$$

We may calculate the probability on the right hand side by considering the formal probability generating series

$$\begin{aligned} f(z_1, z_2, \dots) &= \sum_{(k_i)_i \in \mathbb{N}_0^{(\mathbb{N})}} \mathbb{P}(\nu_i = k_i \text{ for all } i) z_1^{k_1} z_2^{k_2} \cdots \\ &= \exp\left(\lambda_C (\sum_{i \geq 1} \mathbb{P}(|\Gamma B'(y)| = i) z_i - 1)\right). \end{aligned}$$

Clearly

$$\begin{aligned} [z_1^{n_1} z_2^{n_2} \cdots z_d^{n_d}] f(z_1, z_2, \dots) &= \exp(-\lambda_C)[z_1^{n_1} z_2^{n_2} \cdots z_d^{n_d}] \prod_{i \geq 1} \exp(z_i [z^i] B'(yz)) \\ &= \exp(-\lambda_C) \prod_{i=1}^{d} [x^{n_i}] \exp(x [z^i] B'(yz)) \\ &= \exp(-\lambda_C) \prod_{i=1}^{d} \frac{([z^i] B'(yz))^{n_i}}{n_i!}. \end{aligned}$$

This concludes the proof. □

Lemma 5.7.7. *The following procedure ΓA is a Boltzmann-Sampler for the class $\mathcal{A}$ of $SET \circ \mathcal{B}'$-enriched trees at the singularity ρ. We let ΓA_n denote the sampler conditioned on output size n.*

ΓA:
  $T \leftarrow \Gamma T^\bullet$
  for each $v \in V(T)$
    $d \leftarrow d^+(v)$
    $M \leftarrow$ *the offspring set of the vertex* v *in the tree* T
    $(\nu_1, \dots, \nu_d) \leftarrow$ *PSeq*(d)
    $(m_1, m_2, \dots, m_d) \leftarrow$ *a uniformly at random chosen ordering of the set* M
    $(M_{i,j})_{1 \leq i \leq d, 1 \leq j \leq \nu_i} \leftarrow$ *the partition of the offspring set* M *given by*
      $M_{i,j} = \{m_{t_{i,j}+1}, \dots m_{t_{i,j}+i}\}$, $t_{i,j} = \sum_{\ell=1}^{i-1} \ell \nu_\ell + i(j-1)$
    $(\sigma_{i,j})_{1 \leq i \leq d, 1 \leq j \leq \nu_i} \leftarrow$ *the sequence of bijections* $\sigma_{i,j} : [i] \to M_{i,j}$, $t \mapsto m_{t_{i,j}+t}$
    $(B_{i,j})_{1 \leq i \leq d, 1 \leq j \leq \nu_i} \leftarrow$ *a sequence of independently u.a.r. drawn blocks* $B_{i,j} \in \mathcal{B}'_i$
    $\alpha(v) \leftarrow \{\sigma_{i,j}.B_{i,j} \mid 1 \leq i \leq d, 1 \leq j \leq \nu_i\}$ *the set of relabeled blocks*
  endfor
  return (T, α)

Proof. Note that by Lemma 5.7.6 the sampler PSeq(d) is well-defined if $\varphi_d \neq 0$. Hence ΓA is almost surely well-defined, since the random tree $\Gamma T^\bullet$ has almost surely no vertices with outdegree d satisfying $\varphi_d = 0$. Let (T, α) be drawn according to the sampler ΓA and $(T, \beta) \in \mathcal{A}$ an enriched tree of size n. We have to show that $\mathbb{P}((\mathsf{T}, \alpha) = (T, \beta)) = y^{-1} \frac{\rho^n}{n!}$. Clearly we have that

$$\mathbb{P}((\mathsf{T}, \alpha) = (T, \beta)) = \mathbb{P}(\mathsf{T} = T)\mathbb{P}(\alpha = \beta \mid \mathsf{T} = T)$$

and Lemma 5.7.1 yields

$$\mathbb{P}(\mathsf{T} = T) = \frac{1}{n!} \prod_{v \in V(T)} d^+(v)!\, \varphi_{d^+(v)}.$$

The sampler ΓA chooses the SET $\circ \mathcal{B}'$-structures on the offspring sets independently, hence

$$\mathbb{P}(\alpha = \beta \mid \mathsf{T} = T) = \prod_{v \in V(T)} \mathbb{P}(\alpha(v) = \beta(v) \mid \mathsf{T} = T).$$

Let $v \in V(T)$ be a vertex and $d = d_T^+(v)$ its outdegree. Let P and Q denote the partition of the offspring set M of v given by α and β, respectively. Clearly

$$\mathbb{P}(\alpha(v) = \beta(v) \mid \mathsf{T} = T) = \mathbb{P}(P = Q \mid \mathsf{T} = T)\mathbb{P}(\alpha(v) = \beta(v) \mid P = Q, \mathsf{T} = T).$$

For all $1 \le i \le d$ let ν_i and n_i denote the number of blocks of size i in the set $\alpha(v)$ and the set $\beta(v)$, respectively. Then

$$\mathbb{P}(P = Q \mid \mathsf{T} = T) = \mathbb{P}(\nu_i = n_i \text{ for all } i \mid \mathsf{T} = T)\mathbb{P}(P = Q \mid \mathsf{T} = T, \nu_i = n_i \text{ for all } i)$$

and

$$\mathbb{P}(\nu_i = n_i \text{ for all } i \mid \mathsf{T} = T) = \frac{1}{\exp(\lambda_C)\varphi_d} \prod_{i=1}^{d} \frac{([z^i]B'(yz))^{n_i}}{n_i!}.$$

Given $\mathsf{T} = T$ and $\nu_i = n_i$ for all $1 \le i \le d$, we have that $P = Q$ if and only if the ordering $(m_1, \ldots, m_d)$ of the offspring set M drawn uniformly at random by the sampler is among one of the $\prod_{i=1}^d n_i!(i!)^{n_i}$ possible choices corresponding to the partition Q. The probability for this is given by

$$\mathbb{P}(P = Q \mid \mathsf{T} = T, \nu_i = n_i \text{ for all } i) = \frac{1}{d!} \prod_{i=1}^{d} n_i!(i!)^{n_i}.$$

It remains to calculate $\mathbb{P}(\alpha(v) = \beta(v) \mid P = Q, \mathsf{T} = T)$. Let $\mathcal{E}$ denote the event $P = Q$ and $\mathsf{T} = T$. Applying the law of total probability yields

$$\mathbb{P}(\alpha(v) = \beta(v) \mid \mathcal{E}) = \sum_{(k_1, \ldots, k_d)} \mathbb{P}(\alpha(v) = \beta(v) \mid \mathcal{E} \text{ and } m_i = k_i \text{ for all } i)\mathbb{P}(m_i = k_i \text{ for all } i \mid \mathcal{E})$$

where $(k_1, ..., k_d)$ ranges over all possible choices for the ordering $(m_1, \ldots, m_d)$. Given any such, suppose that $P = Q$, $\mathsf{T} = T$ and $m_i = k_i$ for all i. Let $\sigma_{i,j}$ denote the corresponding bijections used in the sampler. Then we have $\alpha(v) = \beta(v)$

if and only if the sequence $(B_{i,j})_{1\le i\le d, 1\le j\le n_i}$ of derived blocks drawn uniformly at random by the sampler satisfies

$$\beta(v) = \{\sigma_{i,j}.B_{i,j} \mid 1 \le i \le d, 1 \le j \le n_i\}.$$

There is precisely one possible choice for the blocks since we already fixed the labels and bijections, hence

$$\mathbb{P}(\alpha(v) = \beta(v) \mid \mathcal{E} \text{ and } m_i = k_i \text{ for all } i) = \prod_{i=1}^{d} \frac{1}{|\mathcal{B}_i|^{n_i}}.$$

This holds for all multiindices $(k_1, \dots, k_d)$, thus

$$\mathbb{P}(\alpha(v) = \beta(v) \mid P = Q, \mathsf{T} = T) = \prod_{i=1}^{d} \frac{1}{|\mathcal{B}_i|^{n_i}} = \prod_{i=1}^{d} \frac{1}{(i![z^i]B'(z))^{n_i}}.$$

Combining the equations above yields that the probability $\mathbb{P}((\mathsf{T}, \alpha) = (T, \beta))$ is given by

$$\frac{1}{n!} \prod_{v \in V(T)} \left(d^+(v)!\, \varphi_{d^+(v)} \frac{1}{\exp(\lambda_C)\varphi_{d^+(v)}} \prod_{i=1}^{d^+(v)} \frac{([z^i]B'(yz))^{n_i}}{n_i!} \frac{n_i!(i!)^{n_i}}{d^+(v)!} \frac{1}{(i![z^i]B'(z))^{n_i}} \right).$$

This simplifies to

$$\frac{1}{\exp(n\lambda_C)n!} \prod_{v \in V(T)} \prod_{i=1}^{d^+(v)} y^{in_i} = \frac{1}{\exp(n\lambda_C)n!} y^{\sum_{v\in V(T)} d^+(v)}.$$

Clearly the sum $\sum_{v \in V(T)} d^+(v)$ of all outdegrees of the rooted tree T is equal to $n-1$. Recall that we have $y = \rho \exp(\lambda_C)$. Hence

$$\mathbb{P}((\mathsf{T}, \alpha) = (T, \beta)) = \frac{1}{\exp(n\lambda_C)n!} y^{n-1} = y^{-1} \frac{\rho^n}{n!}.$$

This concludes the proof. □

Corollary 5.7.8. *Let* C *be a random graph drawn from the class* $\mathcal{C}^\bullet$ *according to the Boltzmann distribution at the singularity* ρ *and* (T, α) *be the corresponding enriched tree. Let* $T \in \mathcal{T}^\bullet$ *be a tree with* $\mathbb{P}(\Gamma T^\bullet = T) > 0$. *If we condition on the event* $\mathsf{T} = T$ *then* α *is drawn as an independent family of* $\mathrm{SET} \circ \mathcal{B}'$*-structures on the offspring sets of the tree* T. *Let* $v \in V(T)$ *be a vertex with outdegree* $d = d_T^+(v) \ge 1$ *and* w *one of its offspring. Let* B *denote the unique block of* C *containing the vertices* v *and* w *and* B' *the unique* $\mathcal{B}'$*-structure in* $\alpha(v)$ *containing* w. *Then* $|\mathsf{B}| = |\mathsf{B}'| + 1$ *and for all* $s \ge 1$ *we have that*

$$\mathbb{P}(|\mathsf{B}'| = s \mid \mathsf{T} = T) = \sum_{\substack{(n_i)_i \in \mathbb{N}_0^{(\mathbb{N})} \\ \sum_i in_i = d}} \mathbb{P}(PSeq(d) = (n_1, \dots, n_d)) \frac{sn_s}{d} =: p_{s,d}. \tag{5.7.1}$$

Thus $p_{s,d}$ is defined for all d with $\varphi_d \neq 0$. We have that

$$B'^{\bullet}(yz) = \sum_{s \geq 1} (\sum_{d \geq 1} d\varphi_d p_{s,d}) z^s. \tag{5.7.2}$$

Here we set $p_{s,d} = 0$ whenever $\varphi_d = 0$. For any block $B \in \mathcal{B}'^{\bullet}$ we let $\mathsf{d}(B)$ denote the length of a shortest path connecting the $*$-vertex and the root. The distance $d(v,w)$ is the length of a shortest path connecting the vertices v and w in the block B or, equivalently, in the graph C. Given an integer $s \geq 1$ with $p_{s,d} > 0$ and a uniformly at random chosen block $\mathsf{B}_s'^{\bullet}$ from the class $\mathcal{B}_s'^{\bullet}$, we have that

$$\mathbb{P}(d(v,w) = t \mid |\mathsf{B}'| = s, \mathsf{T} = T) = \mathbb{P}(\mathsf{d}(\mathsf{B}_s'^{\bullet}) = t) \tag{5.7.3}$$

for all integers $t \geq 1$.

Proof. First we prove Equations (5.7.1) and (5.7.3). Let M denote the offspring set of the vertex v in the tree T and $d = d_T^+(v)$ its outdegree. The sampler ΓA generates the SET $\circ\, \mathcal{B}'$-structure on the set M as follows:

1. Draw the partition sequence $\nu_1, \ldots, \nu_d$ according to the distribution PSeq(d).
2. For all $1 \leq i \leq d$, $1 \leq j \leq \nu_i$ choose $\mathsf{B}_{i,j} \in \mathcal{B}_i'$ uniformly at random.
3. Choose a matching of the set M and the disjoint union $\bigsqcup_{i,j}(V(\mathsf{B}_{i,j}) \setminus \{*\}) = \bigsqcup_{i,j}[i]$ uniformly at random.
4. Relabel according to the matching.

If there is no sequence $n_1, \ldots, n_d \in \mathbb{N}_0$ with $\sum_i in_i = d$ and $\mathbb{P}(\text{PSeq}(d) = (n_1, \ldots, n_d)) > 0$ then

$$\mathbb{P}(|B'| = s | \mathsf{T} = T) = 0 = p_{s,d}.$$

Otherwise, let $(n_i)_i$ be such a sequence and suppose that $\mathsf{T} = T$ and $\nu_i = n_i$ for all i. Then w is matched to a uniformly at random chosen vertex from $\bigsqcup_{i,j} V(\mathsf{B}_{i,j}) = \bigsqcup_{i,j}[i]$. Hence we have $|B'| = s$ if and only if w gets matched to a vertex from $\bigsqcup_{1 \leq j \leq n_s}[s]$. The probability for this is given by

$$\mathbb{P}(|B'| = s \mid \mathsf{T} = T, \nu_i = n_i \text{ for all } i) = \frac{sn_s}{d}.$$

It follows that

$$\mathbb{P}(|\mathsf{B}'| = s \mid \mathsf{T} = T) = \sum_{\substack{(n_i)_i \in \mathbb{N}_0^{(\mathbb{N})} \\ \sum_i in_i = d}} \mathbb{P}(\text{PSeq}(d) = (n_1, \ldots, n_d)) \frac{sn_s}{d} = p_{s,d}.$$

Thus Equation (5.7.1) holds. Now, suppose that $p_{s,d} > 0$. Then there are sequences $n_1, \ldots, n_d$ be with $\sum_i n_i = d, n_s > 0$ and $\mathbb{P}(\text{PSeq}(d) = (n_1, \ldots, n_d)) > 0$. Let $(n_i)_i$ be such a sequence and suppose that $\mathsf{T} = T$, $|\mathsf{B}'| = s$ and $\nu_i = n_i$ for all i.

Then the vertex w gets matched to a uniformly at random chosen non-$*$-vertex of $B_{s,1},\dots,B_{s,n_s}$. Let $1\le k\le n_s$ denote the index of the corresponding block. Hence the distance $d(v,w)$ is equal to length of a shortest path from the $*$-vertex of $\mathsf{B}_{s,k}$ to a uniformly at random chosen root r. The rooted graph $(\mathsf{B}_{s,k},r)$ is distributed like a uniformly at random chosen graph $\mathsf{B}_s'^{\bullet}\in\mathcal{B}_s'^{\bullet}$. Thus we have for all $t\ge 1$

$$\mathbb{P}(d(v,w)=t\mid |\mathsf{B}'|=s,\mathsf{T}=T,\nu_i=n_i \text{ for all i})=\mathbb{P}(\mathsf{d}(\mathsf{B}_s'^{\bullet})=t).$$

It follows that Equation (5.7.3) holds. It remains to prove that for all $s\ge 1$ we have that $[z^s]B'(yz)=\sum_{d\ge 1} d\varphi_d p_{s,d}$. For any d with $\varphi_d=0$ we have that

$$d\varphi_d p_{s,d}=s\exp(-\lambda_C)\sum_{\substack{(n_i)_i\in\mathbb{N}_0^{(\mathbb{N})}\\ \sum_i in_i=d}}\prod_{i=1}^{d}\frac{([z^i]B'(yz))^{n_i}}{n_i!}n_s \qquad (\star)$$

Note that

$$\varphi(z)=\exp(B'(yz)-\lambda_C)=\exp(-\lambda_C)\prod_{i\ge 1}\exp(z^i[z^i]B'(yz))$$

implies that for all $d\ge 1$ with $\varphi_d=0$ and $n_1,\dots,n_d\ge 0$ with $\sum_i in_i=d$ we have that

$$\prod_{i=1}^{d}([z^i]B'(yz))^{n_i}=0.$$

Hence equality $(\star)$ also holds for $\varphi_d=0$. It follows that

$$\begin{aligned}\sum_{d\ge 1} d\varphi_d p_{s,d}&=s\exp(-\lambda_C)\sum_{d\ge 0}\sum_{\substack{(n_i)_i\in\mathbb{N}_0^{(\mathbb{N})}\\ \sum_i in_i=d}}\prod_{i=1}^{d}\frac{([z^i]B'(yz))^{n_i}}{n_i!}n_s\\ &=s\exp(-\lambda_C)\left(\sum_{n_s\ge 1}\frac{([z^s]B'(yz))^{n_s}}{n_s!}n_s\right)\prod_{i\in\mathbb{N}\setminus\{s\}}\sum_{n_i\ge 0}\frac{([z^i]B'(yz))^{n_i}}{n_i!}\\ &=s\exp(-\lambda_C)\left([z^s]B'(yz)\right)\prod_{i\in\mathbb{N}}\exp([z^i]B'(yz)).\end{aligned}$$

Since $\lambda_C=B'(y)$ we have that

$$\sum_{d\ge 1} d\varphi_d p_{s,d}=s[z^s]B'(yz)=[z^s]B'^{\bullet}(yz).$$

□

5.7.3 Number of blocks along a path

Given an enriched tree $(T, \alpha) \in \mathcal{A}$ and a path $P = v_1, \dots, v_{l+1}$ in T emanating from the root, we may count the number of blocks of a given size s along that path. We denote it by

$$b(s, P) := |\{1 \le i \le \ell \mid v_{i+1} \text{ lies in a derived block of size } s \text{ in } \alpha(v_i)\}|.$$

Suppose that the class $\mathcal{C}$ is subcritical. Then the probability generating function $\varphi(z) = \exp(B'(yz) - \lambda_C)$ is analytic at the point $z = 1$ and has expected value $\varphi'(1) = 1$. In particular we may apply the results of Section 5.7.1 to random trees drawn by the sampler $\Gamma T^\bullet$.

Lemma 5.7.9. *Suppose that the class $\mathcal{C}$ is subcritical. Let (T_n, α) be an enriched tree of size n drawn according to ΓA_n. Let $\epsilon > 0$ and an integer $S \ge 1$ be given. Then the following holds with high probability as n tends to infinity. For every path $P = v_1 \dots v_{\ell+1}$ in T_n emanating from the root with length $\ell \ge \log^4(n)$ we have that*

$$b(s, P) \in (1 \pm \epsilon)\ell[z^s]B'^\bullet(yz)$$

for all integers $1 \le s \le S$.

Proof. Let (T_n, α) be an enriched tree of size n drawn according to ΓA_n. Clearly it suffices to show that for any given $\epsilon > 0$ and $s \ge 1$ the above holds with high probability. If $[z^s]B'^\bullet(yz) = 0$ then $[z^s]B'(yz) = 0$ and in particular $\mathbb{P}(|\Gamma B'(y)| = s) = 0$. It follows that the random graph $\Gamma C^\bullet(\rho)$ has almost surely no blocks of size $s + 1$ and the claim holds trivially. Hence we may suppose that $[z^s]B'^\bullet(yz) > 0$. Let $0 < \epsilon_1, \epsilon_2, \epsilon_3 < 1$ be some constants depending only on ϵ and s. We will choose convenient values later on. Let $n \ge 1$. For any $D \ge 1$ call a tree $T \in \mathcal{T}_n^\bullet$ D-good, if $\mathbb{P}(\Gamma T_n^\bullet = T) > 0$ and for any path $P = v_1 \dots v_{\ell+1}$ in T emanating from the root with length $\ell \ge \log(n)^4$ we have that

$$|\{1 \le i \le \ell \mid d_T^+(v_i) = d\}| \in (1 \pm \epsilon_1)\ell d\varphi_d \quad \text{and} \quad |\{1 \le i \le \ell \mid d_T^+(v_i) > D\}| \le \epsilon_2 \ell$$

for all $1 \le d \le D$. Since $\mathcal{C}$ is subcritical by assumption, it follows by Corollaries 5.7.4 and 5.7.5 that there is a constant $D_0 \ge 1$ such that for all $D \ge D_0$ we have that

$$\lim_{n \to \infty} \mathbb{P}(\mathsf{T}_n \text{ is } D\text{-good}) = 1.$$

By Corollary 5.7.8 we know that

$$[z^s]B'^\bullet(yz) = \sum_{d \ge 1} d\varphi_d p_{s,d}$$

with

$$p_{s,d} = \sum_{\substack{(n_i)_i \in \mathbb{N}_0^{(\mathbb{N})} \\ \sum_i i\nu_i = d}} \mathbb{P}(\mathrm{PSeq}(d) = (\nu_1, \dots, \nu_d)) \frac{s\nu_s}{d}$$

for $\varphi_d \neq 0$ and $p_{s,d} = 0$ if $\varphi_d = 0$. Choose some constant $D \geq D_0$ such that

$$[z^s]B'^{\bullet}(yz) - \sum_{d=1}^{D} d\varphi_d p_{s,d} < \epsilon_3.$$

We will show that for every D-good tree $T \in \mathcal{T}_n^{\bullet}$ and P a path in T emanating from the root with length $\ell(P) \geq \log(n)^4$ we have that

$$\mathbb{P}(b(s,P) \notin (1 \pm \epsilon)\ell[z^s]B'^{\bullet}(yz) \mid \mathsf{T}_n = T) \leq \exp(-C\log(n)^4) \qquad (\star)$$

for some constant $C > 0$ depending only on ϵ and s. This suffices to prove the claim: The probability

$$\mathbb{P}(b(s,P) \notin (1 \pm \epsilon)\ell[z^s]B'^{\bullet}(yz) \text{ for some path } P \text{ with } \ell(P) \geq \log(n)^4)$$

is bounded by

$$\mathbb{P}(\mathsf{T}_n \text{ is not } D\text{-good}) + \sum_{\substack{T \in \mathcal{T}_n \ D\text{-good} \\ P \text{ path in } T \text{ with } \ell(P) \geq \log(n)^4}} \mathbb{P}(b(s,P) \notin (1 \pm \epsilon)\ell[z^s]B'^{\bullet}(yz) \mid \mathsf{T}_n = T)\mathbb{P}(\mathsf{T}_n = T).$$

A tree of size n has n different paths emanating from the root. By applying inequality $(\star)$ we thus obtain the upper bound

$$\mathbb{P}(\mathsf{T}_n \text{ is not } D\text{-good}) + n\exp(-C\log(n)^4)\mathbb{P}(\mathsf{T}_n \text{ is } D\text{-good}) = o(1).$$

This proves the claim. Hence it remains to show that inequality $(\star)$ holds. Let $T \in \mathcal{T}_n^{\bullet}$ be a D-good tree and suppose that $\mathsf{T}_n = T$. Note that the family $(\alpha(v))_{v \in V(T)}$ of $\mathrm{SET} \circ \mathcal{B}'$-structures on the offspring sets of the tree T is independent with respect to the conditioned probability measure $\mathbb{P}(\cdot \mid \mathsf{T}_n = T)$. Let $P = v_1 \ldots v_{\ell+1}$ be a path in T emanating from the root with length $\ell \geq \log(n)^4$. We have that

$$b(s,P) = \sum_{d \geq 1} X_d \quad \text{with} \quad X_d := \sum_{\substack{1 \leq i \leq \ell \\ d_T^+(v_i) = d}} \mathbb{1}_{\{v_{i+1} \text{ lies in a } \mathcal{B}'\text{-object of size } s \text{ in } \alpha(v_i)\}}.$$

Since the tree T is D-good, it follows by Corollary 5.7.8 that

$$\mathbb{E}[X_d] \in (1 \pm \epsilon_1)\ell d\varphi_d p_{s,d} \quad \text{for } d \leq D$$

and

$$\sum_{d > D} X_d \leq \epsilon_2 \ell.$$

Let $1 \leq d \leq D$. If $\varphi_d p_{s,d} = 0$ then $X_d = 0$ holds $\mathbb{P}(\cdot \mid \mathsf{T} = T)$-almost surely. Suppose that $\varphi_d p_{s,d} \neq 0$. We have that

$$\mathbb{P}(X_d \notin (1 \pm \epsilon_1)\ell d\varphi_d p_{s,d} \mid \mathsf{T}_n = T) \leq \mathbb{P}(X_d \notin (1 \pm \epsilon_1/3)\mathbb{E}[X_d] \mid \mathsf{T}_n = T).$$

Applying the Chernoff bounds yields that this probability is bounded by

$$\exp(-\mathbb{E}[X_d]\epsilon_1^2/36) \leq \exp(-\epsilon_1^2(1-\epsilon_1)\ell d\varphi_d p_{s,d}/36).$$

Since $\ell \geq \log(n)^4$ it follows that

$$\begin{aligned}\mathbb{P}(b(s,P) \notin (1\pm\epsilon_1)\ell\sum_{d=1}^{D} d\varphi_d p_{s,d} + \epsilon_2\ell) &\leq \sum_{\substack{1\leq d\leq D\\ \varphi_d p_{s,d}\neq 0}} \exp(-\epsilon_1^2(1-\epsilon_1)\log(n)^4 d\varphi_d p_{s,d}/36)\\ &\leq \exp(-C\log(n)^4)\end{aligned}$$

with $C > 0$ depending only on ϵ_1 and D. By choice of D we have that

$$0 \leq [z^s]B'^{\bullet}(yz) - \sum_{d=1}^{D} d\varphi_d p_{s,d} < \epsilon_3.$$

Hence

$$\mathbb{P}(b(s,P) \notin (1\pm\epsilon)\ell B'^{\bullet}(yz)) \leq \mathbb{P}(b(s,P) \notin (1\pm\epsilon_1)\ell\sum_{d=1}^{D} d\varphi_d p_{s,d} + \epsilon_2\ell)$$

for a suitable choice of ϵ_1, ϵ_2 and ϵ_3. (For example, we could choose $\epsilon_1 = 2\epsilon$, $\epsilon_2 = \epsilon[z^s]B'^{\bullet}(yz)$ and $\epsilon_3 = \epsilon[z^s]B'^{\bullet}(yz)/(1+2\epsilon)$.) This proves inequality ($\star$). □

5.7.4 Expansion of path length

Given a vertex v of an enriched tree $(T,\alpha) \in \mathcal{A}$ we may consider the distance from the root to the vertex v in the corresponding graph C. We denote this length by $\mathsf{rl}(P)$ where P is the unique path in the tree T connecting the root and v. Let $\mathsf{C}_n \in \mathcal{C}_n^{\bullet}$ be a uniformly at random chosen graph and (T_n, α) the corresponding SET $\circ\, \mathcal{B}'$-enriched tree. Recall that for any $\mathcal{B}'^{\bullet}$-object B the number $\mathsf{d}(B)$ denotes the distance of the $*$-vertex from the root.

Lemma 5.7.10. *Suppose that the graph class $\mathcal{C}$ is subcritical. Then for any $\epsilon > 0$ the following holds with high probability. For every path P in the tree T_n starting from the root with length $\ell \geq \log(n)^4$ we have that*

$$\mathsf{rl}(P) \in (1\pm\epsilon)\ell\mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))]$$

with $\Gamma B'^{\bullet}(y)$ denoting a Boltzmann sampler for the class $\mathcal{B}'^{\bullet}$ at the point y.

Proof. Note that

$$\mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] = \sum_{s\geq 1} p_s \quad \text{with} \quad p_s = \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y)) \mid |(\Gamma B'^{\bullet}(y))| = s][z^s]B'^{\bullet}(yz)$$

is finite, since $p_s \le s[z^s]B'^{\bullet}(yz)$ and the class $\mathcal{C}^{\bullet}$ is subcritical. Let (T_n, α) be an enriched tree drawn according to the sampler ΓA_n and C_n denote the corresponding graph. Let $T \in \mathcal{T}_n$ be a tree with $\mathbb{P}(\mathsf{T}_n = T) > 0$ and for any vertex $v \in V(T)$ let $M(v)$ denote its offspring set. If we condition on $\mathsf{T}_n = T$ then the familiy $(\alpha(v))_{v \in V(T)}$ is independent and each $\mathrm{SET} \circ \mathcal{B}'$ structure $\alpha(v)$ on the set $M(v)$ is generated by the following independent steps.

1. Draw the partition sequence $\nu_1(v), \dots, \nu_d(v)$ according to the distribution $\mathrm{PSeq}(d^+_T(v))$.
2. Choose a bijection $f_v : M(v) \to \bigsqcup_{i,j}[i]$ uniformly at random.
3. For all $i, j \ge 1$ choose a derived block $\mathsf{B}_{i,j}(v) \in \mathcal{B}'_i$ uniformly at random.

The final structure is obtained by selecting the blocks $\mathsf{B}_{i,j}(v)$ with $1 \le i \le d^+(v)$, $1 \le j \le \nu_j(v)$ according to the choices made in step 1 and relabeling them according to the matching chosen in step 2. Let $\beta(v) = ((\nu_i(v))_i, f_v)$ denote the pair of random choices made in steps 1 and 2. Note that for any path P in the tree T emanating from the root the number $b(s, P)$ of derived blocks of size s along that path depends only on the family β. For any possible outcome (T, γ) with $\gamma(v) = ((n_i(v))_i, g_v)$ we have $\mathbb{P}((\mathsf{T}_n, \beta) = (T, \gamma)) > 0$ if and only if $\mathbb{P}(\mathsf{T}_n = T) > 0$ and $n_i(v) = 0$ whenever $[z^i]B'(z) = 0$. Now, let $\epsilon > 0$ be given and $0 < \epsilon_1, \epsilon_2, \epsilon_3 < 1$ be some constants depending only on ϵ. We will choose convenient values later on. By Corollary 5.7.5 there exists $S \ge 1$ such that with high probability all paths $P = v_1, \dots, v_{l+1}$ in $\Gamma T_n^{\bullet}$ that start from the root and have length $\ell \ge \log(n)^2$ satisfy

$$\sum_{\substack{1 \le i \le \ell \\ d^+(v_i) > S}} d^+(v_i) \le \epsilon_1 \ell. \tag{$\star$}$$

This also holds for all constants bigger than S, hence according to Corollary 5.7.8 we may choose S large enough such that additionally

$$\sum_{s > S} p_s \le \epsilon_2.$$

We say the pair (T, γ) is S-good, if $\mathbb{P}((\mathsf{T}_n, \beta) = (T, \gamma)) > 0$ and for all paths $P = v_1, \dots, v_{\ell+1}$ emanating from the root with length $\ell \ge \log(n)^4$ we have that equation $(\star)$ holds and additionally $b(s, P) \in (1 \pm \epsilon_3) l[z^s]B'^{\bullet}(yz)$ for all $1 \le s \le S$. By Corollary 5.7.5 and Lemma 5.7.9 it follows that the pair (T_n, β) is S-good with high probability as n tends to infinity. We will show that for every S-good pair (T, γ) and P a path in T emanating from the root with length $\ell(P) \ge \log(n)^4$ we have that

$$\mathbb{P}(\mathsf{rl}(P) \notin (1 \pm \epsilon)\ell(P)\mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] \mid (\mathsf{T}_n, \beta) = (T, \gamma)) \le \exp(-C \log(n)^4) \tag{$\star\star$}$$

for some constant $C > 0$ depending only on ϵ. This suffices to prove the claim: The probability

$$\mathbb{P}(\mathsf{rl}(P) \notin (1 \pm \epsilon) l \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] \text{ for some path } P \text{ with } l(P) \geq \log(n)^4)$$

is bounded by the sum of the probability $\mathbb{P}((\mathsf{T}_n, \beta)$ is not S-good) and

$$\sum_{\substack{(T,\gamma)\ S\text{-good} \\ P \text{ path in } T \text{ with } \ell(P) \geq \log(n)^4}} \mathbb{P}(\mathsf{rl}(P) \notin (1 \pm \epsilon) \ell \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] \mid (\mathsf{T}_n, \beta) = (T, \gamma)) \mathbb{P}((\mathsf{T}_n, \beta) = (T, \gamma)).$$

A tree of size n has n different paths emanating from the root. By applying inequality $(\star\star)$ we thus obtain the upper bound

$$\mathbb{P}((\mathsf{T}_n, \beta) \text{ is not } S\text{-good}) + n \exp(-C \log(n)^4) \mathbb{P}((\mathsf{T}_n, \beta) \text{ is } S\text{-good}) = o(1).$$

This proves the claim. Hence it remains to show that inequality $(\star\star)$ holds. Let (T, γ) be S-good and suppose that $(\mathsf{T}_n, \beta) = (T, \gamma)$. Let $P = v_1, \ldots, v_{\ell+1}$ be a path in the tree T emanating from the root with length $\ell \geq \log(n)^4$. For all $1 \leq i \leq \ell$ let $d(v_i, v_{i+1})$ denote the length of a shortest path connecting the vertices v_i and v_{i+1} in the graph C_n. Then the distances $d(v_i, v_{i+1})$ are independent and

$$\mathsf{rl}(P) = \sum_{i=1}^{\ell} d(v_i, v_{i+1}). \qquad (\star\star\star)$$

Given an index $1 \leq i \leq \ell$ let B denote the derived block containing the vertex v_{i+1} and s its size. Then s is determined by (more precisely, $\mathbb{P}(\cdot \mid (\mathsf{T}_n, \beta) = (T, \gamma))$-almost surely equal to a constant determined by) the pair (T, γ). The derived block B is generated by drawing a block B_s uniformly at random from the set $\mathcal{B}'_s$ and relabeling by a fixed bijective function $\sigma : [s] \to M$ determined by γ. Hence the distance $d(v_i, v_{i+1})$ is equal to the length of a shortest path from the $*$-vertex to the vertex $v := \sigma^{-1}(v_{i+1}) \in [s]$ in the derived block B_s. Since B_s was chosen uniformly at random and the set $\mathcal{B}_s$ is closed under relabeling, this distance $d(*, v)$ is distributed like the distance $d(*, r)$ from the $*$-vertex to a independently and uniformly at random chosen non-$*$-vertex $r \in [s]$. In particular, it is distributed like the distance from the $*$-vertex to the root in a uniformly chosen block $\mathsf{B}_s^\bullet \in \mathcal{B}'^{\bullet}_s$. See Lemma 5.7.11 below for details. Hence

$$d(v_i, v_{i+1}) \overset{(d)}{=} \mathsf{d}(\Gamma B'^{\bullet}_s(y)).$$

For each $1 \leq s \leq S$ let I_s denote the set of all indices $1 \leq i \leq \ell$ such that v_{i+1} lies in a derived block of size s. Since $d(v_i, v_{i+1}) \leq d_T^+(v_i)$ it follows by equations $(\star)$ and $(\star\star\star)$ that

$$\mathsf{rl}(P) = \sum_{s=1}^{S} \sum_{i \in I_s} d(v_i, v_{i+1}) + R\ell$$

with $0 \leq R \leq \epsilon_1$. Clearly $|I_s| = b(s, P)$ for all s. Since the pair (T, γ) is S-good, we have that $b(s, P) \in (1 \pm \epsilon_3)\ell[z^s]B'^{\bullet}(yz)$ for all $s \leq S$. In particular, $b(s, P) \neq 0$ if and only if $[z^s]B'^{\bullet}(yz) \neq 0$. Suppose that $b(s, P) \neq 0$. Then

$$\mathbb{E}[\sum_{i \in I_s} d(v_i, v_{i+1})] \in (1 \pm \epsilon_3)\ell p_s.$$

For convenience, let $\mathcal{E}$ denote the event $(\mathsf{T}_n, \beta) = (T, \gamma)$. By monotonicity we have that

$$\mathbb{P}(\sum_{i \in I_s} d(v_i, v_{i+1}) \notin (1 \pm \epsilon_3)\, \ell p_s \mid \mathcal{E}) \leq \mathbb{P}(\sum_{i \in I_s} d(v_i, v_{i+1}) \notin (1 \pm \epsilon_3/3)\, \mathbb{E}[\sum_{i \in I_s} d(v_i, v_{i+1})] \mid \mathcal{E}).$$

We have that $d(v_i, v_{i+1}) \in [s]$ for all $i \in I_s$. Hence we may apply Hoeffding's inequality to bound this probability by

$$2 \exp\left(-2 \frac{(\mathbb{E}[\sum_{i \in I_s} d(v_i, v_{i+1})]\epsilon_3/3)^2}{b(s, P)s^2}\right) \leq 2 \exp\left(-\frac{2}{9} \frac{\epsilon_3^2 (1 - \epsilon_3)^2}{(1 + \epsilon_3)^2} \frac{p_s \mathbb{E}[\mathsf{d}(\mathsf{B}_s^\bullet)]}{s^2} \ell\right).$$

Since $l \geq \log(n)^4$, it follows that

$$\mathbb{P}(\sum_{s=1}^{S} \sum_{i \in I_s} d(v_i, v_{i+1}) \notin (1 \pm \epsilon_3) l \sum_{s=1}^{S} p_s \mid \mathcal{E}) \leq \exp(-C \log(n)^4)$$

for some constant $C > 0$ depending only on the ϵ_i and S. By choice of S we have that

$$0 \leq \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] - \sum_{s=1}^{S} p_s \leq \epsilon_2.$$

Hence

$$(1 \pm \epsilon_3)\ell \sum_{s=1}^{S} p_s + Rl \subset (1 \pm \epsilon)\ell \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))]$$

for a suitable choice for the ϵ_i. (For example, we could choose $\epsilon_3 = \frac{\epsilon}{2}$ and $\epsilon_1 = \epsilon_2 = \min(\frac{1}{2}, \frac{\epsilon}{2}\mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))]^{-1})$.) By monotonicity we get

$$\mathbb{P}(\mathsf{rl}(P) \notin (1 \pm \epsilon) l \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] \mid \mathcal{E}) \leq \exp(-C \log(n)^4).$$

This proves inequality ($\star\star$). □

Lemma 5.7.11. *Let $s \geq 1$ and $v \in [s]$. Let $\mathsf{B} \in \mathcal{B}'_s$ and $\mathsf{B}^\bullet \in \mathcal{B}'^{\bullet}_s$ be drawn uniformly at random. Then*

$$d_{\mathsf{B}}(*, v) \overset{(d)}{=} d_{\mathsf{B}^\bullet}(*, \bullet),$$

i.e. the length of a shortest path connecting the $*$-vertex and the vertex v in the derived block B is distributed like the distance from the $*$-vertex to the root in the block $\mathsf{B}^\bullet$.

Proof. Given $v, w \in [s]$ take a permuation $\tau \in S_s$ with $\tau(v) = w$. Then $\tau.\mathsf{B} \overset{(d)}{=} \mathsf{B}$ and hence

$$d_{\mathsf{B}}(*, v) = d_{\tau.\mathsf{B}}(*, w) \overset{(d)}{=} d_{\mathsf{B}}(*, w).$$

If we choose a vertex $r \in [s]$ independently and uniformly at random, then the rooted block $(\mathsf{B}, r) \in \mathcal{B}'^{\bullet}$ is uniformly distributed, i.e $(\mathsf{B}, r) \overset{(d)}{=} \mathsf{B}^{\bullet}$. We have that

$$\mathbb{P}(d_{\mathsf{B}}(*, r) = t) = \sum_{w \in [s]} \mathbb{P}(d_{\mathsf{B}}(*, w) = t)\mathbb{P}(r = w) = \mathbb{P}(d_{\mathsf{B}}(*, v) = t).$$

Hence $d_{\mathsf{B}}(*, v) \overset{(d)}{=} d_{\mathsf{B}^{\bullet}}(*, \bullet)$. □

Lemma 5.7.12. *Suppose that the graph class $\mathcal{C}$ is subcritical. Then $\mathsf{lb}(\mathsf{C}_n) = O(\log(n))$ with high probability.*

Proof. Clearly we have that

$$\mathsf{lb}(\mathsf{C}_n) \leq \Delta(\mathsf{T}_n) + 1$$

where $\Delta(\mathsf{T}_n)$ denotes the maximum out-degree of the tree T_n. By assumption, the probability generating function $\varphi(z)$ of the offspring distribution is analytic at the point 1. By [MM90, Jan12] the maximum outdegree satisfies

$$\Delta(\mathsf{T}_n) = O(\log(n))$$

with high probability. This proves the claim. □

Proofs for the logarithmic bound of the largest block are also given in [PS10, DN13].

Theorem 5.7.13. *Let $\mathsf{C}_n \in \mathcal{C}_n^{\bullet}$ be a uniformly at random chosen rooted graph of size n from a subcritical class and (T_n, α) the corresponding $\mathrm{SET} \circ \mathcal{B}'$-enriched tree. Then for any $\epsilon > 0$ the following holds with high probability. For every path P in the tree T_n we have that*

$$\mathsf{rl}(P) \in (1 \pm \epsilon)\ell(P)\mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] + O(\log(n)^5)$$

with $\Gamma B'^{\bullet}(y)$ denoting a Boltzmann sampler for the class $\mathcal{B}'^{\bullet}$ at the point y.

Proof. Let $\epsilon > 0$ be given. Let P be a path in the tree T_n connecting the vertices x and y. We let r denote the root of T_n and a the last common ancestor of the vertices x and y. Clearly we have that

$$\ell(P) = d_{\mathsf{T}_n}(r, x) + d_{\mathsf{T}_n}(r, y) - 2d_{\mathsf{T}_n}(r, a).$$

A shortest path connecting the vertices x and y in the graph C_n might take a shortcut in a single block to avoid the vertex a. Thus the corresponding path lengths in the graph C_n satisfy

$$\mathsf{rl}(P) = d_{\mathsf{C}_n}(r, x) + d_{\mathsf{C}_n}(r, y) - 2d_{\mathsf{C}_n}(r, a) + R$$

with $|R| \leq 2\mathsf{lb}(\mathsf{C}_n)$. By Lemmata 5.7.12 and 5.7.10 we have with high probability for any vertex z

$$d_{\mathsf{C}_n}(r,z) \in (1 \pm \epsilon)\mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))]d_{\mathsf{T}_n}(r,z)$$

if $d_{\mathsf{T}_n}(r,z) \geq \log(n)^4$ and otherwise

$$d_{\mathsf{C}_n}(r,z) = O(\log(n)^5) = \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))]d_{\mathsf{T}_n}(r,z) + O(\log(n)^5).$$

Thus it follows that in this case we have that

$$\mathsf{rl}(P) \in (1 \pm \epsilon)\ell(P)\mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))] + O(\log(n)^5).$$

□

5.7.5 The scaling limit

Let C_n and $\mathsf{C}_n^\bullet$ denote the labeled unrooted and rooted random graph drawn uniformly from the graphs of size n of the subcritical class of connected graphs $\mathcal{C}$. Recall that ρ denotes the radius of convergence of the generating series $C(z)$. Let $\kappa = \mathbb{E}[\mathsf{d}(\Gamma B'^{\bullet}(y))]$ denote the expected distance between the two roots of a doubly rooted block drawn from the class $\mathcal{B}'^{\bullet}$ according to the Boltzman distribution with parameter $y = C^\bullet(\rho)$. We may thus obtained an alternative proof of Theorem 1.4.1.

Theorem 5.7.14. *The rescaled graph $\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{C}_n^\bullet$ converges in distribution to the continuum random tree $\mathcal{T}_{\mathsf{e}}$ with respect to the (pointed) Gromov-Hausdorff metric.*

Since $\mathsf{C}_n^\bullet$ and C_n are identically distributed as metric spaces, the same holds for labeled unrooted graphs.

Proof. Consider the coupling with the conditioned GWT T_n. Given any bounded Lipschitz-continuous function $f : \mathbb{K} \to \mathbb{R}$ with upper bound M and Lipschitz-constant L we have that

$$|\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{C}_n^\bullet)] - \mathbb{E}[f(\mathcal{T}_{\mathsf{e}})]| \leq |\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{C}_n^\bullet)] - \mathbb{E}[f(\frac{\sigma}{2\sqrt{n}}\mathsf{T}_n^\bullet))]| + o(1)$$

By Theorem 5.7.13, and considering the distortion of the natural correspondence between the vertices of T_n and $\mathsf{C}_n^\bullet$, we know that with high probability

$$d_{\mathrm{GH}}(\mathsf{C}_n^\bullet/(\kappa\sqrt{n}), \mathsf{T}_n/\sqrt{n}) = o(1).$$

Call this event $\mathcal{E}_n$. Then

$$|\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathsf{C}_n^\bullet)] - \mathbb{E}[f(\frac{\sigma}{2\sqrt{n}}\mathsf{T}_n^\bullet))]| \leq L\,o(1) + M\mathbb{P}(\mathcal{E}_n^{\mathsf{c}}) = o(1).$$

This concludes the proof. □

Bibliography

[ABDJ13] Louigi Addario-Berry, Luc Devroye, and Svante Janson. Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees. *Ann. Probab.*, 41(2):1072–1087, 2013.

[Ald91a] David Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[Ald91b] David 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.

[Ald93] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[AM08] Marie Albenque and Jean-François Marckert. Some families of increasing planar maps. *Electron. J. Probab.*, 13:no. 56, 1624–1671, 2008.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BBY06] Jason P. Bell, Stanley N. Burris, and Karen 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.

[Bet15] Jérémie Bettinelli. Scaling limit of random planar quadrangulations with a boundary. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):432–477, 2015.

[BF08] Nicolas Broutin and Philippe Flajolet. The height of random binary unlabelled trees. In *Fifth Colloquium on Mathematics and Computer Science*, Discrete Math. Theor. Comput. Sci. Proc., AI, pages 121–134. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[BF12] Nicolas Broutin and Philippe Flajolet. The distribution of height and diameter in random non-plane binary trees. *Random Structures Algorithms*, 41(2):215–252, 2012.

[BFKV07] Manuel Bodirsky, Éric Fusy, Mihyun Kang, and Stefan Vigerske. Enumeration and asymptotic properties of unlabeled outerplanar graphs. *Electron. J. Combin.*, 14(1):Research Paper 66, 24, 2007.

[BFKV11] Manuel Bodirsky, Éric Fusy, Mihyun Kang, and Stefan Vigerske. Boltzmann samplers, Pólya theory, and cycle pointing. *SIAM J. Comput.*, 40(3):721–769, 2011.

[BGH05] Nicolas Bonichon, Cyril Gavoille, and Nicolas Hanusse. Canonical decomposition of outerplanar maps and application to enumeration, coding and generation. *J. Graph Algorithms Appl.*, 9(2):185–204 (electronic), 2005.

[BLL98] François Bergeron, Gilbert Labelle, and Pierre 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.

[BPS09] Nicla Bernasconi, Konstantinos Panagiotou, and Angelika Steger. The degree sequence of random graphs from subcritical classes. *Combin. Probab. Comput.*, 18(5):647–681, 2009.

[BPS10] Nicla Bernasconi, Konstantinos Panagiotou, and Angelika Steger. On properties of random dissections and triangulations. *Combinatorica*, 30(6):627–654, 2010.

[BPY01] Philippe Biane, Jim Pitman, and Marc Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bull. Amer. Math. Soc. (N.S.)*, 38(4):435–465 (electronic), 2001.

[Car] Alessandra Caraceni. The scaling limit of random outerplanar maps. *To appear in Annales de l'Institut Henri Poincaré.*

[CHK14] Nicolas Curien, Bénédicte Haas, and Igor Kortchemski. The crt is the scaling limit of random dissections. *Random Structures & Algorithms*, pages n/a–n/a, 2014.

[Chu76] Kai Lai Chung. Excursions in Brownian motion. *Ark. Mat.*, 14(2):155–177, 1976.

[DFK+11] Michael Drmota, Éric Fusy, Mihyun Kang, Veronika Kraus, and Juanjo Rué. Asymptotic study of subcritical graph classes. *SIAM J. Discrete Math.*, 25(4):1615–1651, 2011.

[DFLS02] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles 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.

[DFLS04] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combin. Probab. Comput.*, 13(4-5):577–625, 2004.

[DG10] Michael Drmota and Bernhard Gittenberger. The shape of unlabeled rooted random trees. *European J. Combin.*, 31(8):2028–2063, 2010.

[Die10] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Heidelberg, fourth edition, 2010.

[DLG05] Thomas Duquesne and Jean-François Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005.

[DLL82] H. Décoste, G. Labelle, and P. Leroux. Une approche combinatoire pour l'itération de Newton-Raphson. *Adv. in Appl. Math.*, 3(4):407–416, 1982.

[DN13] Michael Drmota and Marc Noy. Extremal parameters in sub-critical graph classes. In *ANALCO13—Meeting on Analytic Algorithmics and Combinatorics*, pages 1–7. SIAM, Philadelphia, PA, 2013.

[Duq03] Thomas Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003.

[FFP07] Philippe Flajolet, Éric Fusy, and Carine 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.

[FS09] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

[GMN13] O. Gimenez, D. Mitsche, and M. Noy. Maximum degree in minor-closed classes of graphs. *ArXiv e-prints*, April 2013.

[HM12] Bénédicte Haas and Grégory Miermont. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.*, 40(6):2589–2666, 2012.

[HP73] Frank Harary and Edgar M. Palmer. *Graphical enumeration.* Academic Press, New York-London, 1973.

[Jan12] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012.

[Joy81] André Joyal. Une théorie combinatoire des séries formelles. *Adv. in Math.*, 42(1):1–82, 1981.

[JS15] Svante Janson and Sigurdur Örn Stefánsson. Scaling limits of random planar maps with a unique large face. *Ann. Probab.*, 43(3):1045–1081, 2015.

[Kor13] Igor Kortchemski. A simple proof of Duquesne's theorem on contour processes of conditioned Galton-Watson trees. In *Séminaire de Probabilités XLV*, volume 2078 of *Lecture Notes in Math.*, pages 537–558. Springer, Cham, 2013.

[Lab81] Gilbert Labelle. Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange. *Adv. in Math.*, 42(3):217–247, 1981.

[Lab10] Gilbert Labelle. On extensions of the Newton-Raphson iterative scheme to arbitrary orders. In *22nd International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2010)*, Discrete Math. Theor. Comput. Sci. Proc., AN, pages 845–856. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010.

[LG10] Jean-François Le Gall. Itô's excursion theory and random trees. *Stochastic Process. Appl.*, 120(5):721–749, 2010.

[LGM12] Jean-François Le Gall and Grégory 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.

[LPP95] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.

[ML14] Miguel A. Méndez and Jean Carlos Liendo. An antipode formula for the natural Hopf algebra of a set operad. *Adv. in Appl. Math.*, 53:112–140, 2014.

[MM90] A. Meir and J. W. Moon. On the maximum out-degree in random trees. *Australas. J. Combin.*, 2:147–156, 1990. Combinatorial mathematics and combinatorial computing, Vol. 2 (Brisbane, 1989).

[MM11] Jean-François Marckert and Grégory Miermont. The CRT is the scaling limit of unordered binary trees. *Random Structures Algorithms*, 38(4):467–501, 2011.

[MSW06] Colin McDiarmid, Angelika Steger, and Dominic J. A. Welsh. Random graphs from planar and other addable classes. In *Topics in discrete mathematics*, volume 26 of *Algorithms Combin.*, pages 231–246. Springer, Berlin, 2006.

[Ott48] Richard Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.

[Pól37] George Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.*, 68(1):145–254, 1937.

[PS10] Konstantinos Panagiotou and Angelika Steger. Maximal biconnected subgraphs of random planar graphs. *ACM Trans. Algorithms*, 6(2):Art. 31, 21, 2010.

[PS15] K. Panagiotou and B. Stufler. Scaling limits of random P\'{o}lya trees. *ArXiv e-prints*, February 2015.

[PSW14] Konstantinos Panagiotou, Benedikt Stufler, and Kerstin Weller. Scaling Limits of Random Graphs from Subcritical Classes. *ArXiv e-prints*, November 2014.

[RS67] A. Rényi and G. Szekeres. On the height of trees. *J. Austral. Math. Soc.*, 7:497–507, 1967.

[Stu14] Benedikt Stufler. The continuum random tree is the scaling limit of unlabelled unrooted trees. *ArXiv e-prints*, December 2014.

[Stu15a] B. Stufler. Random enriched trees with applications to random graphs. *ArXiv e-prints*, April 2015.

[Stu15b] B. Stufler. Scaling limits of random outerplanar maps with independent link-weights. *ArXiv e-prints*, May 2015.

[Sze83] 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.

[Wan15] M. Wang. Height and diameter of brownian tree. *ArXiv e-prints*, March 2015.

[Wil06] Herbert S. Wilf. *generatingfunctionology.* A K Peters, Ltd., Wellesley, MA, third edition, 2006.