The structure of large random graphs

Serte Donderwinkel

Department of Statistics
University of Oxford

This dissertation is submitted for the degree of
Doctor of Philosophy

St Hugh's College
Trinity Term 2022

Voor mama,

for building us the safest nest and giving us the freedom to fly out.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original. This thesis is submitted to the University of Oxford for the degree Doctor of Philosophy, and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Serte Donderwinkel
Trinity Term 2022

Acknowledgements

I could not start this thesis in any other way than by thanking Christina Goldschmidt and Louigi Addario-Berry. Your mentorship and guidance for all aspects of this job have made me a better researcher, teacher and colleague. Whether the supervisions took place in an office, in a garden, in a park, under a gazebo or in an online meeting room, I always left with more maths in my brain and a bigger smile on my face (and sometimes also with Christina's produce in my bag or a running appointment with Louigi in my diary). I hope that one day I can mean to my students what you have meant to me.

James Martin and Minmin Wang, thank you for being the best pair of examiners that I could have wished for. Your comments on both the big picture and the nitty-gritty have improved the quality of this thesis by a lot and have made me extra excited about the maths ahead. Maybe we will even be able to answer some of those intriguing questions together.

The Oxford probability group is a wonderful workplace thanks to its members (and the cake they provide). I want to especially thank Aaron, Dominic, Mickaël, Rivka, Zheneng, Kim, Alessandra, James, Romek, Alexandra, Maria, Robin, David and Freddy for the companionship along the way. I also want to thank Alison Etheridge for her continuous efforts to keep an entire department (and specifically the third floor) in good spirits during the pandemic.

Carla, Thomas, Tom and Joost, work gets better when you do it with friends and friendship gets better when you work together. Let us stick around in this positive feedback loop to retirement and beyond.

Being part of the team of maths tutors at St Hugh's College has been tremendously rewarding. I want to especially thank Tom Sanders and James Martin for being great colleagues and my lovely students for their clever questions and answers, enthusiasm and legible handwriting.

Niels Wensink, Adam Laasri, Wim den Ouden, Bert Tromp, Paul Wiewel, Kanat Camlibel, Daniel Valesin, Aernout van Enter and Perla Sousi, thank you for your encouragement during the different stages of my mathematical career. This thesis would probably not exist without

you.

Jet, Margreeth, Victoria, Fieneke, Leonie, Reza, Annie, Vicky, Josh, Annelies, Isadora, Lotte, Laura, Rob, Jolien, Nikki, Kristel, Janneke, Willeke, Jasper, Steven, Hilbert, Friso, Dominique, Erika, Fernanda, Claire and Nicky, thank you for the joy and warmth you bring to my life and for your patience with my puns.

Mama, Jan-Willem en Hervé, waar al die wiskunde me ook brengt, jullie blijven altijd mijn thuis. Bedankt voor alles.

Abstract

We obtain universality results on the structure of random graphs from two different angles. Naturally, there is a trade-off between the generality of the assumptions on a model and the specificity of the results one may obtain. The strength of the first work is its generality: we obtain non-asymptotic universal tail bounds on the height of random trees that do not require any assumptions on e.g. the tail behaviour of the degrees. In the second work, we need stronger assumptions, but our description of the structure is more specific: we do not study a real-valued statistic of the random graph model, but show convergence in distribution under rescaling of the graph itself, which entails the convergence of a whole panoply of such statistics.

Firstly, we obtain new non-asymptotic universal tail bounds for the height of uniformly random trees with a given degree sequence. We also obtain universal tail bounds for the height of simply generated trees and conditioned Bienaymé trees (the family trees of branching processes) that settle several conjectures from the literature [68, 2, 4, 8]. Moreover, we define a partial ordering on degree sequences and show that it induces a stochastic ordering on the heights of uniformly random trees with given degree sequences. The latter result implies that sub-binary random trees are stochastically the tallest trees with a given number of vertices and leaves.

Secondly, we consider the strongly connected components (SCCs) of a uniform directed graph on n vertices with i.i.d. in- and out-degree pairs distributed as (D^-, D^+), with $\mathbb{E}[D^+] = \mathbb{E}[D^-] = \mu$. We condition on equal total in- and out-degree. A phase transition for the emergence of a giant SCC is known to occur at the critical point where $\mathbb{E}[D^-D^+] = \mu$. We study the model at this critical point and show that, under some additional finite moment conditions, the SCCs ranked by decreasing number of edges with distances rescaled by $n^{-1/3}$ converge in distribution to a sequence of finite strongly connected directed multigraphs with edge lengths, and that these are either 3-regular or loops almost surely. This is the first universality result for the scaling limit of a critical directed graph model and the first quantitative result on the directed configuration model at criticality.

Contents

3 The directed configuration model 85

Chapter 1

Introduction

In this chapter, we introduce the models and techniques that play a central rôle in this thesis. First, we introduce the relevant graph theoretic objects and introduce the encoding of trees by sequences and excursions that are key to the results in Chapters 2 and 3 respectively. Thereafter, we discuss several models for random trees and the configuration model, in both the undirected and directed set-ups. Then, we introduce the results that are presented in Chapters 2 and 3. We conclude with a review of earlier work on the height of random trees and the configuration model.

1.1 Random trees, random graphs and random directed graphs

1.1.1 About graphs and trees

1.1.1.1 Graphs and directed graphs

We start by introducing multigraphs, graphs, directed multigraphs and directed graphs. A *multigraph* is a tuple $M = (V, E)$ where V is a set of vertices and E is a multiset of multisets of size 2 of vertices. A *graph* $G = (V, E)$ is a multigraph that contains neither loops nor multiple edges, so now E is a set of sets of size 2 of vertices in V. A *directed multigraph* is a tuple $\vec{M} = (V, E, r)$ where V is a set of vertices, E is a set of edges, and $r : E \to V \times V$ is a function mapping each edge to its *tail* and *head*; associated with r are two functions $r_1 : E \to V$ and $r_2 : E \to V$ such that $r(e) = (r_1(e), r_2(e))$ for all $e \in E$. $r_1(e)$ is the tail of the edge e and $r_2(e)$ is the head of the edge e. A *directed graph* $\vec{G} = (V, E, r)$ is a directed multigraph with neither loops nor multiple edges so now r is injective and all edges $e \in E$ satisfy $r_1(e) \neq r_2(e)$.

For v a vertex in a (multi-)graph, its degree $d(v)$ equals $\sum_{w \in V} \#\{e \in E : e = \{v, w\}\}$.

For v a vertex in a directed (multi-)graph, its out-degree $d^+(v)$ equals $\sum_{e\in E} \mathbb{1}_{\{r_1(e)=v\}}$ and its in-degree $d^-(v)$ equals $\sum_{e\in E} \mathbb{1}_{\{r_2(e)=v\}}$.

In a (multi-)graph, the *components* are the maximal connected submultigraphs. In a directed (multi-)graph (V, E, r), we say a vertex v *leads to* a vertex w, written $v \to w$, if there exists a directed path from v to w, i.e. a $k \in \mathbb{N}$ and $e_1, \ldots, e_k \in E$ such that $r_1(e_1) = v$, $r_2(e_k) = w$ and $r_2(e_i) = r_1(e_{i+1})$ for each $1 \leq i < k$. We say v is *strongly connected to* w, written $v \leftrightarrow w$, if v leads to w and w leads to v. By convention, v leads to itself. A graph is *strongly connected* if all pairs of vertices in the graph are strongly connected. The relation $v \leftrightarrow w$ is an equivalence relation; the digraphs induced by the equivalence classes of $\leftrightarrow$ are referred to as the *strongly connected components* (SCCs).

1.1.1.2 Trees and their encodings

A *rooted, labeled tree* $\mathrm{t} = (V, E, \rho)$ is an acyclic graph (V, E) with a distinguished vertex $\rho \in V$ that we call the *root*. The *size* of t is $|\mathrm{t}| := |V|$. For $v \in V$, the *height* $|v|$ of v is the number of edges on the unique path from v to ρ and the *height of* t equals $ht(\mathrm{t}) = \max_{v\in V} |v|$. For $v \neq \rho$, the *parent* of v, denoted by $p(v)$, is the first vertex distinct from v on the unique path from v to ρ. For each $v \in V$, we let the *children* of v be $C(v) = \{w \in V : p(w) = v\}$ and we let its *degree* equal $d_{\mathrm{t}}(v) = |C(v)|$. (This is often called the *out-degree* of vertex v.) We say that a vertex v is a *leaf* if $d_{\mathrm{t}}(v) = 0$ and we say it is a *branchpoint* if $d_{\mathrm{t}}(v) \geq 2$. For S a connected subset of V, let the *path from S to x in* t be the unique path in t which starts at a vertex of S, does not visit any other vertex of S, and ends at x.

We now discuss how to encode rooted, labeled trees with integer labels by sequences. Let $\mathcal{T}(n)$ denote the set of rooted, labeled trees for which $V = [n]$. There are several bijections between $\mathcal{T}(n)$ and $[n]^{n-1}$. In particular, this implies Cayley's formula, which states that

$$|\mathcal{T}(n)| = n^{n-1}.$$

In Chapter 2 we will make particular use of the following bijection that was introduced by Addario-Berry, the author, Maazoun and Martin in [11] and by Blanc-Renaudie in independent work [25].

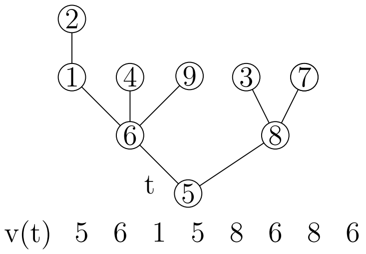


Fig. 1.1: A tree t and the corresponding sequence v(t).

Bijection

For a tree t on $[n]$ with root ρ:

- Let $\ell_1 < \ell_2 < \cdots < \ell_k$ be the leaves of t.
- Let $S_0 = \{\rho\}$.
- Recursively, for $i = 1, \ldots, k$, let P_i be the path in t from S_{i-1} to ℓ_i, and let $S_i = S_{i-1} \cup P_i$. Let P_i^* be P_i omitting its final point.
- Let v(t) be the concatenation of P_1^*, $P_2^*, \ldots, P_k^*$.

It is not hard to see that $\mathrm{v(t)} \in [n]^{n-1}$; a rigorous proof can be found in [11].

We now describe the inverse of the bijection. For a sequence $\mathrm{v} = (v_1, v_2, \ldots, v_m)$, we say that $j \in \{2, \ldots, m\}$ is the location of a repeat of v if $v_j = v_i$ for some $i < j$.

Inverse of the bijection

Given a sequence $\mathrm{v} = (v_1, v_2, \ldots, v_{n-1}) \in [n]^{n-1}$:

- Let $j(0) = 1$, let $j(1) < j(2) < \cdots < j(k-1)$ be the locations of the repeats of the sequence v, and let $j(k) = n$.
- Let $\ell_1 < \ell_2 < \cdots < \ell_k$ be the elements of $[n]$ not occurring in v.
- For $i = 1, \ldots, k$, let P_i be the path $(v_{j(i-1)}, \ldots, v_{j(i)-1}, \ell_i)$ with $j(i) - j(i-1)$ edges.
- Let t(v) be the graph with vertex set $[n]$, with root v_1 and with edge set given by the union of the edges of the paths $P_1, P_2, \ldots, P_k$.

It is easy to convince oneself that this is indeed the inverse of the bijection; a formal proof can be found in [11].

We now introduce the Ulam–Harris notation for the space of ordered, rooted trees.

Definition 1.1.1. *Let* $\mathcal{U} := \{\emptyset\} \cup \bigcup_{k\geq 1} \mathbb{N}^k$*. For* $u = (u_1, \dots, u_k) \in \mathcal{U}\backslash\{\emptyset\}$*, let the* parent *of* u *be* $p(u) := (u_1, \dots, u_{k-1})$*. A* plane tree *(also called an* ordered, rooted tree*)* t *is a subset of* $\mathcal{U}$ *such that* $\emptyset \in \mathrm{t}$ *and for any* $u = (u_1, \dots, u_k) \in \mathrm{t}$ *we have* $p(u) \in \mathrm{t}$ *and* $(u_1, \dots, u_{k-1}, i) \in \mathrm{t}$ *for each* $1 \leq i < u_k$*.*

In a plane tree t, if we add an edge between each $u \in \mathrm{t}\backslash\{\emptyset\}$ and its parent and we forget about the partial ordering that the vertex labels induce on the vertices, we obtain a tree. If we let $\emptyset$ be the root of the tree, we obtain a rooted, labeled tree $\tilde{\mathrm{t}}$. The parent of any vertex in t is preserved under this projection. For any $u \in \mathrm{t}$ we let its *height* $|u|$, *children* $C(u)$ and *degree* $d_{\mathrm{t}}(u)$ equal the height, children and degree of u in $\tilde{\mathrm{t}}$ respectively. There is also a natural map from rooted, labeled trees with vertices in $\mathbb{N}$ (or, in fact, any ordered set) to the space of ordered, rooted trees. Indeed, for any vertex v, we can order the elements of $C(v)$ by increasing label, which naturally gives rise to an ordered, labeled tree. If the vertex set of a rooted, labeled tree is unordered, we can, for any vertex v, arbitrarily order the elements of $C(v)$ to obtain an ordered, labeled tree.

The *depth-first order* on the vertices of an ordered, rooted tree is given by the lexicographic order. Let t be an ordered, rooted tree of size n and let $(v_0, \dots, v_{n-1})$ be its vertices in depth-first order. Then, the *height process* $(h_{\mathrm{t}}(i), 0 \leq i \leq n-1)$ is defined by

$$h_{\mathrm{t}}(i) = |v_i|.$$

The height process characterizes the tree t.

Another process that characterizes t is its *Łukasiewicz path* $(y_{\mathrm{t}}(i), 0 \leq i \leq n)$ that is defined by

$$y_{\mathrm{t}}(i) = \sum_{j=1}^{i} (d_{\mathrm{t}}(v_{j-1}) - 1).$$

The Łukasiewicz path satisfies

1. $y_{\mathrm{t}}(0) = 0$ and $y_{\mathrm{t}}(n) = -1$;

2. $y_{\mathrm{t}}(i) \geq 0$ for each $i < n$;

3. $y_{\mathrm{t}}(i) - y_{\mathrm{t}}(i-1) \geq -1$ for each $1 \leq i \leq n$.

Then, the following proposition from Le Gall's survey paper relates the height process to the Łukasiewicz path.

Proposition 1.1.2 (Proposition 1.2 [76]). *For any ordered, rooted tree* t, *for any* $i \in \{1, \dots, |\mathrm{t}|\}$, *it holds that*

$$h_{\mathrm{t}}(i) = \#\left\{ j \in \{0, 1, \dots, i-1\} : y_{\mathrm{t}}(j) = \inf_{j \le l \le i} y_{\mathrm{t}}(l) \right\}. \tag{1.1}$$

Similarly, we can encode a forest of ordered, rooted trees by concatenating the height processes of the trees in the forest. To obtain the Łukasiewicz path of the forest, we define its sequence of differences to be the concatenation of the sequences of differences of the Łukasiewicz paths of the trees.

1.1.2 Uniform trees with a given degree sequence, Bienaymé trees and simply generated trees

We now introduce the random trees for which we will study the tail-behaviour of their height in Chapter 2. For $\mathcal{S}$ a finite set, we write $S \in_u \mathcal{S}$ to denote that S is a uniformly random element of $\mathcal{S}$.

A *degree sequence* is a sequence of non-negative integers $\mathrm{d} = (d_1, \dots, d_n)$ with $\sum_{i \in [n]} d_i = n - 1$. For $p > 0$ and a degree sequence $\mathrm{d} = (d_1, \dots, d_n)$ write $|\mathrm{d}|_p = (\sum_{i=1}^{n} d_i^p)^{1/p}$ and let $(\sigma_{\mathrm{d}})^2 = \frac{1}{n} \sum_{i=1}^{n} d_i(d_i - 1)$ be its *variance*. We write $\mathcal{T}_{\mathrm{d}}$ for the subset of $\mathcal{T}(n)$ consisting of all trees t for which for each $i \in [n]$ it holds that $d_{\mathrm{t}}(i) = d_i$. For $\mathrm{t} \in \mathcal{T}(n)$, we say that $(d_{\mathrm{t}}(1), \dots, d_{\mathrm{t}}(n))$ is the degree sequence of t. We write T_{d} for a uniformly random element of $\mathcal{T}_{\mathrm{d}}$.

We also introduce *Bienaymé trees*, which constitute the most elementary and well-studied random graph model and the oldest stochastic model for studying population growth. Let $\mu = (\mu_k, k \ge 0)$ be a probability distribution on $\mathbb{N}$. Then, a Bienaymé tree with offspring distribution μ, denoted T_μ, is the random ordered rooted tree that is the family tree of a branching process with offspring distribution μ.[1] Formally, it is defined as follows. Sample independent random variables $d(u)$ distributed as μ, for every $u \in \mathcal{U}$. Then let T_μ be the unique ordered, rooted tree in which for each vertex u the degree of u in T_μ is equal to $d(u)$. (An example appears in Figure 1.2.) To avoid degenerate cases, we will from now on only consider μ with $\mu_0 + \mu_1 < 1$.

---

[1] Bienaymé trees are often referred to as Galton–Watson trees, but we adopt the change in terminology suggested in [10].

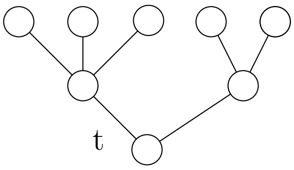


Fig. 1.2: In the tree t, there are 5 vertices with no children, two vertices with two children and one vertex with 3 children, so for $\mu = (\mu_k, k \geq 0)$ a probability distribution on $\mathbb{N}$, we have that $\mathbb{P}(\mathrm{T}_\mu = \mathrm{t}) = \mu_0^5 \mu_2^2 \mu_3$.

Note that the Łukasiewicz path of a Bienaymé tree is a random walk with steps distributed as $X - 1$, with X a random variable with law μ, stopped when it first hits -1, and this hitting time is the size of the tree. This last observation implies the following phase transition in the size of a Bienaymé tree, which is a standard result (see for example [19, Chapter I.5]). For $p > 0$, write $|\mu|_p = \left(\sum_{k=0}^{\infty} k^p \mu_k\right)^{1/p}$.

Proposition 1.1.3. *Consider a probability distribution μ on $\mathbb{N}$. If $|\mu|_1 \leq 1$ then $\mathbb{P}(|T_\mu| < \infty) = 1$. Otherwise, the probability generating function of μ has a unique fixed point $\rho \in [0, 1)$ and $\mathbb{P}(|T_\mu| < \infty) = \rho$. Finally, $\mathbb{E}[|T_\mu|]$ is finite if and only if $|\mu|_1 < 1$.*

We call $\mathbb{P}(|T_\mu| < \infty)$ the *extinction probability* of a μ-Bienaymé tree. Proposition 1.1.3 motivates the following definition.

Definition 1.1.4. *We say a Bienaymé tree is* subcritical, critical *or* supercritical *when $|\mu|_1$ is, respectively, less than 1, equal to 1, or greater than 1.*

For $n \in \mathbb{N}$ such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$, we write $\mathrm{T}_{\mu,n}$ to denote a Bienaymé tree with offspring distribution μ conditioned to have size n.

There is a duality relation between subcritical and supercritical Bienaymé trees; to be precise, it holds that any supercritical Bienaymé tree conditioned to be finite is a subcritical Bienaymé tree, which is the content of the following proposition. This is a standard result, see for example [19, Theorem I.12.3].

Proposition 1.1.5. *Let μ be a probability distribution on $\mathbb{N}$ with $|\mu|_1 > 1$ and $\mu_0 > 0$, and let $\rho \in (0, 1)$ be the extinction probability of a μ-Bienaymé tree. Let $\mu^* = (\rho^{k-1}\mu_k, k \geq 0)$. Then, μ^* is a probability distribution with $|\mu^*|_1 < 1$ and for any finite plane tree t it holds that*

$$\mathbb{P}\left(\mathrm{T}_\mu = \mathrm{t} \,\middle|\, |\mathrm{T}_\mu| < \infty\right) = \mathbb{P}\left(\mathrm{T}_{\mu^*} = \mathrm{t}\right)$$

and in particular, for any n such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$ it holds that

$$\mathrm{T}_{\mu,n} \overset{d}{=} \mathrm{T}_{\mu^*,n}.$$

We also introduce simply generated trees, which are a generalization of conditioned Bienaymé trees for which the branching behaviour is not determined by an offspring distribution but by an offspring weight sequence that is not necessarily normalizable.

Fix non-negative real weights $\mathrm{w} = (w_k, k \geq 0)$ with $w_0 > 0$. Given a plane tree t, we define the weight of t to be

$$\mathrm{w}(\mathrm{t}) = \prod_{v \in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)}.$$

Next, for positive integers n, let

$$Z_n = Z_n(\mathrm{w}) = \sum_{\{\text{plane trees } \mathrm{t}:|\mathrm{t}|=n\}} \mathrm{w}(\mathrm{t}).$$

If $Z_n > 0$, then we define a random tree $\mathrm{T}_{\mathrm{w},n}$ by setting

$$\mathbb{P}(\mathrm{T}_{\mathrm{w},n} = \mathrm{t}) = \frac{\mathrm{w}(\mathrm{t})}{Z_n}$$

for plane trees t with $|v(\mathrm{t})| = n$. The random tree $\mathrm{T}_{\mathrm{w},n}$ is called a *simply generated tree* of size n with weight sequence w. Note that if $\sum_{k\geq 0} w_k = 1$, then w is an offspring distribution, and $\mathrm{T}_{\mathrm{w},n}$ is indeed distributed as a Bienaymé tree with offspring distribution w, conditioned to have size n, so this notation agrees with (but generalizes) that above. From this we also see that conditioned Bienaymé trees are a subclass of simply generated trees; we will use this in Chapter 2. To avoid degenerate cases, we will from now on only consider $\mathrm{w} = (w_k, k \geq 0)$ with $\sum_{k>1} w_k > 0$.

A useful property of simply generated trees is that, given a weight sequence w and an $a, b > 0$, for $\mathrm{w}_{a,b} := (ab^k w_k, k \geq 0)$, we have that w is *equivalent* to $\mathrm{w}_{a,b}$, namely, it holds that for any n, $\mathrm{T}_{\mathrm{w},n} \overset{d}{=} \mathrm{T}_{\mathrm{w}_{a,b},n}$. Indeed, given a plane tree t with size n,

$$\mathrm{w}_{a,b}(\mathrm{t}) = \prod_{v \in v(\mathrm{t})} ab^{d_{\mathrm{t}}(v)} w_{d_{\mathrm{t}}(v)} = a^n b^{n-1} \prod_{v \in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)} \propto \mathrm{w}(\mathrm{t}).$$

In particular, for $\Phi(z) = \Phi_{\mathrm{w}}(z) = \sum_{k\geq 0} w_k z^k$ the generating function of the weight sequence

w, and for $\rho = \rho_{\mathrm{w}} = \sup(s \geq 0 : \Phi(s) < \infty)$ its radius of convergence, we see that for any $a > 0$ and $s < \rho$ it holds that $\mathrm{w}_{a,s}$ is summable and that $\mathrm{w}^{(s)} := \mathrm{w}_{\Phi(s)^{-1},s}$ is a probability distribution. Therefore, for any n such that $Z_n > 0$ and any $s < \rho$, it holds that $\mathrm{T}_{\mathrm{w},n}$ has the same law as a $\mathrm{w}^{(s)}$-Bienaymé tree conditioned to have size n. For $s > 0$ such that $\Phi(s) < \infty$, we let

$$\Psi(s) = \Psi_{\mathrm{w}}(s) = \frac{s\Phi'(s)}{\Phi(s)} = \frac{\sum_{k\geq 0} k w_k s^k}{\sum_{k\geq 0} w_k s^k}$$

so that $\Psi(s) = |\mathrm{w}^{(s)}|_1$. If $\Phi(\rho) = \infty$ then also define

$$\Psi(\rho) = \Psi_{\mathrm{w}}(\rho) = \lim_{s\uparrow\rho} \Psi(s) \in [0, \infty];$$

by Lemma 3.1(i) in [68], Ψ is strictly increasing on $[0, \rho)$ so this limit exists.

Then, note that in any tree of size n, the average number of offspring of a vertex equals $(n-1)/n \approx 1$ for large n, so it is natural to expect that if there is an s' such that $\Phi(s') = 1$, the empirical offspring distribution in $\mathrm{T}_{\mathrm{w},n}$ will approximate $\mathrm{w}^{(s')}$ as n tends to infinity. This is indeed true by the next theorem. Furthermore, it turns out that if $\Phi(\rho) < 1$, so that all probability distributions in the family $\{\mathrm{w}^{(s)} : \Phi(s) < \infty\}$ are subcritical, the empirical offspring distribution will approximate $\mathrm{w}^{(\rho)}$; in this case, the missing contribution to the mean offspring size will come from a negligible proportion of vertices with degree tending to ∞.

Let $\nu = \nu_{\mathrm{w}} = \Psi_{\mathrm{w}}(\rho)$ and define

$$\tau = \tau_{\mathrm{w}} := \begin{cases} \rho & \text{if } \nu_{\mathrm{w}} \leq 1 \\ \Psi^{-1}(1) & \text{if } \nu_{\mathrm{w}} > 1\,. \end{cases} \tag{1.2}$$

and let $\hat{\mu} = \mathrm{w}^{(\tau)}$. Let $N_k(\mathrm{w}, n)$ be distributed as the number of vertices with degree k in $\mathrm{T}_{\mathrm{w},n}$.

Theorem 1.1.6 ([68], Theorem 11.4; [10], Theorem 5.2)**.** *It holds that $\hat{\mu} = (\hat{\mu}_k, k \geq 0)$ is a probability distribution with mean $m = 1 \wedge \nu$ and variance $\hat{\sigma}^2 = \tau\Psi'(\tau)$. Moreover, for every $\epsilon > 0$ there exists $c(\epsilon) > 0$ such that for all n sufficiently large, for every integer $k \geq 0$,*

$$\mathbb{P}\left(\left|\frac{N_k(\mathrm{w}, n)}{n} - \hat{\mu}_k\right| > \epsilon\right) < e^{-cn}.$$

The following proposition, that appears as Proposition 2.1.9 in Chapter 2, states that if we assign arbitrary labels to the vertices of a simply generated tree and forget about the plane structure, we obtain a mixture of uniform trees with a prescribed degree sequence. In other words, a simply generated tree projected onto the space of labelled rooted trees is an average over uniform trees with a given degree sequence, and we understand the asymptotics of its degree distribution by Theorem 1.1.6. Therefore, we can transfer results on uniform trees with a given degree sequence to simply generated trees and conditioned Bienaymé trees, a fact that we will exploit in Chapter 2.

Proposition 1.1.7. *Fix non-negative weights* $(w_i, i \geq 0)$. *Let* T *be a random plane tree of size* n *with distribution given by*

$$\mathbb{P}(\mathrm{T} = \mathrm{t}) \propto \prod_{v \in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)},$$

for plane trees t *of size* n. *Conditionally given* T, *let* $\hat{\mathrm{T}}$ *be the random tree in* $\mathcal{T}(n)$ *obtained as follows: label the vertices of* T *by a uniformly random permutation of* $[n]$, *then forget about the plane structure. For* $i \in [n]$ *let* D_i *be the degree of vertex* i *in* $\hat{\mathrm{T}}$ *of vertex* i. *Then for any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$, *conditionally given that* $(D_1, \ldots, D_n) = (d_1, \ldots, d_n)$, *we have* $\hat{\mathrm{T}} \in_u \mathcal{T}_{\mathrm{d}}$.

1.1.3 The configuration model

The configuration model is a method to construct a multigraph with a given degree sequence that was introduced by Bollobás in [27]. We like to remind the reader that, in the setting of (multi-)graphs, the degree of vertex v is the number of edges containing the vertex v.

> Consider n vertices labelled by $[n]$ and a sequence $\mathrm{d} = (d_i)_{i \in [n]} \in \mathbb{N}^n$ such that $\sum_{i \in [n]} d_i$ is even. We will sample a multigraph such that the degree of vertex i is equal to d_i for every $i \in [n]$. The configuration model on n vertices having degree sequence d is constructed as follows. Equip vertex j with d_j half-edges. Sample a uniformly random pairing of all half-edges and let each pair of half-edges create an edge. Call the resulting multigraph $\mathrm{CM}(\mathrm{d})$.

Observe that $\mathrm{CM}(\mathrm{d})$ may contain self-loops or multiple edges. It can be shown that for any multigraph G with degree sequence d,

$$\mathbb{P}\left(\mathrm{CM}(\mathrm{d}) = G\right) = \frac{1}{(l_n - 1)!!} \frac{\prod_{i=1}^{n} d_i!}{2^{\mathrm{sl}(G)} \prod_{e \in E(G)} \mathrm{mult}(e)!},$$

where $l_n = \sum_{i=1}^n d_i$, $\mathrm{sl}(G)$ is the number of self-loops in G and $\mathrm{mult}(e)$ is the multiplicity of the edge $e \in E(G)$. Hence, if there exists a simple graph with degree sequence d, conditionally on the event that CM(d) is simple, we have that CM(d) is a uniformly random simple graph with degree sequence d.

Now, consider a progression of sequences of non-negative integers $(d_{1,n}, \dots, d_{n,n})_{n\geq 1}$ such that $\sum_{k=1}^n d_{k,n}$ is even for each n, $\max_{k\leq n} d_{k,n} = o(\sqrt{n})$ and such that there exist positive numbers a_1, a_2 for which $\frac{1}{n}\sum_{k=1}^n d_{k,n} \to a_1$ and $\frac{1}{n}\sum_{k=1}^n d_{k,n}^2 \to a_2$. Then, by [101, Theorem 7.12], the probability that $\mathrm{CM}(d_{1,n}, \dots, d_{n,n})$ is simple is asymptotically equal to e^{-a-a^2} with $a = \frac{a_2-a_1}{2a_1}$, and in particular, it is bounded away from 0. For a detailed discussion on the configuration model and many standard results, see [101, Chapter 7].

There is an obvious directed version of the configuration model that is used to construct a directed multigraph with a given degree sequence. The *directed configuration model* was introduced by Cooper and Frieze [38].

> Consider n vertices labelled by $[n]$ and a sequence $\mathrm{d} = (d_i^-, d_i^+)_{i\in[n]} \in (\mathbb{N}\times\mathbb{N})^n$ such that $\sum_{i\in[n]} d_i^- = \sum_{i\in[n]} d_i^+$. We will sample a directed multigraph such that for every $i \in [n]$ the in-degree of vertex i is equal to d_i^- and the out-degree of vertex i is equal to d_i^+. The directed configuration model on n vertices having degree sequence d is constructed as follows. For each $i \in [n]$ equip vertex i with d_i^- in-half-edges and d_i^+ out-half-edges. Sample a uniformly random matching of the in-half-edges with the out-half-edges and let any pair of an out-half-edge of a vertex v and an in-half-edge of a vertex w create a directed edge e with $r_1(e) = v$ and $r_2(e) = w$. Call the resulting multigraph $\vec{\mathrm{CM}}(\mathrm{d})$.

Cooper and Frieze [38, Sec. 2.1] proved that, just like in the undirected setting, if we condition on the resulting multigraph being simple, we obtain a uniformly chosen random digraph with the prescribed degree sequence. Moreover, in the proof of Proposition 3.4.16, we show the counterpart of [101, Theorem 7.12] in the directed setting. To be precise, let $(\mathbf{d}_{1,n}, \dots, \mathbf{d}_{n,n})_{n\geq 1}$ be a progression of sequences with $\mathbf{d}_{k,n} = (d_{k,n}^-, d_{k,n}^+) \in \mathbb{N}\times\mathbb{N}$ for each k, n. Assume that for each n it holds that, firstly, $\sum_{k=1}^n d_{k,n}^- = \sum_{k=1}^n d_{k,n}^+$, secondly, $\max_{k\leq n} d_{k,n}^- \vee d_{k,n}^+ = o(\sqrt{n})$, and, finally, for all non-negative integers i, j such that $1 \leq i + j \leq 2$ there exist positive $a_{i,j}$ such that

$$\frac{1}{n}\sum_{k=1}^n (d_{k,n}^-)^i (d_{k,n}^-)^j \to a_{i,j}.$$

(Observe that the condition that $\sum_{k=1}^{n} d_{k,n}^{-} = \sum_{k=1}^{n} d_{k,n}^{+}$ implies that $a_{1,0} = a_{0,1}$.) Then, we show that the probability that $\vec{\mathrm{CM}}(\mathbf{d}_{1,n}, \dots, \mathbf{d}_{n,n})$ is simple is asymptotically equal to

$$\exp\left(-\frac{a_{1,1}}{a_{1,0}} - \frac{(a_{2,0} - a_{1,0})(a_{0,2} - a_{0,1})}{a_{1,0}^2}\right)$$

and in particular, it is bounded away from 0.

1.2 Limits of (directed) graphs

In the past few decades, there has been much interest in showing convergence of random trees and graphs themselves instead of their local structure or real-valued statistics such as their width or diameter. To do so, we view graphs as metric spaces and show convergence in an appropriate topology. We obtain such results for the directed configuration model in Chapter 3 and for the undirected configuration model in two of the works described in Chapter 4. We will first introduce the description of trees and graphs as measured metric spaces.

1.2.1 Trees and graphs as measured metric spaces

Let T be an ordered rooted finite tree, say $|T| = n$. Let $v_0, \dots, v_{n-1}$ denote the vertices of the tree visited in depth-first order, so that v_0 is the root of the tree. We can view T as a metric space by forgetting the ordering and regarding all edges as line segments of length 1 that connect the vertices at their endpoints. The distance d_T between points a_1 and a_2 on line segments l_1 and l_2 respectively is then defined to be the length of the unique non-self-intersecting path between a_1 and a_2 that traverses the line segments of the tree. Let $\mathcal{B}(T)$ be the Borel σ-algebra on T and define a measure μ on $\mathcal{B}(T)$ by assigning all vertices mass $1/n$. Denote the triplet (T, d_T, μ) by T.

We can also view a finite connected (multi-)graph G with $|G| = n$ as a metric space by regarding all edges as line segments of length 1 that connect the vertices at their endpoints.. The distance d_G between points a_1 and a_2 on line segments l_1 and l_2 respectively is then defined as the length of the shortest path between a_1 and a_2 that traverses the line segments of the graph. Let $\mathcal{B}(G)$ denote the Borel σ-algebra on G, and define a measure μ on $\mathcal{B}(G)$ by assigning all vertices mass $1/n$. Denote the triplet (G, d_G, μ) by G.

1.2.1.1 $\mathbb{R}$-trees and graphs

We will now introduce the notion of an $\mathbb{R}$-tree, which is a generalization of finite trees considered as measured metric spaces.

A compact metric space $(\mathcal{T}, d)$ is an $\mathbb{R}$-*tree* (also referred to as a *real tree*) if the following two properties hold.

- $(\mathcal{T}, d)$ is *acyclic*; i.e. there is no continuous injective function $f : \mathbb{S}_1 \to \mathcal{T}$.
- $(\mathcal{T}, d)$ is *geodesic*; i.e. for all $x, y \in \mathcal{T}$ there is a geodesic between x and y. A *geodesic* is defined as an isometric embedding $f : [0, d(x, y)] \to \mathcal{T}$ such that $f(0) = x$ and $f(d(x, y)) = y$. Note that this implies that $\mathcal{T}$ is connected. Define $[\![x, y]\!] = f([0, d(x, y)])$.

An element $x \in \mathcal{T}$ is called a *vertex*. The *degree* $\deg_{\mathcal{T}}(x)$ of x is the number of connected components of $\mathcal{T}\backslash\{x\}$. A *leaf* is a vertex of degree 1, and a *branchpoint* is a point with degree strictly greater than 2. Let $\mathcal{L}(\mathcal{T})$ denote the set of leaves of $\mathcal{T}$. A *rooted* $\mathbb{R}$-*tree* is an $\mathbb{R}$-tree $(\mathcal{T}, d)$ with a distinguished vertex ρ called the *root*. The *height* of a vertex v is $d(\rho, v)$. We can also endow an $\mathbb{R}$-tree with a *genealogical order*: say $x \preceq y$ if $x \in [\![\rho, y]\!]$.

We can then endow $(\mathcal{T}, d)$ with a Borel probability measure μ. If μ satisfies the following properties, we call the triplet $(\mathcal{T}, d, \mu)$ a *continuum tree*.

1. μ is non-atomic (i.e. there are no vertices v with $\mu(\{v\}) > 0$);
2. $\mu(\mathcal{L}(\mathcal{T})) = 1$;
3. For a vertex $v \in \mathcal{T}$ of degree $k \geq 2$, let $\mathcal{T}_1, \ldots, \mathcal{T}_k$ be the connected components of $\mathcal{T}\backslash\{v\}$. Then, $\mu(\mathcal{T}_i) > 0$ for all i.

Note that conditions 1 and 2 together imply that continuum trees have uncountably many leaves and, in particular, that finite trees with the metric structure defined in the last subsection cannot be equipped with a measure to be viewed as continuum trees.

The conditions on μ were introduced by Aldous [14] to ensure that, for the k-dimensional distribution of a tree T defined as the law of the random subtree spanned by k uniform samples from μ, the finite-dimensional distributions of a continuum tree T characterize T.

A *continuous excursion* is a continuous function $h : [0, \sigma] \to \mathbb{R}_+$ such that $h(0) = h(\sigma) = 0$. These can be used to encode $\mathbb{R}$-trees, and can be viewed as the continuous equivalent of the height process of a tree. Define a pseudo-metric on $[0, \sigma]$ via

$$d_h(x, y) = h(x) + h(y) - 2 \inf_{x \wedge y \leq x \leq x \vee y} h(x).$$

To turn this into a metric, define an equivalence relation $x \sim y$ iff $d_h(x, y) = 0$. Define $\mathcal{T}_h$ to be the quotient $[0, \sigma]/\sim$. Note that d_h corresponds to the metric on finite trees as metric spaces (i.e. the distance between x and y on the tree is given by the length of $[\![x, \rho]\!]$ plus the length of $[\![y, \rho]\!]$, minus twice the length of the path from ρ to the point z where $[\![x, \rho]\!]$ and $[\![y, \rho]\!]$ meet). This turns local minima of the function into branchpoints. Then, Duquesne and Le Gall showed in [51] that for any continuous excursion h, $(\mathcal{T}_h, d_h)$ is an $\mathbb{R}$-tree.

There is also a natural probability measure on $(\mathcal{T}_h, d_h)$, denoted by μ_h, induced by scaling the pushforward of the Lebesgue measure on $[0, \sigma]$ by $1/\sigma$.

Just as in the discrete case, we would like to derive a continuous excursion that serves as the height process of the tree from a continuous time analogue of a Łukasiewicz path. Ideally, for a positive excursion $x : [0, \sigma] \to \mathbb{R}_+$, we would want to define its height process by the continuous time analogue of (1.1), namely by

$$h(t) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \int_0^t 1_{\{x(s) < \inf\{x(r) + \epsilon : s \leq r \leq t\}\}} ds. \tag{1.3}$$

However, it is not clear what the conditions on $x : [0, \sigma] \to \mathbb{R}_+$ are in order for this limit to exist and for h to be a continuous excursion, so that x encodes a $\mathbb{R}$-tree. Because we are interested in $\mathbb{R}$-trees that are the metric space limits of a sequence of finite trees, it is natural to at least require that $x(0) = x(\sigma) = 0$ and that x has no negative jumps, corresponding to the property that discrete Łukasiewicz paths are downwards skip-free. We will further discuss the results in [77] and [50] on the existence of the height process corresponding to random Łukasiewicz paths in Subsection 1.2.2. If the limit in (1.3) exists for all t and h is a continuous excursion, we call x and h, respectively, the Łukasiewicz path and height process of $\mathcal{T}_h$.

We can encode a forest of $\mathbb{R}$-trees by concatenating the Łukasiewicz paths and height processes of the trees in the forest.

An $\mathbb{R}$-*graph* is a compact geodesic metric space $(\mathcal{G}, d)$ that is locally an $\mathbb{R}$-tree. To be precise, $(\mathcal{G}, d)$ is an $\mathbb{R}$-graph if for every $x \in \mathcal{G}$, there exists $\epsilon > 0$ such that $(B_\epsilon(x), d|_{B_\epsilon(x)})$ is an $\mathbb{R}$-tree. For $x \in \mathcal{G}$, its degree $d_{\mathcal{G}}(x)$ is defined to be the degree of x in $B_\epsilon(x)$ for any ϵ small enough that $B_\epsilon(x)$ is an $\mathbb{R}$-tree. This is well-defined, since the degree of a vertex v in an $\mathbb{R}$-tree can be observed on any open set containing v.

1.2.1.2 A topology on measured metric spaces

Given a metric space (X, d), write $[X, d]$ for the isometry class of (X, d). Write $\hat{\mathcal{M}}$ for the space of isometry classes of compact metric spaces. We will define the Gromov–Hausdorff distance on this space.

Let $\mathrm{X} = (X, d)$ and $\mathrm{X}' = (X', d')$ be metric spaces. A *correspondence* C between X and X' is a measurable subset of $\mathrm{X} \times \mathrm{X}'$ such that for every $x \in \mathrm{X}$ there exists an $x' \in \mathrm{X}'$ such that $(x, x') \in C$ and vice versa. Write $C(\mathrm{X}, \mathrm{X}')$ for the set of correspondences between X and X'.

For C a subset of $\mathrm{X} \times \mathrm{X}'$, the *distortion* $\mathrm{dis}(C)$ is defined by

$$\mathrm{dis}(C) = \sup\{|d(x, y) - d(x', y')| : (x, x') \in C, (y, y') \in C\}.$$

Then, the Gromov–Hausdorff distance $d_{\mathrm{GH}}(\mathrm{X}, \mathrm{X}')$ between $[X, d]$ and $[X', d']$ is defined as

$$d_{\mathrm{GH}}(\mathrm{X}, \mathrm{X}') = \frac{1}{2} \inf\{\mathrm{dis}(C) : C \in C(\mathrm{X}, \mathrm{X}')\}.$$

A *compact measured metric space* is a triple $\mathsf{X} = (X, d, \mu)$ where $\mathrm{X} = (X, d)$ is a compact metric space and μ is a finite Borel-measure on X. Given measured metric spaces $\mathsf{X} = (X, d, \mu)$ and $\mathsf{X}' = (X', d', \mu')$, and a measurable function $\psi : \mathsf{X} \to \mathsf{X}'$, write $\psi_*\mu$ for the pushforward of the measure μ to $\mathrm{X}' := (X', d')$. Call X and X' *isometry-equivalent* if there exists an isometry $\psi : \mathsf{X} \to \mathsf{X}'$ such that $\psi_*\mu = \mu'$. Denote the isometry-equivalence class of (X, d, μ) by $[X, d, \mu]$, and define $\mathcal{M}$ for the set of measured isometry-equivalence classes of compact measured metric spaces. We will now define a metric on $\mathcal{M}$.

If $\mathrm{X} = (X, d)$ and $\mathrm{X}' = (X', d')$ are two metric spaces, let $M(\mathrm{X}, \mathrm{X}')$ be the set of finite nonnegative Borel measures on $\mathrm{X} \times \mathrm{X}'$. For ν a finite signed measure on $\mathrm{X} \times \mathrm{X}'$ and $\nu = \nu^+ - \nu^-$ its Hahn–Jordan decomposition, its total variation is given by $||\nu|| := \nu^+(X \times X') + \nu^-(X \times X')$. Let p and p' denote the canonical projections from $\mathrm{X} \times \mathrm{X}'$ onto X and X' respectively. Then, the *discrepancy* $D(\pi; \mu, \mu')$ of $\pi \in M(\mathrm{X}, \mathrm{X}')$ with respect to μ and μ' is defined as

$$D(\pi; \mu, \mu') = ||\mu - p_*\pi|| + ||\mu' - p'_*\pi||.$$

Then, the Gromov–Hausdorff–Prokhorov distance d_{GHP} between X and X' is defined as

$$d_{\mathrm{GHP}}(\mathsf{X}, \mathsf{X}') = \inf \left\{ \frac{1}{2} \operatorname{dis}(C) \vee D(\pi; \mu, \mu') \vee \pi(C^c) \right\},$$

with the infimum taken over correspondences $C \in C(\mathrm{X}, \mathrm{X}')$ and measures $\pi \in M(\mathrm{X}, \mathrm{X}')$. Then, d_{GHP} is a metric and $(\mathcal{M}, d_{\mathrm{GHP}})$ is a Polish space by Theorem 2.5 of [1]. (The presentation of the metric as considered in [1] is different, but it induces the same topology, as discussed in [9].)

Note that we have the following upper bound for the GHP-distance between $\mathbb{R}$-trees encoded by continuous excursions. For g and h continuous excursions with excursion length σ_g and σ_h respectively, we have that

$$\begin{aligned} & d_{\mathrm{GHP}}((\mathcal{T}_g, d_g, \mu_g), (\mathcal{T}_h, d_h, \mu_h)) \\ & \leq 2 \max \left\{ \sup_{0 \leq x \leq \sigma_g \wedge \sigma_h} |g(x) - h(x)|, \sup_{\sigma_g \wedge \sigma_h < x \leq \sigma_g} g(x) + \sup_{\sigma_g \wedge \sigma_h < x \leq \sigma_h} h(x), \frac{1}{2} |\sigma_g - \sigma_h| \right\}. \end{aligned} \tag{1.4}$$

Namely, this is an upper bound for the value of $\frac{1}{2} \operatorname{dis}(C) \vee D(\pi; \mu_h, \mu_g) \vee \pi(C^c)$ for the combination of the following correspondence C' and measure π', where without loss of generality we assume that $\sigma_h \geq \sigma_g$. Let p_h denote the projection of $[0, \sigma_h]$ onto $\mathcal{T}_h$ and define p_g similarly. Let ρ_g be the root of $\mathcal{T}_g$. Then set

$$C' = \{(v, w) \in \mathcal{T}_h \times \mathcal{T}_g \text{ s.t. } p_h^{-1}(v) \cap p_g^{-1}(w) \neq \emptyset\} \cup \{(v, \rho_g) : p_h^{-1}(v) \subseteq (\sigma_g, \sigma_h]\},$$

and for $\hat{\lambda}$ being the pushforward of the Lebesgue measure on $[0, \sigma_g]$ under mapping $[0, \sigma_g] \to [0, \sigma_h] \times [0, \sigma_g] : x \mapsto (x, x)$, let π be the pushforward of $\hat{\lambda}$ under $p_h \times p_g$.

Then, (1.4) suggests that if we have suitable convergence of the height processes encoding a sequence of trees, we get convergence of the sequence of trees in the topology induced by d_{GHP}. We will use this in Chapter 3 and in two of the works described in Chapter 4. Convergence of the Łukasiewicz paths does not imply convergence of the trees in the d_{GHP} distance; informally, the embedding of the tree in the Łukasiewicz path distorts distances.

1.2.2 Examples of scaling limits of random trees and graphs

One of the pioneering universality results in the field of random graphs is by Aldous, who constructed the Brownian Continuum Random Tree (CRT) [13, 14]. This is the random tree $\mathcal{T}_{2e}$, for e a Brownian excursion. It almost surely has Hausdorff dimension 2 and all its branchpoints have degree 3. Apart from via its height process e, the CRT can also be defined via a stick-breaking construction, as shown in [13]. In [14], Aldous showed that for μ a probability distribution on $\mathbb{N}$ with $|\mu|_2 < \infty$, it holds that $n^{-1/2}\mathrm{T}_{\mu,n}$ converges in distribution in the Gromov–Hausdorff topology to a constant multiple of the Brownian CRT (over all n such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$) by showing that its rescaled height process converges in distribution to a Brownian excursion. Then, in [80], Marckert and Mokkadem show that the rescaled Łukasiewicz path and height process in fact converge to the same Brownian excursion under the additional assumption that $\sum_{k=1}^{\infty} \mu_k e^{ak}$ is finite for some $a > 0$. Since then, the Brownian CRT has been shown to be the metric space scaling limit of a wide variety of random graph models, such as random uniform unordered trees [65], the uniform spanning tree of the high-dimensional torus [18], random dissections [40], random graphs from subcritical classes [89] and others.

In [77], Le Gall and Le Jan consider a wide class of spectrally positive Lévy processes L that do not drift to $+\infty$, for which they construct a continuous version of the height process that encodes a sequence of random continuum trees with L as their Łukasiewicz path. In [50], Duquesne and Le Gall show that for such a Lévy process L, the height process of a forest of i.i.d. Bienaymé trees, for which the Łukasiewicz path is in the domain of attraction of L and the extinction times of the trees satisfy an additional regularity condition, converges under rescaling to the height processes as constructed in [77].

In [49], Duquesne proves a conditioned version of the result in [50], which is a generalization of Aldous' result [14]. To be precise, he shows that if the rescaled Łukasiewicz path of an unconditioned forest of μ-Bienaymé trees converges to an α-stable process with $\alpha \in (1, 2]$, it holds that $n^{-1+1/\alpha}\mathrm{T}_{\mu,n}$ converges in distribution in the Gromov–Hausdorff–Prokhorov topology to a random continuum tree that has a constant multiple of a normalized excursion of the α-stable process as its Łukasiewicz path. Again, this convergence is shown by showing convergence of the height process under rescaling. Such trees are called *α-stable trees*. (Note that the 2-stable tree is just the Brownian CRT up to an unimportant constant factor.) An α-stable tree with $\alpha \in (1, 2)$ is structurally very different from the Brownian CRT: almost surely all branchpoints have infinite degree and the tree has Hausdorff dimension $\alpha/(\alpha-1)$ [64]. Just like the Brownian CRT, it can be constructed via a stick-breaking procedure [58].

The convergence results on rescaled Bienaymé trees in [14, 77, 49, 50] do not use the formalism of $\mathbb{R}$-trees. The restatement of there results in the framework of $\mathbb{R}$-trees is due to Le Gall in [56] and has since then become standard.

There are also several works that show convergence of random graph models to random non-acyclic $\mathbb{R}$-graphs, the first being [3] by Addario-Berry, Broutin and Goldschmidt. They study the *Erdős–Rényi random graph*, $G(n, p)$, which is the random graph model with vertex set $[n]$ in which each edge is present with probability p. They consider the parameter regime $p = p_{n,\lambda} = 1/n + \lambda n^{-4/3}$ for $\lambda \in \mathbb{R}$ in which the sequence of the ordered component sizes, rescaled by $n^{-2/3}$, has a non-degenerate limit, as shown in [15]. This regime is called the *critical window*, and the case $\lambda = 0$ is called *criticality*. The authors of [9] show that in the critical window, the large components converge under rescaling in the Gromov–Hausdorff topology to a sequence of random $\mathbb{R}$-graphs that they call the *Brownian graph*. Each component of the Brownian graph consist of an $\mathbb{R}$-tree that is absolutely continuous to a randomly rescaled Brownian CRT with a finite number of random identifications to create the cyclic structure. The components are independent conditionally on their total mass; almost surely all branchpoints have degree 3 and the $\mathbb{R}$-graph has Hausdorff dimension 2. In [22], Bhamidi, Sen and Wang show that the limit under rescaling of the components of the critical inhomogeneous random graph under suitable conditions have the same law as rescaled components of the Brownian graph, showing that the Brownian graph is also a universal scaling limit.

We will review the results on metric space scaling limits of the configuration model in various set-ups in Section 1.4.

1.2.3 Metric directed multigraphs

There is no natural way to view a directed strongly connected (multi-)graph $\vec{G}$ with $|\vec{G}| = n$ as a metric space, because for two vertices v and w, the distance from v to w might differ from the distance from w to v. Therefore, we need to consider them as *metric directed multigraphs*, which is a class that was introduced by Goldschmidt and Stephenson in [59]. A metric directed multigraph (MDM) is a tuple $M = (V, E, r, l)$ where (V, E, r) is a directed multigraph and $l : E \to [0, \infty)$. An *isomorphism* between two MDMs $M = (V, E, r, l)$ and $M' = (V', E', r', l')$ is a pair of functions (i_V, i_E) where $i_V : V \to V'$ and $i_E : E \to E'$ are bijections satisfying the relation

$$r'(i_E(e)) = (i_V(r_1(e)), i_V(r_2(e)))$$

for all $e \in E$. We say two MDMs are *isomorphic* if there exists an isomorphism between them. In other words, isomorphic MDMs have the same graph structures for their underlying directed multigraphs up to a relabelling of the edges and vertices. Write $\mathrm{Iso}(M, M')$ for the set of all isomorphisms between M and M'. Let $\mathfrak{L}$ be the MDM that consists of a single vertex with a self-loop of length 0.

We see that $\vec{G}$ can be viewed as an MDM by assigning length 1 to each edge, but it is more natural to view any maximal path of which all internal vertices have in- and out-degree 1 as one long edge.

We will now make this idea formal. Consider an MDM M and a vertex $w \in M$ with in-degree 1 and out-degree 1 which is not a self-loop. Let u and v be the unique in-neighbour and out-neighbour of w respectively. The MDM obtained by *smoothing* w is obtained by deleting the edges e_1 and e_2 such that $r(e_1) = (u, w)$ and $r(e_2) = (w, v)$, then adding an edge e such that $r(e) = (u, v)$ and assigning it length $l(e) = l(e_1) + l(e_2)$. This is illustrated in Figure 1.3.

(a) The graph before smoothing w

(b) The graph after smoothing w

Fig. 1.3: Smoothing a vertex w

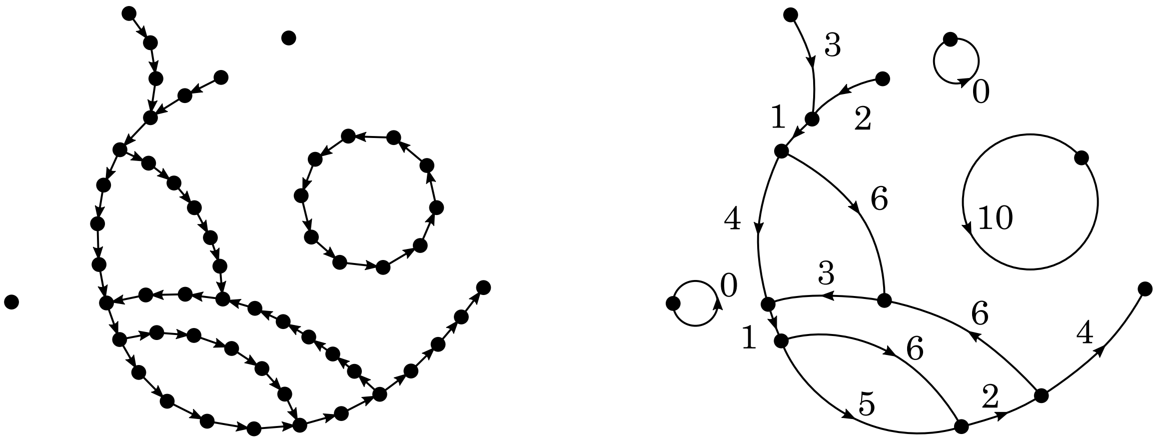
Then the kernel of a digraph $\vec{G}$ is obtained by the following procedure:

1. Assign length 1 to each edge.
2. Iteratively smooth vertices with in-degree 1 and out-degree 1 that are not self-loops until there are none remaining or we have obtained a MDM consisting of one vertex with a single self-loop.
3. Replace all singletons by $\mathfrak{L}$.

An example is shown in Figure 1.4.

1.2.3.1 A topology on metric directed multigraphs

We now define a distance $d_{\vec{G}}$ between two MDMs M and M' that was introduced in [59]. Any isomorphism between M and M' gives a correspondence between the edges of M and the edges of M'. We can then take an ℓ_∞ distance between the lengths of the edges and finally take the isomorphism which minimizes this distance. If M and M' are not isomorphic, we set the



(a) $\vec{G}$ (b) Kernel of $\vec{G}$

Fig. 1.4: An example of a digraph $\vec{G}$ and its kernel. The numbers indicate the edge lengths.

distance to be infinite. Formally,

$$d_{\vec{G}}(M, M') = \begin{cases} \inf_{(i_V, i_E) \in \mathrm{Iso}(M,M')} \sup_{e \in E} |l(e) - l'(i_E(e))| & \text{if } M \text{ and } M' \text{ are isomorphic,} \\ \infty & \text{otherwise.} \end{cases}$$

1.2.4 An example of a scaling limit of random strongly connected directed graph

There are only two random directed graph models for which the strongly connected components have been shown to converge in distribution under rescaling as MDMs: the directed Erdős–Rényi model and the directed configuration model as considered by us in Chapter 3. The *directed Erdős–Rényi model* on n vertices with parameter p, denoted by $\vec{G}(n, p)$, is a random digraph with vertex set $[n]$ in which each of the $n(n-1)$ possible directed edges is included with probability p independently. The cases $p = p_{n,\lambda} = 1/n + \lambda n^{-4/3}$ for $\lambda \in \mathbb{R}$ are referred to as *the critical window*, and the case $p = 1/n$ is called *criticality*. In [59], the authors show that in this regime, the strongly connected components converge under rescaling to a random sequence of MDMs that are all 3-regular or loops. Notably, the total length in the limit object is finite. As we will discuss in Section 1.3, the limit object for $\lambda = 0$ is part of the family of limit objects of the strongly connected components of the configuration model under rescaling that we obtain in Chapter 3.

1.3 Results in this thesis

1.3.1 The height of random trees

In this section, we introduce the theorems that we prove in Chapter 2.

We show that, for any degree sequence d of length n, a uniform tree with degree sequence d will have height $O(n^{1/2})$ with sub-Gaussian tails, with parameters not depending on the degree sequence other than through the number of vertices with degree 1. This is the content of the following theorem, that appears as Theorem 2.1.7 in Chapter 2.

Theorem 1.3.1. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*, and write* $\delta = (n - n_1(\mathrm{d}))/n$*. Then for all* $x > 0$*,*

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}) > xn^{1/2}\right) < 5\exp(-\delta x^2/2^{13}).$$

The dependence on $n_1(\mathrm{d})$ is to be expected because vertices with degree 1 make the tree tall. Indeed, given a uniform tree with degree sequence d, say T, adding a vertex with degree 1 to d corresponds (under a suitable coupling) to splitting a uniform edge of T into two edges (or including an additional edge below the root). So, including vertices of degree 1 into a degree sequence makes the paths between branchpoints in a uniform tree with the degree sequence longer and thus makes the tree taller.

Our second theorem on trees with a given degree sequence, which appears as Theorem 2.1.8 in Chapter 2, implies that if the variance of a degree sequence d is large, the height of a uniform tree with degree sequence d is in fact smaller than implied by Theorem 2.1.7.

Theorem 1.3.2. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*, and define* $(\sigma')^2 = \frac{n}{n-n_1(\mathrm{d})}\sigma_{\mathrm{d}}^2$*. Then for all* $x \geq 2^{15}$*,*

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}) > xn^{1/2}\frac{\log(\sigma'+1)}{\sigma_{\mathrm{d}}}\right) \leq 4\exp\left(-x\log(\sigma'+1)/2^{14}\right).$$

The bounds in this theorem are not always tight. For example, if the degrees in a degree sequence d_n are i.i.d. samples from a heavy-tailed distribution with polynomial tails, say $\mu([k,\infty)) \sim k^{-\alpha}$ for large k, conditioned to be a degree sequence, the variance is $\Theta(n^{2/\alpha-1})$ in probability, so the theorem shows that a uniform tree with degree sequence d_n has height $O(n^{1-1/\alpha}\log(n))$. However, the annealed tree is just a conditioned μ-Bienaymé tree projected onto the space of rooted, labeled trees, which has height $\Theta(n^{1-1/\alpha})$ in probability, so the asymptotic upper bound is off by a factor of $\log(n)$.

On the other hand, consider a degree sequence of length n of the form $(n^{1/\alpha}, 2, \ldots, 2, 0, \ldots, 0)$. Then the variance is again $\Theta(n^{2/\alpha-1})$, but it is easy to show that the height of the tree is $\Omega(n^{1-1/\alpha}\log(n))$ in probability, implying that the order of the height in the theorem is correct. So although the order in Theorem 1.3.2 is not tight in all cases, it is the smallest order that one can hope for in a *universal* bound that works for all degree sequences.

We also prove some results for simply generated trees and Bienaymé trees. By Proposition 1.1.7, which states that simply generated trees projected onto the space of rooted, labeled trees are an average over uniform trees with a given degree sequence, and by Theorem 1.1.6, which identifies the asymptotic degree distribution of a simply generated tree, we see that these results are consequences of our theorems above.

The following theorem resolves Conjecture 21.5 and Problem 21.7 from [68]. We refer the reader to the statement of Theorem 1.1.6 for the definition of $\hat{\sigma}$.

Theorem 1.3.3. *Let* $\mathrm{w} = (w_k, k \geq 0)$ *be a weight sequence with* $w_0 > 0$ *and with* $w_k > 0$ *for some* $k \geq 2$. *If either* $\hat{\sigma}^2 = \infty$ *or* $\nu < 1$ *then* $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) \to 0$, *where the convergence is in both probability and expectation, as* $n \to \infty$ *along integers* n *such that* $Z_n(\mathrm{w}) > 0$.

By Theorem 1.3.2, to prove Theorem 1.3.3, it is sufficient to show that the empirical variance of the degree sequence of $\mathrm{T}_{\mathrm{w},n}$ goes to infinity in probability as $n \to \infty$ and that the number of vertices with degree 1 does not get too big.

We also prove the following, more quantitative theorem. The definition of τ can be found before the statement of Theorem 1.1.6.

Theorem 1.3.4. *Fix* $\epsilon \in [0, 1)$. *Then, there exist constants* $c_1 = c_1(\epsilon)$ *and* $c_2 = c_2(\epsilon)$ *such that the following holds. Let* $\mathrm{w} = (w_k, k \geq 0)$ *be a weight sequence with* $w_0 > 0$ *and* $w_k > 0$ *for some* $k \geq 2$ *and* $\frac{w_1\tau}{\Phi(\tau)} < 1 - \epsilon$. *For any* $t > 1$ *and any* n *large enough with* $Z_n(\mathrm{w}) > 0$,

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > tn^{1/2}\right) \leq c_1 \exp(-c_2 t^2).$$

This theorem is an averaged version of Theorem 1.3.1 because, by Theorem 1.1.6, the proportion of vertices with degree 1 in $\mathrm{T}_{\mathrm{w},n}$ tends to $\frac{w_1\tau}{\Phi(\tau)}$ in probability.

These theorems have the following consequences for Bienaymé trees.

Theorem 1.3.5. *Fix a probability distribution* μ *supported on* $\mathbb{N}$ *with* $|\mu|_1 \leq 1$ *and* $|\mu|_2 = \infty$*. Then,* $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ *as* $n \to \infty$ *along values* n *such that* $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$*, both in probability and in expectation.*

Theorem 1.3.5 answers an open question from [8].

Theorem 1.3.6. *Fix a probability distribution* $\mu = (\mu_k, k \geq 0)$ *supported on* $\mathbb{N}$ *with* $|\mu|_1 < 1$ *and with* $\sum_{k\geq 0} e^{tk}\mu_k = \infty$ *for all* $t > 0$*. Then,* $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ *as* $n \to \infty$ *over all* n *such that* $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$*, both in probability and in expectation.*

It might seem restrictive that we only consider critical and subcritical offspring distributions in Theorems 1.3.5 and 1.3.6, but by Proposition 1.1.5, a supercritical Bienaymé tree conditioned to have size n has the same law as a particular subcritical Bienaymé tree conditioned to have size n.

Theorem 1.3.7. *Fix* $\epsilon \in [0, 1)$*. Then, there exist constants* $c = c(\epsilon)$ *and* $C = C(\epsilon)$ *such that the following holds. For any probability distribution* $\mu = (\mu_k, k \geq 0)$ *on* $\mathbb{N}$ *with* $\mu_0 + \mu_1 < 1 - \epsilon$ *and* $\mu_0/(\mu_0 + \mu_1) > \epsilon$*, for all* n *sufficiently large and for which* $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$*, for all* $x > 0$*,*

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mu,n}) > xn^{1/2}\right) \leq C\exp(-cx^2).$$

We need the conditions $\mu_0 + \mu_1 < 1 - \epsilon$ and $\mu_0/(\mu_0 + \mu_1) > \epsilon$ for the following reasons. If $\mu_0 + \mu_1$ is large, then branchpoints appear with low probability, so the tree is likely to have long paths between branchpoints and is therefore likely to be tall. To see why $\mathrm{T}_{\mu,n}$ is tall if $\mu_0/(\mu_0 + \mu_1)$ is small, note that every additional branchpoint creates at least 1 additional leaf. If leaves are unlikely compared to vertices of degree 1, then branchpoints are unlikely compared to vertices of degree 1, so again, the tree is likely to have many degree 1 vertices and is therefore tall. This reasoning is made formal in Chapter 2.

We want to point out that any non-trivial universal tail bound for the height of Bienaymé trees (and therefore for simply generated trees) must be asymptotic and one cannot hope to obtain a non-asymptotic tail bound analogous to Theorems 1.3.1 and 1.3.2. Indeed, for any m and any probability distribution $\mu = (\mu_k, k \geq 0)$ such that $\mu_k > 0$ if and only if $k \in \{0, 1, m\}$, we see that $\mathrm{T}_{\mu,m}$ cannot contain vertices of degree m and will therefore be a path of length m with height $m - 1$, so the asymptotic height of $O(n^{1/2})$ in probability can definitely not be observed for $n \leq m$. Since m was arbitrary, we see that a universal tail-bound for the height of

Bienaymé trees must be asymptotic.

Finally, in Chapter 2, we define a partial ordering $\prec$ on degree sequences and show that $\mathrm{d} \prec \mathrm{d}'$ implies that the height of T_{d} is stochastically dominated by the height of $\mathrm{T}_{\mathrm{d}'}$. This in particular implies that binary trees are the tallest trees with a given number of leaves and vertices. We refer the reader to Chapter 2 for the statement of these results.

1.3.2 The metric space scaling limit of the strongly connected components of the directed configuration model

In Chapter 3 we consider the directed configuration model in which the degree sequence consists of n i.i.d. random variables conditioned on the total in-degree being equal to the total out-degree. Let ν be a distribution on $\mathbb{N} \times \mathbb{N}$, and let $\mathbf{D}_1, \ldots, \mathbf{D}_n$ be a sequence of i.i.d. random variables with distribution ν. We condition on the event

$$\left\{\textstyle\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+\right\},$$

observing that this is an asymptotically singular event as $n \to \infty$. Let $\vec{G}_n(\nu)$ be a digraph chosen uniformly at random from all digraphs with degree sequence $\mathbf{D}_1, \ldots, \mathbf{D}_n$. We are interested in the limit under rescaling of the SCCs of $\vec{G}_n(\nu)$ as $n \to \infty$.

Suppose (D^-, D^+) has law ν. We will require the following assumptions to hold:

1. $\mathbb{E}[(D^-)^i (D^+)^j] < \infty$ for integers $1 \le i + j \le 3$, $(i, j) = (1, 3)$ and $(i, j) = (3, 1)$.

2. $\mathbb{E}[D^-] = \mathbb{E}[D^+]$.

3. $D^- - D^+$ is strongly aperiodic. This means that for all $p > 1$, there does not exist $k \in \mathbb{Z}$ such that
$$\mathbb{P}(D^- - D^+ \in k + p\mathbb{Z}) = 1.$$

4. $\mathbb{E}[D^- D^+] = \mathbb{E}[D^\pm]$.

The first condition is required to ensure that the steps of a random process, akin to the Łukasiewicz path of a forest, have finite variance, so that the random process will convergence under rescaling to a random continuous process that is locally absolutely continuous to Brownian motion. It also ensures similar regularity of other random variables that we use to encode the directed graph.

The second and third conditions make sure the event $\{\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+\}$ is well-behaved. The second condition ensures that it is not a large deviation event. Using a result from Spitzer [96, Page 42, P1], the third condition ensures that the event has positive probability for all sufficiently large $n \geq 1$. This condition can be relaxed to assuming that $D^- - D^+$ is non-constant by taking limits for $n \in p\mathbb{N}$ rather than $n \in \mathbb{N}$ where p is the periodicity of $D^- - D^+$. However, for simplicity of presentation, we keep it as an assumption.

The fourth assumption is the criticality condition. To understand how this arises, consider the directed configuration model and let (V_n, W_n) be a uniformly chosen edge. For now, ignore the conditioning on the total in- and out-degrees being equal. We consider the distribution of the in- and out-degree of W_n. Because the degree sequence is an i.i.d. sequence, W_n is equally likely to be any vertex i. Thus for any $\mathbf{k} = (k^-, k^+)$,

$$\begin{aligned}\mathbb{P}(d^-(W_n) = k^-, d^+(W_n) = k^+) &= n\mathbb{P}(W_n = 1, \mathbf{D}_1 = \mathbf{k}) \\ &= n\mathbb{E}[\mathbb{P}(W_n = 1 \mid \mathbf{D}_1 = \mathbf{k}, \mathbf{D}_2, \ldots, \mathbf{D}_n)]\mathbb{P}(\mathbf{D}_1 = \mathbf{k})\end{aligned}$$

Conditionally on the degree sequence, we have that $W_n = i$ with probability proportional to D_i^- since we used an uniform pairing of the in- and out-half-edges. Therefore

$$\mathbb{P}(W_n = 1 \mid \mathbf{D}_1 = \mathbf{k}, \mathbf{D}_2, \ldots, \mathbf{D}_n) = \frac{k^-}{k^- + \sum_{i=2}^n D_i^-}.$$

Thus

$$\mathbb{P}(d^-(W_n) = k^-, d^+(W_n) = k^+) = \mathbb{E}\left[\frac{k^-}{\frac{1}{n}\left(k^- + \sum_{i=2}^n D_i^-\right)}\right]\mathbb{P}\left[D^- = k^-, D^+ = k^+\right].$$

Using the law of large numbers, the above will converge to

$$\frac{k^-}{\mathbb{E}[D^-]}\mathbb{P}\left[D^- = k^-, D^+ = k^+\right].$$

Let (Z^-, Z^+) be such that $P(Z^- = k^-, Z^+ = k^+)$ is given by the above expression. We say (Z^-, Z^+) has the law of the *degree distribution size-biased by in-degree*. For large n, any other fixed out-edge of W_n is then also distributed approximately like a uniformly chosen edge (here we are ignoring the fact that we have already sampled an edge) since we chose the in- and out-edge pairing uniformly at random. Therefore the out-degree of the head will have approximately the same distribution as Z^+. Thus if we were to look at the graph of all vertices leading from W_n, it would look approximately like a Bienaymé tree with offspring distributed as Z^+. By Proposition 1.1.3 these trees exhibit critical behaviour in whether or not the tree is finite at $\mathbb{E}[Z^+] = 1$. This is equivalent to assuming $\mathbb{E}[D^-D^+] = E[D^-]$. This argument can be formalised as a local limit theorem for the graph.

We define the following parameters that will determine the behaviour of the SCCs in the limit.

1. $\mu := \mathbb{E}[D^-] = \mathbb{E}[D^+] = \mathbb{E}[D^-D^+]$
2. $\nu_- := \mathbb{E}[Z^-] - 1 = \frac{\mathbb{E}[(D^-)^2] - \mu}{\mu}$
3. $\sigma_-^2 := \mathrm{Var}(Z^-) = \frac{\mu\mathbb{E}[(D^-)^3] - \mathbb{E}[(D^-)^2]^2}{\mu^2}$
4. $\sigma_+^2 := \mathrm{Var}(Z^+) = \frac{\mathbb{E}[D^-(D^+)^2] - \mu}{\mu}$
5. $\sigma_{-+} := \mathrm{Cov}(Z^-, Z^+) = \frac{\mathbb{E}[(D^-)^2D^+] - \mathbb{E}[(D^-)^2]}{\mu}$

We will now state the main result from Chapter 3. For M an MDM and $c \in (0, \infty)$, let cM be equal to M with all lengths multiplied by c. Let $C_i(n)$ for $i \geq 1$ be the kernels of the SCCs of $\vec{G}_n(\nu)$, listed in decreasing order of number of edges, breaking ties arbitrarily. Complete the list with an infinite repeat of $\mathfrak{L}$. Then, we prove the following theorem.

Theorem 1.3.8. *There exists a sequence $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ of random strongly connected MDMs such that*

$$\left(n^{-1/3}C_i(n), i \in \mathbb{N}\right) \xrightarrow{\text{(d)}} (\mathcal{C}_i, i \in \mathbb{N})$$

as $n \to \infty$, with respect to the product $d_{\vec{\mathcal{G}}}$-topology. The law of $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ depends only on the parameters μ, σ_+, and $(\sigma_{-+} + \nu_-)/\mu$. Further, for each $i \geq 1$, $\mathcal{C}_i$ is either 3-regular or a loop.

Notably, the limit under rescaling of the directed Erdős–Rényi model at criticality obtained in [59] is part of the family of limit objects that we obtain, which is the content of the following corollary.

Corollary 1.3.9. *Consider $\vec{G}_n(\nu)$, with ν such that*

$$\mu = \sigma_+ = \sigma_{-+} + \nu_- = 1.$$

Let $(C_i^{\nu}(n), i \geq 1)$ *be the kernels of the SCCs of* $\vec{G}_n(\nu)$*. Furthermore, let* $(C_i^{ER}(n), i \geq 1)$ *be the kernels of the SCCs of* $\vec{G}(n, 1/n)$*. Then,* $\left(n^{-1/3}C_i^{\nu}(n), i \in \mathbb{N}\right)$ *and* $\left(n^{-1/3}C_i^{ER}(n), i \in \mathbb{N}\right)$ *have the same limit in distribution in the product-*$d_{\vec{\mathcal{G}}}$*-topology as* $n \to \infty$*.*

Note that the condition in Corollary 1.3.9 is satisfied by $\nu(k^-, k^+) = \nu_1(k^-)\nu_2(k^+)$, with ν_1 and ν_2 the law of a Poisson(1) random variable.

Moreover, Theorem 1.3.8 has the following trivial corollaries, which were previously unknown.

Corollary 1.3.10. *Let* E_n^i *and* V_n^i *be the number of edges and vertices in* $C_i(n)$ *respectively, both appended with infinite repeats of* 0*. Then there exists a random sequence* $(E_i, i \in \mathbb{N}) \in \mathbb{R}_+^{\infty}$*, such that*

$$\left(n^{-1/3}E_i^n, n^{-1/3}V_i^n, i \in \mathbb{N}\right) \xrightarrow{\text{(d)}} (E_i, E_i, i \in \mathbb{N})$$

as $n \to \infty$ *in the product topology on* $(\mathbb{R}^2)^{\infty}$*.*

In particular, note that, in the above corollary, the number of vertices and number of edges have exactly the same scaling limit.

Corollary 1.3.11. *For* $v, w \in \vec{G}_n(\nu)$ *such that* $v \to w$*, let* $d(v, w)$ *denote the length of the shortest directed path from* v *to* w*, and let*

$$\text{Diam}\left(\vec{G}_n(\nu)\right) = \max_{v,w \in V}\{d(v, w) : v \to w\}$$

be the diameter *of* $\vec{G}_n(\nu)$*. Then,* $\text{Diam}\left(\vec{G}_n(\nu)\right) = \Omega(n^{1/3})$ *in probability.*

All these results will be proved in Chapter 3.

1.4 Brief literature review

1.4.1 The height of random trees

We start by reviewing previous work on the height of random trees. We will closely follow the presentation in Chapter 2.

The study of the heights of trees spans decades, beginning with work of Harary and Prins [66], Riordan [91] and Rényi and Szekeres [90]. The work [66] developed generating functions for the number of rooted, labeled trees with given height (which they called *root diameter*); their work was extended to more general models of trees, including partially labeled trees, in [91]. Rényi and Szekeres [90] analyzed the generating functions developed in [66, 91] to prove that for a uniformly random rooted labeled tree $\mathrm{T}_n \in_u \mathcal{T}(n)$, the height satisfies $n^{-1/2}\mathrm{ht}(\mathrm{T}_n) \overset{\mathrm{d}}{\to} H$ for a random variable H with $\mathbb{E}H = (2\pi)^{1/2}$. Most other early work also focussed on heights and diameters of specific random tree models, such as random labeled trees [99] or random plane trees [42], and on distances between typical pairs of points in such models [83, 82].

A number of works then investigated the height (and width) of somewhat more general families of trees, such as the so-called "simple" trees [54, 35]. (Simple trees may be thought of as random coloured plane trees, where nodes with c children may receive any colour from $\{1, \ldots, \kappa(c)\}$; the values $\{\kappa(c), c \geq 0\}$ must all be bounded by some fixed constant M.) In all these models, for a random tree T_n of size n, the height $\mathrm{ht}(\mathrm{T}_n)$ is of order $\sqrt{n}$ in probability, for n large. Of particular note in the context of the current work are the papers of Flajolet *et al* [55], which stated the first explicit non-asymptotic tail bounds on the heights of uniform random binary trees; and of Łuczak [79], which established uniform sub-Gaussian tail bounds for the rescaled height of random labeled trees. Łuczak showed in particular that there exists an absolute constant $C > 0$ such that for $\mathrm{T}_n \in_u \mathcal{T}(n)$, for all $k \geq \sqrt{n}$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_n) = k) \leq C \frac{n!}{n^k (n-k)!} \frac{k^3}{n^3}\,.$$

A sub-Gaussian tail bound follows on observing that $n!/(n^k(n-k)!) = \prod_{i=0}^{k-1}(1 - i/n) \leq \exp(-k(k-1)/(2n))$.

Kolchin [71, Theorem 2.4.3] proved a far-reaching generalization of the above distributional results, showing that $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n})$ converges in distribution, to a random variable with sub-Gaussian tails, whenever $\sum_{i\geq 1} i\mu_i = 1$ and $\sum_{i\geq 1} i^2\mu_i \in (0, \infty)$. All the models described in the previous paragraphs are handled by Kolchin's results. This bound plays an essential rôle in Aldous's proof that critical branching processes conditioned to have n vertices converge to the Brownian CRT, whenever the offspring distribution has finite variance, as discussed in Subsection 1.2.2.

In contrast to our work, almost all recent research on non-asymptotic bounds for heights of random trees has been focussed on showing that, for sequences of trees which converge in distribution (after rescaling) to limiting continuum random trees, tail bounds of the same form as the ones that hold for the height of the limit object can already be observed in the finite setting. In the "finite variance" case, where the limiting object is the Brownian CRT, this is accomplished in [8, 2]. Building on [8], such tail bounds have also been proved for random graph ensembles which are *not* trees, but which have the Brownian CRT as their scaling limit; see [89, 98]. In the case of *heavy-tailed* degree distributions, for which the associated random trees converge in distribution to the α-stable trees that we discussed in Subsection 1.2.2, non-asymptotic tail bounds for the height (which match those of the limiting objects) have been obtained by Kortchemski [72].

One of the main points of Chapter 2 is to show that the assumption of convergence under rescaling, a feature of all the above works, is not necessary in order to obtain strong, non-asymptotic bounds on the height. Indeed, in the setting of conditioned branching processes $\mathrm{T}_{\mu,n}$, our main results precisely describe what information is required in order to obtain sub-Gaussian tail bounds for $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n})$, and also under precisely what conditions $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ in probability; in both cases, convergence is not a necessary ingredient. In the setting of simply generated trees and random trees with fixed degree sequences, our results likewise do not depend on any assumptions about asymptotic behaviour.

1.4.2 The configuration model

We now review some previous work on the configuration model, with a focus on the phase transition for the component sizes and scaling limits. We will closely follow the presentation in Chapter 3. As previously mentioned, the configuration model was introduced by Bollobás [27] to sample a uniformly random undirected graph with a given degree sequence. (For a discussion of the configuration model and proofs of standard results, we refer the reader to [101, Chapter 7].)

Most results on the configuration model are proved for models with a deterministic degree sequence. The phase transition for the undirected setting was shown in [85, 86]. The convergence in distribution of the component sizes at criticality and in the critical window were obtained by Riordan [92] under the assumption that the degrees are bounded. Dhara, van der Hofstad, van Leeuwaarden and Sen showed convergence of the size and number of surplus edges in the critical window with a finite third moment [44] and in the heavy-tailed regime [45]. Bhamidi, Dhara, van der Hofstad and Sen obtained metric space convergence in the critical window in [24], a result that the authors later improved to a stronger topology in [23].

Configuration models with a random degree sequence are considered in [70], [37], [46], and

[47]. Joseph [70] showed convergence of the component sizes and surpluses of the large components under rescaling at criticality, both for degree distributions with finite third moments and for the heavy-tailed regime. Conchon-Kerjan and Goldschmidt [37] show Gromov–Hausdorff–Prokhorov convergence of the rescaled components ordered by decreasing size at criticality in these two regimes. The results in [37] in the heavy-tailed regime are extended to the critical window related to bond percolation by the author in [46] and to the critical window related to inhomogeneous site percolation by the author in forthcoming work [47] that is discussed in Chapter 4.

The directed configuration model was first considered by Cooper and Frieze [38]. They prove the existence of a phase transition in the size of the largest strongly-connected component for the directed configuration model with a deterministic degree sequence $\mathbf{d}_1, \ldots, \mathbf{d}_n$. They define the parameter

$$d = \frac{\sum_{i=1}^{n} d_i^+ d_i^-}{\sum_{i=1}^{n} d_i^-}$$

which is the counterpart of $\mathbb{E}[Z^-]$ for deterministic degree sequences. They then show that, under additional assumptions, the existence of a giant SCC depends on whether d is strictly greater than or less than 1. Our work in this paper shows our corresponding condition, $\mathbb{E}[Z^-] = 1$, is also the correct criticality condition to take for i.i.d. random degree sequences. They show that for $d < 1$, with high probability, all SCCs contain $O(\Delta \log(n))$ vertices, for Δ the maximal degree. On the other hand, for $d > 1$, there is a unique SCC that contains a positive proportion of the vertices and edges. Their conditions are restrictive, and include finite second moments for both the in- and out-degree of a uniformly chosen vertex, and a bound of size $n^{1/12}/\log(n)$ on the largest degree. Their proofs are based on an algorithm to explore the directed graph. The condition on the largest degree was later relaxed to $O(n^{1/4})$ by Graf [60]. These results are in contrast with the critical case; see Corollary 1.3.10, which says that in our set-up the number of vertices and edges in the largest strongly connected components are $\Theta(n^{1/3})$ in probability.

Recently, Cai and Perarnau have obtained a number of results on the directed configuration model with deterministic degrees. In [30], they show, under first and second moment conditions of the degree of a uniformly picked vertex, for $d \neq 1$ (i.e. not at criticality), that the diameter of the model on n vertices, rescaled by $\log(n)$ converges to a constant that they identify. This is in contrast with Corollary 3.1.4, which says that in our set-up the diameter is $\Omega(n^{1/3})$ in probability at criticality. Then, in [31], they show a law of large numbers for the number of vertices and edges in the largest SCC, under slightly stronger moment conditions, and again away from the critical point. In [32], they study the behaviour of a random walk on a directed configuration model.

The directed configuration model with random in- and out-degrees is also considered by Chen and Olvera-Cravioto [36] although, importantly, they do not allow for the in- and out-degree of a vertex to be dependent. The authors consider a model in which the in- and out-degrees are two independent sequences of i.i.d. random variables drawn from different probability distributions. They propose an algorithm to sample degree sequences that correspond to a simple graph and show the limiting distribution of the degrees generated by this algorithm.

Chapter 2

Random trees have height $O(\sqrt{n})$

This chapter is joint work with Louigi Addario-Berry and previously appeared as a preprint [6].

We obtain new non-asymptotic tail bounds for the height of uniformly random trees with a given degree sequence, simply generated trees and conditioned Bienaymé trees (the family trees of branching processes), in the process settling two conjectures of Janson [68] and answering several other questions from the literature. Moreover, we define a partial ordering on degree sequences and show that it induces a stochastic ordering on the heights of uniformly random trees with given degree sequences. The latter result can also be used to show that sub-binary random trees are stochastically the tallest trees with a given number of vertices and leaves (and thus that random binary trees are the stochastically tallest random homeomorphically irreducible trees [66] with a given number of vertices).

Our proofs are based in part on a bijection introduced by Addario-Berry, Donderwinkel, Maazoun and Martin [11], which can be recast to provide a line-breaking construction of random trees with given vertex degrees.

2.1 Introduction

This paper proves optimal asymptotic and non-asymptotic tail bounds on the heights of random trees from several natural and well-studied classes, in the process proving conjectures and answering questions from [68, 8, 4]. We elaborate on connections with previous literature in Section 2.1.2, after the presentation of our results. In Section 2.5, we discuss directions for future research opened by the current work.

We begin by stating our results for the *Bienaymé* trees, which constitute the most elementary and well-studied random graph model and the oldest stochastic model for studying population growth. For $\mu = (\mu_k, k \geq 0)$ a probability distribution on $\mathbb{N} = \mathbb{Z} \cap [0, \infty)$, a Bienaymé tree with offspring distribution μ, denoted T_μ, is the family tree of a branching process with offspring distribution μ.[1] (An example appears in Figure 2.1.) We write $|\mathrm{T}_\mu|$ for the size (number of vertices) of T_μ. For $n \in \mathbb{N}$ such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$, we write $\mathrm{T}_{\mu,n}$ to denote a Bienaymé tree with offspring distribution μ conditioned to have size n. Bienaymé trees are random *plane* trees: rooted, unlabeled trees in which the set of children of each node is endowed with a (left-to-right) total order.

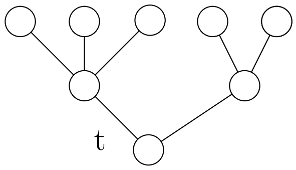


Fig. 2.1: In the tree t, there are 5 vertices with no children, two vertices with two children and one vertex with 3 children, so for $\mu = (\mu_k, k \geq 0)$ a probability distribution on $\mathbb{N}$, we have that $\mathbb{P}(\mathrm{T}_\mu = \mathrm{t}) = \mu_0^5 \mu_2^2 \mu_3$.

For $p > 0$, write $|\mu|_p = (\sum_{k=0}^{\infty} k^p \mu_k)^{1/p}$.

Theorem 2.1.1. *Fix a probability distribution μ supported on $\mathbb{N}$ with $|\mu|_1 \leq 1$ and $|\mu|_2 = \infty$. Then, $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ as $n \to \infty$ along values n such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$, both in probability and in expectation.*

Theorem 2.1.1 answers an open question from [8].

Theorem 2.1.2. *Fix a probability distribution $\mu = (\mu_k, k \geq 0)$ supported on $\mathbb{N}$ with $|\mu|_1 < 1$ and with $\sum_{k\geq 0} e^{tk}\mu_k = \infty$ for all $t > 0$. Then, $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ as $n \to \infty$ over all n such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$, both in probability and in expectation.*

Theorem 2.1.2 solves a slight variant of Problem 21.7 from [68]. (That problem is stated for simply generated trees, which are a slight generalization of Bienaymé trees. We in fact solve the problem from [68] in full; see Theorem 2.3.1, below.)

A collection $(X_i)_{i\in I}$ of random variables is *sub-Gaussian* if there exist constants $c, C > 0$ such that $\mathbb{P}(X_i \geq x) \leq C\exp(-cx^2)$ for all $i \in I$ and all $x > 0$. Our third theorem states that for any offspring distribution $\mu = (\mu_k, k \geq 0)$, the family of random variables $(n^{-1/2}\mathrm{ht}(T_{\mu,n}))$ has sub-Gaussian tails, with constants c, C that only depend on μ_0 and μ_1.

---

[1] Bienaymé trees are often referred to as Galton–Watson trees, but we adopt the change in terminology suggested in [10].

Theorem 2.1.3. *Fix $\epsilon \in [0,1)$. Then, there exist constants $c = c(\epsilon)$ and $C = C(\epsilon)$ such that the following holds. For any probability distribution $\mu = (\mu_k, k \geq 0)$ on $\mathbb{N}$ with $\mu_0 + \mu_1 < 1 - \epsilon$ and $\mu_0/(\mu_0 + \mu_1) > \epsilon$, for all n sufficiently large and for which $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$, for all $x > 0$,*

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mu,n}) > xn^{1/2}\right) \leq C \exp(-cx^2).$$

This theorem has the following immediate corollary.

Corollary 2.1.4. *For any probability distribution $\mu = (\mu_k, k \geq 0)$ on $\mathbb{N}$ with $\mu_0 + \mu_1 < 1$, $\mathbb{E}[\mathrm{ht}(\mathrm{T}_{\mu,n})] = O(n^{1/2})$, and, more generally, for any fixed $r < \infty$, $\mathbb{E}[\mathrm{ht}(\mathrm{T}_{\mu,n})^r] = O(n^{r/2})$ as $n \to \infty$ over all n such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$.* ☐

Theorem 2.1.3 and Corollary 2.1.4 strengthen Theorem 1.2 and Corollary 1.3 in [8], as they are not restricted to critical offspring distributions with finite variance and the bounds only depend on μ via μ_0 and μ_1. This resolves conjectures stated in [4] and [8].

Remark 2.1.5. *The requirement in Theorem 2.1.3 that n is sufficiently large (where "sufficiently large" depends on μ) is necessary. To see this, note that if μ has support $\{0, 1, N\}$ then for any $n < N + 1$, with probability one $\mathrm{T}_{\mu,n}$ is a path (so has height $n - 1$). The requirement in Theorem 2.1.3 that $1 - \mu_0 - \mu_1$ and $\mu_0/(\mu_0 + \mu_1)$ be bounded from below is also necessary. To see this, it suffices to consider probability distributions μ of the form*

$$\mu_0 = q(1-p) \qquad \mu_1 = (1-q)(1-p) \qquad \mu_2 = p\,.$$

For any $x > 1$, it is possible to make $\liminf_{n\to\infty} \mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mu,n}) > xn^{1/2})$ arbitrarily close to one by either taking p fixed and q sufficiently small or q fixed and p sufficiently small; this fact is proved as Claim 2.3.5, below.

For random variables X, Y, we write $X \preceq_{\mathrm{st}} Y$ to mean that Y stochastically dominates X, which is to say that $\mathbb{P}(X \leq t) \geq \mathbb{P}(Y \leq t)$ for all $t \in \mathbb{R}$. We also write $X \prec_{\mathrm{st}} Y$, and say that Y strictly stochastically dominates X, if $X \preceq_{\mathrm{st}} Y$ and X and Y are not identically distributed. Our final result for Bienaymé trees states that, among all conditioned Bienaymé trees which almost surely have no vertices with exactly 1 child, the binary Bienaymé trees have the stochastically largest heights.

Theorem 2.1.6. *Fix any probability distribution $\mu = (\mu_k, k \geq 0)$ on $\mathbb{N}$ with $\mu_0 \in (0,1)$ and $\mu_1 = 0$, and let ν be the probability distribution on $\mathbb{N}$ with $\nu(0) = \nu(2) = 1/2$. Then for all*

$n \geq 1$ *such that* $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$,

$$\mathrm{ht}(\mathrm{T}_{\mu,n}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\nu,n}) \text{ if } n \text{ is odd}$$

$$\mathrm{ht}(\mathrm{T}_{\mu,n}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\nu,n+1}) \text{ if } n \text{ is even.}$$

The probability measure ν in the preceding theorem could be replaced by any nondegenerate probability measure ν' with $\nu'(0) + \nu'(2) = 1$, as for any such measure and any $m \in \mathbb{N}$, the tree $\mathrm{T}_{\nu',2m+1}$ is uniformly distributed over rooted binary plane trees with m internal nodes and $m+1$ leaves. The parity requirement is due to the fact that binary trees always have an odd number of vertices. Also, the requirement that $\mu_1 = 0$ is necessary, since if $\mu_1 > 0$, then with positive probability $\mathrm{T}_{\mu,2m}$ (resp. $\mathrm{T}_{\mu,2m+1}$) is a path of length $2m-1$ (resp. $2m$), whereas $\mathrm{T}_{\nu,2m+1}$ has height at most $m+1$.

The preceding theorem has the following consequence for tail bounds on the height of conditioned Bienaymé trees. Fix any constants c and C such that the inequality in Theorem 2.1.3 holds for uniform binary trees $\mathrm{T}_{\nu,n}$. Then, the same inequality also holds for $\mathrm{T}_{\mu,n}$ for any probability distributions μ with $\mu(1) = 0$. In particular, by the remark above, we see that if we restrict the theorem to such probability distributions, the constants need not depend on ϵ.

2.1.1 Trees with fixed degree sequences

The above results for Bienaymé trees are consequences of more refined results, presented in this subsection, on the heights of random trees with fixed degree sequences.

For a rooted tree t and a vertex v of t, the *degree* $d_{\mathrm{t}}(v)$ of v is the number of children of v in t. A *leaf* of t is a vertex of t with degree zero.

A *degree sequence* is a sequence of non-negative integers $\mathrm{d} = (d_1, \ldots, d_n)$ with $\sum_{i \in [n]} d_i = n-1$. We write $\mathcal{T}_{\mathrm{d}}$ for the set of finite rooted labeled trees t with vertex set labeled by $[n]$ and such that for each $i \in [n]$, the vertex with label i has degree d_i. Also write $\mathcal{T}(n)$ for the set of all finite rooted labeled trees with vertex set labeled by $[n]$. For $\mathrm{t} \in \mathcal{T}(n)$, we write $d_{\mathrm{t}}(i)$ for the degree of the vertex with label i and say that $(d_{\mathrm{t}}(1), \ldots, d_{\mathrm{t}}(n))$ is the degree sequence of t.

For $p > 0$ and a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$ write $|\mathrm{d}|_p = (\sum_{i=1}^n d_i^p)^{1/p}$ and let $(\sigma_{\mathrm{d}})^2 = \frac{1}{n} \sum_{i=1}^n d_i(d_i - 1)$. Also, for $i \geq 0$ write $n_i(\mathrm{d}) = |\{j \in [n] : d_j = i\}|$ for the number of entries of d which equal i.

Given a finite set S, we write $X \in_u S$ to mean that X is a uniformly random element of S.

Theorem 2.1.7. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*, and write* $\delta = (n - n_1(\mathrm{d}))/n$*. Then for all* $x > 0$*,*

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}) > xn^{1/2}\right) < 5\exp(-\delta x^2/2^{13}).$$

Theorem 2.1.8. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*, and define* $(\sigma')^2 = \frac{n}{n-n_1(\mathrm{d})}\sigma_{\mathrm{d}}^2$*. Then for all* $x \geq 2^{15}$*,*

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}) > xn^{1/2}\frac{\log(\sigma'+1)}{\sigma_{\mathrm{d}}}\right) \leq 4\exp\left(-x\log(\sigma'+1)/2^{14}\right).$$

The following proposition is the tool which allows us to transfer results from the setting of random trees with given degree sequences to that of Bienaymé trees (and that of simply generated trees, which are introduced in Section 2.4, below).

Proposition 2.1.9. *Fix non-negative weights* $(w_i, i \geq 0)$*. Let* T *be a random plane tree of size* n *with distribution given by*

$$\mathbb{P}(\mathrm{T} = \mathrm{t}) \propto \prod_{v \in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)}, \tag{2.1}$$

for plane trees t *of size* n*. Conditionally given* T*, let* $\hat{\mathrm{T}}$ *be the random tree in* $\mathcal{T}(n)$ *obtained as follows: label the vertices of* T *by a uniformly random permutation of* $[n]$*, then forget about the plane structure. For* $i \in [n]$ *let* D_i *be the degree of vertex* i *in* $\hat{\mathrm{T}}$ *of vertex* i*. Then for any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, conditionally given that* $(D_1, \ldots, D_n) = (d_1, \ldots, d_n)$*, we have* $\hat{\mathrm{T}} \in_u \mathcal{T}_{\mathrm{d}}$*.*

We provide the proof of Proposition 2.1.9 right away since it is short, but it is not necessary to read the proof in order to understand what follows.

Proof of Proposition 2.1.9. In this proof, by an *ordered labeled tree* we mean a finite tree $\mathbf{t}$ whose vertices are labeled by the integers $[n]$ (for some n) and such that the children of each node are endowed with a left-to-right order. From an ordered labeled tree, we may obtain a plane tree by ignoring the vertex labels, and we may obtain a labeled tree lying in $\mathcal{T}(n)$ by ignoring the orderings.

Let $\mathbf{T}$ be the random ordered labeled tree obtained from T by labeling the vertices of T by a uniformly random permutation of $[n]$.

Now fix any labeled tree $\hat{\mathrm{t}}$ with vertices labeled by $[n]$. For $j \geq 0$ write $n_j = \{i \in [n] : d_{\hat{\mathrm{t}}}(i) = j\}$. Then

$$\mathbb{P}(\hat{\mathrm{T}} = \hat{\mathrm{t}}) = \sum_{\mathbf{t}} \mathbb{P}(\mathbf{T} = \mathbf{t}),$$

where the sum is over ordered labeled trees $\mathbf{t}$ with vertices labeled by $[n]$ whose underlying (unordered) labeled tree is $\hat{\mathrm{T}}$. This sum has

$$\prod_{v \in \hat{\mathrm{t}}} d_{\hat{\mathrm{t}}}(v)! = \prod_{j \geq 0} (j!)^{n_j}$$

summands, corresponding to the number of ways to choose an ordering of the children of each vertex of $\hat{\mathrm{t}}$. Moreover, for any tree $\mathbf{t}$ included in the sum, writing t for the (unlabeled) plane tree underlying $\mathbf{t}$, we have

$$\mathbb{P}(\mathbf{T} = \mathbf{t}) = \frac{1}{n!}\mathbb{P}(\mathrm{T} = \mathrm{t}) \propto \prod_{v \in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)} = \prod_{j \geq 0} w_j^{n_j}.$$

This value is the same for all summands, so it follows that

$$\mathbb{P}(\hat{\mathrm{T}} = \hat{\mathrm{t}}) \propto \prod_{j \geq 0} (j!)^{n_j} \prod_{j \geq 0} w_j^{n_j}.$$

The right-hand side depends on $\hat{\mathrm{t}}$ only through the number of vertices of each degree. In particular, for any degree sequence d, it is constant over trees in $\mathcal{T}_{\mathrm{d}}$. The result follows. ☐

Using this proposition, we can transfer results from the setting of random trees with given degree sequences to that of random plane trees, provided that we can obtain sufficiently precise control of the degree statistics of the random plane trees. (By "degree statistics" we mean the number of nodes of given degrees.) Once such control is established, Theorems 2.1.1, 2.1.2, and 2.1.3 follow straightforwardly from Theorems 2.1.7 and 2.1.8. Similarly, the stochastic inequality of Theorem 2.1.6 follows from a finer stochastic ordering on the heights of random trees with given degree sequences. We in fact obtain stochastic domination results for the heights of trees under two different partial orders on degree sequences We define these partial orders by specifying their corresponding covering relations[2].

Theorem 2.1.10. *Let $\prec_1$ be the partial order on degree sequences of length n defined by the following covering relation:* $\mathrm{d} = (d_1, \ldots, d_n)$ *covers* $\mathrm{d}' = (d'_1, \ldots, d'_n)$ *if there exist distinct*

---

[2] For a partially ordered set $(\mathcal{P}, \prec)$, $y \in \mathcal{P}$ covers $x \in \mathcal{P}$ if $x \prec y$ and for all $z \in \mathcal{P}$, if $x \preceq z \preceq y$ then $z \in \{x, y\}$.

$i, j \in [n]$ and a permutation $\nu : [n] \to [n]$ such that

1. *$d_i \geq d_j$;*
2. *$d'_{\nu(i)} = d_i + 1$;*
3. *$d'_{\nu(j)} = d_j - 1$; and*
4. *$d'_{\nu(k)} = d_k$ for any $k \in [n] \backslash \{i, j\}$.*

Then with $\mathrm{T}_{\mathrm{d}} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}_{\mathrm{d}'} \in_u \mathcal{T}_{\mathrm{d}'}$,

$$\mathrm{d}' \prec_1 \mathrm{d} \implies \mathrm{ht}(\mathrm{T}_{\mathrm{d}'}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}}).$$

Moreover, the stochastic domination of $\mathrm{ht}(\mathrm{T}_{\mathrm{d}})$ over $\mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$ is strict if $\mathrm{d}' \prec_1 \mathrm{d}$ and d contains at least 3 non-leaf vertices.

In words, to obtain d' from d in the definition of $\prec_1$, for $a \leq b$, we replace one vertex with a children and one vertex with b children by a vertex with $a - 1$ children and one with $b + 1$ children and potentially relabel the vertices; informally, the degrees in d' are more skewed than the degrees in d. Then, $\mathrm{d}' \prec_1 \mathrm{d}$ and $\mathrm{ht}(\mathrm{T}_{\mathrm{d}'}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}})$. So, informally, Theorem 2.1.10 states that more skewed degrees yield shorter trees. This theorem has the following corollary, which resolves a conjecture from [11].

Corollary 2.1.11. *Let $\prec_2$ be the partial order on degree sequences of length n defined by the following covering relation: $\mathrm{d} = (d_1, \ldots, d_n)$ covers $\mathrm{d}' = (d'_1, \ldots, d'_n)$ if there is a permutation $\nu : [n] \to [n]$ and an $i, j \in [n]$ such that*

1. *$d_i, d_j \geq 1$;*
2. *$d'_{\nu(i)} = d_i + d_j$;*
3. *$d'_{\nu(j)} = 0$; and*
4. *$d'_{\nu(k)} = d_k$ for any $k \in [n] \backslash \{i, j\}$.*

Then with $\mathrm{T}_{\mathrm{d}} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}_{\mathrm{d}'} \in_u \mathcal{T}_{\mathrm{d}'}$,

$$\mathrm{d}' \prec_2 \mathrm{d} \implies \mathrm{ht}(\mathrm{T}_{\mathrm{d}'}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}}).$$

Corollary 2.1.11 follows immediately from Theorem 2.1.10 by observing that if $\mathrm{d}' \prec_2 \mathrm{d}$ then $\mathrm{d}' \prec_1 \mathrm{d}$. Arthur Blanc-Renaudie told us a proof of the stochastic domination presented in Corollary 2.1.11 (which he does not intend to write up), using a bijection presented in [25]. Our proof of Theorem 2.1.10 proceeds differently, but was inspired by Blanc-Renaudie's approach.

Before concluding the introduction, we state and prove two further corollaries of Theorem 2.1.10, and use the second of them to prove Theorem 2.1.6. The first states that height of any random tree with a fixed degree sequence is stochastically dominated by that of a random sub-binary tree with the same number of leaves and degree-one vertices. (We say d is sub-binary if $n_i(\mathrm{d}) = 0$ for all $i \geq 3$, and in this case also say that T_d is sub-binary.) The second essentially says that among all random trees with a given size and given number of vertices of degree 1, the sub-binary trees have the stochastically largest heights (with a very minor *caveat* that essentially addresses a potential parity issue).

Corollary 2.1.12. *Let* b *be a sub-binary degree sequence, let* d *be a degree sequence with* $n_0(\mathrm{d}) = n_0(\mathrm{b})$ *and* $n_1(\mathrm{d}) = n_1(\mathrm{b})$*, and let* $\mathrm{T}_\mathrm{d} \in_u \mathcal{T}_\mathrm{d}$ *and* $\mathrm{T}_\mathrm{b} \in_u \mathcal{T}_\mathrm{b}$*. Then* $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_\mathrm{b})$*.*

Proof. Note that necessarily $n_2(\mathrm{b}) = n_0(\mathrm{b}) - 1$, so b has length $2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$ Write $\mathrm{d} = (d_1, \ldots, d_n)$. If $n = 2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$ then d is a permutation of b and there is nothing to prove. So suppose that $n < 2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$; in this case necessarily d contains at least one entry which is three or greater. We construct a degree sequence $\mathrm{d}' = (d'_1, \ldots, d'_{n+1})$ with $n_0(\mathrm{d}) = n_0(\mathrm{d}')$ and $n_1(\mathrm{d}) = n_1(\mathrm{d}')$ and $\mathrm{ht}(T_\mathrm{d}) \preceq \mathrm{ht}(T_{\mathrm{d}'})$. This proves the corollary (by induction on $2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1 - n$), as we can then transform d into a degree sequence of length sub-binary degree sequence $2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$ without changing the number of zeros or of ones, and while stochastically increasing the height of the associated random tree.

To construct d′, first let $\mathrm{d}^+ = (d^+_1, \ldots, d^+_{n+1}) = (d_1, \ldots, d_n, 1)$. The trees T_d and $\mathrm{T}_{\mathrm{d}^+} \in_u \mathcal{T}_{\mathrm{d}^+}$ may be coupled as follows. If $n+1$ is not the root of $\mathrm{T}_{\mathrm{d}^+}$ then form T_d from $\mathrm{T}_{\mathrm{d}^+}$ by replacing the two-edge path containing $n+1$ in $\mathrm{T}_{\mathrm{d}^+}$ by a single edge connecting its endpoints. If $n+1$ *is* the root, then instead form T_b by deleting $n+1$ and rerooting at its unique child. (The reverse operation is to add $n+1$ as the parent of a uniformly random vertex of T_d.) Under this coupling, the height of $\mathrm{T}_{\mathrm{d}^+}$ is at least that of T_d, so it follows that

$$\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}^+})\,.$$

Now choose $k \in [n]$ such that $d_k^+ = d_k \geq 3$ and define $\mathrm{d}' = (d'_1, \ldots, d'_{n+1})$ by

$$d'_i = \begin{cases} d_i^+ & \text{if } i \notin \{k, n+1\} \\ d_k^+ - 1 & \text{if } i = k \\ 2 & \text{if } i = n+1\,. \end{cases}$$

Then $n_0(\mathrm{d}) = n_0(\mathrm{d}')$ and $n_1(\mathrm{d}) = n_1(\mathrm{d}')$. Moreover, $\mathrm{ht}(\mathrm{T}_{\mathrm{d}^+}) \preceq \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$ by Theorem 2.1.10, and so $\mathrm{ht}(\mathrm{T}_{\mathrm{d}}) \preceq \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$ as required. □

Corollary 2.1.13. *Let* $\mathrm{d} = (d_1, \ldots, d_n)$ *be any degree sequence and let* $\mathrm{T}_{\mathrm{d}} \in_u \mathcal{T}_{\mathrm{d}}$. *Let*

$$n^+ = \begin{cases} n & \textit{if } n_1(\mathrm{d}) \geq 1 \textit{ or if } n \textit{ is odd} \\ n+1 & \textit{if } n_1(\mathrm{d}) = 0 \textit{ and } n \textit{ is even.} \end{cases}$$

Then there is a sub-binary degree sequence $\mathrm{b} = (b_1, \ldots, b_{n^+})$ *with* $n_1(\mathrm{b}) \leq n_1(\mathrm{d})$, *such that with* $\mathrm{T}_{\mathrm{b}} \in_u \mathcal{T}_{\mathrm{b}}$, *then* $\mathrm{ht}(\mathrm{T}_{\mathrm{d}}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{b}})$.

Proof. Suppose that d has at least one entry $d_k \geq 4$. Choose $j \in [n]$ with $d_j = 0$. Then the degree sequence $\mathrm{d}' = (d'_1, \ldots, d'_n)$ given by

$$d'_i = \begin{cases} d_i - 2 & \text{if } i = k \\ 2 & \text{if } i = j \\ d_i & \text{otherwise} \end{cases}$$

has the same length as d and satisfies $n_1(\mathrm{d}') = n_1(\mathrm{d})$. Moreover, by Theorem 2.1.10, $\mathrm{ht}(\mathrm{T}_{\mathrm{d}}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$. We have reduced the number of vertices of degree 4 while stochastically increasing the height, and without changing the number of vertices of degree 1. It follows that to prove the corollary we may restrict our attention to degree sequences corresponding to trees with maximum degree three.

Next suppose that d has $n_3(\mathrm{d}) \geq 2$ and choose distinct $k, \ell \in [n]$ with $d_k = d_\ell = 3$ and

$j \in [n]$ with $d_j = 0$. Then the sequence $\mathrm{d}' = (d_1, \ldots, d'_n)$ given by

$$d'_i = \begin{cases} d_i - 1 & \text{if } i \in \{k, \ell\} \\ 2 & \text{if } i = j \\ d_i & \text{otherwise} \end{cases}$$

has the same length as d and satisfies $n_1(\mathrm{d}') = n_1(\mathrm{d})$ and has two fewer vertices of degree three, and two applications of Theorem 2.1.10 give that $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$. This shows that we may restrict our attention to degree sequences corresponding to trees of maximum degree three and with at most one vertex of degree three.

Among such degree sequences, if $n_3(\mathrm{d}) = 0$ there is nothing to prove – the tree T_d is already sub-binary. If $n_3(\mathrm{d}) = 1$ and $n_1(\mathrm{d}) \geq 1$ then let k be such that $d_k = 3$ and choose $j \in [n]$ with $d_j = 1$. Then the sequence $\mathrm{d}' = (d'_1, \ldots, d'_n)$ given by

$$d'_i = \begin{cases} 2 & \text{if } i \in \{j, k\} \\ d_i & \text{otherwise} \end{cases}$$

is sub-binary and has $n_1(\mathrm{d}') < n_1(\mathrm{d})$, and $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$.

Finally, if $n_3(\mathrm{d}) = 1$ and $n_1(\mathrm{d}) = 0$ then necessarily n is even. In this case, choose $k \in [n]$ such that $d_k = 3$, and define $\mathrm{d}' = (d'_1, \ldots, d'_{n+1})$ by

$$d'_i = \begin{cases} d_i & \text{if } i \notin \{k, n+1\} \\ d_k - 1 & \text{if } i = k \\ 2 & \text{if } i = n+1\,. \end{cases}$$

Then d' is sub-binary. Moreover, the same argument as that in the final paragraph of the proof of Corollary 2.1.12 shows that with $\mathrm{T}_{\mathrm{d}'} \in_u \mathcal{T}_{\mathrm{d}'}$, then $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$. This completes the proof. □

Proof of Theorem 2.1.6. For any $n \in \mathbb{N}$ with $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$ and any plane tree t of size n,

$$\mathbb{P}(\mathrm{T}_{\mu,n} = \mathrm{t}) = \frac{1}{\mathbb{P}(|\mathrm{T}_\mu| = n)} \prod_{v \in \mathrm{t}} \mu_{d_\mathrm{t}(v)}$$

and likewise if $\mathbb{P}(|\mathrm{T}_\nu| = n) > 0$ then $\mathbb{P}(\mathrm{T}_{\nu,n} = \mathrm{t}) = (\mathbb{P}(|\mathrm{T}_\mu| = n))^{-1} \prod_{v\in\mathrm{t}} \nu_{d_\mathrm{t}(v)}$. Thus, the laws of $\mathrm{T}_{\mu,n}$ and $\mathrm{T}_{\nu,n}$ both have the product structure required by Proposition 2.1.9.

Now fix n for which $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$ and let $n^+ = n + \mathbf{1}_{[n \text{ even}]}$, so that $\mathbb{P}(|\mathrm{T}_\nu| = n^+) > 0$. Label the vertices of $\mathrm{T}_{\mu,n}$ (resp. T_{ν,n^+}) by a uniformly random permutation of $[n]$ and write D (resp. B) for the resulting degree sequence. Then D is a degree sequence of length n with $n_1(\mathrm{D}) = 0$ and B is a binary degree sequence of length n^+. Corollary 2.1.13 yields that with $\mathrm{T}_\mathrm{B} \in_u \mathcal{T}_\mathrm{B}$ and $\mathrm{T}_\mathrm{D} \in_u \mathcal{T}_\mathrm{D}$, then

$$\mathrm{ht}(\mathrm{T}_\mathrm{D}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_\mathrm{B})\,.$$

But Proposition 2.1.9 implies that $\mathrm{ht}(\mathrm{T}_{\mu,n}) \overset{\mathrm{d}}{=} \mathrm{ht}(\mathrm{T}_\mathrm{D})$ and $\mathrm{ht}(\mathrm{T}_{\nu,n^+}) \overset{\mathrm{d}}{=} \mathrm{ht}(\mathrm{T}_\mathrm{B})$; the result follows. □

2.1.2 Related work

The study of the heights of trees spans decades, beginning with work of Harary and Prins [66], Riordan [91] and Rényi and Szekeres [90]. The work [66] developed generating functions for the number of labeled rooted trees with given height (which they called *root diameter*); their work was extended to more general models of trees, including partially labeled trees, in [91]. Rényi and Szekeres [90] analyzed the generating functions developed in [66, 91] to prove that for a uniformly random rooted labeled tree $\mathrm{T}_n \in_u \mathcal{T}(n)$, the height satisfies $n^{-1/2}\mathrm{ht}(\mathrm{T}_n) \xrightarrow{\mathrm{d}} H$ for a random variable H with $\mathbb{E}H = (2\pi)^{1/2}$. Most other early work also focussed on heights and diameters of specific random tree models, such as random labeled trees [99] and random plane trees [42], and on distances between typical pairs of points in such models [83, 82].

A number of works then investigated the height (and width) of somewhat more general families of trees, such as the so-called "simple" trees [54, 35]. (Simple trees may be thought of as random coloured plane trees, where nodes with c children may receive any colour from $\{1, \ldots, \kappa(c)\}$; the values $\{\kappa(c), c \geq 0\}$ must all be bounded by some fixed constant M.) In all these models, for a random tree T_n of size n, the height $\mathrm{ht}(\mathrm{T}_n)$ is of order $\sqrt{n}$ in probability, for n large. Of particular note in the context of the current work are the work of Flajolet *et al* [55], which stated the first explicit non-asymptotic tail bounds on the heights of uniform random binary trees; and of Łuczak [79], which established uniform sub-Gaussian tail bounds for the rescaled height of random labeled trees. Łuczak showed in particular that there exists an absolute constant $C > 0$ such that for $\mathrm{T}_n \in_u \mathcal{T}(n)$, for all $k \geq \sqrt{n}$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_n) = k) \leq C \frac{n!}{n^k (n-k)!} \frac{k^3}{n^3}\,.$$

A sub-Gaussian tail bound follows on observing that $n!/(n^k(n-k)!) = \prod_{i=0}^{k-1}(1 - i/n) \leq \exp(-k(k-1)/(2n))$.

Kolchin [71, Theorem 2.4.3] proved a far-reaching generalization of the above distributional results, showing that $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n})$ converges in distribution, to a random variable with sub-Gaussian tails, whenever $\sum_{i\geq 1} i\mu_i = 1$ and $\sum_{i\geq 1} i^2\mu_i \in (0,\infty)$. All the models described in the previous paragraphs are handled by Kolchin's results. This bound plays an essential role in Aldous's proof that critical branching processes conditioned to have n vertices converge to the *Brownian continuum random tree (CRT)*, whenever the offspring distribution has finite variance.

In contrast to our work, almost all recent research on non-asymptotic bounds for heights of random trees has been focussed on showing that, for sequences of trees which converge in distribution (after rescaling) to limiting continuum random trees, tail bounds of the same form as the ones that hold for the height of the limit object can already be observed in the finite setting. In the "finite variance" case, where the limiting object is the Brownian CRT, this is accomplished in [8, 2]. Building on [8], such tail bounds have also been proved for random graph ensembles which are *not* trees, but which have the Brownian CRT as their scaling limit; see [89, 98]. In the case of *heavy-tailed* degree distributions, for which the associated random trees converge in distribution to the so-called *α-stable trees*, non-asymptotic tail bounds for the height (which match those of the limiting objects) have been obtained by Kortchemski [72].

One of the main points of this work is to show that the assumption of convergence under rescaling, a feature of all the above works, is not necessary in order to obtain strong, non-asymptotic bounds on the height. Indeed, in the setting of conditioned branching processes $\mathrm{T}_{\mu,n}$, our main results precisely describe what information is required in order to obtain sub-Gaussian tail bounds for $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n})$, and also under precisely what conditions $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ in probability; in both cases, convergence is not a necessary ingredient. In the setting of simply generated trees and random trees with fixed degree sequences, our results likewise do not depend on any assumptions about asymptotic behaviour.

2.1.3 Outline

The remainder of the paper is structured as follows. Section 2.2 presents the proofs of our results on random trees with fixed degree sequences, Theorems 2.1.7 and 2.1.8. Section 2.3 uses these results to prove our results on Bienaymé trees, Theorems 2.1.1, 2.1.2 and 2.1.3. This section also introduces simply generated trees, and presents our results on their heights. Section 2.4, which can be read independently of the rest of the paper, contains the proof of Theorem 2.1.10. Finally, in Section 2.5, we discuss applications and extensions of our results and techniques to convergence of random trees under rescaling; height bounds and convergence in other random tree models; convergence of the diameter of the configuration model at criticality; and non-asymptotic bounds on distances in the components of the configuration model.

2.2 Proofs of Theorems 2.1.7 and 2.1.8

In this section we present the proofs of Theorems 2.1.7 and 2.1.8. (To streamline the presentation, a few of the longer proofs are postponed to Section 2.2.1.) Our approach is based on a bijection between rooted trees on $[n]$ and sequences in $[n]^{n-1}$ that was very recently introduced by Addario-Berry, Donderwinkel, Maazoun and Martin in [11]. We use the version of the bijection presented in [11, Section 3], specialized to trees with a fixed degree sequence.

To explain the bijection, it is convenient to focus on degree sequences with a particular form. Given a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$, define another degree sequence $\mathrm{d}' = (d'_1, \ldots, d'_n)$ as follows. Let m be the number of non-zero entries of d; necessarily $1 \leq m \leq n-1$. List the non-zero entries of d in order of appearance as $d'_1, \ldots, d'_m$, and then set $d'_{m+1} = \ldots = d'_n = 0$. So, for example, if $\mathrm{d} = (1, 0, 3, 0, 0, 2, 0)$ then $\mathrm{d}' = (1, 3, 2, 0, 0, 0, 0)$. Say that the degree sequence d' is *compressed*. (So a degree sequence is compressed if all of its non-zero entries appear before all of its zero entries.)

There is a natural bijection between $\mathcal{T}_{\mathrm{d}}$ and $\mathcal{T}_{\mathrm{d}'}$: from a tree $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$, construct a tree $\mathrm{t}' \in \mathcal{T}_{\mathrm{d}'}$ by relabeling the non-leaf vertices of t as $1, \ldots, m$ and the leaves of t as $m+1, \ldots n$, in both cases in increasing order of their original labels. Using this bijection provides a coupling $(\mathrm{T}, \mathrm{T}')$ of uniformly random elements of $\mathcal{T}_{\mathrm{d}}$ and $\mathcal{T}_{\mathrm{d}'}$, respectively, such that T and T' have the same height. It follows that any tail bound for the height of a uniformly random tree in $\mathcal{T}_{\mathrm{d}'}$ applies *verbatim* to the height of a uniformly random tree in $\mathcal{T}_{\mathrm{d}}$.

Now, let $\mathrm{d} = (d_1, \ldots, d_n)$ be a compressed degree sequence, so there is $1 \leq m \leq n-1$ such that $d_i = 0$ if and only if $i > m$. Write $n_0 = n_0(\mathrm{d}) = n - m$ for the number of leaves in a tree with degree sequence d. Then define

$$\mathcal{S}_{\mathrm{d}} := \{(v_1, \ldots, v_{n-1}) : |\{k : v_k = i\}| = d_i \text{ for all } i \in [n]\} .$$

For example, if $\mathrm{d} = (1, 3, 2, 0, 0, 0, 0)$ then $\mathcal{S}_{\mathrm{d}}$ is the set of all permutations of the vector $(1, 3, 3, 3, 2, 2)$, so has size $\binom{6}{1,3,2} = 60$.

The following bijection between $\mathcal{S}_{\mathrm{d}}$ and $\mathcal{T}_{\mathrm{d}}$ appears in [11, Section 3]. For $\mathrm{v} = (v_1, \ldots, v_{n-1}) \in \mathcal{S}_{\mathrm{d}}$, we say that $j \in \{2, \ldots, n-1\}$ is the location of a repeat of v if $v_j = v_i$ for some $i < j$.

Bijection t between $\mathcal{S}_{\mathrm{d}}$ and $\mathcal{T}_{\mathrm{d}}$.

- Let $j(0) = 1$, let $j(1) < j(2) < \cdots < j(n_0 - 1)$ be the locations of the repeats of the sequence v, and let $j(l) = m + n_0 = n$.
- For $i = 1, \ldots, n_0$, let P_i be the path $(v_{j(i-1)}, \ldots, v_{j(i)-1}, m + i)$.
- Let $\mathrm{t}(\mathrm{v}) \in \mathcal{T}_{\mathrm{d}}$ have root v_1 and edge set given by the union of the edge sets of the paths $P_1, P_2, \ldots, P_{n_0}$.

The inverse of the bijection works as follows. Fix a tree $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$. Let $S_0 = \{r(\mathrm{t})\}$ consist of the root of t. Recursively, for $1 \leq i \leq n_0$ let P_i be the path from S_{i-1} to $m + i$ in t, and let P_i^* be P_i excluding its final point $m + i$. Then let $\mathrm{v} = \mathrm{v}(\mathrm{t})$ be the concatenation of $P_1^*, \ldots, P_{n_0}^*$. For later use, we observe that this bijection implies that

$$|\mathcal{T}_{\mathrm{d}}| = \binom{n-1}{d_1, \ldots, d_n} = \frac{(n-1)!}{\prod_{i \in [n]} d_i!}. \tag{2.2}$$

This formula (which appears as Theorem 5.3.4 in [97]) holds for all degree sequences, not just compressed ones, by the observation from earlier in the section.

We now discuss how to use the bijection to bound the height of $\mathrm{t}(\mathrm{v})$. For this, we think of the bijection as constructing t from $\mathrm{v}(\mathrm{t}) = (v_1, \ldots, v_{n-1})$ by adding vertices one-at-a-time, in order of their first appearance in the concatenation of $P_1, \ldots, P_{n_0}$. Formally, define a permutation $(w(1), \ldots, w(n)) = (w_{\mathrm{t}}(1), \ldots, w_{\mathrm{t}}(n))$ of $[n]$ as follows. For $1 \leq k \leq n$, let $w(k) = v_k$ if k is not the location of a repeat, and let $w(k) = m + r$ if $j(r) = k$ (so either $r < n_0$ and k is the location of the r'th repeat in v, or $r = n_0$ and $k = n$). Then let t_k be the subtree of t with vertices $w(1), \ldots, w(k)$. An example appears in Figure 2.2.

$$\mathrm{v} = (4, 4, 3, 1, 2, 1, 2)$$
$$w = (4, 5, 3, 1, 2, 6, 7, 8)$$

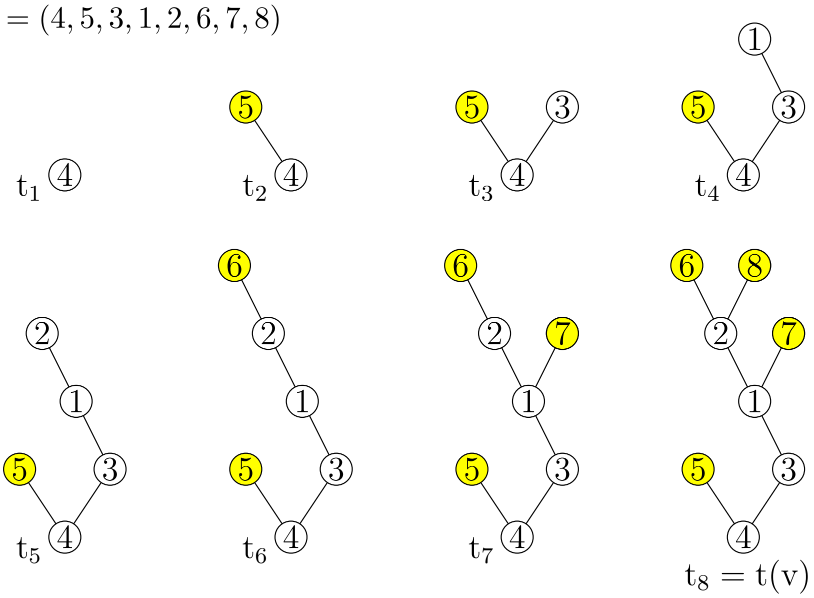


Fig. 2.2: This figure illustrates the bijection and the sequential construction. In this example, we have $\pi(2) = 5$ since t_5 is the first tree in the sequence containing vertex 2. We also have $\rho(2) = 4$, since 3 is the minimal k such that $\sum_{1 \le j \le k}(d_{i(j)} - 1)$ is at least 2 and t_4 is the first tree in the sequence to contain vertices $\{i(1), i(2), i(3)\} = \{4, 3, 1\}$.

Let $(\pi(i), i \in [n]) = (\pi_{\mathrm{t}}(i), i \in [n])$ be the inverse permutation of w, so $\pi(i) = w^{-1}(i)$ is the step at which the vertex with label i is added when constructing t from v(t). List the non-leaf vertices $\{1, \ldots, m\}$ of t in the order they appear when constructing t as $(i(1), \ldots, i(m)) = (i_{\mathrm{t}}(1), \ldots, i_{\mathrm{t}}(m))$. In other words, $(i(1), \ldots, i(m))$ is the permutation of $\{1, \ldots, m\}$ such that $(\pi(i(1)), \ldots, \pi(i(m)))$ is increasing. Then for $0 \le x \le n_0 - 1$, let $k = k(x) = k_{\mathrm{t}}(x)$ be minimal so that

$$\sum_{j=1}^{k}(d_{i(j)} - 1) \ge x,$$

and let $\rho(x) = \rho_{\mathrm{t}}(x) = \pi(i(k(x)))$, so that $v_{\rho(x)} = i(k(x))$. Equivalently, $\rho(x)$ is the smallest integer ρ such that the subtree of t consisting of the nodes of t_ρ and all their children has strictly more than $\lceil x \rceil$ leaves; this is also the smallest integer ρ such that t_ρ contains vertices $i(1), \ldots, i(k(x))$. Note that $\rho(x)$ is non-decreasing, and since

$$\sum_{j=1}^{m}(d_{i(j)} - 1) = (\sum_{j=1}^{m} d_j) - m = n - 1 - m = n_0 - 1\,,$$

the tree $\mathrm{t}_{\rho(n_0-1)}$ has n_0 leaves. It follows that all vertices of t that do not belong to $\mathrm{t}_{\rho(n_0-1)}$ have degree either 0 or 1 in t. This in particular implies that if t has no vertices of degree exactly 1 then $\rho(n_0-1)=\pi(i(m))$ and $\mathrm{t}_{\rho(n_0-1)}$ contains all vertices of t except a subset of its leaves, so $\mathrm{ht}(\mathrm{t})\le \mathrm{ht}(\mathrm{t}_{\rho(n_0-1)})+1$.

For any real sequence $0\le y_0<y_1<\ldots<y_N=n_0-1$, writing $\mathrm{t}=\mathrm{t}(\mathrm{v})$ and $\mathrm{t}_k=\mathrm{t}_k(\mathrm{v})$, we may now bound ht(t) via the telescoping sum

$$\mathrm{ht}(\mathrm{t})\le \mathrm{ht}(\mathrm{t}_{\rho(y_0)})+\Big(\sum_{i=1}^{N}\big(\mathrm{ht}(\mathrm{t}_{\rho(y_i)})-\mathrm{ht}(\mathrm{t}_{\rho(y_{i-1})})\big)\Big)+(\mathrm{ht}(\mathrm{t})-\mathrm{ht}(\mathrm{t}_{\rho(n_0-1)})), \tag{2.3}$$

and the final term in the sum is at most 1 if t has no vertices of degree 1. Here is the key lemma which makes such a decomposition useful. Given $\mathrm{t}\in\mathcal{T}_{\mathrm{d}}$, let $\mathrm{v}=\mathrm{v}(\mathrm{t})=(v_1,\ldots,v_{n-1})$.

Lemma 2.2.1. *Fix any degree sequence* $\mathrm{d}=(d_1,\ldots,d_n)$. *Then for all* $0\le x\le y\le n_0-1$, *if* $\mathrm{T}\in_u\mathcal{T}_{\mathrm{d}}$ *then*

$$\mathbb{P}(\mathrm{dist}(v_{\rho(y)},\mathrm{T}_{\rho(x)})\ >b)\le\left(1-\frac{x}{n-1}\right)^b$$

for all $b\ge 1$.

The proof of Lemma 2.2.1 appears in Section 2.2.1, below. This lemma is not quite enough to control an increment of the form $\mathrm{ht}(\mathrm{T}_{\rho(y)})-\mathrm{ht}(\mathrm{T}_{\rho(x)})$, since for that we need to control the distance from $v_{\rho_{\mathrm{t}}(z)}$ to $\mathrm{t}_{\rho(x)}$ for *all* $x\le z\le y$ – i.e. we need a "maximal inequality" version of Lemma 2.2.1. We achieve this in the following corollary. (Although in this section we defer the proofs of most of the supporting results, we prove the corollary immediately, as its proof is quite short.)

Corollary 2.2.2. *Fix any degree sequence* $\mathrm{d}=(d_1,\ldots,d_n)$ *with* $n_1(\mathrm{d})=0$, *and let* $\mathrm{T}\in_u\mathcal{T}_{\mathrm{d}}$. *Then for any integers* $0\le x\le y$ *and any* $p\in(0,1)$ *and* $b>0$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\rho(y)})-\mathrm{ht}(\mathrm{T}_{\rho(x)})>b+1)\le\frac{y-x}{(1-p)b}\left(1-\frac{x}{n-1}\right)^{\lfloor pb\rfloor}.$$

Proof. By relabeling the vertices, we may assume without loss of generality that d is compressed. Write $m=\max(i:d_i\neq 0)$.

Since T has no vertices of degree 1, for all $k\in[m]$ we have $d_{i(k)}\ge 2$ and so $\sum_{j=1}^{k}(d_{i(j)}-1)>\sum_{j=1}^{k-1}(d_{i(j)}-1)$. It follows that the non-leaf vertices of $\mathrm{T}_{\rho(y)}$ lying outside of $\mathrm{T}_{\rho(x)}$ are a subset of $\{i(k(x+1)),\ldots,i(k(y))\}=\{v_{\rho(x+1)},\ldots,v_{\rho(y)}\}$, and so

$$\mathrm{ht}(\mathrm{T}_{\rho(y)})-\mathrm{ht}(\mathrm{T}_{\rho(x)})\le 1+\max\big(\mathrm{dist}(v_{\rho(z)},\mathrm{T}_{\rho(x)}):z\in\{x+1,\ldots,y\}\big).$$

Now, if for a given $z \in \{x+1, \ldots, y\}$ we have

$$\mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)})) > b$$

then by considering the path in T from $v_{\rho(z)}$ to $\mathrm{T}_{\rho(x)}$, we see that there must be at least $(1-p)b$ distinct vertices $v \in \{v_{\rho(x+1)}, \ldots, v_{\rho(y)}\}$ for which $\mathrm{dist}(v, \mathrm{T}_{\rho(x)}) > pb$. It follows that

$$\begin{aligned}
&\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\rho(y)}) - \mathrm{ht}(\mathrm{T}_{\rho(x)}) > b+1) \\
&\le \mathbb{P}\Big(\max\big(\mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) : z \in \{x+1, \ldots, y\}\big) > b\Big) \\
&\le \mathbb{P}\Big(\big|\{z \in \{x+1, \ldots, y\} : \mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) > pb\}\big| > (1-p)b\Big) \\
&\le \frac{1}{(1-p)b}\mathbb{E}\big|\{z \in \{x+1, \ldots, y\} : \mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) > pb\}\big| \\
&= \frac{1}{(1-p)b}\sum_{z=x+1}^{y} \mathbb{P}\big(\mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) > pb\big),
\end{aligned}$$

from which the corollary follows by the bound from Lemma 2.2.1. □

Combining the telescoping sum bound (2.3) with Corollary 2.2.2 will allow us to prove tail bounds for the heights of random trees $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ for degree sequences d with $n_1(\mathrm{d}) = 0$. The next lemma allows us to transfer tail bounds from this setting to that of general trees with fixed degree sequences.

Lemma 2.2.3. *Fix a degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$ *and write* $n_0 = n_0(\mathrm{d})$ *and* $n_1 = n_1(\mathrm{d})$. *Let* d' *be obtained from* d *by removing all entries which equal* 1. *Let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ *and* $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$. *Then for any* $h \ge 2$ *and* $y \ge 8hn/(n-n_1)$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > y) \le \mathbb{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + n_0 \exp(-y/4).$$

Lemma 2.2.3, whose proof appears in Section 2.2.1, is the last tool we need to prove Theorem 2.1.7.

Proof of Theorem 2.1.7. Note that if $n \le 64$ then the result is trivially true, since $5e^{-\delta x^2/2^{13}} > 1$ for $x < 8$, and $\mathbb{P}(\mathrm{ht}(\mathrm{t}) > xn^{1/2}) = 0$ for $x \ge 8$. We thus hereafter assume that $n > 64$.

As described above, we shall bound $\mathbb{P}(\mathrm{ht}(\mathrm{T}) > xn^{1/2})$ by bounding the height via the telescoping sum (2.3), which for the random tree T becomes

$$\mathrm{ht}(\mathrm{T}) \le \mathrm{ht}(\mathrm{T}_{\rho(y_0)}) + \Big(\sum_{i=1}^{N} \big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})})\big)\Big) + (\mathrm{ht}(\mathrm{T}) - \mathrm{ht}(\mathrm{T}_{\rho(n_0-1)})), \tag{2.4}$$

for a suitably chosen sequence $0 \le y_0 < y_1 < \ldots < y_N = n_0 - 1$, and then controlling the contribution of each summand.

We first treat the case that d does not contain any entries equal to 1. In this case, the final term in (2.4) is at most 1, since when there are no vertices of degree one, all vertices of T not lying in $\mathrm{T}_{\rho(n_0-1)}$ are leaves. It follows that for any positive constants $b_0, \ldots, b_N$ such that $\sum_{i=0}^{N} b_i \le x$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > \lceil x n^{1/2} \rceil) \le \sum_{i=0}^{N} \mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big)\,, \tag{2.5}$$

where we use the convention that $\mathrm{ht}(\mathrm{T}_{\rho(y_{-1})}) = 0$. We choose $b_0 = x/2$ and $b_i = x \cdot i/2^{i+2}$, so that $\sum_{i\ge 0} b_i = x$ and $\sum_{i=0}^{N} b_i < x$ for all $N \in \mathbb{N}$.

Provided that $b_i n^{1/2} \ge 4$, Corollary 2.2.2 applied with $b = b_i n^{1/2} - 1$ and $p = (b-1)/2b$ yields

$$\begin{aligned}
\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) &\le \frac{y_i - y_{i-1}}{(1-p)b}\left(1 - \frac{y_{i-1}}{n-1}\right)^{\lfloor pb \rfloor} \\
&= \frac{2(y_i - y_{i-1})}{b+1}\left(1 - \frac{y_{i-1}}{n-1}\right)^{(b-1)/2} \\
&\le \frac{2(y_i - y_{i-1})}{b_i n^{1/2}}\left(1 - \frac{y_{i-1}}{n-1}\right)^{b_i n^{1/2}/4} .
\end{aligned}$$

Now set $y_i = \min(n_0 - 1, 2^{i-1} x n^{1/2})$. For $i \ge 1$, if $y_{i-1} = n_0 - 1$ then $\mathrm{T}_{\rho(y_i)} = \mathrm{T}_{\rho(y_{i-1})}$ so $\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}) = 0$. On the other hand, if $2^{i-1} x n^{1/2} \le n_0 - 1$, then $y_{i-1} = 2^{i-2} x n^{1/2} < n_0 - 1$. In this case we have $2(y_i - y_{i-1})/(b_i n^{1/2}) \le 2^{2i+1}/i$ and $1 - y_{i-1}/(n-1) \le e^{-2^{i-2}x/n^{1/2}}$, so if also $b_i n^{1/2} \ge 4$ then this gives

$$\begin{aligned}
\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) &\le \frac{2^{2i+1}}{i} \exp\left(-\frac{2^{i-2}x}{n^{1/2}} \frac{b_i n^{1/2}}{4}\right) \\
&= \frac{2^{2i+1}}{i} \exp\left(-\frac{i x^2}{2^6}\right) .
\end{aligned} \tag{2.6}$$

But if $i \ge 2$ is small enough that $2^{i-1} x n^{1/2} \le n_0 - 1$, then since $n_0 < n$, we also have

$n^{1/2}b_i = n^{1/2}xi/2^{i+2} \geq x^2 i/8$ and so $b_i n^{1/2} \geq 4$ provided $x \geq 4$. If $i = 1$ then we also have $n^{1/2}b_i = n^{1/2}x/8 \geq 4$ when $x \geq 4$, since we assumed that $n \geq 64$. It follows that (2.6) holds for all $i \geq 1$.

To handle the $i = 0$ term, note that by the definition of $\rho(y_0)$, the subtree of T consisting of all nodes of $\mathrm{T}_{\rho(y_0)-1}$ and their children contains at most $\lceil y_0 \rceil$ leaves. Since there are no vertices of degree exactly 1, this also implies that $\mathrm{T}_{\rho(y_0)-1}$ has fewer than $\lceil y_0 \rceil$ non-leaf vertices. Therefore, $\mathrm{ht}(\mathrm{T}_{\rho(y_0)-1}) \leq \lceil y_0 \rceil - 1$, so $\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) \leq \lceil y_0 \rceil \leq b_0 n^{1/2}$ and thus

$$\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{-1})}) > b_0 n^{1/2}\big) = \mathbb{P}(\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) > b_0 n^{1/2}) = 0\,.$$

Using this bound and (2.6) in the telescoping sum bound (2.5), with $N = \min(i : 2^i x n^{1/2} \geq n_0 - 1) = \min(i : y_i = n_0 - 1)$, we obtain that for all $x \geq 4$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2} \rceil) \leq \sum_{i \geq 1} \frac{2^{2i+1}}{i} \exp\left(-\frac{ix^2}{2^6}\right) = \sum_{i \geq 1} \exp\left((2i+1)\log 2 - \log i - \frac{ix^2}{2^6}\right)\,.$$

Provided that $x^2/2^6 > 4 \log 2$, the exponent in the sum is less than $\log 2 - ix^2/2^7 - \log i$, so for such x the probability we aim to bound is at most

$$\sum_{i \geq 1} 2\exp(-ix^2/2^7) = 2e^{-x^2/2^7}\big(1 - e^{-x^2/2^7}\big)^{-1} < 4e^{-x^2/2^7}\,.$$

On the other hand, if $x^2/2^6 \leq 4 \log 2$ then $4e^{-x^2/2^7} \geq 4e^{-2\log 2} = 1$, which is an upper bound on any probability. It follows that for all $x > 0$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2} \rceil) \leq 4e^{-x^2/2^7}\,. \tag{2.7}$$

This proves (in fact, is stronger than) the necessary bound when T has no vertices of degree 1.

Now suppose that $n_1 > 0$ and write $\delta = (n - n_1(\mathrm{d}))/n$; so $\delta < 1$. By permuting the vertex labels we may assume that all the ones in d are at the end, i.e., $d_i \neq 1$ for $i \leq n - n_1$ and $d_i = 1$ for $n - n_1 < i \leq n$. Let $\mathrm{d}' = (d_1, \ldots, d_{n-n_1})$ be obtained by removing all entries equal to 1 from d, and let $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$. By Lemma 2.2.3, for any $h \geq 2$ and any $y \geq 8hn/(n - n_1)$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > y) \leq \mathbb{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + n_0 e^{-y/4} \leq \mathbb{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + ne^{-y/4}\,.$$

Take $y = xn^{1/2}$ and $h = \lceil y(n-n_1)/(8n)\rceil = \lceil x'(n-n_1)^{1/2}\rceil$, where we have set $x' = x(n-n_1)^{1/2}/(8n^{1/2})$. Then, if $h \geq 2$, the above bound and (2.7) together yield

$$\begin{aligned}\mathbb{P}(\mathrm{ht}(\mathrm{T}) > xn^{1/2}) &\leq \mathbb{P}(\mathrm{ht}(\mathrm{T}') > \lceil x'(n-n_1)^{1/2}\rceil) + ne^{-xn^{1/2}/4}\\ &\leq 4e^{-(x')^2/2^7} + ne^{-xn^{1/2}/4}\\ &= 4e^{-\delta x^2/2^{13}} + ne^{-xn^{1/2}/4}\\ &\leq 5e^{-\delta x^2/2^{13}},\end{aligned}$$

where the last inequality holds for $8 \leq x \leq n^{1/2}$, as can be straightforwardly checked. But for $x < 8$ we have $5e^{-\delta x^2/2^{13}} > 1$, and for $x > n^{1/2}$ we have $\mathbb{P}(\mathrm{ht}(\mathrm{T}) > xn^{1/2}) = 0$, so this bound holds in those cases as well. If $h < 2$, then $x' < 2$ as well, so the first term in the second line exceeds 1 and the inequality holds trivially. □

In the preceding proof, the first term in the telescoping sum could be essentially "given away" - we just used the deterministic bound $\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) \leq \lceil y_0\rceil$. In the setting of Theorem 2.1.8, where we aim for stronger bounds when σ_{d} is large than when it is small, we can not be so casual about this first term. The additional tool we require, to bound the height of the random tree constructed during the early stages of the bijection in the case when σ_{d} is large, is given in the following proposition.

Proposition 2.2.4. *Fix a degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$ *and let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$.

Then taking $\alpha = (1-e^{-2})/24$, *with* $\sigma = \sigma_{\mathrm{d}}$, *we have that for all* $b \geq 1$,

$$\mathbb{P}\left(\mathrm{ht}\left(\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}\right) > \tfrac{bn^{1/2}}{2\sigma}\right) \leq \exp\left(-\tfrac{3}{32}\tfrac{bn^{1/2}}{\sigma}\right) + \exp\left(-\tfrac{\alpha b}{2}\right).$$

The proof of Proposition 2.2.4 appears in Section 2.2.1. With this proposition in hand, the proof of Theorem 2.1.8 proceeds quite similarly to that of Theorem 2.1.7 (though it is necessarily somewhat more involved as the choices of the values y_i and b_i for the telescoping sum bound must depend on σ_{d}).

Proof of Theorem 2.1.8. In the proof we write $n_i = n_i(\mathrm{d})$ and $\sigma = \sigma_{\mathrm{d}}$ for succinctness. We first assume that $n_1 = 0$, so T contains no degree-1 vertices. In the proof of Theorem 2.1.7 we showed that for such a tree, $\mathbb{P}(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2}\rceil) \leq 4e^{-x^2/2^7}$; see inequality (2.7). If $\sigma < e$ then this implies that for $x \geq 2^{12}$,

$$\mathbb{P}\left(\mathrm{ht}(T) \geq \left\lceil x n^{1/2} \frac{\log(1+\sigma)}{\sigma} \right\rceil\right) \leq 4\exp\left(-\frac{(x\log(1+\sigma))^2}{2^7\sigma^2}\right)$$
$$< 4\exp\left(-\frac{2^5\log(1+\sigma)}{\sigma^2}(x\log(1+\sigma))\right) < 4e^{-x\log(1+\sigma)}\,.$$

Continuing to assume that $n_1 = 0$, we now turn our attention to the case that $\sigma \geq e$. Like in the proof of Theorem 2.1.7, we begin by using the telescoping sum inequality (2.3), and obtain the bound

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > \lceil x n^{1/2} \tfrac{\log\sigma}{\sigma}\rceil) \leq \sum_{i=0}^{N} \mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\tfrac{\log\sigma}{\sigma}\big)\,, \tag{2.8}$$

which holds for any sequence $0 \leq y_0 < y_1 < \ldots < y_N = n_0 - 1$ and positive constants $b_0, \ldots, b_N$ with $\sum_{i=0}^N b_i \leq x$, and then controlling the contribution of each summand. (We again take $\mathrm{ht}(\mathrm{T}_{\rho(y_{-1})}) := 0$ for convenience.) Also as in the proof of Theorem 2.1.7, we take $b_0 = x/2$ and $b_i = x \cdot i/2^{i+2}$, so that $\sum_{i\geq 0} b_i = x$ and $\sum_{i=0}^N b_i < x$ for all $N \in \mathbb{N}$.

For $x \geq 2^{12}$, since $\sigma \geq e$ we have $(b_0 n^{1/2}\log\sigma)/\sigma \geq n^{1/2}/(2\sigma)$, so taking $y_0 = \sigma n^{1/2}(1 - e^{-2})/24$, by Corollary 2.2.4 applied with $b = x\log\sigma$ we obtain

$$\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_0)} > b_0 n^{1/2}\tfrac{\log\sigma}{\sigma}\big) < \exp\left(-\tfrac{3}{32}x n^{1/2}\tfrac{\log\sigma}{\sigma}\right) + \exp\left(-\tfrac{1-e^{-2}}{48}x\log\sigma\right) \leq 2e^{-(x\log\sigma)/64}\,, \tag{2.9}$$

the last inequality holding since $\sigma \leq n^{1/2}$ and $\min(3/32, (1-e^{-2})/48) > 1/64$.

For the remaining summands, we take $y_i = \min(n_0 - 1, 2^i y_0) = \min(n_0 - 1, 2^i\sigma n^{1/2}(1 - e^{-2})/24)$. Provided that $b_i n^{1/2}\frac{\log\sigma}{\sigma} \geq 4$, then Proposition 2.2.2 applied with $b = b_i n^{1/2}\frac{\log\sigma}{\sigma} - 1$ and $p = (b-1)/2b$ yields

$$\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\tfrac{\log\sigma}{\sigma}\big) \leq \frac{y_i - y_{i-1}}{(1-p)b}\left(1 - \frac{y_{i-1}}{n-1}\right)^{\lfloor pb\rfloor}$$
$$= \frac{2(y_i - y_{i-1})}{b+1}\left(1 - \frac{y_{i-1}}{n-1}\right)^{(b-1)/2}$$
$$\leq \frac{2(y_i - y_{i-1})}{b_i n^{1/2}}\frac{\sigma}{\log\sigma}\left(1 - \frac{y_{i-1}}{n-1}\right)^{b_i n^{1/2}(\log\sigma)/(4\sigma)}\,.$$

To simplify this upper bound, note that

$$\frac{2(y_i - y_{i-1})}{b_i n^{1/2}}\frac{\sigma}{\log\sigma} \leq \frac{2^{2i+2}(1-e^{-2})}{24ix}\frac{\sigma^2}{\log\sigma} < \frac{2^{2i-2}\sigma^2}{ix\log\sigma}\,,$$

and provided $y_{i-1} < n_0 - 1$ we also have $1 - y_{i-1}/(n-1) \le e^{-2^i\sigma(1-e^{-2})/(24n^{1/2})}$, so

$$\left(1 - \frac{y_{i-1}}{n-1}\right)^{b_i n^{1/2}(\log\sigma)/(4\sigma)} \le \exp\left(-\frac{2^i\sigma(1-e^{-2})}{24n^{1/2}}\frac{b_i n^{1/2}\log\sigma}{4\sigma}\right) < \exp(-ix\log\sigma/2^9)\,.$$

Combining the three preceding bounds, it follows that when $b_i n^{1/2}\frac{\log\sigma}{\sigma} \ge 4$ we have

$$\begin{aligned}\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\tfrac{\log\sigma}{\sigma}\big) &\le \frac{2^{2i-2}\sigma^2}{ix\log\sigma}\exp\left(-\frac{ix\log\sigma}{2^9}\right)\\ &= \frac{\sigma^2}{4x\log\sigma}\exp\left((2\log 2)i - \log i - \frac{ix\log\sigma}{2^9}\right)\,.\end{aligned}$$

(If $y_{i-1} \ge n_0 - 1$ then $\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}) = 0$ so the bound holds in this case as well.) By the assumption that $x \ge 2^{12}$ and since $\sigma \ge e$, we have $(x\log\sigma)/2^9 > 8 > 2(2\log 2)$, so this yields

$$\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\tfrac{\log\sigma}{\sigma}\big) < \frac{\sigma^2}{2^{14}}\exp\left(-\frac{ix\log\sigma}{2^{10}}\right)\,.$$

For the above bound we needed that $b_i n^{1/2}\frac{\log\sigma}{\sigma} \ge 4$, or in other words that $ixn^{1/2}\log\sigma/(2^{i+4}\sigma) \ge 1$. But since $n_0 - 1 < n$, the condition $y_{i-1} < n_0 - 1$ implies that

$$\frac{2^{i-1}\sigma n^{1/2}(1-e^{-2})}{24} < n\,,$$

so $ix\log\sigma n^{1/2}/(2^{i+4}\sigma) \ge 1$ provided that $x\log\sigma > 768/(1-e^{-2})$, which holds since $x \ge 2^{12}$ and $\sigma \ge e$. It follows that

$$\begin{aligned}\sum_{i\ge 1}\mathbb{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\tfrac{\log\sigma}{\sigma}\big) &\le \frac{\sigma^2}{2^{14}}\exp\left(-\frac{x\log\sigma}{2^{10}}\right)\left(1 - e^{-(x\log\sigma)/2^{10}}\right)^{-1}\\ &\le \frac{1}{2^{13}}\exp\left(-\frac{x\log\sigma}{2^{11}}\right)\,,\end{aligned}$$

the last bound holding since $x\log\sigma \ge x \ge 2^{12}$ and thus $\sigma^2 e^{-(x\log\sigma)/2^{10}} \le e^{-(x\log\sigma)/2^{11}}$ and $(1 - e^{-(x\log\sigma)/2^{10}})^{-1} < 2$.

Using this bound and (2.9) in (2.8), we thus obtain

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2}\tfrac{\log\sigma}{\sigma}\rceil) < 3\exp\left(-\frac{x\log\sigma}{2^{11}}\right)\,.$$

Combining this with the bound from the start of the proof, which handles the case $\sigma < e$, it follows that when there are no vertices of degree one, for all $x \geq 2^{12}$ we have

$$\mathbb{P}\big(\mathrm{ht}(\mathrm{T}) > \lceil x n^{1/2} \tfrac{\log(\sigma+1)}{\sigma} \rceil\big) \leq 4\exp(-x\log(\sigma+1)/2^{11}). \tag{2.10}$$

This proves the theorem (in fact, something slightly stronger) when T has no vertices of degree one.

Now suppose that d contains vertices of degree 1. Like in the proof of Theorem 2.1.7, we let d' be obtained from d by removing all entries equal to 1. Let $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and let $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$. Write $\sigma = \sigma_{\mathrm{d}}$. Also write $n' = n - n_1$ for the number of vertices of T' so that we have that $\sigma_{\mathrm{d}'}^2 = \frac{n}{n'}\sigma^2 = (\sigma')^2$.

By Lemma 2.2.3, for any $h \geq 2$ and any $y \geq 8hn/n'$,

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}) > y\right) \leq \mathbb{P}\left(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil\right) + n_0 e^{-y/4} \leq \mathbb{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + n e^{-y/4}.$$

We take

$$y = x n^{1/2} \frac{\log(\sigma'+1)}{\sigma}, \qquad h = \frac{y}{8}\frac{n'}{n} = \frac{x}{8}(n')^{1/2}\frac{\log(\sigma'+1)}{\sigma'}.$$

Then, if $x/8 \geq 2^{12}$, (2.10) gives that

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil\right) \leq 4\exp\left(-x\log(\sigma'+1)/2^{14}\right).$$

Also, since $\sigma' \leq (n')^{1/2} \leq n^{1/2}$ and $\sigma \leq \sigma'$, we have $n^{1/2}\frac{\log(\sigma'+1)}{\sigma} \geq (\log n)/2$, so $y/4 - \log n \geq y/8$ if $x \geq 16$, but this constraint is weaker than $x/8 \geq 2^{12}$. It follows that for $x/8 \geq 2^{12}$,

$$n e^{-y/4} \leq e^{-y/8} \leq \exp\left(-\frac{x}{8}\log(\sigma'+1)\right),$$

so the above bounds together yield that for all $x \geq 2^{15}$,

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}) > x n^{1/2}\frac{\log(\sigma'+1)}{\sigma}\right) \leq 4\exp\left(-x\log(\sigma'+1)/2^{14}\right). \qquad \square$$

2.2.1 The postponed proofs

This section contains the proofs of the results that were stated without proof earlier in Section 2.2: namely, Lemmas 2.2.1 and 2.2.3 and Proposition 2.2.4.

Proof of Lemma 2.2.1. It is useful to assume d is compressed (and in particular that $d_n = 0$).

Write $\mathrm{par}^1_{\mathrm{t}}(u) = \mathrm{par}_{\mathrm{t}}(u)$ for the parent of vertex u in tree t; we define the parent of the root to be the root itself. Inductively set $\mathrm{par}^{b+1}_{\mathrm{t}}(u) = \mathrm{par}_{\mathrm{t}}(\mathrm{par}^b_{\mathrm{t}}(u))$ for $b \geq 1$. Throughout the proof, we take $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and let $\mathrm{V} = \mathrm{v}(\mathrm{T})$, so $\mathrm{V} \in_u \mathcal{S}_{\mathrm{d}}$.

Let $j \leq m$ and fix any vector $(i(1), \ldots, i(j))$ of distinct elements of $[m]$ with $\sum_{k=1}^{j}(d_{i(k)}-1) \geq y$, and let $\mathcal{S}_{\mathrm{d}}(i(1), \ldots, i(j))$ be the set of vectors $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}$ such that, writing $\mathrm{t} = \mathrm{t}(\mathrm{v})$,

$$(i_{\mathrm{t}}(1), \ldots, i_{\mathrm{t}}(j)) = (i(1), \ldots, i(j)).$$

Let $r \in [j]$ be such that

$$\sum_{k=1}^{r-1}(d_{i(k)} - 1) < y \leq \sum_{k=1}^{r}(d_{i(k)} - 1),$$

Note that if $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}(i(1), \ldots, i(j))$ and $\mathrm{t} = \mathrm{t}(\mathrm{v})$ then $v_{\rho_{\mathrm{t}}(y)} = i(r) = i_{\mathrm{t}}(r)$. This allows us to rewrite

$$\begin{aligned}
&\mathbb{P}\left(D_{\mathrm{T}}(y, x) > k \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j))\right) \\
&= \mathbb{P}\left(\mathrm{par}^k_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho_{\mathrm{T}}(x)} \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j))\right).
\end{aligned}$$

For $\mathrm{V} \in_u \mathcal{S}_{\mathrm{d}}$, the ordering of the integers $\{1, \ldots, n\}$ by their first appearance in V is degree-biased, so

$$\mathbb{P}(i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)) = \prod_{k=1}^{j} \frac{d_{i(k)}}{n - 1 - d_{i(1)} - \ldots - d_{i(k-1)}},$$

or equivalently

$$|\mathcal{S}_{\mathrm{d}}(i(1), \ldots, i(j))| = \binom{n-1}{d_1, \ldots, d_n} \cdot \prod_{k=1}^{j} \frac{d_{i(k)}}{n - 1 - d_{i(1)} - \ldots - d_{i(k-1)}}.$$

Fix $\ell < r$ and let $\mathrm{d}^{\ell} = (d^{\ell}_1, \ldots, d^{\ell}_{n-1})$ where $d^{\ell}_{i(\ell)} = d_{i(\ell)} - 1$ and $d^{\ell}_i = d_i$ for all $i \in [n-1] \setminus \{i(\ell)\}$. Since $d_n = 0$ we have $\sum_{i \in [n-1]} d^{\ell}_i = n - 2$, so d^{ℓ} is another degree sequence. Now consider the subset $\mathcal{S}^{\ell}_{\mathrm{d}}(i(1), \ldots, i(j))$ of $\mathcal{S}_{\mathrm{d}}(i(1), \ldots, i(j))$ consisting of those vectors v where the first instance of $i(r)$ in v is the immediate successor of some instance of $i(\ell)$ other than the first. The set $\mathcal{S}^{\ell}_{\mathrm{d}}(i(1), \ldots, i(j))$ is in bijection with the set of vectors $\mathrm{v}' \in \mathcal{S}_{\mathrm{d}^{\ell}}(i(1), \ldots, i(j))$, i.e., the set of vectors $\mathrm{v}' \in \mathcal{S}_{\mathrm{d}^\ell}$ such that

$$(i(1,\mathrm{v}'),\ldots,i(j,\mathrm{v}')) = (i(1),\ldots,i(j)).$$

To see this, fix $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}^{\ell}(i(1),\ldots,i(j))$, and let v' be the vector obtained from v by deleting the entry with value $i(\ell)$ immediately preceding the first instance of $i(r)$. Then $(i(1,\mathrm{v}),\ldots,i(j,\mathrm{v})) = (i(1,\mathrm{v}'),\ldots,i(j,\mathrm{v}'))$, since the deleted entry was not the first instance of $i(\ell)$ in v, so v' is an element of $\mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))$. To recover v from v', simply insert an entry with value $i(\ell)$ immediately before the first instance of $i(r)$ in v'.

The same computation as for the size of $\mathcal{S}_{\mathrm{d}}(i(1),\ldots,i(j))$ now yields the formula

$$|\mathcal{S}_{\mathrm{d}}^{\ell}(i(1),\ldots,i(j))| = |\mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))| = \binom{n-2}{d_1^\ell,\ldots,d_n^\ell}\prod_{k=1}^{j}\frac{d_{i(k)}^\ell}{n-2-d_{i(1)}^\ell-\ldots-d_{i(k-1)}^\ell}.$$

Writing $E(r,\ell)$ for the event that the first instance of $i(r,\mathrm{V})$ in V is an immediate successor of some instance of $i(\ell,\mathrm{V})$ other than the first, it follows that

$$\begin{aligned}
&\mathbb{P}\left(E(r,\ell) \mid (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j)) = (i(1),\ldots,i(j))\right)\\
&= \frac{|\mathcal{S}_{\mathrm{d}}^{\ell}(i(1),\ldots,i(j))|}{|\mathcal{S}_{\mathrm{d}}(i(1),\ldots,i(j))|}\\
&= \frac{d_{i(\ell)}}{n-1}\cdot\prod_{k=1}^{j}\frac{d_{i(k)}^\ell}{d_{i(k)}}\cdot\prod_{k=1}^{j}\frac{n-1-d_{i(1)}-\ldots-d_{i(k-1)}}{n-2-d_{i(1)}^\ell-\ldots-d_{i(k-1)}^\ell}\\
&= \frac{d_{i(\ell)}-1}{n-1}\cdot\prod_{k=1}^{j}\frac{n-1-d_{i(1)}-\ldots-d_{i(k-1)}}{n-2-d_{i(1)}^\ell-\ldots-d_{i(k-1)}^\ell}\\
&> \frac{d_{i(\ell)}-1}{n-1}.
\end{aligned}$$

Now note that if $E(r,\ell)$ occurs then $i_{\mathrm{T}}(\ell)$ is the parent of $i_{\mathrm{T}}(r)$ in $\mathrm{T} = \mathrm{t}(\mathrm{V})$. Moreover, letting $q \in [j]$ be such that

$$\sum_{k=1}^{q-1}(d_{i(k)}-1) < x \le \sum_{k=1}^{q}(d_{i(k)}-1),$$

then on the event $\{(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j)) = (i(1),\ldots,i(j))\}$, the tree $\mathrm{T}_{\rho(x)}$ has vertices $i(1),\ldots,i(q)$.

It follows that

$$
\begin{aligned}
&\mathbb{P}\left(\mathrm{par}_{\mathrm{T}}(v_{\rho_{\mathrm{T}}(y)}) \in \mathrm{t}_{\rho(x)}(\mathrm{V}) \ \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j))\right) \\
&\mathbb{P}\left(\mathrm{par}_{\mathrm{T}}(i_{\mathrm{T}}(r)) \in \mathrm{t}_{\rho(x)}(\mathrm{V}) \ \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j))\right) \\
&\geq \sum_{\ell=1}^{q} \mathbb{P}\left(E(r, \ell) \ \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)\right) \\
&> \sum_{\ell=1}^{q} \frac{d_{i(\ell)} - 1}{n-1} \geq \frac{x}{n-1}\,.
\end{aligned}
$$

This proves the case $b = 1$ of the lemma by summing over $(i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j))$. In fact, we have proved something slightly stronger: in the case $b = 1$, the necessary bound holds even conditionally given the values of $(i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j))$. From this we will prove the full lemma by arguing inductively, by conditioning on the label of the parent of $i_{\mathrm{T}}(r)$.

There are two cases to consider. In either case, we continue to take $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{V} = \mathrm{v}(\mathrm{T})$, and fix a sequence $(i(1), \ldots, i(j))$ of distinct elements of $[m]$. For the first case, fix $\ell \in \{q + 1, \ldots, r-1\}$, let $W \in_u \mathcal{S}_{\mathrm{d}^\ell}$, where d^ℓ is as defined above, and suppose that $(i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j))$ and that $E(r, \ell)$ occurs. Under this conditioning, we have $\mathrm{V} \in_u \mathcal{S}_{\mathrm{d}}^{\ell}(i(1), \ldots, i(j))$, and in particular the parent of $i(r) = i_{\mathrm{T}}(r)$ in T is $i(\ell)$. Now let V' be obtained from V by deleting the entry immediately preceding the first instance of $i(r)$ in V and let $\mathrm{T}' = \mathrm{t}(\mathrm{V}')$. Under this conditioning, $\mathrm{V}' \in_u \mathcal{S}_{\mathrm{d}^\ell}(i(1), \ldots, i(j))$, which in particular implies that $(i_{\mathrm{T}'}(1), \ldots, i_{\mathrm{T}'}(j)) = (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j))$. Moreover, the sequences V and V' agree until after the first instance of $i(\ell)$, so the ancestral line of $i(\ell)$ is the same in both T and T'. Since $q < \ell$, it likewise holds that $\rho_{\mathrm{T}}(x) = \rho_{\mathrm{T}'}(x)$ and, writing $\rho(x)$ for their common value, that $\mathrm{T}_{\rho(x)} = \mathrm{T}'_{\rho(x)}$. It follows that

$$
\begin{aligned}
&\mathbb{P}\left(\mathrm{par}_{\mathrm{T}}^{b}(i(r)) \notin \mathrm{T}_{\rho(x)} \ \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)), E(r, \ell)\right) \\
&= \mathbb{P}\left(\mathrm{par}_{\mathrm{T}'}^{b-1}(i(\ell)) \notin \mathrm{T}'_{\rho(x)} \ \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)), E(r, \ell)\right) \\
&= \mathbb{P}\left(\mathrm{par}_{\mathrm{t}(\mathrm{W})}^{b-1}(i(\ell)) \notin \mathrm{t}(\mathrm{W})_{\rho(x)} \ \Big| \ (i_{\mathrm{t}(\mathrm{W})}(1), \ldots, i_{\mathrm{t}(\mathrm{W})}(j)) = (i(1), \ldots, i(j))\right) \\
&\leq \left(1 - \frac{x}{n-2}\right)^{b-1}\,.
\end{aligned}
\tag{2.11}
$$

The second equality holds since under the second conditioning $\mathrm{V}' \in_u \mathcal{S}_{\mathrm{d}^\ell}(i(1), \ldots, i(j))$ and under the third conditioning $\mathrm{W} \in_u \mathcal{S}_{\mathrm{d}^\ell}(i(1), \ldots, i(j))$ and $\rho_{\mathrm{t}(\mathrm{W})}(x) = \rho_{\mathrm{T}}(x)$. The last inequality holds by induction (on b or n or both) and since d^ℓ has length $n - 1$.

For the second case, consider the subset $\mathcal{S}_{\mathrm{d}}^{*}$ of $\mathcal{S}_{\mathrm{d}}(i(1), \ldots, i(j))$ consisting of those vectors v where the first instance of $i(r)$ in v is the immediate successor of the first instance of $i(r-1)$. Let $\mathrm{d}^* = (d_1^*, \ldots, d_{n-1}^*)$, where

$$d_{i(k)}^* = \begin{cases} d_{i(k)} & \text{if } k < r-1 \\ d_{i(r-1)} + d_{i(r)} - 1 & \text{if } k = r-1 \\ d_{i(k+1)} & \text{if } r \le k < n\,, \end{cases}$$

and let $(i^*(1), \ldots, i^*(j-1)) = (i(1), \ldots, i(r-1), i(r+1), \ldots, i(j-1))$. From a vector $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}^*$, we can obtain a vector $\mathrm{v}' \in \mathcal{S}_{\mathrm{d}^*}(i^*(1), \ldots, i^*(j-1))$ by deleting the first instance of $i(r-1)$, then changing all other instances of $i(r)$ to $i(r-1)$. Conversely, given a vector $\mathrm{v}' \in \mathcal{S}_{\mathrm{d}^*}(i^*(1), \ldots, i^*(j-1))$, there are $\binom{d_{i(r-1)}+d_{i(r)}-2}{d_{i(r)}-1}$ ways to reconstruct a vector $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}^*$: first insert an entry with value $i(r)$ just after the first instance of $i(r-1)$, then replace $d_{i(r)} - 1$ of the other instances of $i(r-1)$ by $i(r)$'s. It follows that for $\mathrm{V} \in_u \mathcal{S}_{\mathrm{d}}$ as above, conditionally given that $\mathrm{V} \in \mathcal{S}_{\mathrm{d}}^*$, then $\mathrm{V}' \in_u \mathcal{S}_{\mathrm{d}^*}(i^*(1), \ldots, i^*(j-1))$. Moreover, the ancestral lines of $i(r-1)$ are the same in both $\mathrm{t}(\mathrm{V})$ and $\mathrm{t}(\mathrm{V}')$, and if $\mathrm{par}_{\mathrm{t}(\mathrm{V})}(i(r)) \notin \mathrm{t}_{\rho(x)}(\mathrm{V})$ then we must have $q < r-1$, so also $\rho_{\mathrm{t}(\mathrm{V})}(x) = \rho_{\mathrm{t}(\mathrm{V}')}(x) =: \rho(x)$, and $\mathrm{t}_{\rho(x)}(\mathrm{V}) = \mathrm{t}_{\rho(x)}(\mathrm{V}')$.

Let $F(r)$ be the event that the first instance of $i_{\mathrm{T}}(r)$ is an immediate successor of the first instance of $i_{\mathrm{T}}(r-1)$, and let $W \in_u \mathcal{S}_{\mathrm{d}^*}$. By the conclusions of the preceding paragraph, it follows that

$$\begin{aligned} &\mathbb{P}\left(\mathrm{par}_{\mathrm{T}}^b(i(r)) \notin \mathrm{T}_{\rho(x)} \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)), F(r)\right) \\ &= \mathbb{P}\left(\mathrm{par}_{\mathrm{T}'}^b(i(r)) \notin \mathrm{T}'_{\rho(x)} \mid (i_{\mathrm{T}'}(1), \ldots, i_{\mathrm{T}'}(j)) = (i(1), \ldots, i(j)), F(r)\right) \\ &= \mathbb{P}\left(\mathrm{par}_{\mathrm{t}(\mathrm{W})}^{b-1}(i^*(r-1)) \notin \mathrm{t}(\mathrm{W})_{\rho(x)} \;\Big|\; (i_{\mathrm{t}(\mathrm{W})}(1), \ldots, i_{\mathrm{t}(\mathrm{W})}(j-1) = (i^*(1), \ldots, i^*(j-1))\right) \\ &\le \left(1 - \frac{x}{n-2}\right)^{b-1}, \end{aligned} \tag{2.12}$$

the second equality holding since $i^*(r-1) = i(r-1)$ and $\rho_{\mathrm{t}(\mathrm{W})}(x) = \rho(x)$, and the final bound again holding by induction.

For $V \in_u \mathcal{S}_{\mathrm{d}}$, on the event that $(i(1,\mathrm{V}), \ldots, i(j,\mathrm{V})) = (i(1), \ldots, i(j))$, if $\mathrm{par}_{\mathrm{t}(\mathrm{V})}(i(r,\mathrm{V})) \notin \mathrm{t}_{\rho(x)}(\mathrm{V})$ then either $F(r)$ occurs or else $E(r,\ell)$ occurs for some $\ell \in \{q+1, \ldots, r-1\}$, so (2.11) and (2.12) together imply that

$$\begin{aligned}&\mathbb{P}\left(\mathrm{par}^b_{\mathrm{T}}(v_{\rho_{\mathrm{T}}(y)}) \notin \mathrm{T}_{\rho(x)} \;\middle|\; (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j)) = (i(1),\ldots,i(j)), \mathrm{par}_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)}\right)\\ &\mathbb{P}\left(\mathrm{par}^b_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)} \;\middle|\; (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j)) = (i(1),\ldots,i(j)), \mathrm{par}_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)}\right)\\ &\le \left(1-\frac{x}{n-2}\right)^{b-1}.\end{aligned}$$

Summing over possible values for $(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))$ yields that

$$\mathbb{P}\left(\mathrm{par}^b_{\mathrm{T}}(v_{\rho_{\mathrm{T}}(y)}) \notin \mathrm{T}_{\rho(x)} \;\middle|\; \mathrm{par}_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)}\right) \le \left(1-\frac{x}{n-2}\right)^{b-1},$$

which combined with the bound for the case $b=1$ completes the proof. □

Proof of Lemma 2.2.3. The bound is obvious for $n_1 = n-1$ so we assume $n_1 < n-1$. We begin with some fairly elementary combinatorics. Given a set $S = \{s_1,\ldots,s_m\}$, by a *labeled composition of S with k parts* we mean a weak composition $m_1,\ldots,m_k$ of m together with a permutation $(s_{\sigma(1)},\ldots,s_{\sigma(m)})$ of S. There are $(m+k-1)!/(k-1)!$ labeled compositions of S with k parts.

We now return to the setting of the lemma. By permuting the entries of d, we may assume that the n_1 entries which equal 1 appear at the end; this does not affect the law of the height of T. In this case, we have $\mathrm{d}' = (d_1,\ldots,d_{n-n_1})$.

Next, given a tree $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$, let $\mathrm{t}^- \in \mathcal{T}_{\mathrm{d}'}$ be obtained from t by suppressing all degree-one vertices. More precisely, t^- is formed from t by replacing any path γ in t all of whose internal vertices have exactly one child by a single edge. If this results in the root having degree 1, we remove the root and its adjacent edge and let its only child be the new root; see Figure 2.3.

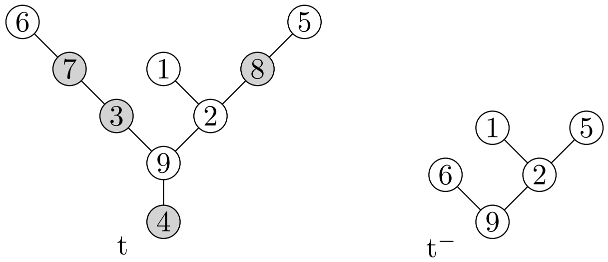


Fig. 2.3: Left: a tree t. Right: the tree t^- obtained from t by suppressing degree-one vertices. Considering the vertices in the order $(1,2,5,6,9)$, the corresponding labeled composition of the set $\{3,4,7,8\}$ of degree-one vertices is $(),(),(8),(7,3),(4)$.

For each tree $\mathrm{t}' \in \mathcal{T}_{\mathrm{d}'}$, there are

$$\frac{(n-1)!}{(n-n_1-1)!}$$

trees $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$ with $\mathrm{t}^- = \mathrm{t}'$. This is the case since each such tree t is uniquely determined by t' together with a labeled composition of the n_1 degree-one vertices $\{n-n_1+1,\ldots,n\}$ into $n-n_1$ parts, as follows (see Figure 2.3). Fix an ordering of the vertices of t' as $v_1,\ldots,v_{n-n_1}$. Then for each $i \in [n-n_1]$, assign the vertices of the i'th part of the composition to v_i's ancestral edge, in the same order they appear in the composition. (When v_i is the root, these vertices are simply attached as ancestors of v_i.)

It follows that if $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ then $\mathrm{T}^- \in_u \mathcal{T}_{\mathrm{d}'}$. Moreover, given that $\mathrm{T}^- = \mathrm{t}'$, the conditional distribution of T may be described as follows. Choose any fixed order of the vertices of t'. Then choose a uniformly random labeled composition of $\{n-n_1+1,\ldots,n\}$ with $n-n_1$ parts, and then assign degree-one vertices to the ancestral edges of vertices of t' as specified by the composition, as above.

For any *root-to-leaf path* $U = (u_1,\ldots,u_m)$ in t', it follows from the above construction that conditionally given $\mathrm{T}^- = \mathrm{t}'$, the total number of vertices lying along the corresponding root-to-leaf path in T has the same distribution as the random variable X_m described in the following experiment. Consider an urn with $n-1-n_1$ white balls and n_1 black balls. Repeatedly sample without replacement from the urn until it is empty. For $0 \le i \le n-1$ write W_i for the total number of white balls drawn after the i'th sample, and set $W_n = n-n_1$. Then let $X_m = \min(i : W_i = m)$.

For $0 \le i \le n-1$ let W_i^* be Binomial$(i,(n-1-n_1)/(n-1))$; so W_i^* can be thought of as the number of white balls drawn after i samples *with* replacement from the urn of the previous paragraph. By [16, Proposition 20.6], W_i^* is a dilation of W_i, which is to say that there is a coupling (W,W^*) of W_i and W_i^* so that $\mathbb{E}(W^*|W) = W$. This implies that $\mathbb{E}(\phi(W)) \le \mathbb{E}(\phi(W^*))$ for all continuous convex functions of $\phi : \mathbb{R} \to \mathbb{R}$. In particular, concentration inequalities for W_i^* which are proved by bounding exponential moments $\mathbb{E}(e^{\lambda W})$ of W apply without change to W_i. By a Chernoff bound (see, e.g., [81, Theorem 2.1]). It follows that that for $x \geq 2$,

$$\begin{aligned}\mathbb{P}(X_m > xm(n-1)/(n-n_1-1)) &= \mathbb{P}(W_{\lceil xm(n-1)/(n-n_1-1)\rceil} < m) \\ &\leq \mathbb{P}(W_{\lceil xm(n-1)/(n-n_1-1)\rceil} \leq \mathbb{E}W_{\lceil xm(n-1)/(n-n_1-1)\rceil}/2) \\ &\leq e^{-xm(n-1)/(2(n-n_1-1))}\end{aligned}$$

Since $n_1 < n-1$ we have $x(n-1)/(n-n_1-1) \leq 2xn/(n-n_1)$, so the preceding bound implies that for $y \geq 4mn/(n-n_1)$,

$$\mathbb{P}(X_m > y) \leq e^{-y/4}\,.$$

Write r' for the root of t' and r for the root of T. Since the number of vertices on a fixed root-to-leaf path is one larger than the distance of that leaf from the root, it follows that for any leaf v of t', for any $y \geq 4(\mathrm{dist}_{\mathrm{t}'}(v,r')+1)n/(n-n_1)$,

$$\mathbb{P}\left(\mathrm{dist}_{\mathrm{T}}(v,r) > y \;\middle|\; \mathrm{T}^- = \mathrm{t}'\right) \leq e^{-(y+1)/4}.$$

If t' has height at most $\lceil h\rceil$ then since $h \geq 2$ we always have $4(\mathrm{dist}_{\mathrm{t}'}(v,r')+1)n/(n-n_1) \leq 8hn/(n-n_1)$, so if T' has n_0 leaves and height at most $\lceil h\rceil$ then, applying the above inequality to all root-to-leaf paths in t', a union bound yields that for any $y \geq h \cdot 8n/(n-n_1)$,

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}) > y \mid \mathrm{T}^- = \mathrm{t}') \leq n_0 e^{-(y+1)/4} < n_0 e^{-y/4}\,.$$

Finally, since $\mathrm{T}^- \overset{\mathrm{d}}{=} \mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$, it follows that

$$\begin{aligned}\mathbb{P}(\mathrm{ht}(\mathrm{T}) > y) &\leq \mathbb{P}(\mathrm{ht}(\mathrm{T}^-) > \lceil h\rceil) + \sum_{\{\mathrm{t}'\in\mathcal{T}_{\mathrm{d}'}:\mathrm{ht}(\mathrm{t}')\leq\lceil h\rceil\}} \mathbb{P}(\mathrm{ht}(\mathrm{T}) > y \mid \mathrm{T}^- = \mathrm{t}')\mathbb{P}(\mathrm{T}^- = \mathrm{t}') \\ &\leq \mathbb{P}(\mathrm{ht}(\mathrm{T}') > \lceil h\rceil) + n_0 e^{-y/4}\,.\end{aligned}$$

□

The proof of Proposition 2.2.4 requires an auxiliary lemma, which itself has an auxiliary lemma.[3] We first state and prove these two lemmas, then proceed to the proof of Proposition 2.2.4.

The auxiliary auxiliary lemma, which is proved by a Chernoff bound-type argument, is similar to [25, Lemma 15].

---

[3] "So, Nat'ralists observe, a Flea/Hath smaller Fleas that on him prey/And these have smaller yet to bite 'em/And so proceed *ad infinitum.*" From *On Poetry: A Rapsody*, Johnathan Swift, 1733.

Lemma 2.2.5. *Fix positive constants* $x_1, \dots, x_m$ *and let* $E_1, \dots, E_m$ *be independent with* $E_m \sim \mathrm{Exp}(x_i)$*. Fix* $t > 0$ *and write* $S = x_1 \mathbf{1}_{[E_1 \le t]} + \dots + x_m \mathbf{1}_{[E_m \le t]}$*. Then*

$$\mathbb{P}(S < \mathbb{E}S/2) \le \exp(-t\mathbb{E}S/4).$$

Proof. For readability we take $t = 1$, and explain how to handle general $t > 0$ at the end of the proof. Let $X_i = x_i \mathbf{1}_{[E_i \le 1]}$, so $\mathbb{E}X_i = x_i(1 - e^{-x_i})$ and

$$\overline{X}_i := X_i - \mathbb{E}X_i = x_i e^{-x_i} - x_i \mathbf{1}_{[E_i > 1]}\,,$$

so that $\overline{S} := S - \mathbb{E}S = \sum_{i=1}^m \overline{X}_i$. An easy computation gives

$$\mathbb{E}e^{-\overline{X}_i} = e^{-x_i e^{-x_i}} \mathbb{E}e^{x_i \mathbf{1}_{[E_i > 1]}} = e^{-x_i e^{-x_i}} (2 - e^{-x_i})\,,$$

and a delicate but elementary computation shows that this quantity is at most $\exp\left(x_i(1 - e^{-x_i})/4\right) = \exp(\mathbb{E}X_i/4)$. Markov's inequality then gives

$$\mathbb{P}(S < \mathbb{E}S/2) \le \mathbb{E}e^{-\overline{S}} e^{-\mathbb{E}S/2} = \prod_{i=1}^m \mathbb{E}e^{-\overline{X}_i} e^{-\mathbb{E}X_i/2} \le \prod_{i=1}^m e^{-\mathbb{E}X_i/4} = e^{-\mathbb{E}S/4}.$$

This proves the lemma in the case $t = 1$. For general t, the proof works identically by instead showing that $\mathbb{E}e^{-t\overline{X}_i} \le e^{t\mathbb{E}X_i/4}$. □

We use Lemma 2.2.5 to prove a lower tail bound for sums of the form $\sum_{1 \le j \le k} (d_{i_{\mathrm{T}}(j)} - 1)$.

Lemma 2.2.6. *Fix a degree sequence* $\mathrm{d} = (d_1, \dots, d_n)$ *and let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*. Then taking* $c = (1 - e^{-2})/8$*, for all* $t \in [0, 1]$*, with* $\sigma = \sigma_{\mathrm{d}} = (n^{-1} \sum_{j \in [n]} d_j(d_j - 1))^{1/2}$*, we have*

$$\mathbb{P}\Big(\sum_{1 \le j \le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le \frac{c\sigma^2 nt}{1 + \sigma n^{1/2} t} \Big) \le \exp(-\tfrac{3}{16} tn) + \exp\left(-\frac{c}{4} \frac{n\sigma^2 t^2}{1 + \sigma n^{1/2} t} \right).$$

Proof. We assume d is compressed, so that $m := |\{i \in [n] : d_i \ne 0\}| = \max(i \in [n] : d_i \ne 0)$ is the number of non-leaf vertices of T. Let $E_1, \dots, E_m$ be independent with $E_i \sim \mathrm{Exp}(d_i)$, and let $(I_1, \dots, I_m)$ be the permutation of $(1, \dots, m)$ for which $E_{I_1} < \dots < E_{I_m}$. Then $d_{I_1}, \dots, d_{I_m}$ is a size-biased permutation of $(d_1, \dots, d_m)$. In other words, $(I_1, \dots, I_m)$ has the same distribution as $(i_{\mathrm{T}}(1), \dots, i_{\mathrm{T}}(m))$ for $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$. For the remainder of the proof we may therefore assume that

T and $E_1, \dots, E_m$ are coupled so that

$$(I_1, \dots, I_m) = (i_{\mathrm{T}}(1), \dots, i_{\mathrm{T}}(m)).$$

Now let $s = t/2$ and define

$$\begin{aligned} N_s &:= |\{i \in [m] : E_i \le s\}| = \sum_{i \in [m]} \mathbf{1}_{[E_i \le s]} = \max(j : E_{I_j} \le s)\,, \\ X_s &:= \sum_{i \in [m]} (d_i - 1)\mathbf{1}_{[E_i \le s]} = \sum_{j \in [N_s]} (d_{I_j} - 1) \end{aligned}$$

and

$$M = \frac{1 - e^{-2}}{4} \frac{n\sigma^2 s}{1 + \sigma n^{1/2} s} > \frac{1 - e^{-2}}{8} \frac{n\sigma^2 t}{1 + \sigma n^{1/2} t} = \frac{cn\sigma^2 t}{1 + \sigma n^{1/2} t}$$

Under the above coupling,

$$\sum_{1 \le j \le nt} (d_{i_{\mathrm{T}}(j)} - 1) = \sum_{1 \le j \le nt} (d_{I_j} - 1).$$

It follows that if $\sum_{1 \le j \le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le M$ then either $nt \le N_s$ or $X_s \le M$, so

$$\begin{aligned} \mathbb{P}\Big(\sum_{1 \le j \le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le \frac{c\sigma^2 nt}{1 + \sigma n^{1/2} t} \Big) &\le \mathbb{P}\Big(\sum_{1 \le j \le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le M \Big) \\ &\le \mathbb{P}(N_s \ge nt) + \mathbb{P}(X_s \le M), \end{aligned} \tag{2.13}$$

where in the first line we have used our lower bound on M.

To make use of this inequality, we first bound $\mathbb{E}X_s = \sum_{i \in [m]} (d_i - 1)(1 - e^{-d_i s})$ from below. We consider two cases, depending on whether $\sigma^2 = n^{-1} \sum_{i \in [m]} d_i(d_i - 1)$ is dominated by the contributions from small or large summands.

Fix $r \in (0, 1)$. First suppose that $\sum_{i \in [m]: d_i s \le 2} d_i(d_i - 1) \ge rn\sigma^2$. For $d_i s \le 2$ we have $(1 - e^{-d_i s}) \ge (1 - e^{-2}) d_i s/2$, so

$$\sum_{i \in [m]: d_i s \le 2} (d_i - 1)(1 - e^{-d_i s}) \ge (1 - e^{-2}) \frac{s}{2} \sum_{i \in [m]: d_i s \le 2} d_i(d_i - 1) \ge (1 - e^{-2}) \frac{s}{2} rn\sigma^2\,.$$

Next suppose that $\sum_{i \in [m]: d_i s \le 2} d_i(d_i - 1) < rn\sigma^2$. For i such that $d_i s > 2$ we have $1 - e^{-d_i s} > 1 - e^{-2}$ and $(d_i - 1)/d_i > (2 - s)/2$, so using that $\sum_i x_i > (\sum_i x_i^2)^{1/2}$ for non-negative x_i, we obtain

$$\begin{aligned}\sum_{i\in[m]:d_is>2}(d_i-1)(1-e^{-d_is}) &\ge (1-e^{-2})\Big(\sum_{i\in[m]:d_is>2}(d_i-1)^2\Big)^{1/2}\\ &\ge (1-e^{-2})\Big(\frac{2-s}{2}\sum_{i\in[m]:d_is>2}d_i(d_i-1)\Big)^{1/2}\\ &> (1-e^{-2})\frac{2-s}{2}(1-r)n^{1/2}\sigma\,.\end{aligned}$$

Taking $r=(2-s)/(2-s+\sigma n^{1/2}s)$ makes the lower bounds on $\sum_{i\in[m]:d_is\le 2}d_i(d_i-1)$ and on $\sum_{i\in[m]:d_is>2}d_i(d_i-1)$ in the two cases equal, and yields the bound

$$\mathbb{E}X_s=\sum_{i\in[m]}(d_i-1)(1-e^{-d_is})\ge\frac{1-e^{-2}}{2}\frac{(2-s)n\sigma^2s}{2-s+\sigma n^{1/2}s}\ge\frac{1-e^{-2}}{2}\frac{n\sigma^2s}{1+\sigma n^{1/2}s}=2M\,,$$

the last inequality holding since $s=t/2\le 1/2$. By Lemma 2.2.5 it follows that

$$\mathbb{P}(X_s\le M)\le\mathbb{P}(X_s\le\mathbb{E}X_s/2)\le\exp(-s\mathbb{E}X_s/4)\le\exp(-tM/4)\,. \tag{2.14}$$

Next, note that

$$\mathbb{E}N_s=\sum_{i\in[m]}\mathbb{P}(E_i\le s)=\sum_{i\in[m]}(1-e^{-d_is})\le\sum_{i\in[m]}d_is=s(n-1)<sn\,.$$

Since N_s is a sum of $[0,1]$-valued random variables, Bernstein's inequality [21, 39] then gives

$$\mathbb{P}(N_s\ge tn)=\mathbb{P}(N_s\ge 2sn)\le\exp(-\tfrac{3}{8}sn)=\exp(-\tfrac{3}{16}tn)\,. \tag{2.15}$$

Using the bounds (2.14) and (2.15) in (2.13) now yields

$$\mathbb{P}\Big(\sum_{1\le j\le nt}(d_{i_{\mathrm{T}}(j)}-1)\le\frac{c\sigma^2nt}{1+\sigma n^{1/2}t}\Big)\le\exp(-\tfrac{3}{16}tn)+\exp(-tM/4).$$

In view of the lower bound on M, this proves the bound claimed in the lemma. □

Proof of Proposition 2.2.4. First recall that the non-leaf vertices contained in $\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}$ are precisely $i(1),\dots,i(k)$, with k minimal such that $\sum_{j=1}^{k}(d_{i(j)}-1)\ge\alpha\sigma n^{1/2}$. Now, if $\mathrm{ht}\left(\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}\right)>\frac{bn^{1/2}}{2\sigma}$, then $\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}$ must contain more than $\frac{1}{2}bn^{1/2}/\sigma$ non-leaf vertices. This implies that $k>\frac{1}{2}bn^{1/2}/\sigma$, which by the definition of k yields that $\sum_{1\le j\le\frac{1}{2}bn^{1/2}/\sigma}\left(d_{i_{\mathrm{T}}(j)}-1\right)<\alpha\sigma n^{1/2}$. It follows from this inclusion of events that

$$\mathbb{P}\left(\mathrm{ht}\left(\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}\right) > \tfrac{bn^{1/2}}{2\sigma}\right) \le \mathbb{P}\left(\sum_{1\le j\le \frac{1}{2}bn^{1/2}/\sigma} \left(d_{i_{\mathrm{T}}(j)} - 1\right) < \alpha\sigma n^{1/2}\right).$$

Now fix $b \ge 1$ and set $t = b/(2\sigma n^{1/2})$ so that $tn = bn^{1/2}/(2\sigma)$. (We see that if $t > 1$ then $\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}) > \frac{bn^{1/2}}{2\sigma}) = 0$, so we assume $t \le 1$.) Then with $c = (1-e^{-2})/8$ as in the statement of Lemma 2.2.6, we have $c = 3\alpha$, so since $b \ge 1$,

$$\alpha\sigma n^{1/2} = \frac{c\sigma n^{1/2}}{3} \le c\sigma n^{1/2}\frac{b/2}{1+b/2} = \frac{c\sigma^2 nt}{1+\sigma n^{1/2}t}.$$

It follows that

$$\mathbb{P}\left(\sum_{1\le j\le \frac{1}{2}bn^{1/2}/\sigma} \left(d_{i_{\mathrm{T}}(j)} - 1\right) < \alpha\sigma n^{1/2}\right) \le \mathbb{P}\left(\sum_{1\le j\le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le \frac{c\sigma^2 nt}{1+\sigma n^{1/2}t}\right),$$

and Lemma 2.2.6 now yields the result. □

2.3 Bienaymé trees and simply generated trees: proofs of Theorems 2.1.1, 2.1.2 and 2.1.3.

In this section we use Theorems 2.1.7 and 2.1.8 to prove Theorems 2.1.1, 2.1.2 and 2.1.3, and additionally to prove two conjectures from [68] on *simply generated trees*; this class of trees is defined as follows. Fix non-negative real weights $\mathrm{w} = (w_k, k \ge 0)$ with $w_0 > 0$. Given a finite plane tree t, we define the weight of t to be

$$\mathrm{w}(\mathrm{t}) = \prod_{v\in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)}.$$

Next, for positive integers n, let

$$Z_n = Z_n(\mathrm{w}) = \sum_{\{\text{plane trees t}:|v(\mathrm{t})|=n\}} \mathrm{w}(\mathrm{t}).$$

If $Z_n > 0$, then we define a random tree $\mathrm{T}_{\mathrm{w},n}$ by setting

$$\mathbb{P}(\mathrm{T}_{\mathrm{w},n} = \mathrm{t}) = \frac{\mathrm{w}(\mathrm{t})}{Z_n}$$

for plane trees t with $|v(\mathrm{t})| = n$. The random tree $\mathrm{T}_{\mathrm{w},n}$ is called a simply generated tree of size n with weight sequence w. Note that if $\sum_{k\ge 0} w_k = 1$, then w is an offspring distribution, and $\mathrm{T}_{\mathrm{w},n}$ is indeed distributed as a Bienaymé tree with offspring distribution w, conditioned to have size n – so this notation agrees with (but generalizes) that from earlier in the paper. From this we also see that conditioned Bienaymé trees are a subclass of simply generated trees; we will use this fact below.

Let $\Phi(z) = \Phi_{\mathrm{w}}(z) = \sum_{k\ge 0} w_k z^k$ be the generating function of the weight sequence w, and let $\rho = \rho_{\mathrm{w}} = \sup(s \ge 0 : \Phi(s) < \infty)$ be its radius of convergence. For $s > 0$ such that $\Phi(s) < \infty$, we let

$$\Psi(s) = \Psi_{\mathrm{w}}(s) = \frac{s\Phi'(s)}{\Phi(s)} = \frac{\sum_{k\ge 0} k w_k s^k}{\sum_{k\ge 0} w_k s^k}.$$

If $\Phi(\rho) = \infty$ then also define

$$\Psi(\rho) = \Psi_{\mathrm{w}}(\rho) = \lim_{s\uparrow\rho} \Psi(s);$$

by Lemma 3.1(i) in [68], Ψ is strictly increasing on $[0, \rho)$ so this limit exists. Let $\nu = \nu_{\mathrm{w}} = \Psi_{\mathrm{w}}(\rho)$, and define

$$\tau = \tau_{\mathrm{w}} := \begin{cases} \rho & \text{if } \nu_{\mathrm{w}} \le 1 \\ \Psi^{-1}(1) & \text{if } \nu_{\mathrm{w}} > 1\,. \end{cases} \tag{2.16}$$

Note that if $\nu_{\mathrm{w}} > 1$ then $\tau \in [0, \rho)$. Define $\hat\sigma^2 = \hat\sigma^2_{\mathrm{w}} = \tau\Psi'(\tau)$. For later use, we note that if $\mathrm{w} = (w_k, k \ge 0)$ is a probability distribution with finite mean, so $\sum_{k\ge 0} w_k = 1$ and $|\mathrm{w}|_1 = \sum_{k\ge 0} k w_k < \infty$, then we always have $\rho \ge 1$, so

$$\nu = \Psi_{\mathrm{w}}(\rho) \ge \Psi_{\mathrm{w}}(1) = \sum_{k\ge 0} k w_k = |\mathrm{w}|_1, \text{ and } \Psi'_{\mathrm{w}}(1) = \sum_{k\ge 0} k^2 w_k - \left(\sum_{k\ge 0} k w_k\right)^2 = |\mathrm{w}|_2 - |\mathrm{w}|_1^2\,. \tag{2.17}$$

The following theorem resolves Conjecture 21.5 and Problem 21.7 from [68].

Theorem 2.3.1. *Let* $\mathrm{w} = (w_k, k \ge 0)$ *be a weight sequence with* $w_0 > 0$ *and with* $w_k > 0$ *for some* $k \ge 2$. *If either* $\hat\sigma^2 = \infty$ *or* $\nu < 1$ *then* $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) \to 0$, *where the convergence is in both probability and expectation, as* $n \to \infty$ *along integers* n *such that* $Z_n(\mathrm{w}) > 0$.

We also prove the following, more quantitative theorem.

Theorem 2.3.2. *Fix* $\epsilon \in [0,1)$*. Then, there exist constants* $c_1 = c_1(\epsilon)$ *and* $c_2 = c_2(\epsilon)$ *such that the following holds. Let* $\mathrm{w} = (w_k, k \geq 0)$ *be a weight sequence with* $w_0 > 0$ *and* $w_k > 0$ *for some* $k \geq 2$ *and* $\frac{w_1 \tau}{\Phi(\tau)} < 1 - \epsilon$*. For any* $t > 1$ *and any* n *large enough with* $Z_n(\mathrm{w}) > 0$ *it holds that*

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > tn^{1/2}\right) \leq c_1 \exp(-c_2 t^2).$$

This theorem has the following immediate corollary.

Corollary 2.3.3. *For any weight sequence* w *on* $\mathbb{N}$ *with* $w_0 > 0$ *and* $w_k > 0$ *for some* $k \geq 2$*,* $\mathbb{E}[\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})] = O(n^{1/2})$*, and, more generally, for any fixed* $r < \infty$*,* $\mathbb{E}[\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})^r] = O(n^{r/2})$ *as* $n \to \infty$ *over all* n *such that* $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$*.* □

In this section we will make use of the following notation. For w a weight sequence, let $\mathrm{D} = \mathrm{D}(\mathrm{w}, n) = (D_1, \ldots, D_n)$ be a random degree sequence with the following law. Let T be a simply generated tree of size n with weight sequence w. Conditionally given T, let $\hat{\mathrm{T}}$ be the random tree obtained as follows: label the vertices of T by a uniformly random permutation of $[n]$, then forget about the plane structure. Let D_i be equal to the degree of i in $\hat{\mathrm{T}}$. Then, for $k \in \mathbb{N}$, let $N_k = N_k(\mathrm{w}, n) = |\{i \in [n] : D_i = k\}|$ be the number of entries of D which equal k, so $N_k = n_k(\mathrm{D}(\mathrm{w}, n))$ and $\sum_{k=0}^{\infty} N_k = n$.

Our proofs will exploit two distributional identities. The first is the following. Let $\mathrm{T}_{\mathrm{D}(\mathrm{w},n)} \in_u \mathcal{T}_{\mathrm{D}(\mathrm{w},n)}$, by which we mean that, conditionally given that $\mathrm{D}(\mathrm{w}, n) = \mathrm{d}$, we have $\mathrm{T}_{\mathrm{D}(\mathrm{w},n)} \in_u \mathcal{T}_{\mathrm{d}}$. Then $\hat{\mathrm{T}} \overset{d}{=} \mathrm{T}_{\mathrm{D}(\mathrm{w},n)}$ by Proposition 2.1.9, so $\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) \overset{d}{=} \mathrm{ht}(\mathrm{T}_{\mathrm{D}(\mathrm{w},n)})$. The second is the fact that for any weight sequence $\mathrm{w} = (w_k, k \geq 0)$ and any constants $a, b > 0$, the weight sequence $\hat{\mathrm{w}}$ with $\hat{w}_k = ab^k w_k$ is *equivalent* to w, i.e., it satisfies that $\mathrm{T}_{\mathrm{w},n} \overset{\mathrm{d}}{=} \mathrm{T}_{\hat{\mathrm{w}},n}$ for all n for which either (and thus both) of the random trees are defined. This is an immediate consequence of the formula (2.1) for the distribution of $\mathrm{T}_{\mathrm{w},n}$.

These two equalities in distribution imply that if we obtain good control over the asymptotic behaviour of $(N_k(\hat{w}, n), k \geq 0)$ for some weight sequence $\hat{w}$ which is equivalent to w, then we can use Theorems 2.1.7 and 2.1.8, on the heights of trees with given degree sequences, to prove tail bounds for $\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})$. To obtain such control, we rely on the following result.

Theorem 2.3.4 ([68], Theorem 11.4; [10], Theorem 5.2)**.** *Let* $\mathrm{w} = (w_k, k \geq 0)$ *be a weight* *sequence with* $w_0 > 0$ *and* $w_k > 0$ *for some* $k \geq 2$. *Write* $\tau = \tau_{\mathrm{w}}$ *and define*

$$\hat{\mu}_k = \hat{\mu}_k(\mathrm{w}) = \frac{w_k \tau^k}{\Phi(\tau)}$$

for $k \in \mathbb{N}$. *Then* $\hat{\mu} = (\hat{\mu}_k, k \geq 0)$ *is a probability distribution with mean* $m = 1 \wedge \nu$ *and variance* $\hat{\sigma}^2$. *For* $k \geq 0$ *define* $N_k = n_k(\mathrm{D}(\mathrm{w}, n))$, *as above. Then for every* $\epsilon > 0$ *there exists* $c(\epsilon) > 0$ *such that for all* n *sufficiently large, for every integer* $k \geq 0$,

$$\mathbb{P}\left(\left| \frac{N_k(\mathrm{w}, n)}{n} - \hat{\mu}_k \right| > \epsilon \right) < e^{-cn}.$$

Theorem 11.4 of [68] handles the case $\nu > 0$ in the above statement. The exponentially small error bounds stated above are not made explicit in the statement of [68, Theorem 11.4], but are recorded in the course of its proof (see [68, pages 163-164]). The case $\nu = 0$ was handled in [10].

Before we prove Theorems 2.3.1, 2.3.2, 2.1.1 and 2.1.3, we first illustrate that the requirements in Theorem 2.1.3, that $1-\mu_0-\mu_1$ and $\mu_0/(\mu_0+\mu_1)$ are bounded from below, are necessary. To accomplish this we will consider probability distributions $\mu = \mu^{p,q}$ of the form

$$\mu_0 = q(1-p) \qquad \mu_1 = (1-q)(1-p) \qquad \mu_2 = p\,.$$

Claim 2.3.5. *For any* $x > 0$ *we have that for any* $q > 0$, $\lim_{p \downarrow 0} \liminf_{n \to \infty} \mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mu^{p,q},n}) > xn^{1/2}) = 1$ *and for any* $p > 0$, $\lim_{q \downarrow 0} \liminf_{n \to \infty} \mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mu^{p,q},n}) > xn^{1/2}) = 1$

Proof. We apply Theorem 2.3.4 with $\mathrm{w} = \mathrm{w}^{p,q} = \mu^{p,q}$. Elementary computation shows that the probability distribution $\hat{\mu} = \hat{\mu}^{p,q}$ from Theorem 2.3.4 is given by

$$\hat{\mu}_1 = \hat{\mu}_1^{p,q} = \frac{(1-q)\sqrt{1-p}}{(1-q)\sqrt{1-p} + 2\sqrt{p}\sqrt{q}}$$

and $\hat{\mu}_0 = \hat{\mu}_2 = (1-\hat{\mu}_1)/2$. Since $\hat{\mu}$ is equivalent to μ, it follows that that $T_{\mu^{p,q},n} \overset{\mathrm{d}}{=} T_{\hat{\mu}^{p,q},n}$ for all n. Write $\sigma^{p,q} = |\hat{\mu}^{p,q}|_2 - 1$ for the standard deviation of $\hat{\mu}^{p,q}$. Using this equivalence in distribution together with the convergence of the search-depth process for large critical random

Bienaymé trees [14, Theorem 23], it follows that

$$\frac{2}{n^{1/2}}\mathrm{ht}(T_{\mu^{p,q}},n) \overset{\mathrm{d}}{=} \frac{2}{n^{1/2}}\mathrm{ht}(T_{\hat{\mu}^{p,q}},n) \xrightarrow{\mathrm{d}} \frac{1}{\sigma^{p,q}} \sup_{0\le t\le 1} B(t)\,,$$

where B is a standard Brownian excursion. Finally, since $\hat{\mu}_1^{p,q} \to 1$ (and thus $\sigma^{p,q} \to 0$) as either $p \to 0$ or $q \to 0$, and $\mathbb{P}(\sup_{0\le t\le 1} B(t) > 0) = 1$, it follows that $(\sigma^{p,q})^{-1}\sup_{0\le t\le 1} B(t) \to \infty$ in probability as either $p \to 0$ or $q \to 0$. The result follows. □

In our proof of Theorem 2.3.1, we make use of the following consequence of Theorem 2.3.4.

Corollary 2.3.6. *If $\nu < 1$, or if $\hat{\sigma}^2 = \infty$, then for each $C > 0$ there exists $c = c(\mathrm{w}, C) > 0$ such that for all n sufficiently large,*

$$\mathbb{P}\left(\sigma_{\mathrm{D}(\mathrm{w},n)} \le C\right) \le e^{-cn}\,.$$

Proof. First suppose that $\nu < 1$; in this case $\sum_{i=1}^{\infty} k\hat{\mu}_k = \nu$. Let $n \ge \frac{4}{1-\nu}$. Suppose without loss of generality that $C \ge 1$ and set $K = \frac{4C^2}{1-\nu}$. Then, $\sum_{k=1}^{K-1} k\hat{\mu}_k \le \nu$, so by Theorem 2.3.4, there is $c = c(\mathrm{w}, C)$ such that with $N_k = n_k(\mathrm{D}(\mathrm{w}, n))$,

$$\mathbb{P}\left(\sum_{k=0}^{K-1} kN_k > \frac{1+\nu}{2}n\right) \le e^{-cn}$$

for all n sufficiently large. But on the event that $\sum_{k=0}^{K-1} kN_k \le \frac{1+\nu}{2}n$, using that $\sum_{k=1}^{\infty} kN_k = n-1$, we have $\sum_{k=K+1}^{\infty} kN_k \ge \frac{1-\nu}{2}n - 1 \ge \frac{1-\nu}{4}n$ by our assumption on n. This implies that

$$\sum_{i=1}^{n} D_i(D_i-1) \ge \tfrac{1-\nu}{4}Kn \ge C^2 n,$$

so $\sigma_{\mathrm{D}(\mathrm{w},n)} > C$. This proves the claim in the case that $\nu < 1$.

Next suppose that $\hat{\sigma}^2 = \infty$. In this case there exists $K \in \mathbb{N}$ such that $\sum_{k=0}^{K} \hat{\mu}_k k(k-1) > 2C$. Since $\sigma_{\mathrm{D}(\mathrm{w},n)} = n^{-1}\sum_{i=1}^{n} D_i(D_i-1) \ge n^{-1}\sum_{k=0}^{K} N_k k(k-1)$, Theorem 2.3.4 then implies that there exists $c = c(\mathrm{w}, C)$ such that

$$\mathbb{P}\left(\sigma_{\mathrm{D}(\mathrm{w},n)} \le C\right) \le \mathbb{P}\left(\sum_{k=0}^{K} N_k k(k-1) \le Cn\right) \le e^{-cn}$$

for all n sufficiently large. This proves the claim in the case that $\hat{\sigma}^2 = \infty$. □

In the next proof we write $x \vee y := \max(x, y)$.

Proof of Theorem 2.3.1. Fix $0 < \epsilon \le 2^{-14}$. Let w be a weight sequence with $\nu \ge 1$ and $\hat{\sigma}^2 = \infty$, or with $\nu < 1$, and fix n such that $Z_n(\mathrm{w}) > 0$. Let $\hat{\mu}_1 = \hat{\mu}_1(\mathrm{w})$ and let K be large enough such that $2\log(k+1)/k < \epsilon^2$ and such that $\log(k+1) \ge (\frac{1}{2}\log\frac{2}{(1-\hat{\mu}_1)}) \vee 2^{14}$ for all $k \ge K$. Then let

$$\mathcal{D}_n = \Big\{\text{degree sequences } \mathrm{d} = (d_1, \ldots, d_n) : n_1(\mathrm{d}) \le \frac{n(1+\hat{\mu}_1)}{2}, \sigma_{\mathrm{d}} \ge K\Big\}.$$

By Theorem 2.3.4 and Corollary 2.3.6, there exists $c = c(\mathrm{w}) > 0$ such that for all n sufficiently large,

$$\mathbb{P}(\mathrm{D} \notin \mathcal{D}_n) \le e^{-cn}.$$

Moreover, for any $\mathrm{d} \in \mathcal{D}_n$, with $(\sigma')^2 = \sigma_{\mathrm{d}}^2 n/(n - n_1(\mathrm{d}))$ as in Theorem 2.1.8, we have $\log(\sigma' + 1) \le \log(\sigma_{\mathrm{d}} + 1) + \frac{1}{2}\log\frac{2}{(1-\hat{\mu}_1)} < 2\log(\sigma_{\mathrm{d}} + 1)$, so $\log(\sigma'+1)/\sigma_{\mathrm{d}} \le \epsilon^2$. Also, $\sigma' \ge \sigma_{\mathrm{d}}$, so $\log(\sigma'+1) \ge 2^{14}$.

Fix any $t \ge \epsilon$. We apply Theorem 2.1.8 with $x = \epsilon^{-2}t \ge 2^{14}$ to obtain that for any n and any $\mathrm{d} \in \mathcal{D}_n$,

$$\begin{aligned}
\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > tn^{1/2} \,\middle|\, \mathrm{D} = \mathrm{d}\right) &= \mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > x\epsilon^2 n^{1/2} \,\middle|\, \mathrm{D} = \mathrm{d}\right) \\
&\le \mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > xn^{1/2}\frac{\log(\sigma'+1)}{\sigma_{\mathrm{d}}} \,\middle|\, \mathrm{D} = \mathrm{d}\right) \\
&= \mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{d}}) > xn^{1/2}\frac{\log(\sigma'+1)}{\sigma_{\mathrm{d}}}\right) \\
&\le 4\exp(-x\log(\sigma'+1)/2^{14}) \le 4\exp(-\epsilon^{-2}t).
\end{aligned}$$

Since

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2} \,\middle|\, \mathrm{D} \in \mathcal{D}_n\right) = \sum_{\mathrm{d} \in \mathcal{D}_n} \mathbb{P}\left(\mathrm{ht}(\mathrm{T}_n) > tn^{1/2} \,\middle|\, \mathrm{D} = \mathrm{d}\right)\mathbb{P}(\mathrm{D} = \mathrm{d} | \mathrm{d} \in \mathcal{D}_n),$$

it follows that for all n sufficiently large,

$$\begin{aligned}
\mathbb{E}\left[\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})\mathbb{1}_{\{\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > \epsilon n^{1/2}\}}\right] &\le \mathbb{E}\left[\mathrm{ht}(\mathrm{T}_{\mathrm{D}})\mathbb{1}_{\{\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > \epsilon n^{1/2}\}} \,\middle|\, \mathrm{D} \in \mathcal{D}_n\right] + n\mathbb{P}(\mathrm{D} \notin \mathcal{D}_n) \\
&\le n^{1/2}\left[\epsilon + \int_{\epsilon}^{\infty} \mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2} \,\middle|\, \mathrm{D} \in \mathcal{D}_n\right) dt\right] + o(1) \\
&\le n^{1/2}\left[\epsilon + 4\int_{\epsilon}^{\infty} \exp(-\epsilon^{-2}t) dt\right] + o(1)
\end{aligned}$$

$$= n^{1/2}\left[\epsilon + 4\epsilon^2\exp(-\epsilon^{-1}) + o(1)\right] .$$

We can pick ϵ arbitrary small, so the statement follows. □

Proof of Theorem 2.3.2. Let $\mathrm{D} = \mathrm{D}(\mathrm{w}, n)$ and, as above, write $N_k = n_k(\mathrm{D})$. Then, by Theorem 2.3.4, there is $c = c(\epsilon) > 0$ such that for all n sufficiently large,

$$\mathbb{P}\left(1 - \frac{N_1}{n} < \epsilon/2\right) < e^{-cn}.$$

Let $\mathcal{D}_n$ be the set of degree sequences d with $\mathbb{P}(\mathrm{D} = \mathrm{d}) > 0$ such that $1 - n_1(\mathrm{d})/n \geq \epsilon/2$; so $\mathbb{P}(\mathrm{D} \notin \mathcal{D}_n) < e^{-cn}$. By Theorem 2.1.7, there are $c_1 = c_1(\epsilon)$ and $c_2 = c_2(\epsilon)$ such that for any $\mathrm{d} \in \mathcal{D}_n$, for all $t > 1$,

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_n) > tn^{1/2} \middle| \mathrm{D} = \mathrm{d}\right) \leq c_1 \exp\left(-c_2 t^2\right).$$

The theorem now follows from the observation that

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > tn^{1/2}\right) \leq \mathbb{P}(\mathrm{D} \notin \mathcal{D}_n) + \mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2} \middle| \mathrm{D} \in \mathcal{D}_n\right)$$

and the fact that

$$\mathbb{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2} \middle| \mathrm{D} \in \mathcal{D}_n\right) = \sum_{\mathrm{d} \in \mathcal{D}_n} \mathbb{P}\left(\mathrm{ht}(\mathrm{T}_n) > tn^{1/2} \middle| \mathrm{D} = \mathrm{d}\right) \mathbb{P}(\mathrm{D} = \mathrm{d} | \mathrm{d} \in \mathcal{D}_n).$$ □

Proof of Theorems 2.1.1 and 2.1.2. We use the fact that conditioned Bienaymé trees are a special case of simply generated trees.

First, if $|\mu|_1 \leq 1$ and $|\mu|_2 = \infty$ then $\mu_0 > 0$ and $\mu_k > 0$ for some $k \geq 2$. Next, since $|\mu|_2 = \infty$, for all $t > 0$ we have $\sum_{k \geq 0} e^{tk}\mu_k = \infty$. This implies that the generating function $\Phi = \Phi_\mu$ has radius of convergence $\rho = \rho_\mu = 1$, so $\nu_\mu = \Psi(\rho) = \Psi(1) = |\mu|_1 \leq 1$. This implies that $\tau = \tau_\mu$ defined by (2.16) satisfies $\tau = \rho = 1$ and so by (2.17) we have

$$\hat{\sigma}^2 = \hat{\sigma}_\mu^2 := \tau\Psi'(\tau) = \Psi'(1) = |\mu|_2^2 - |\mu|_1 = \infty.$$

Theorem 2.3.1 now implies that $n^{1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ in probability and expectation along integers n such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$. This proves Theorem 2.1.1.

Similarly, if $|\mu|_1 < 1$ and $\sum_{k\geq 0} e^{tk}\mu_k = \infty$ for all $t > 0$ then $\mu_0 > 0$ and $\mu_k > 0$ for some $k \geq 2$. Moreover, Φ_μ again has radius of convergence $\rho_\mu = 1$, and so $\nu = \Psi(1) = |\mu|_1 < 1$. In this case Theorem 2.3.1 also implies that $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ in probability and expectation along integers n such that $\mathbb{P}(|\mathrm{T}_\mu| = n) > 0$. This proves Theorem 2.1.2. □

Proof of Theorem 2.1.3. We again use the fact that Bienaymé trees are special cases of simply generated trees. We aim to apply Theorem 2.3.2, so proceed to verify the assumptions of that result. The assumptions on μ_0 and μ_1 in particular imply that $\mu_0 > 0$ and that $\mu_k > 0$ for some $k \geq 2$, so that requirement of the theorem is satisfied.

Recall from (2.16) that $\tau = \rho$ if $\Psi(\rho) \leq 1$, and $\tau = \Psi^{-1}(1)$ if $\Psi(\rho) > 1$. With $\hat{\mu}_k = \mu_k\tau^k/\Phi(\tau)$ as in Theorem 2.3.4, we then have

$$\hat{\mu}_1 = \frac{\mu_1\tau}{\Phi(\tau)} = \frac{\mu_1\tau}{\sum_{k\geq 0}\mu_k\tau^k}.$$

If $\tau \geq 1$ then the denominator is at least $\mu_1\tau + (1-\mu_1-\mu_0)\tau > \mu_1\tau + \epsilon\tau$, since by assumption $(1-\mu_1-\mu_0) > \epsilon$. This yields that

$$\hat{\mu}_1 < \frac{\mu_1\tau}{\mu_1\tau + \epsilon\tau} < \frac{(1-\epsilon)\tau}{(1-\epsilon)\tau + \epsilon\tau} = 1-\epsilon\,,$$

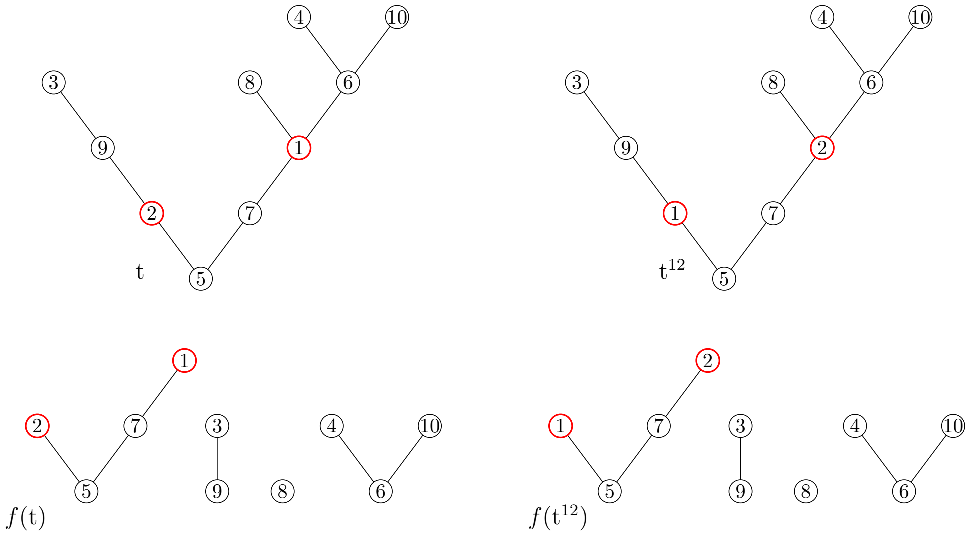
the second inequality holding since $\mu_1 < 1-\epsilon$ and $x/(x+\epsilon\tau)$ is increasing in x. On the other hand, if $\tau < 1$ then

$$\hat{\mu}_1 = \frac{\mu_1\tau}{\Phi(\tau)} < \frac{\mu_1\tau}{\mu_0 + \mu_1\tau} < \frac{\mu_1}{\mu_0+\mu_1} < 1-\epsilon\,,$$

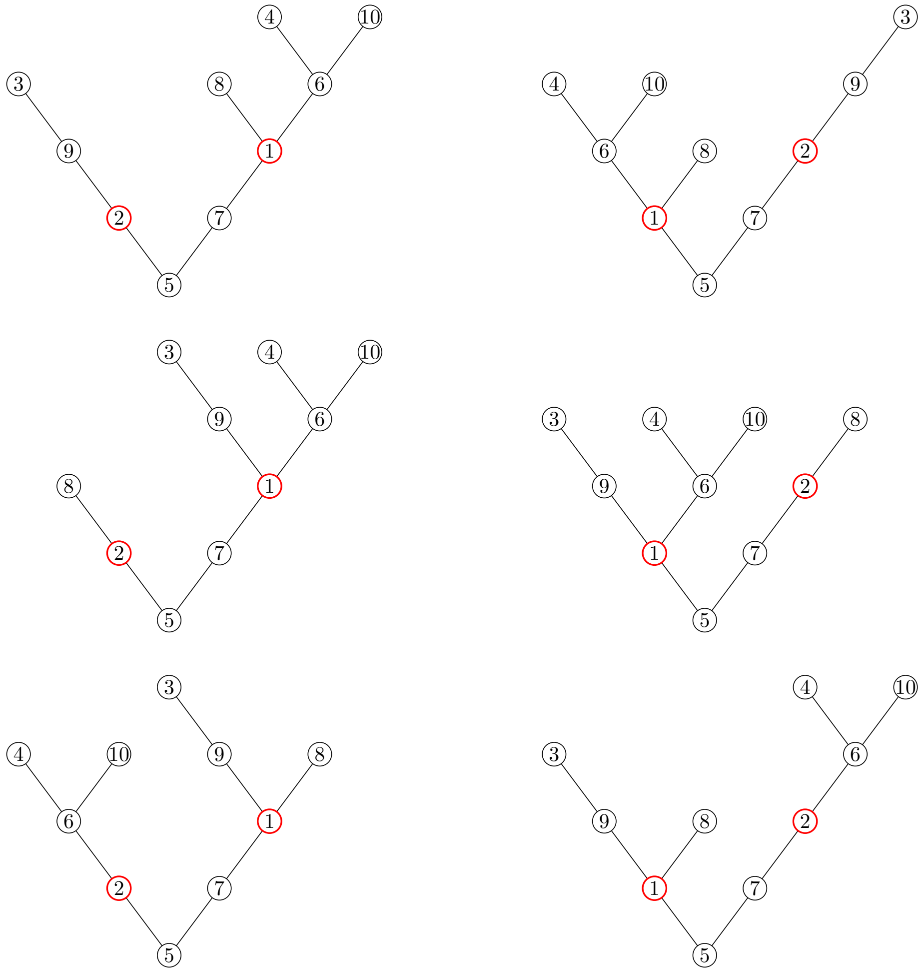
the last inequality holding by the assumptions of the theorem. In either case we have $\hat{\mu}_1 < 1-\epsilon$, so the result follows by Theorem 2.3.2. □

2.4 Stochastic domination results

This section presents the proof of Theorem 2.1.10. The following decomposition is a key input to the proof. Given a tree t, let $f(\mathrm{t})$ be the unordered set of rooted trees obtained from t by removing all edges from vertices 1 and 2 to their children. Also, write t^{12} for the tree obtained from t by swapping the labels of vertices 1 and 2. Then we say that $\mathrm{t} \sim \mathrm{t}'$ if either t and t' have the same root and $f(\mathrm{t}) = f(\mathrm{t}')$ or t^{12} and t' have the same root and $f(\mathrm{t}^{12}) = f(\mathrm{t}')$. (See Figures 2.4 and 2.5) for examples.)

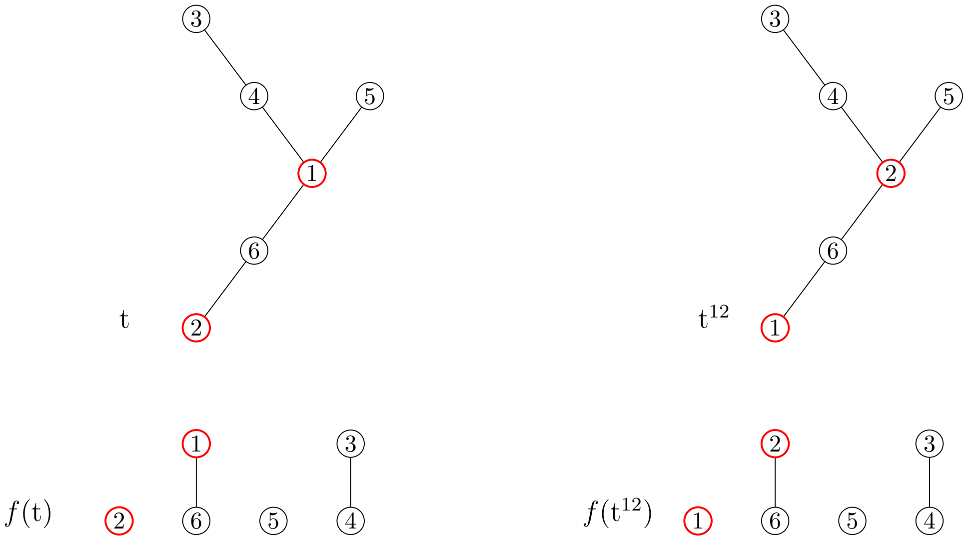


(a) For a tree t, we obtain $f(\mathrm{t})$ by removing the edges to the children of vertex 1 and 2. We obtain t^{12} by swapping the labels of 1 and 2 in t.

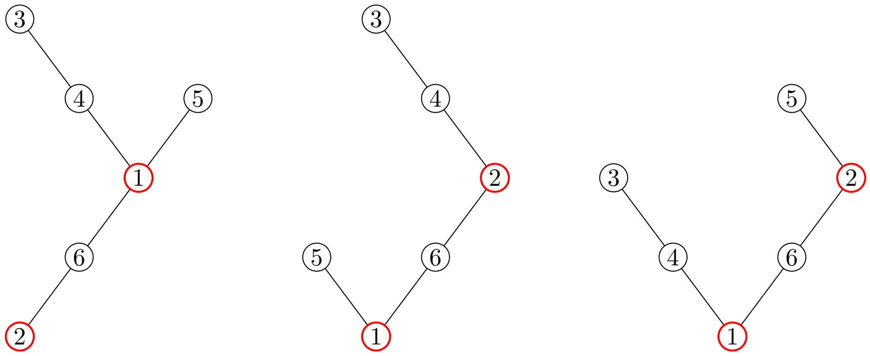


(b) The trees t' such that $\mathrm{t}' \sim \mathrm{t}$ and $d_{\mathrm{t}'}(1) = 2$ and $d_{\mathrm{t}'}(2) = 1$. For each tree t' on the left, $f(\mathrm{t}') = f(\mathrm{t})$. For each tree t' on the right, $f(\mathrm{t}') = f(\mathrm{t}^{12})$.

Fig. 2.4: We show a subset of the equivalence class of tree t. In total, t is equivalent to 2^4 trees: to specify a tree t' such that $\mathrm{t}' \sim \mathrm{t}$ we must choose, for each element in $\{6, 8, 9\}$, whether its parent in t' is 1 or 2, and we must choose whether or not to swap the labels of 1 and 2.



(a) For a tree t, we obtain $f(\mathrm{t})$ by removing the edges to the children of vertex 1 and 2. We obtain t^{12} by swapping the labels of 1 and 2 in t.



(b) The trees in the equivalence class of t in which $d_1 = 2$ and $d_2 = 1$. For the leftmost tree t', we have that $f(\mathrm{t}') = f(\mathrm{t})$ and t' and t have the same root. For each other tree t'', we have that $f(\mathrm{t}'') = f(\mathrm{t}^{12})$ and t'' and t^{12} have the same root .

Fig. 2.5: We show a subset of the equivalence class of tree t. In total, t is equivalent to 2^3 trees: to specify a tree t' such that $\mathrm{t}' \sim \mathrm{t}$ we must choose, for each element in $\{4, 5\}$, whether its parent in t' is 1 or 2, and we must choose whether or not to swap the labels of 1 and 2.

We prove Theorem 2.1.10 using the following two propositions.

Proposition 2.4.1. *Let $\mathcal{C}$ be an equivalence class for the equivalence relation $\sim$. Fix a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$ with $d_2 \geq 1$ and let $\mathrm{d}' = (d_1 + 1, d_2 - 1, \ldots, d_n)$. Then with $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$,*

$$\mathbb{P}(\mathrm{T} \in \mathcal{C}) = \mathbb{P}(\mathrm{T}' \in \mathcal{C}).$$

Proposition 2.4.2. *Fix a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$ with $d_2 \geq 1$ and let $\mathrm{d}' = (d_1 + 1, d_2 - 1, \ldots, d_n)$. Then for any $\sim$-equivalence class $\mathcal{C}$ with $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}} \neq \emptyset$, letting $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}} \cap \mathcal{C}$ and $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}'} \cap \mathcal{C}$, we have*

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}})' \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})\,.$$

Moreover, if d *contains at least three non-zero entries then there exists at least one $\sim$-equivalence class $\mathcal{C}$ for which the preceding stochastic domination is strict.*

Before proving the propositions, we show how they straightforwardly imply Theorem 2.1.10.

Proof of Theorem 2.1.10. First, by relabeling the vertices it suffices to show that for any degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$ with $d_1 \geq d_2 \geq 1$, if $\mathrm{d}' = (d_1 + 1, d_2 - 1, d_3, \ldots, d_n)$ then for $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$ we have $\mathrm{ht}(\mathrm{T}') \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T})$, and that this stochastic domination is strict if d contains at least three non-zero entries.

Fix degree sequences d and d′ related as in the previous paragraph, and let $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$. For a $\sim$-equivalence class $\mathcal{C}$ such that $\mathcal{T}_{\mathrm{d}} \cap \mathcal{C} \neq \emptyset$ we will also use the notation $\mathrm{T}_{\mathcal{C}}$ and $\mathrm{T}'_{\mathcal{C}}$ to denote uniformly random elements of $\mathcal{T}_{\mathrm{d}} \cap \mathcal{C}$ and $\mathcal{T}_{\mathrm{d}'} \cap \mathcal{C}$, respectively.

For any $x > 0$, writing $\sum_{\mathcal{C}}$ to denote a summation over all $\sim$-equivalence classes $\mathcal{C}$, we now have

$$\begin{aligned}
\mathbb{P}(\mathrm{ht}(\mathrm{T}) \leq x) &= \sum_{\mathcal{C}} \mathbb{P}\left(\mathrm{ht}(\mathrm{T}) \leq x \;\middle|\; \mathrm{T} \in \mathcal{C}\right) \mathbb{P}(\mathrm{T} \in \mathcal{C}) \\
&= \sum_{\mathcal{C}} \mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) \leq x)\mathbb{P}(\mathrm{T}' \in \mathcal{C}) \\
&\leq \sum_{\mathcal{C}} \mathbb{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \leq x)\mathbb{P}(\mathrm{T}' \in \mathcal{C}) \\
&= \sum_{\mathcal{C}} \mathbb{P}\left(\mathrm{ht}(\mathrm{T}') \leq x \;\middle|\; \mathrm{T}' \in \mathcal{C}\right) \mathbb{P}(\mathrm{T}' \in \mathcal{C}) \\
&= \mathbb{P}(\mathrm{ht}(\mathrm{T}') \leq x)
\end{aligned}$$

The second equality holds by Proposition 2.4.1 and since the conditional distribution of T given that $\mathrm{T} \in \mathcal{C}$ is precisely that of $\mathrm{T}_{\mathcal{C}}$. The inequality holds by Proposition 2.4.2. The third equality again holds since the conditional distribution of T' given that $\mathrm{T}' \in \mathcal{C}$ is precisely that of $\mathrm{T}'_{\mathcal{C}}$. This shows that $\mathrm{ht}(\mathrm{T}') \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T})$.

Finally, if d has at least three non-zero entries then by Proposition 2.4.2 there exists at least one equivalence class $\mathcal{C}$ and some $x > 0$ for which $\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) < \mathbb{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x)$. For such x the above chain of inequalities then yields that $\mathbb{P}(\mathrm{ht}(\mathrm{T}) \le x) < \mathbb{P}(\mathrm{ht}(\mathrm{T}') \le x)$, so in this case in fact $\mathrm{ht}(\mathrm{T}') \prec_{\mathrm{st}} \mathrm{ht}(\mathrm{T})$. □

Proof of Proposition 2.4.1. Let t be a tree in $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We first suppose that neither vertex 1 nor vertex 2 is an ancestor of the other in t.

The forest $f(\mathrm{t})$ contains $d_1 + d_2 + 1$ trees; list their roots as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2+1}$ is the root of t. Then both 1 and 2 lie in the tree rooted at $r_{d_1+d_2+1}$. In this case, there are $\binom{d_1+d_2}{d_1}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t})$: these are precisely the trees obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subseteq [d_1 + d_2]$ of size d_1.
- Add edges from vertex 1 to the vertices in the set $\{r_i, i \in S\}$, and add edges from vertex 2 to the vertices in the set $\{r_i, i \in [d_1 + d_2] \setminus S\}$.

Likewise, there are $\binom{d_1+d_2}{d_2}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t}^{12})$; these are obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subseteq [d_1 + d_2]$ of size d_2.
- Add edges from vertex 1 to the vertices in the set $\{r_i, i \in S\}$, add edges from vertex 2 to the vertices in the set $\{r_i, i \in [d_1 + d_2] \setminus S\}$, then swap the labels of vertices 1 and 2.

It follows that $|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}| = 2\binom{d_1+d_2}{d_1}$. Likewise, we have $|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}| = 2\binom{d_1+d_2}{d_1+1}$, since any element of $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}$ may be constructed by selecting a size-$(d_1 + 1)$ subset S' of $\{r_i, i \in [d_1 + d_2]\}$, then either (a) attaching the roots in S' to 1 and the remaining roots to 2, or (b) attaching the roots in S' to 2 and the remaining roots to 1 and switching the labels of vertices 1 and 2.

Recalling the formula (2.2) for $|\mathcal{T}_{\mathrm{d}}|$, the preceding computations yield that

$$\frac{|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}|}{|\mathcal{T}_{\mathrm{d}}|} = 2\binom{d_1 + d_2}{d_1}\binom{n-1}{d_1, \ldots, d_n}^{-1} = 2\binom{d_1 + d_2}{d_1 + 1}\binom{n-1}{d_1 + 1, d_2 - 1, d_3, \ldots, d_n}^{-1} = \frac{|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}|}{|\mathcal{T}_{\mathrm{d}'}|},$$

so in this case $\mathbb{P}(\mathrm{T} \in \mathcal{C}) = \mathbb{P}(\mathrm{T}' \in \mathcal{C})$, as required.

Next suppose that vertex 1 is an ancestor of vertex 2 in t. The forest $f(\mathrm{t})$ contains d_1+d_2+1 trees; list their roots as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2}$ and $r_{d_1+d_2+1}$ are the roots of the trees containing vertices 2 and 1, respectively. (This also means that $r_{d_1+d_2+1}$ is the root of t, and that $r_{d_1+d_2}$ is a child of 1 in t.) In this case there are $\binom{d_1+d_2-1}{d_2}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t})$: these are precisely the trees obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subseteq [d_1+d_2-1]$ of size d_2.
- Add edges from vertex 2 to the vertices in the set $\{r_i, i \in S\}$, and add edges from vertex 1 to the vertices in the set $\{r_i, i \in [d_1+d_2] \setminus S\}$.

Similarly, there are $\binom{d_1+d_2-1}{d_1}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t}^{12})$; these are obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subseteq [d_1+d_2-1]$ of size d_1.
- Add edges from vertex 2 to the vertices in the set $\{r_i, i \in S\}$, add edges from vertex 2 to the vertices in the set $\{r_i, i \in [d_1+d_2] \setminus S\}$, then swap the labels of vertices 1 and 2.

It follows that

$$|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}| = \binom{d_1+d_2-1}{d_2} + \binom{d_1+d_2-1}{d_1} = \binom{d_1+d_2}{d_1},$$

and likewise $|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}| = \binom{d_1+d_2}{d_1+1}$. (We omit the details for this last identity as they are so similar to the previous arguments.) It follows that in this case we also have $|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}|/|\mathcal{T}_{\mathrm{d}}| = |\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}|/|\mathcal{T}_{\mathrm{d}'}|$, so again $\mathbb{P}(\mathrm{T} \in \mathcal{C}) = \mathbb{P}(\mathrm{T}' \in \mathcal{C})$.

Finally, if vertex 2 is an ancestor of vertex 1 in t, then in t^{12} vertex 1 is an ancestor of vertex 2, so since $\mathrm{t} \sim \mathrm{t}^{12}$, this situation is already handled by the previous case. □

For the proof of Proposition 2.4.2 we require an additional lemma, which although fairly straightforward we find independently pleasing. Write $\binom{[n]}{k} = \{S \subseteq [n] : |S| = k\}$. Below we use the convention that $\max \emptyset = 0$.

Lemma 2.4.3 (Eggs-in-one-basket lemma). *Fix non-negative real numbers* $0 < a_1 \leq \ldots \leq a_n$ *and integers* k, ℓ *with* $n/2 \leq k < \ell \leq n$.

1. *Let* $\mathrm{A} \in_u \binom{[n]}{k} \cup \binom{[n]}{n-k}$ *and* $\mathrm{A}' \in_u \binom{[n]}{\ell} \cup \binom{[n]}{n-\ell}$. *Then* $\max(a_i, i \in \mathrm{A}') \preceq_{\mathrm{st}} \max(a_i, i \in \mathrm{A})$.
2. *Let* $\mathrm{B} \in_u \binom{[n-1]}{k} \cup \binom{[n-1]}{n-k}$ *and* $\mathrm{B}' \in_u \binom{[n-1]}{\ell} \cup \binom{[n-1]}{n-\ell}$. *Then* $\max(a_i, i \in \mathrm{B}') \preceq_{\mathrm{st}} \max(a_i, i \in \mathrm{B})$.

The proverb “don’t put all your eggs in one basket” means “don’t put all your resources into a single endeavour” or, more pithily, “diversify your portfolio”. (Its origins are obscure but an Italian equivalent, “non mettere tutte le uova in un solo cesto”, has been traced to at least 1666 [100].) To understand our use of this phrase, note that if the “portfolio” is the random set A or the set B from the lemma, and the payoff of a portfolio is the value of its largest element, then the lemma implies that larger-entropy portfolios have stochastically higher payoffs. To our knowledge, this lemma is the first rigorous proof of of the wisdom of the proverb.

Proof of Lemma 2.4.3. If $a_1, \ldots, a_\ell$ are all equal then the result is obvious so we hereafter assume that this is not the case. It suffices to prove the lemma with $\ell = k+1$; the general case follows by induction. It is useful to set $a_0 = 0$. Then for any $0 \le i < n$ and $a_i < x \le a_{i+1}$,

$$\begin{aligned}\mathbb{P}\left(\max\{a_j : j \in \mathrm{A}\} < x\right) &= \mathbb{P}\left(\mathrm{A} \cap \{i+1, \ldots, n\} = \emptyset\right) \\ &= \frac{1}{2\binom{n}{k}}\left[\binom{i}{k} + \binom{i}{n-k}\right].\end{aligned}$$

To prove the first claim of the lemma, that $\max(a_i, i \in \mathrm{A}') \preceq_{\mathrm{st}} \max(a_i, i \in \mathrm{A})$, it thus suffices to show that

$$\frac{1}{2\binom{n}{k}}\left[\binom{i}{k} + \binom{i}{n-k}\right] \le \frac{1}{2\binom{n}{k+1}}\left[\binom{i}{k+1} + \binom{i}{n-k-1}\right].$$

It is possible that some of the binomial coefficients above are zero; regardless, multiplying through by $2\binom{n}{n-i}$ and rearranging terms yields that this is equivalent to showing that

$$\binom{n-k}{n-i} - \binom{n-k-1}{n-i} \le \binom{k+1}{n-i} - \binom{k}{n-i},$$

which by the addition rule of binomials reduces to

$$\binom{n-k-1}{n-i-1} \le \binom{k}{n-i-1}.$$

This holds, because $k > n-k-1$.

For the second claim of the lemma, note that B has the law of A conditional on the event $\{n \notin \mathrm{A}\}$, which, by symmetry, has probability $1/2$. This implies that for any $x \le a_n$,

$$\mathbb{P}\left(\max\{a_i : i \in \mathrm{A}\} < x\right) = \mathbb{P}\left(n \notin \mathrm{A}\right)\mathbb{P}\left(\max\{a_i : i \in \mathrm{A}\} < x \mid n \notin \mathrm{A}\right) = \frac{1}{2}\mathbb{P}\left(\max\{a_i : i \in \mathrm{B}\} < x\right),$$

and similarly,

$$\mathbb{P}\left(\max\{a_i : i \in \mathrm{A}'\} < x\right) = \frac{1}{2}\mathbb{P}\left(\max\{a_i : i \in \mathrm{B}'\} < x\right).$$

Therefore,

$$\begin{aligned}&\mathbb{P}\left(\max\{a_i : i \in \mathrm{A}\} < x\right) - \mathbb{P}\left(\max\{a_i : i \in \mathrm{A}'\} < x\right)\\ &= \frac{1}{2}\left(\mathbb{P}\left(\max\{a_i : i \in \mathrm{B}\} < x\right) - \mathbb{P}\left(\max\{a_i : i \in \mathrm{B}'\} < x\right)\right),\end{aligned}$$

so the second claim of the lemma follows from the first. □

Proof of Proposition 2.4.2. Fix a $\sim$-equivalence class $\mathcal{C}$ and a tree $\mathrm{t} \in \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We first suppose that neither vertex 1 nor vertex 2 is an ancestor of the other in t. The forest $f(\mathrm{t})$ contains $d_1 + d_2 + 1$ trees; list their roots as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2+1}$ is the root of t, and let t_i be the tree with root r_i. Then both 1 and 2 lie in the tree $\mathrm{t}_{d_1+d_2+1}$ rooted at $r_{d_1+d_2+1}$. Write h_1 and h_2 for the distance from $r_{d_1+d_2+1}$ to 1 and 2, respectively.

By the definition of $\mathcal{C}$, starting from $f(\mathrm{t})$ we may sample $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ as follows.

- Let $(\mathrm{C}, \mathrm{A}) \in_u \left\{(1, S) : S \in \binom{[d_1+d_2]}{d_1}\right\} \cup \left\{(2, S) : S \in \binom{[d_1+d_2]}{d_2}\right\}$.
- Add edges from vertex 1 to the roots $\{r_i, i \in \mathrm{A}\}$ and from vertex 2 to the roots $\{r_i, i \in [d_1 + d_2] \setminus A\}$.
- If $\mathrm{C} = 2$ then swap the labels of vertices 1 and 2.

Note that $\mathrm{A} \in_u \binom{[d_1+d_2]}{d_1} \cup \binom{[d_1+d_2]}{d_2}$. For $1 \le i \le d_1 + d_2$ letting $a_i = 1 + \mathrm{ht}(\mathrm{t}_i)$, with the above construction of $\mathrm{T}_{\mathcal{C}}$ we then have

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + \max(a_i, i \in \mathrm{A}), h_2 + \max(a_i, i \in [d_1 + d_2] \setminus \mathrm{A})).$$

Next, again starting from $f(\mathrm{t})$, apply the same procedure (with d_1 and d_2 replaced by $d_1 + 1$ and $d_2 - 1$, respectively) to sample $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We obtain

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + \max(a_i, i \in \mathrm{A}'), h_2 + \max(a_i, i \in [d_1 + d_2] \setminus \mathrm{A}')).$$

where $\mathrm{A}' \in_u \binom{[d_1+d_2]}{d_1+1} \cup \binom{[d_1+d_2]}{d_2-1}$.

Since A and $[d_1 + d_2] \setminus \mathrm{A}$ have the same distribution, as do A' and $[d_1 + d_2] \setminus \mathrm{A}'$, we may assume without loss of generality that $h_1 \ge h_2$. It then follows that both the above maxima are at least $M^- := \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_2 + \max(a_i, i \in [d_1+d_2]))$ and at most $M^+ := \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + \max(a_i, i \in [d_1+d_2]))$, so for $x \le M^-$ we have

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) = 0 = \mathbb{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x)$$

while for $x > M^+$ we have

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) = 1 = \mathbb{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x).$$

For $M^- < x \le M^+$, we have

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) = \mathbb{P}(h_1 + \max(a_i, i \in \mathrm{A}) < x)$$

and

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x) = \mathbb{P}(h_1 + \max(a_i, i \in \mathrm{A}') < x)\,,$$

so the first part of the eggs-in-one-basket lemma yields that

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) \le \mathbb{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x).$$

This establishes that $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$ when neither 1 nor 2 is an ancestor of the other for trees in $\mathcal{C}$.

We next suppose that either 1 is an ancestor of 2 in t or vice-versa. Note that $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ contains a tree in which 1 is an ancestor of 2 if and only if it contains a tree in which 2 is an ancestor of 1. It follows that, by replacing t by another element of $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ if necessary, we may assume that in fact 1 is an ancestor of 2 in t.

List the roots of the trees in $f(\mathrm{t})$ as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2}$ and $r_{d_1+d_2+1}$ are the roots of the trees containing vertices 2 and 1, respectively, and write t_i for the tree of $f(\mathrm{t})$ with root r_i. Necessarily $r_{d_1+d_2+1}$ is also the root of t, and $r_{d_1+d_2}$ is a child of 1 in t. Write h_1 and h_2 for the distance from $r_{d_1+d_2+1}$ to 1 and from $r_{d_1+d_2}$ to 2, respectively.

By the definition of the equivalence class $\mathcal{C}$, starting from the forest $\mathrm{t}_0, \ldots, \mathrm{t}_{d_1+d_2}$, we may sample $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ as follows.

- Let

$$(\mathrm{C},\mathrm{B}) \in_u \left\{(1,S) : S \in \binom{[d_1+d_2-1]}{d_2}\right\} \cup \left\{(2,S) : S \in \binom{[d_1+d_2-1]}{d_1} : \right\}.$$

- Add edges from vertex 2 to the roots $\{r_i, i \in \mathrm{B}\}$ and from vertex 1 to the roots $\{r_i, i \in [d_1+d_2] \setminus \mathrm{B}\}$.

- If $\mathrm{C} = 2$ then swap the labels of vertices 1 and 2.

Note that $\mathrm{B} \in_u \{S \subseteq [d_1+d_2-1] : |S| \in \{d_1, d_2\}\}$. Moreover, letting $a_i = 1 + \mathrm{ht}(\mathrm{t}_i)$ for $i \in [d_1+d_2-1]$, and letting $H = \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + 1 + \mathrm{ht}(\mathrm{t}_{d_1+d_2}))$, with the above construction of $\mathrm{T}_{\mathcal{C}}$ we then have

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(H, h_1 + \max(a_i, i \in [d_1+d_2] \setminus \mathrm{B}), h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}))\,.$$

The term H accounts for the possibility that the height of $\mathrm{T}_{\mathcal{C}}$ is achieved by a vertex of either $\mathrm{t}_{d_1+d_2}$ or $\mathrm{t}_{d_1+d_2+1}$.

Next, again starting from $f(\mathrm{t})$, apply the same procedure (with d_1 and d_2 replaced by d_1+1 and d_2-1, respectively) to sample $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We obtain

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(H, h_1 + \max(a_i, i \in [d_1+d_2] \setminus \mathrm{B}'), h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}'))\,.$$

where $\mathrm{B}' \in_u \{S \subseteq [d_1+d_2-1] : |S| \in \{d_1+1, d_2-1\}\}$.

The heights of $\mathrm{T}_{\mathcal{C}}$ and $\mathrm{T}'_{\mathcal{C}}$ both lie between

$$M^- := \max(H, h_1 + \max(a_i, i \in [d_1+d_2-1]))$$

and

$$M^+ := \max(H, h_1 + 1 + h_2 + \max(a_i, i \in [d_1+d_2-1]))$$

so for $x \notin (M^-, M^+]$ we have $\mathbb{P}(\mathrm{ht}(\mathrm{T}_C) < x) = \mathbb{P}(\mathrm{ht}(\mathrm{T}'_C) < x)$. For $M^- < x \leq M^+$, we have $\mathrm{ht}(\mathrm{T}_C) < x$ if and only if $h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}) < x$, and likewise $\mathrm{ht}(\mathrm{T}'_C) < x$ if and only if $h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}') < x$. It thus follows by the second part of the eggs-in-one-basket lemma that

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) \leq \mathbb{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x).$$

This establishes that $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$ when 1 is an ancestor of 2 in t. (As already noted, this also handles the case where 2 is an ancestor of 1.)

It remains to establish strict stochastic inequality when there are at least three non-leaf vertices. We accomplish this by showing that in this case there exists $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$ such that for $\mathcal{C}$ the $\sim$-equivalence class of t, if $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}'}$ then $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \prec_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$.

By relabeling we may assume that vertex 3 is not a leaf. (We also still assume that $d_1 \geq d_2 \geq 1$.) Consider a tree $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$ with root 1, such that 3 is a child of 2 and 2 is a child of 1, and such that all other children of vertices 1 and 2 are leaves (see Figure 2.6). Then the $\sim$-equivalence class $\mathcal{C}$ of t contains $\binom{d_1+d_2-1}{d_1}$ trees with root 2 and $\binom{d_1+d_2-1}{d_2}$ trees with root 1, so $\binom{d_1+d_2}{d_1}$ trees in total.

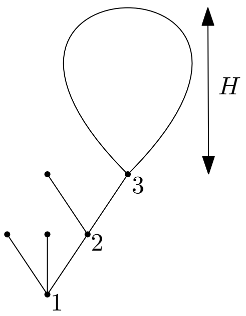


Fig. 2.6: A schematic depiction of a tree in an equivalence class $\mathcal{C}$ for which, if $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}'}$ then $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \prec_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$.

Let H be the height of the subtree of t rooted at vertex 3 (this is at least 1 by the assumption that 3 is not a leaf). Then the height of $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ is either $H+1$ or $H+2$, and is $H+2$ precisely if either 1 is the root and 3 is a child of 2, or if 2 is the root and 3 is a child of 1. The total number of trees in $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ with height $H+2$ is thus $2\binom{d_1+d_2-2}{d_1-1}$, so

$$\mathbb{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = H+2) = 2\binom{d_1+d_2-2}{d_1-1}\binom{d_1+d_2}{d_1}^{-1} = \frac{2d_1d_2}{(d_1+d_2-1)(d_1+d_2)}.$$

This probability decreases if d_1 and d_2 are replaced by d_1+1 and d_2-1, respectively, which establishes the required strict stochastic domination. □

2.5 Future directions

Although part of our point in this work is to show that tail bounds for the height do not rely on being in a setting where there is convergence to a limiting tree (or limiting space), our results and techniques can be useful for *proving* such convergence. For example, note that for $k \in \mathbb{N}$

and d a degree sequence corresponding to a tree with at least k leaves, for $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}$, the subtree of t(v) spanned by the k smallest labeled leaves is encoded by the first $j(k)$ elements of v, where $j(k)$ is the index of the k'th repeat in v. Since the law of $\mathrm{T}_{\mathrm{d}} \in_u \mathcal{T}_{\mathrm{d}}$ is invariant under relabeling of the leaves by an independent uniform permutation, our sampling technique thus gives direct access to the law of the subtree spanned by a uniformly random set of k leaves, also called the *k-dimensional distribution* of the tree. Furthermore, our approach to bounding the height in Theorem 2.1.7 in fact yields a stronger result: we show that, when the tree is grown sequentially according to the bijection (as illustrated in Figure 2.2), *all* vertices in the tree are close (on the scale of $\Theta(\sqrt{n})$) to the tree built in the first $\Theta(\sqrt{n})$ steps. The counterpart of this fact and convergence of the k-dimensional distributions are the two ingredients of Aldous' proof that a uniformly random tree $\mathrm{T}_n \in_u \mathcal{T}(n)$ converges in distribution to the Brownian continuum random tree after rescaling [13, Theorem 8]. Therefore, our techniques give immediate access to comparable results for suitable sequences of trees with a given degree sequence, simply generated trees and conditioned Bienaymé trees. This idea is exploited for various laws on trees with a fixed degree sequence by Arthur Blanc-Renaudie in [25], to prove rescaled convergence of such trees toward inhomogeneous continuum random trees.

Our approach is also useful for the study of other random tree models. Indeed, via the bijection of Section 2.2, any distribution on labeled rooted trees gives rise to a distribution on sequences. If one can understand the law of the first k repeated entries in such random sequences, one thereby obtains the law of the subtree spanned by the k smallest labeled leaves. Our arguments for controlling the height can then in principle be combined with this spanning subtree information in order to prove both global height bounds and rescaled convergence in distribution (when the trees are viewed as measured metric spaces). On the other hand, by considering any natural distribution on sequences, the bijection yields a random tree model. Here is one natural example: consider a multiset $\mathcal{M}$ of elements of $[n]$ with size (counted with multiplicity) at least $n-1$. Then construct a sequence of length $n-1$ by uniform sampling without replacement from $\mathcal{M}$, and consider the tree T on $[n]$ encoded by this sequence via the bijection. (If $\mathcal{M}$ itself has $n-1$ elements then T has the law of a uniform tree with a given degree sequence, where for $i \in [n]$, the multiplicity of i in $\mathcal{M}$ is the degree of i in T. More generally, the number of copies of i in the subsample is the degree of i in T, and T is a uniform tree with this degree sequence.) It turns out that T has the law of a uniform spanning tree of a random *tree-rooted graph* with degree sequence defined by $\mathcal{M}$. Tree-rooted graphs are the non-planar analogues of tree-rooted maps [102, 88, 28, 20], whose random instantiations are active objects of study in the planar probability and statistical physics community [78, 63, 61, 67, 84, 62]. In [5], a different sampling method is used to show that in the finite variance regime, the spanning trees of large random tree-rooted graphs converge after rescaling to the Brownian continuum random tree. In future work, we plan to build on the current work to study distances in and convergence of random tree-rooted graphs for other degree regimes.

Our results can also be applied to many models of random graphs that contain cycles, because the tree models that we study are important building blocks for many such sparse random graphs. One family of examples is provided by the *configuration model*, which is used to sample graphs with a given degree sequence. In a number of recent works [24, 37, 23], it is shown that, under conditions which ensure that the resulting random graphs are in some sense "critical", the large components in the configuration model converge in distribution to random compact measured metric spaces, after rescaling. In all aforementioned papers, the techniques rely on studying spanning trees of the large components, and connecting these to the models for random trees that we study. However, none of those works achieve control over inter-vertex distances in smaller components. The reason for this is precisely that these components do not necessarily have scaling limits (in particular because some of them will have "atypical" degree distributions), and so previous results bounding the diameters of random trees do not apply. This lack of control prohibits the authors from proving convergence of the ordered sequence of components in any topology stronger than the product topology. As a consequence, important statistics such as the diameter and the length of the longest path cannot be shown to converge under rescaling. We believe that, with the results and techniques of the current work in hand, it is now possible to bound distances in all components, thereby proving convergence in a stronger topology (and deducing convergence in distribution for the rescaled diameter), in all the above works.

Finally, we believe that our results can be used to obtain non-asymptotic tail bounds on distances in uniform connected graphs with a given degree sequence and fixed surplus, which occur as the components of the configuration model. Such graphs can be related to trees with a fixed degree sequence by uniform cycle breaking, and twice the height of the resulting tree is an upper bound for the diameter in the graph. The law on trees with a given degree sequence induced by this cycle-breaking procedure has a non-trivial bias relative to the uniform distribution , whose form we believe is predicted by analogous results for components in the

Erdős–Rényi random graph [7]. By studying the bias it should be possible to translate our results to this set-up.

Chapter 3

Universality for the directed configuration model: metric space convergence of the strongly connected components at criticality

This chapter is joint work with Zheneng Xie and previously appeared as a preprint [48][1]

We consider the strongly connected components (SCCs) of a uniform directed graph on n vertices with i.i.d. in- and out-degree pairs distributed as (D^-, D^+), with $\mathbb{E}[D^+] = \mathbb{E}[D^-] = \mu$. We condition on equal total in- and out-degree. A phase transition for the emergence of a giant SCC is known to occur at the critical value $\mathbb{E}[D^- D^+] = \mu$. We study the model at this critical value and, additionally, require that $\mathbb{E}[(D^-)^i (D^+)^j] < \infty$ for all $i + j \leq 3$, and for $(i, j) = (1, 3)$ and $(i, j) = (3, 1)$. We show that, under these conditions, the SCCs ranked by decreasing number of edges with distances rescaled by $n^{-1/3}$ converge in distribution to a sequence of finite strongly connected directed multigraphs with edge lengths, and that these are either 3-regular or loops. The limit objects lie in a 3-parameter family, which contains the scaling limit of the SCCs in the directed Erdős-Rényi model at criticality as found by Goldschmidt and Stephenson (2019). This is the first universality result for the scaling limit of a critical directed graph model and the first quantitative result on the directed configuration model at criticality. As an immediate consequence, the largest SCCs at criticality contain $\Theta(n^{1/3})$ vertices and edges

---

[1] In [48], Proposition 3.4.16, in which we show that the configuration yields a simple graph with probability bounded away from 0, is missing.

in probability, and the diameter of the directed graph at criticality is $\Omega(n^{1/3})$ in probability. We use a metric on the space of weighted multigraphs in which two multigraphs are close if there are compatible isomorphisms between their vertex and edge sets which roughly preserve the edge lengths. We use the product topology on the sequence of multigraphs. Our method of proof involves a depth-first exploration of the directed graph, resulting in a spanning forest with additional identifications, of which we study the limit under rescaling.

3.1 Introduction

3.1.1 Overview

Edges in real-world networks are often directed, such as links on the world wide web, "follows" on Twitter, financial transactions or disease transmission in a social network. When analysing networks, the first quantity that is often considered is the distribution of the degrees of nodes in the network. In this paper we will consider sampling an i.i.d. sequence of in- and out-degrees, conditional on the total in-degree being equal to the total out-degree. We will then sample a uniform directed graph (digraph) with the given degree sequence. Results on such graphs are a useful benchmark, exposing additional underlying structure of a real-world network compared to a uniformly random graph with its degree sequence.

When considering such models, previous work by Cooper and Frieze [38] (which we will discuss in more detail in Section 3.1.6) shows that there exists a phase transition in the strong directed connectivity of the graph. Two vertices are part of the same *strongly connected component* (SCC) if and only if there exists a directed cycle that contains both of them. Above some threshold, there will exist a unique giant SCC that occupies a positive proportion of the vertices whereas, below the threshold no SCC will occupy a positive proportion of the vertices. In Figure 3.1, a directed graph and its strongly connected components are depicted. In this paper we will prove the first detailed results about the critical case - specifically, that there exists a sequence of random weighted directed multigraphs that can be understood as the scaling limit of the SCCs when viewed in decreasing order of size.

3.1.2 Directed graphs

There are two notions of connectivity when working with a directed graph: weak and strong connectivity. We will be working with the strong notion. We say a vertex v leads to a vertex

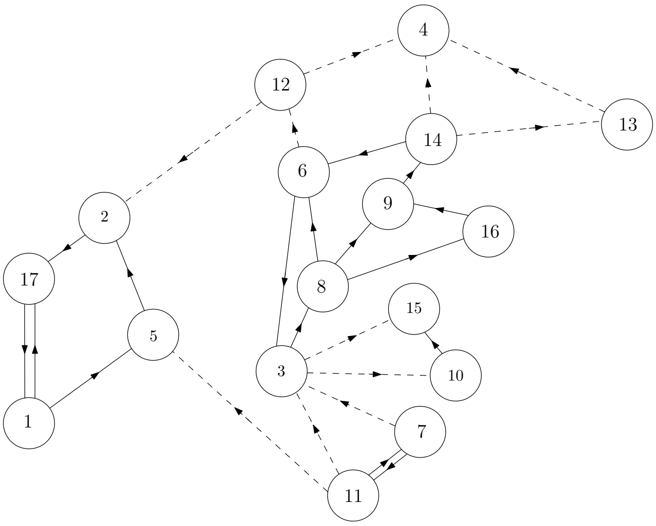


Fig. 3.1: A directed graph on [17]. The strongly connected components have vertex sets $\{1, 2, 5, 17\}$, $\{3, 6, 8, 9, 14, 16\}$, $\{7, 11\}$, $\{4\}$, $\{10\}$, $\{12\}$, $\{13\}$, and $\{15\}$. Edges that are not part of an SCC are depicted as dashed arrows. Taken from [59] with permission of the authors.

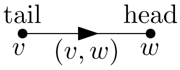


Fig. 3.2: An edge (v, w) will be depicted as an arrow from v to w.

w, written $v \to w$, if there exists a directed path from v to w in the graph. We say v is *strongly connected to* w, written $v \leftrightarrow w$, if v leads to w and w leads to v. By convention, v leads to itself. A graph is *strongly connected* if all pairs of vertices in the graph are strongly connected. The relation $v \leftrightarrow w$ is an equivalence relation; the digraphs induced by the equivalence classes of $\leftrightarrow$ are referred to as the *strongly connected components* (SCCs). For each vertex v in a directed graph $\vec{G}$, we will use the notation $d^-(v)$ for the in-degree of v and $d^+(v)$ for the out-degree of v. Moreover, a directed edge (v, w) has *tail* v and *head* w (see Figure 3.2).

3.1.3 Description of the model

First consider a deterministic degree sequence $\mathbf{d}_1, \ldots, \mathbf{d}_n$ where $\mathbf{d}_i = (d_i^-, d_i^+) \in \mathbb{N} \times \mathbb{N}$ for $i = 1, \ldots, n$. We say a directed graph with vertex set $[n]$, where $[n] = \{1, \ldots, n\}$, has degree sequence $\mathbf{d}_1, \ldots, \mathbf{d}_n$ if $(d^-(i), d^+(i)) = (d_i^-, d_i^+)$ for $i = 1, \ldots, n$.

In order to sample a uniformly random graph with a given degree sequence, we first consider the *directed configuration model* introduced by Cooper and Frieze [38]. Take n vertices $v_1, \ldots, v_n$ such that v_i has d_i^- in-half-edges and d_i^+ out-half-edges. Then construct a multigraph by choosing a uniformly random pairing of the in-half-edges with the out-half-edges. Cooper and Frieze [38, Sec. 2.1] proved that if we condition on the resulting multigraph being simple, we obtain a uniformly chosen random digraph with the given degree sequence.

In this paper we will consider the case where the degree sequence consists of n i.i.d. random variables conditioned on the total in-degree being equal to the total out-degree. Let ν be a distribution on $\mathbb{N} \times \mathbb{N}$, and let $\mathbf{D}_1, \dots, \mathbf{D}_n$ be a sequence of i.i.d. random variables with distribution ν. We condition on the event

$$\left\{\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+\right\},$$

observing that this is an asymptotically singular event as $n \to \infty$. We also condition on the existence of a digraph with the degree sequence. Let $\vec{G}_n(\nu)$ be a digraph chosen uniformly at random from all digraphs with degree sequence $\mathbf{D}_1, \dots, \mathbf{D}_n$. We are interested in the limit under rescaling of the SCCs of $\vec{G}_n(\nu)$ as $n \to \infty$.

Suppose (D^-, D^+) has law ν. We will require the following assumptions to hold:

1. $\mathbb{E}[(D^-)^i (D^+)^j] < \infty$ for $1 \leq i + j \leq 3$, $(i, j) = (1, 3)$ and $(i, j) = (3, 1)$.
2. $\mathbb{E}[D^-] = \mathbb{E}[D^+]$.
3. $D^- - D^+$ is strongly aperiodic. This means that for all $p > 1$, there does not exist $k \in \mathbb{Z}$ such that
$$\mathbb{P}(D^- - D^+ \in k + p\mathbb{Z}) = 1.$$
4. $\mathbb{E}[D^- D^+] = \mathbb{E}[D^-]$.

The first condition is required to ensure that the steps of a random walk used in the proof have finite variance, so that the random walk will convergence under rescaling to a Brownian motion. It also ensures similar regularity of other random variables that we use to encode the directed graph. (We discuss relaxing the moment conditions in Subsection 3.6.)

The second and third conditions make sure the event $\{\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+\}$ is well-behaved. The second condition ensures that it is not a large deviation event. Using a result from Spitzer [96, Page 42, P1], the third condition ensures that the event has positive probability for all sufficiently large $n \geq 1$. This condition can be relaxed to assuming that $D^- - D^+$ is non-constant by taking limits for $n \in p\mathbb{N}$ rather than $n \in \mathbb{N}$ where p is the periodicity of $D^- - D^+$. However, for simplicity of presentation, we will keep it as an assumption.

The fourth assumption is the criticality condition. To understand how this arises, consider the directed configuration model and let (V_n, W_n) be a uniformly chosen edge. For now, ignore the conditioning on the total in- and out-degrees being equal. We consider the distribution of the in- and out-degree of W_n. Because the degree sequence is an i.i.d. sequence, W_n is equally likely to be any vertex i. Thus for any $\mathbf{k} = (k^-, k^+)$,

$$\begin{aligned}\mathbb{P}(d^-(W_n) = k^-, d^+(W_n) = k^+) &= n\mathbb{P}(W_n = 1, \mathbf{D}_1 = \mathbf{k}) \\ &= n\mathbb{E}[\mathbb{P}(W_n = 1 \mid \mathbf{D}_1 = \mathbf{k}, \mathbf{D}_2, \dots, \mathbf{D}_n)]\mathbb{P}(\mathbf{D}_1 = \mathbf{k})\end{aligned}$$

Conditionally on the degree sequence, we have that $W_n = i$ with probability proportional to D_i^- since we used an uniform pairing of the in- and out-half-edges. Therefore

$$\mathbb{P}(W_n = 1 \mid \mathbf{D}_1 = \mathbf{k}, \mathbf{D}_2, \dots, \mathbf{D}_n) = \frac{k^-}{k^- + \sum_{i=2}^n D_i^-}.$$

Thus

$$\mathbb{P}(d^-(W_n) = k^-, d^+(W_n) = k^+) = \mathbb{E}\left[\frac{k^-}{\frac{1}{n}\left(k^- + \sum_{i=2}^n D_i^-\right)}\right]\mathbb{P}\left[D^- = k^-, D^+ = k^+\right].$$

Using the strong law of large numbers and the bounded convergence theorem, the above will converge to

$$\frac{k^-}{\mathbb{E}[D^-]}\mathbb{P}\left[D^- = k^-, D^+ = k^+\right].$$

Let (Z^-, Z^+) be such that $P(Z^- = k^-, Z^+ = k^+)$ is given by the above expression. We say (Z^-, Z^+) has the law of the *degree distribution size-biased by in-degree.* For large n, any fixed out-edge of W_n is then also distributed approximately like a uniformly chosen edge (here we are ignoring the fact that we have already sampled an edge) since we chose the in- and out-edge pairing uniformly at random. Therefore the out-degree of the head will have approximately the same distribution as Z^+. Thus if we were to look at the graph of all vertices leading from W_n, it would look approximately like a Bienaymé tree[2] with offspring distribution Z^+. It is well known that such trees exhibit critical behaviour in whether or not the tree is finite at $\mathbb{E}[Z^+] = 1$. This is equivalent to assuming $\mathbb{E}[D^-D^+] = E[D^-]$.

Cooper and Frieze [38] studied this phase transition for a deterministic degree sequence

---

[2] For μ a probability distribution on $\mathbb{N}$, a Bienaymé tree with offspring distribution μ is the family tree of a branching process with offspring distribution μ. Bienaymé trees are often referred to as Galton–Watson trees, but we decide to follow the name change suggested by Addario-Berry et al. [10].

$\mathbf{d}_1, \ldots, \mathbf{d}_n$. They defined the parameter

$$d = \frac{\sum_{i=1}^n d_i^+ d_i^-}{\sum_{i=1}^n d_i^-}$$

which is a counterpart of $\mathbb{E}[Z^+]$ for deterministic degree sequences. They then showed that, under additional assumptions, there exists a phase transition for the existence of a giant SCC depending on whether d is strictly greater than or less than 1. Our work in this paper shows our corresponding condition, $\mathbb{E}[Z^+] = 1$, is also the correct criticality condition to take for i.i.d. random degree sequences.

We define the following parameters that will determine the behaviour of the SCCs in the limit.

1. $\mu := \mathbb{E}[D^-] = \mathbb{E}[D^+] = \mathbb{E}[D^- D^+]$
2. $\nu_- := \mathbb{E}[Z^-] - 1 = \frac{\mathbb{E}[(D^-)^2] - \mu}{\mu} > 0$
3. $\sigma_-^2 := \mathrm{Var}(Z^-) = \frac{\mu \mathbb{E}[(D^-)^3] - \mathbb{E}[(D^-)^2]^2}{\mu^2}$
4. $\sigma_+^2 := \mathrm{Var}(Z^+) = \frac{\mathbb{E}[D^-(D^+)^2] - \mu}{\mu}$
5. $\sigma_{-+} := \mathrm{Cov}(Z^-, Z^+) = \frac{\mathbb{E}[(D^-)^2 D^+] - \mathbb{E}[(D^-)^2]}{\mu}$

3.1.4 Metric directed multigraphs and kernels

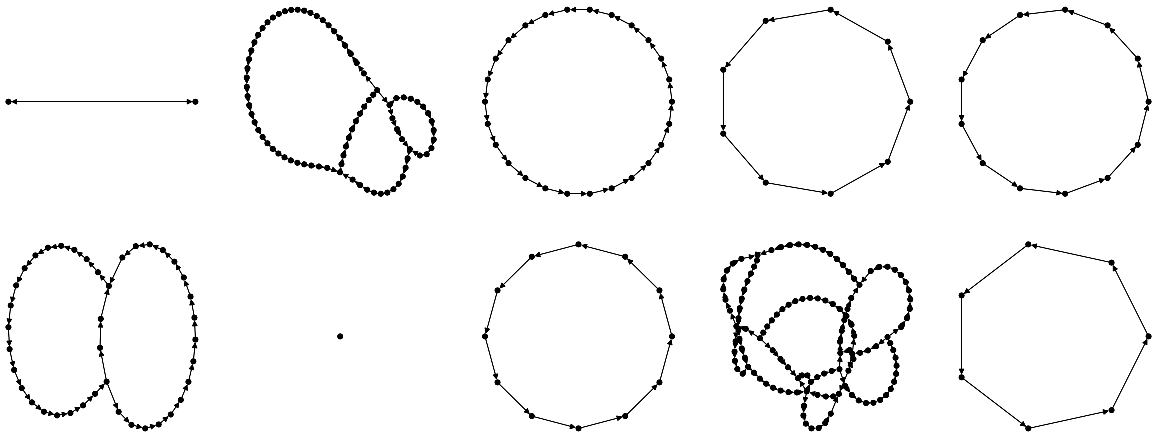


Fig. 3.3: The largest SCC from samples of a directed configuration model with independent Poisson(1) in- and out-degrees

Figure 3.3 shows the largest SCC from samples of a directed configuration model. As can be seen, while the lengths of paths in the SCC are long, the actual structure of the SCC is often quite simple. Previous work by Goldschmidt and Stephenson [59], studying the directed Erdős-Rényi graph, formalised this using *metric directed multigraphs* (MDMs), and we follow the same approach. These are simply weighted directed multigraphs, but in our context it is more appropriate to think of the weights as lengths, which motivates the change in naming. Formally, a *directed multigraph* is a tuple (V, E, r) where

1. V is a set of *vertices*,

2. E is a set of *edges*, and

3. $r : E \to V \times V$ is a function mapping each edge to its *head* and *tail*; associated with r are two functions $r_1 : E \to V$ and $r_2 : E \to V$ such that

   $$r(e) = (r_1(e), r_2(e))$$

   for all $e \in E$. $r_1(e)$ is the tail of the edge e and $r_2(e)$ is the head of the edge e.

Then a *metric directed multigraph (MDM)* is a tuple $M = (V, E, r, l)$ where (V, E, r) is a directed multigraph and $l : E \to [0, \infty)$. Let $\mathfrak{L}$ denote the MDM consisting of a single vertex with a self-loop of length 0.

An *isomorphism* between two MDMs $M = (V, E, r, l)$ and $M' = (V', E', r', l')$ is a pair of functions (i_V, i_E) where $i_V : V \to V'$ and $i_E : E \to E'$ are bijections satisfying the relation

$$r'(i_E(e)) = (i_V(r_1(e)), i_V(r_2(e)))$$

for all $e \in E$. We say two MDMs are *isomorphic* if there exists an isomorphism between them. In other words, isomorphic MDMs have the same graph structures for their underlying directed multigraphs up to a relabelling of the edges and vertices. Write $\mathrm{Iso}(M, M')$ for the set of all isomorphisms between M and M'.

We now define a distance $d_{\vec{G}}$ between two MDMs M and M'. Any isomorphism between M and M' gives a correspondence between the edges of M and the edges of M'. We can then take an ℓ_∞ distance between the lengths of the edges and finally take the isomorphism which minimizes this distance. If M and M' are not isomorphic, we set the distance to be infinite.

Formally,

$$d_{\vec{G}}(M,M') = \begin{cases} \inf_{(i_V,i_E)\in \mathrm{Iso}(M,M')} \sup_{e\in E} |l(e) - l'(i_E(e))| & \text{if } M \text{ and } M' \text{ are isomorphic,} \\ \infty & \text{otherwise.} \end{cases}$$

Consider an MDM M and a vertex $w \in M$ with in-degree 1 and out-degree 1 which is not a self-loop. Let u and v be the unique in-neighbour and out-neighbour of w respectively. The MDM obtained by *smoothing* w is obtained by deleting the edges e_1 and e_2 such that $r(e_1) = (u,w)$ and $r(e_2) = (w,v)$, then adding an edge e such that $r(e) = (u,v)$ and assigning it length $l(e) = l(e_1) + l(e_2)$. This is illustrated in Figure 3.4.

(a) The graph before smoothing w (b) The graph after smoothing w

Fig. 3.4: Smoothing a vertex w

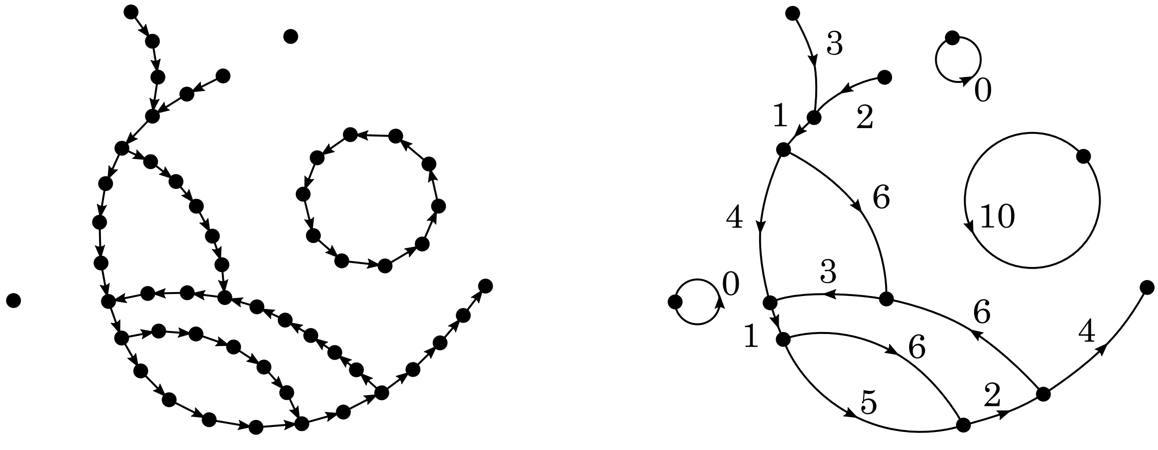
Then the kernel of a digraph $\vec{G}$ is obtained by doing the following:

1. Assign length 1 to each edge.
2. Iteratively smooth vertices with in-degree 1 and out-degree 1 that are not self-loops until there are none remaining or we have obtained a MDM consisting of one vertex with a single self-loop.
3. Replace all singletons by $\mathfrak{L}$.

An example is shown in Figure 3.5. We expect the graph structure of kernels of SCCs in the critical window to remain finite, whereas the lengths assigned to edges will tend to infinity.

3.1.5 Our results

For M an MDM and $c \in (0,\infty)$, let cM be equal to M with all lengths multiplied by c. Let $C_i(n)$ for $i \geq 1$ be the kernels of the SCCs of $\vec{G}_n(\nu)$, listed in decreasing order of number of edges, breaking ties arbitrarily. Complete the list with an infinite repeat of $\mathfrak{L}$. Then, our main theorem is as follows.



(a) $\vec{G}$ (b) Kernel of $\vec{G}$

Fig. 3.5: An example of a digraph $\vec{G}$ and its kernel. The numbers indicate the edge lengths.

Theorem 3.1.1. *There exists a random sequence $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ of strongly connected MDMs such that*

$$\left(n^{-1/3}C_i(n), i \in \mathbb{N}\right) \xrightarrow{(d)} (\mathcal{C}_i, i \in \mathbb{N})$$

as $n \to \infty$, with respect to the product $d_{\vec{\mathcal{G}}}$-topology. The law of $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ depends only on the parameters μ, σ_+, and $(\sigma_{-+} + \nu_-)/\mu$. Further, for each $i \geq 1$, $\mathcal{C}_i$ is either 3-regular or a loop.

We will describe the limit object and some of its further properties in Subsection 3.2.2.

The law of the limit object places some particular cases of our model in the universality class of the directed Erdős–Rényi model as studied by Goldschmidt and Stephenson [59]. This is the content of the following corollary. The *directed Erdős–Rényi model* on n vertices with parameter p, denoted by $\vec{G}(n, p)$, is a random digraph with vertex set $[n]$ in which each of the $n(n-1)$ possible directed edges is included with probability p independently. The cases $p = (1 + \lambda n^{-1/3})/n$ for $\lambda \in \mathbb{R}$ are referred to as *the critical window*, and the case $p = 1/n$ is called *criticality*.

Note however that their result holds in a stronger topology: they use an ℓ_1-like topology on the space of sequences of MDMs, whereas we show our result in the product topology. Due to this, it is important in their paper to consider singletons as loops of length zero. For any fixed k, the kth largest SCC will not be a singleton with high probability as $n \to \infty$. Therefore, no component of our limiting object will be a singleton. Thus they need to pad their SCCs by $\mathfrak{L}$ and consider the kernel of singletons to be $\mathfrak{L}$, to prevent the ℓ_1-distance, as defined by $d_{\vec{\mathcal{G}}}$, between $\left(n^{-1/3}C_i(n), i \in \mathbb{N}\right)$ and $(\mathcal{C}_i, i \in \mathbb{N})$ being infinite. We follow the same convention.

Corollary 3.1.2. *Consider $\vec{G}_n(\nu)$, with ν such that*

$$\mu = \sigma_+ = \sigma_{-+} + \nu_- = 1.$$

Let $(C_i^\nu(n), i \geq 1)$ be the kernels of the SCCs of $\vec{G}_n(\nu)$. Furthermore, let $(C_i^{ER}(n), i \geq 1)$ be the kernels of the SCCs of $\vec{G}(n, 1/n)$. Then, $\left(n^{-1/3} C_i^\nu(n), i \in \mathbb{N}\right)$ and $\left(n^{-1/3} C_i^{ER}(n), i \in \mathbb{N}\right)$ have the same limit in distribution in the product-$d_{\vec{\mathcal{G}}}$-topology as $n \to \infty$.

Note that the condition in Corollary 3.1.2 is satisfied by $\nu(k^-, k^+) = \nu_1(k^-)\nu_2(k^+)$, with ν_1 and ν_2 the law of a Poisson(1) random variable.

Moreover, Theorem 3.1.1 has the following immediate corollaries, which were previously unknown.

Corollary 3.1.3. *Let E_n^i and V_n^i be the number of edges and vertices in $C_i(n)$ respectively, both appended with infinite repeats of 0. Then there exists a random sequence $(E_i, i \in \mathbb{N}) \in \mathbb{R}_+^\infty$, such that*

$$\left(n^{-1/3} E_i^n, n^{-1/3} V_i^n, i \in \mathbb{N}\right) \xrightarrow{\text{(d)}} (E_i, E_i, i \in \mathbb{N})$$

as $n \to \infty$ in the product topology on $(\mathbb{R}^2)^\infty$.

In particular, note that, in the above corollary, the number of vertices and number of edges have the exact same scaling limit.

Corollary 3.1.4. *For $v, w \in \vec{G}_n(\nu)$ such that $v \to w$, let $d(v, w)$ denote the length of the shortest directed path from v to w, and let*

$$\text{Diam}\left(\vec{G}_n(\nu)\right) = \max_{v,w \in V} \{d(v, w) : v \to w\}$$

be the diameter *of $\vec{G}_n(\nu)$. Then, for any $\epsilon > 0$, there is a $\delta > 0$ such that*

$$\mathbb{P}\left(n^{-1/3} \,\text{Diam}\left(\vec{G}_n(\nu)\right) > \delta\right) > 1 - \epsilon$$

for all n large enough. Equivalently, $\text{Diam}\left(\vec{G}_n(\nu)\right) = \Omega_p(n^{1/3})$.

3.1.6 Previous work

The configuration model was introduced by Bollobás [27] to sample a uniformly random undirected graph with a given degree sequence. (For a discussion of the configuration model and proofs of standard results, we refer the reader to [101, Chapter 7].)

Most results on the configuration model are proved for models with a deterministic degree sequence. The phase transition for the undirected setting was shown in [85, 86, 69]. The law of component sizes at criticality and in the critical window were obtained by Riordan [92] under the assumption that the degrees are bounded. Dhara, van der Hofstad, van Leeuwaarden and Sen showed convergence of the size and surplus edges in the critical window with a finite third moment [44] and in the heavy-tailed regime [45]. Bhamidi, Dhara, van der Hofstad and Sen obtained metric space convergence in the critical window in [24], a result that the authors later improved to a stronger topology in [23].

Configuration models with a random degree sequence are considered in [70], [37], and [46]. Joseph [70] showed convergence of the component sizes and surpluses of the large components under rescaling at criticality, both for degree distributions with finite third moments and for the heavy-tailed regime. Conchon-Kerjan and Goldschmidt [37] show Gromov-Hausdorff-Prokhorov convergence of the rescaled components ordered by decreasing size at criticality in these two regimes. The results in [37] in the heavy-tailed regime are extended to the critical window by the first author in [46]. Our techniques are closely related to the techniques introduced in [37].

Some results have been obtained for other directed graph models. Cao and Olvera-Cravioto [33] consider a class of inhomogeneous directed random graphs. Their results include a phase transition for the existence of a giant SCC. This is a generalisation of work by Bloznelis, Götze and Jaworski in [26], in which a smaller class of inhomogeneous directed graphs is considered. Samorodnitsky, Resnick, Towsley, Davis, Willis and Wan [94] studied the tails of the degree distribution in the directed preferential attachment model. As previously mentioned, Goldschmidt and Stephenson [59] have studied the *directed Erdős–Rényi model* in the *critical window*, and were the first to obtain metric space convergence of the SCCs of a directed graph. Our methods build on their techniques.

The directed configuration model was first considered by Cooper and Frieze [38]. They consider a deterministic degree sequence under a number of conditions. As discussed previously in Section 3.1.3, a phase transition for the SCCs occurs when a parameter $d = 1$. They show that for $d < 1$, with high probability, all SCCs contain $O(\Delta \log(n))$ vertices, for Δ the maximal degree. On the other hand, for $d > 1$, there is a unique SCC that contains a positive proportion of the vertices and edges. Their conditions are restrictive, and include finite second moments for both the in- and out-degree of a uniformly chosen vertex, and a bound of size $n^{1/12}/\log(n)$

on the largest degree. Their proofs are based on an algorithm to explore the directed graph. The condition on the largest degree was later relaxed to $O(n^{1/4})$ by Graf [60]. These results are in contrast with the critical case, with Corollary 3.1.3, which says that in our set-up the number of vertices and edges in the largest strongly connected components are $\Theta(n^{1/3})$ in probability.

Recently, Cai and Perarnau have obtained a number of results on the directed configuration model with deterministic degrees. In [30], they show, under first and second moment conditions of the degree of a uniformly picked vertex, for $d \neq 1$ (i.e. not at criticality), that the diameter of the model on n vertices, rescaled by $\log(n)$ converges to a constant that they identify. This is in contrast with Corollary 3.1.4, which says that in our set-up the diameter is $\Omega(n^{1/3})$ in probability at criticality. Then, in [31], they show a law of large numbers for the number of vertices and edges in the largest SCC, under slightly stronger moment conditions, and again away from the critical point. In [32], they study the behaviour of a random walk on a directed configuration model.

A necessary and sufficient condition for the existence of a giant weakly connected component for the directed configuration model with a deterministic degree sequence is discussed in the physics literature by Kryven [73]. He also studies the distribution of the in- and out-components in [74].

The directed configuration model with random in- and out-degrees is also considered by Chen and Olvera-Cravioto [36] although, importantly, they do not allow for the in- and out-degree of a vertex to be dependent. The authors consider a model in which the in- and out-degrees are two independent sequences of i.i.d. random variables drawn from different probability distributions. They propose an algorithm to sample degree sequences that correspond to a simple graph and show the limiting distribution of the degrees generated by this algorithm.

3.1.7 Proof outline

Our techniques use height processes and Łukasiewicz paths, which are standard objects used to encode trees and forests (see for instance [50, Chapter 0]). We will introduce these here. Let $T = (V, E, \rho)$ be an ordered rooted finite tree with vertex set V, edge set E and root vertex ρ; say $|V| = n$. Let $v_0, \ldots, v_{n-1}$ denote the vertices of the tree visited in depth-first order, so that $v_0 = \rho$. We can view T as a metric space by regarding all edges as line segments of length 1 that are connected via the vertices. The distance d_T between points a_1 and a_2 on line segments l_1 and l_2 respectively is then defined as the length of the unique non-self-intersecting path between

a_1 and a_2 that traverses the line segments of the tree. Denote (T, d_T) by T.

We will define the height process and Łukasiewicz path of T. Both of these functions uniquely characterize T. The height process of T, referred to as h, is defined as

$$h(i) = d_T(v_i, v_0),$$

i.e. for all i, $h(i)$ equals the distance from v_i to the root. Moreover, for all $i = 1, \ldots, n$, let y_i be the number of children of v_{i-1}, and set $y_0 = 1$. Then, the Łukasiewicz path of T is defined by

$$s(i) = \sum_{j \leq i} (y_j - 1)$$

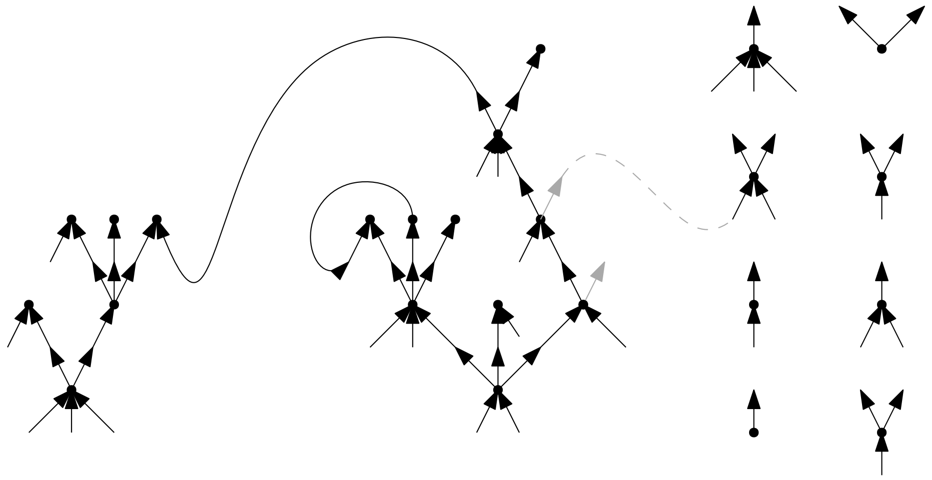
for $i = 0, \ldots, n$. Then, $s(i)$ is the total number of younger siblings of v_i and its ancestors. For a sequence of ordered rooted finite trees, we define its height process by concatenating the height processes of the trees in the sequence. The Łukasiewicz path is defined similarly.

We will study the law of the SCCs of a uniform directed graph with degree sequence $(\mathbf{D}_1, \ldots, \mathbf{D}_n)$, conditional on $\sum_{i=1}^{n} D_i^- = \sum_{i=1}^{n} D_i^+$ by exploring the configuration model in a depth-first manner. This sampling naturally gives rise to a directed subforest of the resulting multigraph, which we call the *out-forest*. The sampling procedure is described in Algorithm 1, and is also illustrated in Figure 3.6a. The definition of the out-forest is illustrated in Figure 3.6b.

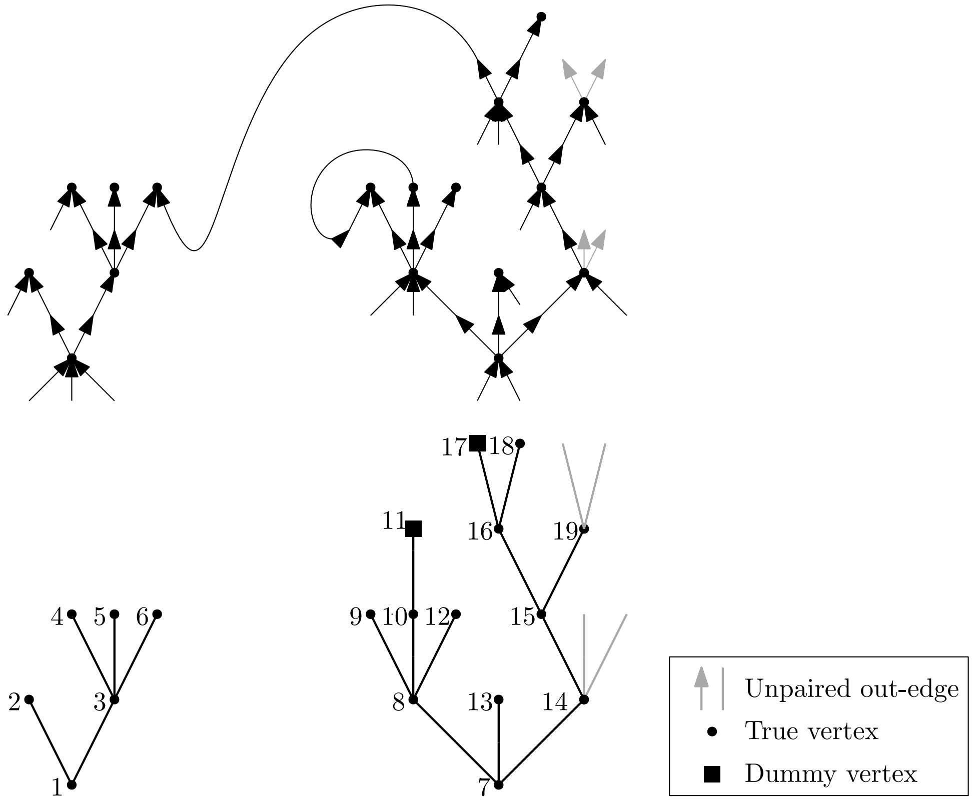
The sampling procedure uses a queue of unpaired out-edges (represented by the label of their corresponding vertex). When the queue is empty, we are at the start of a new out-component and pick a new vertex w with probability proportional to its in-degree if there are vertices with positive in-degree remaining. Else, we pick a new vertex uniformly at random. If the queue is not empty, we pair the first out-edge in the queue to a uniform unpaired in-edge and call the corresponding vertex w. In both cases, if w is not yet in the list of discovered vertices, we add the out-edges from this vertex to the *front* of the queue of edges (this choice is what makes the exploration depth-first) and add w to the list of discovered vertices. The order in which vertices are added to the list of discovered vertices is referred to as their *order of discovery*.

This procedure will discover vertices with in-degree 0 last. This is fine since such vertices form singleton SCCs, so we have discovered the non-trivial SCCs first.

At each step we also track two natural numbers $\hat{s}^-(k)$ and $\hat{s}^+(k)$. The first one, $\hat{s}^-(k)$ keeps track of the number of unpaired in-edges of discovered vertices at time k. The second one,



(a) The gray arrows represent unpaired out-half-edges of vertices that have been discovered. One by one, in depth first order, these are paired to a uniform unpaired in-half-edge.



(b) The out-forest is defined based on the exploration of the digraph. For each surplus edge, we add a dummy leaf. The labels of the vertices correspond to the time step in the exploration at which the vertex is added. The gray edges lead to vertices of which we do not know whether it is a dummy vertex, and if not, what its degree is.

Fig. 3.6: Partial constructions of the configuration model and out-forest

```
Data: A set of vertices $V = \{v_1, \ldots, v_n\}$ with degree pairs
      $(d^-(v_1), d^+(v_1)), \ldots, (d^-(v_n), d^+(v_n))$ satisfying $\sum d^-(v_i) = \sum d^+(v_i)$
$\mathcal{V} \leftarrow$ an empty ordered list of vertices    // the list of discovered vertices;
$\mathcal{Q} \leftarrow$ an empty ordered list of vertices    // the queue;
$(d^-_{\text{unpaired}}(v_1), \ldots, d^-_{\text{unpaired}}(v_n)) \leftarrow (d^-(v_1), \ldots, d^-(v_n))$    // the number of unpaired in-edges per vertex;
$k \leftarrow 0$    // the index of the current step ;
$\hat{s}^- \leftarrow 0$    // the number of unpaired in-edges of discovered vertices;
$\hat{s}^+ \leftarrow 1$    // the queue size minus the number of explored out-components ;
$F \leftarrow$ a directed forest with vertices $V$ and no edges    // current out-forest;
$M \leftarrow$ a directed multigraph with vertices $V$ and no edges    // current di-multigraph;
while there exist undiscovered vertices OR $\mathcal{Q}$ is non-empty do
    if $\mathcal{Q}$ is empty then    // we start a new out-component
        if there exist undiscovered vertices with positive in-degree then
            $w \leftarrow$ a random vertex not in $\mathcal{V}$ chosen with prob. proportional to $d^-(w)$ ;
        else
            $w \leftarrow$ a uniformly random vertex not in $\mathcal{V}$
        end
        $\hat{s}^+ \leftarrow \hat{s}^+ - 1$    // we have explored a component;
    else
        $v \leftarrow$ first entry in $\mathcal{Q}$    // we will pair an unpaired out-edge of $v$;
        remove first entry from $\mathcal{Q}$ ;
        $\hat{s}^+ \leftarrow \hat{s}^+ - 1$    // the queue size decreases by 1;
        $w \leftarrow$ a random vertex chosen with prob. proportional to $d^-_{\text{unpaired}}(w)$ ;
        add $(v, w)$ to $M$    // we pair the out-edge of $v$ with a uniform unpaired in-edge;
        $d^-_{\text{unpaired}}(w) \leftarrow d^-_{\text{unpaired}}(w) - 1$;
        $\hat{s}^- \leftarrow \hat{s}^- - 1$    // we have paired an in-edge;
        if $w \in \mathcal{V}$ then    // we sampled a surplus edge
            add a dummy leaf to $F$ and an edge from $v$ to the leaf;
        else
            add $(v, w)$ to $F$;
        end
    end
    if $w \notin \mathcal{V}$ then
        append $w$ to the end of $\mathcal{V}$    // vertex $w$ is now discovered;
        append $d^+(w)$ repeats of $w$ to the start of $\mathcal{Q}$;
        $\hat{s}^+ \leftarrow \hat{s}^+ + d^+(w)$    // the queue size has increased;
        $\hat{s}^- \leftarrow \hat{s}^- + d^-(w)$    // the number of unpaired in-edges of discovered vertices has increased;
    end
    $k \leftarrow k + 1$;
    $\hat{s}^+_k \leftarrow \hat{s}^+$;
    $\hat{s}^-_k \leftarrow \hat{s}^-$ ;
end
```

Algorithm 1: The edge depth-first configuration model

$\hat{s}^+(k)$ is akin to a Łukasiewicz path. At any given step it is equal to the size of the queue after subtracting the number of fully explored out-components.

We also construct a directed forest for which $\hat{s}^+(k)$ will be the true Łukasiewicz path. At each step of the process we will examine a vertex w. If w has not been discovered yet then either we are at the start of a new out-component, in which case we make w the root of the next out-component, or we added an edge (v, w) to the multigraph with v already discovered, in which case we add the edge (v, w) to the out-forest as well. If w has already been explored we cannot add (v, w) to the out-forest without creating cycles or connecting two different components. We instead add a *dummy leaf* to the out-forest and an edge from v to the dummy leaf. We call any vertex that is not a dummy leaf a *true vertex*. This is illustrated in Figure 3.6b.

Consider an edge (v, w) in the directed multigraph. If (v, w) is not in the out-forest we refer to the edge as *surplus*. Such an edge will instead correspond to an edge (v, d) in the out-forest where d is a dummy leaf.

An important motivation for studying the out-forest is the fact that the vertex set of any SCC is contained in one of the components of the out-forest. This is a straightforward property which we will prove below as part of Lemma 3.2.1. Moreover, we defined the out-forest in such a way that every time step in the exploration corresponds to one vertex in the out-forest.

Our technique relies on dismissing surplus edges that cannot be part of a strongly connected component (for example, surplus edges between two different out-components cannot form a directed cycle and are never part of a strongly connected component). We define a necessary condition for a surplus edge to be part of an SCC (see Definition 3.2.2 and Proposition 3.2.3), and we call dummy leaves that correspond to surplus edges with this property *candidates*. Then, we define a procedure to sample only the out-forest and the edges corresponding to candidates, which allows us to find the SCCs.

A key fact is that the order in which the true vertices are discovered does not depend on the positions of the dummy leaves. Similarly, the positions of the dummy leaves do not depend on the position of the heads of the surplus edges. Finally, whether a dummy leaf is a candidate does not depend on the position of the heads of the surplus edges. This allows us to define the following step-by-step sampling procedure.

1. We sample the order of discovery of the true vertices.

2. We sample at which time steps we add a dummy leaf instead of a true vertex.

3. For each dummy leaf we sample whether it is a candidate.

4. For each candidate we sample the position of the head of the corresponding surplus edge.

For an exact description of the sampling procedure, see Subsection 3.2.1. The analogous sampling procedure for the limit object is described in Subsection 3.2.2.2. Then, our approach to show convergence is as follows.

1. We find the limit under rescaling of the Łukasiewicz path and height process of the out-forest up to time $m_n = \Theta(n^{2/3})$ conditional on the event $\left\{\sum_{i=1}^{n} D_i^- = \sum_{i=1}^{n} D_i^+\right\}$. This is the content of Proposition 3.4.1. Note that we condition on an asymptotically singular event, which causes significant difficulties. Our method relies on a measure change between the sequence of degrees in order of discovery under this conditioning and a sequence of i.i.d. random variables in $\mathbb{N} \times \mathbb{N}$. In Section 3.3 , we show the convergence of the measure change under rescaling.

2. We establish that the positions of the tails of the surplus edges corresponding to the candidates converge. This is the content of Proposition 3.5.2, Lemma 3.5.4, and Proposition 3.5.5.

3. We show that the positions of the heads of the surplus edges corresponding to the candidates converge, which is the content of Proposition 3.5.6.

4. We identify the tails and heads of the surplus edges corresponding to the candidates, and recover the SCCs from the resulting digraph via a cutting procedure. We use a result from [59] to show that the cutting procedure converges. This is summarised in Corollary 3.5.10.

5. We show that with high probability, there exists a simple graph with the degree sequence that we sample and that the probability that the configuration model yields a simple graph is asymptotically bounded away from 0. This is the content of Proposition 3.4.16. Then, we show that conditioning on the resulting multigraph being simple does not affect the sampling procedure on the time scale $O(n^{2/3})$. This is the content of Proposition 3.4.17.

6. We prove that for any $\delta > 0$, with high probability, all SCCs with more than $\delta n^{1/3}$ edges are contained in the exploration up to time $O(n^{2/3})$. Therefore, we can choose m_n such that, with high probability, we do not miss any large SCCs by not considering the

exploration beyond time m_n, which finishes the proof of the convergence in the product topology. This is the content of Lemma 3.5.11.

3.2 Sampling the MDM in the discrete and the continuum

If we forget about the directions of the edges in $\vec{G}_n(\nu)$, the resulting undirected graph is supercritical, and, with high probability, the graph contains a unique giant component with surplus going to infinity as $n \to \infty$ (see e.g. [85, 86, 69] for a discussion of the phase transition in the undirected configuration model). This suggests that if we do not dismiss a large amount of edges, we will not be able to study the digraph in enough detail to find a metric space scaling limit of the SCCs. Therefore, we will not try to sample the entire digraph, but focus on the information that we need to find the SCCs. We start by studying the discrete digraph model, with the goal of identifying which edges can be part of an SCC, and how to sample them. In Subsection 3.2.1.1, we establish necessary conditions for an edge to be part of an SCC. These conditions imply that we only need to study the out-forest, and the surplus edges corresponding to a small subset of the dummy leaves. We call these dummy leaves *candidates*. In Subsections 3.2.1.2 and 3.2.1.3 we study the law of the out-forest and the surplus edges corresponding to the candidates respectively, and we define a procedure to sample them both. This yields a sequence of directed multigraphs with edge lengths in which the SCCs are embedded. In Subsection 3.2.2, we define the continuous counterpart of the sampling procedure. The resulting object will be the limit under rescaling of the sequence of directed multigraphs with edge lengths in which the SCCs are embedded that was constructed in Subsections 3.2.1.2 and 3.2.1.3.

3.2.1 The discrete case

We will discuss the different type of edges that we can encounter in the exploration. Recall from Subsection 3.1.7 that by slight abuse of terminology, we call the dummy leaf that corresponds to a surplus edge its tail.

3.2.1.1 Necessary conditions for an edge to be part of an SCC

Amongst the surplus edges, *ancestral surplus edges*, which are surplus edges that point from a vertex to one of its ancestors, play a special role. All other surplus edges are called *non-ancestral*. This is illustrated in Figure 3.7a. In Figure 3.7b we show how surplus edges affect the structure of the SCCs. This is the content of the next lemma.

Lemma 3.2.1. *The following facts hold for SCCs.*

1. *The vertices of an SCC are contained in precisely one of the components of the out-forest.*
2. *Ancestral surplus edges are always part of an SCC.*
3. *A non-ancestral surplus edge is part of an SCC only if its head is an ancestor of the tail of a surplus edge that is part of an SCC.*
4. *An edge in the out-forest is part of an SCC only if its head is an ancestor of the tail of a surplus edge that is part of an SCC.*
5. *For any non-trivial SCC, the first surplus edge of the SCC that is explored is an ancestral surplus edge, and a component of the out-forest contains an SCC if and only if it contains an ancestral surplus edge.*

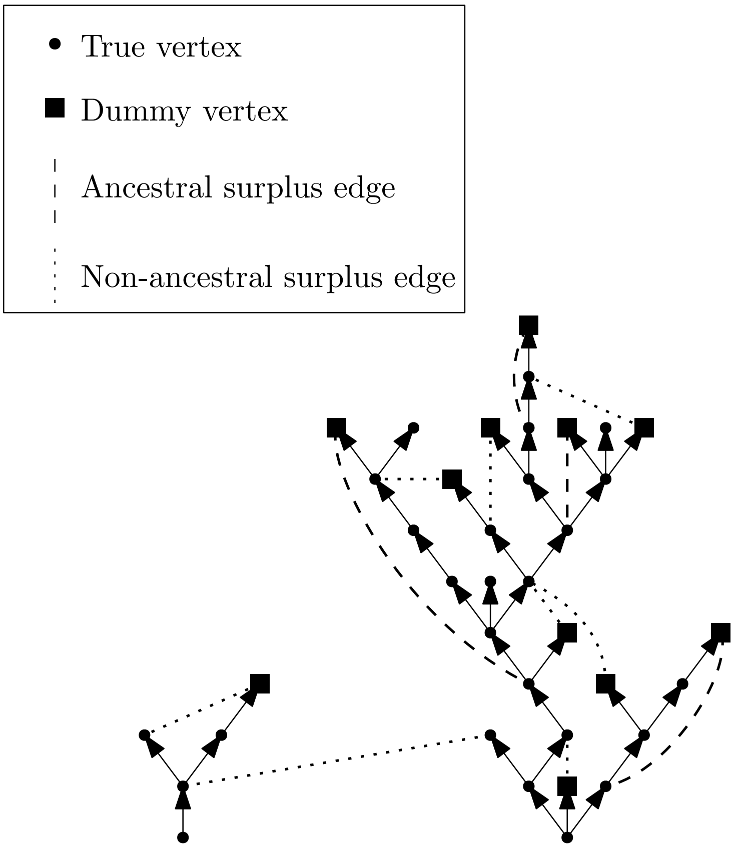
Proof. We start with 1. Let v and w be two vertices in the same SCC. Without loss of generality, v is explored first in depth-first order in the out-direction. Since v and w are part of the same SCC, we know that there is a path from v to w in the out-direction. This implies that w will be part of the out-subtree consisting of the descendants of v. This implies that they are part of the same component of the out-forest.

To prove 2, suppose there is an ancestral surplus edge from v to w. This implies that w is an ancestor of v in an out-component, which implies that there is a path from w to v as well. It follows that w and v are in the same SCC and that the ancestral surplus edge from v to w is in this SCC as well.

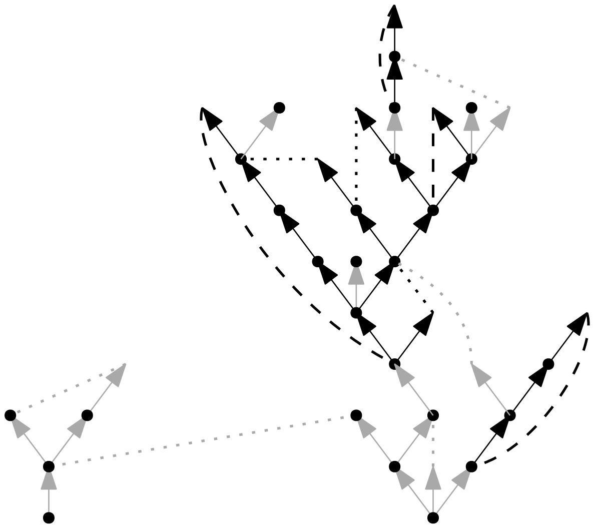
To prove 3 and 4, suppose there is a non-ancestral surplus edge from v to w that is part of an SCC, or that (v, w) is an edge in the out-forest that is part of an SCC. Then, there is some directed path $(x_0, \dots, x_m)$ with $x_0 = w$ and $x_m = v$. Let k be minimal such that x_k is not a descendant of w (such a k exists, because by assumption, v is not a descendant of w). Then, (x_{k-1}, x_k) is a surplus edge that is in the same SCC as v and w, and x_{k-1} is a descendant of w by definition of k.

Finally, 2 and 3 imply 5. □

Lemma 3.2.1 motivates the following definition.



(a) This figure illustrates an example of a depth-first exploration of two out-components with the different type of surplus edges highlighted. The ancestral surplus edges point from a vertex v to one of its ancestors. They are always part of an SCC.



(b) The edges that are part of an SCC are depicted in black. Two vertices are in the same SCC if and only if they are connected by black edges.

Fig. 3.7: We illustrate the different types of surplus edges and how they affect the structure of the SCCs.

Definition 3.2.2. *A dummy vertex is a* candidate *if one of the following statements holds for the surplus edge that it corresponds to.*

- *It is an ancestral surplus edge, or*
- *Its head is the ancestor of a candidate.*

The following proposition is at the core of our strategy to study the SCCs.

Proposition 3.2.3. *Any edge that is part of an SCC is either a surplus edge corresponding to a candidate, or is contained in the subforest of the out-forest that is spanned by the candidates and the roots of the out-components.*

Proof. This follows from Definition 3.2.2 and Lemma 3.2.1. □

Proposition 3.2.3 implies that to sample the SCCs, we do not need to sample the heads corresponding to all dummy leaves. Instead, for every dummy leaf, we only need to know whether it is a candidate, and if so, where its head is.

3.2.1.2 Sampling the out-forest

This subsection discusses how to obtain the out-forest conditional on the order in which the vertices are discovered. We will study the law of the degrees in order of discovery in Section 3.3. The out-forest is obtained in the following way. Let $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,n})$ be the degree pairs in order of discovery (i.e. the order given by $\mathcal{V}$ in Algorithm 1). Up to time-step k, suppose we have discovered the first $m \leq k < n$ elements of $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,n})$. Then, at time $k+1$,

1. If we have finished a component of the out-forest, let the next out-component have a root with out-degree $\hat{D}^+_{n,m+1}$.
2. Otherwise,
   (a) With probability proportional to the total in-degree of the undiscovered vertices, i.e. $\sum_{i=m+1}^{n} \hat{D}^-_{n,i}$, let the next vertex in depth-first order be a true vertex with out-degree $\hat{D}^+_{n,m+1}$.
   (b) With probability proportional to the number of unpaired in-half-edges of the m discovered vertices, let the next vertex in depth-first order be a dummy leaf, and reduce the total number of unpaired in-edges of the m discovered vertices by 1.

We make this rigorous in the following proposition.

Proposition 3.2.4. *Suppose that the sequence of degrees in order of discovery* $(\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,n})$ *is given. Suppose that after time-step* k*, there are still unpaired out-half-edges. Suppose that for* $1 \leq l \leq k$*, that up to time* l*,* $\hat{P}_n(l)$ *surplus edges have been sampled. Then,*

$$\left(\hat{S}_n^+(l), 0 \leq l \leq k\right) := \left(\sum_{i=1}^{l-\hat{P}_n(l)} \hat{D}_{n,i}^+ - l, 0 \leq l \leq k\right)$$

is the Łukasiewicz path of the out-forest up to time k*. Moreover, for*

$$\left(\hat{I}_n^+(l), 1 \leq l \leq k\right) := \left(\min\left\{\hat{S}_n^+(m) : 0 \leq m \leq l\right\}, 1 \leq l \leq k\right),$$

define

$$\left(\hat{S}_n^-(l), 1 \leq l \leq k\right) := \left(\sum_{i=1}^{l-\hat{P}_n(l)} \hat{D}_{n,i}^- - l - \hat{I}_n^+(l) + 1, 1 \leq l \leq k\right),$$

so that $\hat{S}_n^-(k)$ *is equal to the number of unpaired in-half-edges of discovered vertices at time* k*. Then, the probability that we sample a surplus edge at the* $(k+1)$*th time-step is given by*

$$\frac{\hat{S}_n^-(k)}{\sum_{i=1}^n D_i^- - k - \hat{I}_n^+(k) + 1} \mathbb{1}_{\left\{\hat{I}_n^+(k) = \hat{I}_n^+(k-1)\right\}}.$$

We do not need to know the position of the heads of the surplus edges in order to sample the out-forest.

Proof. Note that if up to time k, $\hat{P}_n(k)$ surplus edges have been sampled, this implies that $k - \hat{P}_n(k)$ true vertices have been discovered. Thus, up to time k, the out-forest contains $\hat{P}_n(k)$ dummy leaves, and true vertices with degrees $(\hat{D}_{n,1}^+, \dots, \hat{D}_{n,k-\hat{P}_n(k)}^+)$, so by definition of the Łukasiewicz path, its value is indeed equal to $\hat{S}_n^+(k)$ at time k. Moreover, up to time k, the total in-degree of the discovered true vertices is equal to $\sum_{i=1}^{k-\hat{P}_n(k)} \hat{D}_{n,i}^-$. At every time-step, we pair one in-half-edge of a discovered vertex, unless we start a new component. The value $-\hat{I}_n^+(k)$ corresponds to the number of out-components that are fully explored up to time k, so the total number of unpaired in-half-edges of discovered vertices at time k is equal to $\hat{S}_n^-(k)$. By the same reasoning, the total number of unpaired in-half-edges is equal to $\sum_{i=1}^{k-\hat{P}_n(k)} \hat{D}_{n,i}^- - k - \hat{I}_n^+(k) + 1$. The probability of sampling a surplus edge at step $(k+1)$ follows. We note that this probability does not depend on the positions of the heads of the surplus edges, but only on their number, which implies that we can sample the out-forest without sampling the positions of the heads. □

3.2.1.3 Sampling the candidates

We will now study the law of the candidates and their heads conditional on the out-forest. We will first identify the candidates amongst the dummy leaves, and then we will sample the positions of their heads.

If the vertex discovered at time k is a dummy leaf, the head of the corresponding surplus edge is a uniform pick from the $\hat{S}^-(k)$ unpaired in-half-edges of discovered vertices at time k. Therefore, the probability that a dummy leaf added at time k corresponds to an ancestral surplus edge is given by the number of unpaired in-edges on its path to the root divided by $\hat{S}^-(k)$. This implies that to understand the law of the position of ancestral surplus edges, we need to understand where the unpaired in-edges are.

We will study this by modifying the edge lengths in the out-forest. We extend our definitions in Section 3.1.7 to trees with edge lengths as follows. Suppose $T = (V, E, \rho)$ is an ordered rooted finite tree, and suppose we have a function $\ell : E \to [0, \infty)$. Then, we can view T as a metric space by regarding an e as a line segment with length $\ell(e)$. The distance d_T^ℓ between points a_1 and a_2 on line segments l_1 and l_2 respectively is then defined as the length of the unique non-self-intersecting path between a_1 and a_2 that traverses the line segments of the tree, and we denote the resulting metric space (T, d_T^ℓ) by T^ℓ, and call it a *ordered rooted finite tree with edge lengths.* This gives rise to an alternative height process, referred to as h^ℓ, which is defined

$$h^\ell(i) = d_T^\ell(v_i, v_0),$$

i.e. for all i, $h^\ell(i)$ equals the distance from v_i to the root in T^ℓ. We set the Łukasiewicz path of T^ℓ equal to the Łukasiewicz path of T.

We will now study the positions of the unpaired in-edges by modifying the edge lengths as follows: for a vertex v with in-degree m, the edges connecting it to its children will all have length $m - 1$ (unless v is the root of an out-component, in which case the edges connecting to its children will be assigned length m). The height of vertex w in this forest with modified edge lengths corresponds to the number of in-half-edges that can be used to form an ancestral surplus edge with tail w. We assign lengths to all edges in the out-forest and call the resulting forest with edge lengths *the out-forest with edge lengths.* Denote the height process of the out-forest with edge lengths by $(\hat{H}^{\ell}_n(k), k \geq 1)$. Recall from Lemma 3.2.1 that the surplus edge corresponding to the first candidate in any component of the out-forest is ancestral. The following proposition illustrates the importance of $\hat{H}^{\ell}$ in finding the first ancestral surplus edges in the out-components.

Proposition 3.2.5. *Consider the exploration of the out-forest at time k. If no ancestral surplus edge has been sampled in the current component, then the probability that the kth vertex in depth-first order is a candidate is given by*

$$\frac{\hat{H}^{\ell}(k)}{\hat{S}^{-}_n(k)} \mathbb{1}_{\{\hat{P}_n(k)-\hat{P}_n(k-1)=1\}}.$$

This event is conditionally independent of the positions of the heads of the surplus edges that were found before time k, given that none of them were ancestral in the current component.

Proof. We claim that if no ancestral surplus edge has been sampled in the current component, none of the ancestors of k are the head of a surplus edge. Indeed, for x an ancestor of k, all vertices that are discovered since the discovery of x up to time k are descendants of x, because the out-forest is explored in a depth-first manner. Therefore, any surplus edge with head x sampled up to time k is ancestral. This implies that for d^- the in-degree of x, the number of unpaired in-half-edges of x at time k is equal to $d^- - 1$ (unless x is the root of the out-component, in which case it has d^- unpaired in-half-edges).

Therefore, the number of unpaired in-half-edges corresponding to ancestors of k is equal to $H^{\ell}(k)$. Moreover, note that, by definition of the dummy leaves, k is the tail of a surplus edge if and only if k is a dummy leaf, i.e. if and only if $\hat{P}_n(k) - \hat{P}_n(k-1) = 1$. In that case, the probability that it connects to given unpaired in-half-edge of a discovered vertex is equal to $1/\hat{S}^{-}_n(k)$. The stated probability follows. The independence of the positions of the heads of earlier surplus edges is immediate. □

We now illustrate how to find the other candidates in a component of the out-forest.

Let $T^n_{g_n}$ be a component of the out-forest with root $g_n + 1$ and component size σ_n. Suppose the first ancestral surplus edge with vertices in $T^n_{g_n}$ corresponds to a dummy leaf $V^n_1 \in [g_n + 2, g_n + \sigma_n]$. Let $V^n_1 < k \leq g_n + \sigma_n$, and suppose the candidates found up to time k are given by $V^n_1, \ldots, V^n_m$. Let $T^{n,\mathrm{mk}}_k$ be the subtree of $T^n_{g_n}$ spanned by $\{g_n + 1, V^n_1, \ldots, V^n_m, k\}$, and let $\ell(T^{n,\mathrm{mk}}_k)$ be its total length with edge lengths as encoded by $(\hat{H}^{\ell}(i), i \in [g_n + 1, g_n + \sigma_n])$.

Proposition 3.2.6. *The probability that k is a candidate is given by*

$$\frac{\ell\left(T_k^{n,\mathrm{mk}}\right) - m}{\hat{S}^-(k)} \mathbb{1}_{\{\hat{P}_n(k)-\hat{P}_n(k-1)=1\}}.$$

Proof. Note that if k is a dummy leaf, it gets paired to a uniform pick from the $\hat{S}^-(k)$ as-yet unpaired in-half-edges of discovered vertices. By Definition 3.2.2, in that case, k is a candidate if and only if the head of its corresponding surplus edge is in $T_k^{n,\mathrm{mk}}$. Observe that $\ell\left(T_k^{n,\mathrm{mk}}\right)$ is equal to the number of in-half-edges of T_k that can be used to form surplus edges. By the definition of a candidate, exactly m of those have been paired: one for each element in $\{V_1^n, \ldots, V_m^n\}$. This implies that $\ell\left(T_k^{n,\mathrm{mk}}\right) - m$ of the $\hat{S}^-(k)$ options will cause k to be a candidate. □

Note that the probability that a dummy leaf is a candidate only depends on the out-forest and the number of candidates that have been found in the component so far. The position of the heads of the surplus edges corresponding to candidates can be found as follows.

Let $T_{g_n}^n$ be a component of the out-forest with root $g_n + 1$ and component size σ_n. Suppose its candidates are given by $\{V_1^n, \ldots, V_{N_n}^n\}$. Then, for $1 \leq i \leq N_n$, suppose the heads of the surplus edges corresponding to $V_1^n, \ldots, V_{i-1}^n$ are given by $W_1^n, \ldots, W_{i-1}^n$ respectively.

Proposition 3.2.7. *The in-half-edge that V_i^n gets paired to is a uniform pick from the*

$$\ell\left(T_{V_i^n}^{n,\mathrm{mk}}\right) - (i-1)$$

unpaired in-half-edges of $T_{V_i^n}^{n,\mathrm{mk}}$ that remain.

Proof. Given that V_i^n is a candidate, its head will be in $T_{V_i^n}^{n,\mathrm{mk}}$. Then, the distribution follows. □

Propositions 3.2.4, 3.2.5, 3.2.6, and 3.2.7 justify the following sampling procedure.

1. Sample the out-forest, and suppose it has N vertices.
2. Define a counting process $(A_n(k), k \leq N)$, with the probability of an increment at time k given by
$$\frac{\hat{H}_n^{\ell}(k)}{\hat{S}_n^-(k)} \mathbb{1}_{\{\hat{P}_n(k)-\hat{P}_n(k-1)=1\}}.$$

3. For $i \geq 1$, let $X_i^n = \min\{k : A_n(k) = i\}$ be the time that the ith ancestral surplus edge is sampled. For $i \geq 1$, let G_i^n be the left endpoint of the excursion of $\hat{S}_n^+$ above its running infimum that encodes the out-component that contains the ith ancestral surplus edge, and let Σ_i^n be the length of this excursion, i.e.

$$G_i^n = \min\left\{k \geq 1 : \hat{S}_n^+(k) = \min\{\hat{S}_n^+(l) : l \leq X_i^n\}\right\}$$
$$\Sigma_i^n = \min\left\{k \geq 1 : \min\left\{\hat{S}_n^+(l) : l \leq G_i^n + k\right\} < \min\left\{\hat{S}_n^+(l) : l \leq X_i^n\right\}\right\},$$

so that for each $i \geq 1$, the excursion $\left(\hat{S}^+(k), k \in [G_i^n + 1, G_i^n + \Sigma_i^n]\right)$ encodes the out-tree containing X_i^n. For each $(g_n, \sigma_n) \in \{(G_i^n, \Sigma_i^n)\}$, let $T_{g_n}^n$ be the tree in out-forest with root $g_n + 1$, and do the following.

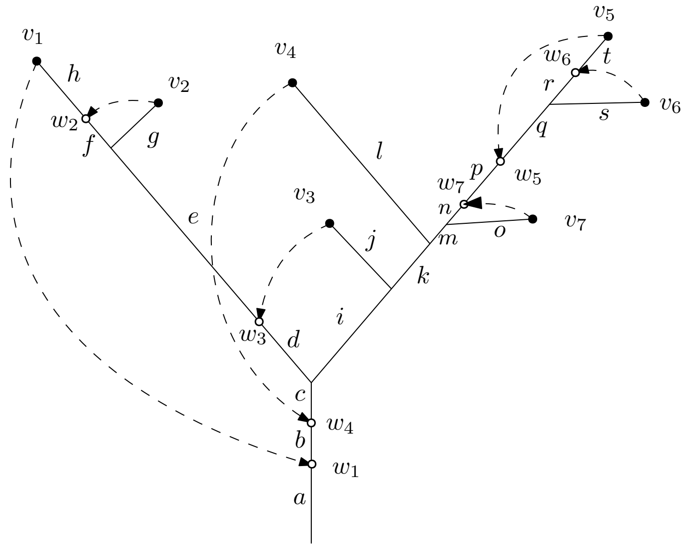
   (a) Set $V_1^n = \min\{m \geq 1 : A_n(m) = A_n(g_n) + 1\}$, and find the other candidates $\{V_2^n, \dots, V_{N_n}^n\}$ according to the procedure described in the statement of Proposition 3.2.6.

   (b) For the tails $V_1^n, \dots, V_{N_n}^n$, sample their corresponding heads $W_1^n, \dots, W_{N_n}^n$ respectively according to the procedure described in the statement of Proposition 3.2.6.

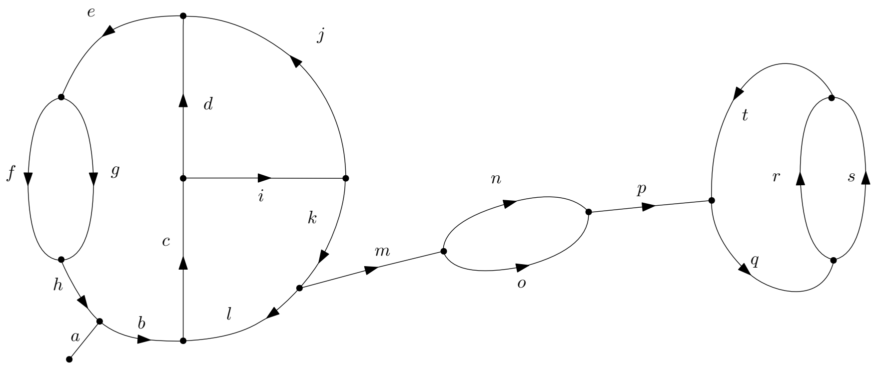
   (c) Let $T^{n,\mathrm{mk}}(g_n)$ be the subtree of $T_{g_n}^n$ spanned by $\{g_n + 1, V_1^n, \dots, V_{N_n}^n\}$. Then, quotient it by the equivalence relation $\sim$ which identifies V_i^n and W_i^n for each $1 \leq i \leq N_n$ to obtain a rooted metric space with surplus N_n

$$M_{g_n}^n = T^{n,\mathrm{mk}}(g_n)/\sim .$$

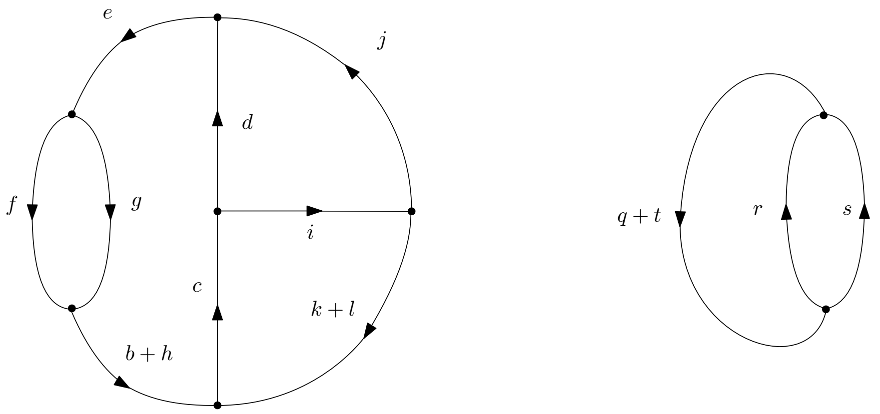
Then, all SCCs of $\vec{G}_n(\nu)$ are sub-digraphs of $\left\{M_{G_i^n}^n, i \geq 1\right\}$. Call the kernels of these SCCs, ordered by decreasing size, $(C_i(n), i \geq 1)$, completed with an infinite repeat of $\mathfrak{L}$. Observe that we may view $M_{G_i^n}^n$ as a finite rooted directed multigraph $M_{G_i^n}^n$ whose edges are endowed with lengths. To be precise, in $M_{G_i^n}^n$, let the vertex set consist of $G_i^n + 1$, W_i^n for $i \leq N_n$, and the branch points $V_i^n \wedge V_j^n$ for $i \neq j \leq N_n$. Then, we obtain $(C_i(n), i \geq 1)$ by ordering the kernels of the non-trivial SCCs in $\left\{M_{G_i^n}^n, i \geq 1\right\}$ by decreasing size, and completing the list with an infinite repeat of $\mathfrak{L}$. See Figures 3.8a, 3.8b and 3.8c for an illustration of this procedure.



(a) This is a subtree of an out-component spanned by the root of the out-component and the candidates $(v_1, \ldots, v_7)$. Call the marked tree T^{mk}. The heads of the surplus edges corresponding to candidates are denoted by $(w_1, \ldots, w_7)$.



(b) Identifying v_i with w_i for $i \in [7]$ gives M.



(c) We find the SCCs that are contained in M.

Fig. 3.8: We illustrate the procedure to find the SCCs in a component of the out-forest after finding the candidates. Taken from [59] with permission of the authors.

3.2.2 The continuum case

We will now define the continuous counterpart of the sampling procedure of the out-forest and the candidates. This is a modification of the procedure defined in Subsection 3.2.2 of [59].

3.2.2.1 $\mathbb{R}$-trees and their encoding

The continuum analogue of discrete trees are given by $\mathbb{R}$-trees. We give the basic definitions here and refer the reader to the survey paper [76] for more details. An $\mathbb{R}$-tree is a compact metric space $(\mathcal{T}, d)$ such that for every $a, b \in \mathcal{T}$ the following two properties hold:

1. There exists a unique isometry
$$i_{a,b} : [0, d(a,b)] \to \mathcal{T}$$
such that $i_{a,b}(0) = a$ and $i_{a,b}(d(a,b)) = b$.
2. If $q : [0,1] \to \mathcal{T}$ is any continuous map such that $q(0) = a$ and $q(1) = b$ then the image of q is the same as the image of $i_{a,b}$.

Let $[\![a,b]\!]$ denote the image of $i_{a,b}$. This is the unique path between a and b.

$\mathbb{R}$-trees are often encoded by continuous excursions which can be seen as a continuous analogue of the height function of a tree. Let $f : [0,\sigma] \to [0,\infty)$ be a continuous excursion, meaning f is continuous, $f(0) = f(\sigma) = 0$ and $f(x) > 0$ for all $x \in (0,\sigma)$. Using f we can define a pseudo-metric
$$d_f(x,y) = f(x) + f(y) - 2 \min_{s \in [x \wedge y, x \vee y]} f(s).$$
This allows us to define the quotient space
$$\mathcal{T}_f = [0,\sigma]/\{d_f = 0\}.$$

The space $\mathcal{T}_f$ equipped with the metric d_f is the $\mathbb{R}$-*tree encoded by the excursion* f. Let $p_f : [0,\sigma] \to \mathcal{T}_f$ be the natural projection function. Then $\mathcal{T}_f$ inherits a distinguished root vertex $\rho = p(0) = p(\sigma)$.

A sequence of $\mathbb{R}$-trees is referred to as an $\mathbb{R}$-forest.

3.2.2.2 The limit object

Let $(B_t, t \geq 0)$ be a standard Brownian motion, and set

$$\left(\hat{B}_t, t \geq 0\right) = \left(B_t - \frac{(\sigma_{-+} + \nu_-)t^2}{2\sigma_+ \mu}, t \geq 0\right).$$

Remark 3.2.8. *We note that the coefficient of the parabolic drift of $\hat{B}$ is negative. Indeed, by definition of σ_{-+} and ν_-, the sign of the parabolic drift is the same as the sign of $\mu - \mathbb{E}[(D^-)^2 D^+]$, and we note that*

$$\frac{\mathbb{E}[(D^-)^2 D^+]}{\mathbb{E}[D^+]} - \left(\frac{\mathbb{E}[D^+ D^-]}{\mathbb{E}[D^+]}\right)^2 = \frac{\mathbb{E}[(D^-)^2 D^+]}{\mu} - 1$$

is the variance of D^- under the law of $\mathbf{D}$ size-biased by D^+, which is positive. Hence $\mathbb{E}[D^+(D^-)^2]/\mu \geq 1$, and the claimed negativity follows.

Define the reflected process

$$(\hat{R}_t, t \geq 0) = \left(\hat{B}_t - \inf\left\{\hat{B}_s : s \leq t\right\}, t \geq 0\right).$$

Then, it follows from the argument in Section 3.4 that $\left(\frac{2}{\sigma_+}\hat{R}_t, t \geq 0\right)$ is the height process corresponding to an $\mathbb{R}$-forest with Łukasiewicz path $\left(\sigma_+ \hat{B}_t, t \geq 0\right)$. Call this forest the out-$\mathbb{R}$-forest.

Conditionally on $\hat{R}$, let $(A_t, t \geq 0)$ be a Cox process of intensity

$$\frac{2(\sigma_{-+} + \nu_-)}{\sigma_+ \mu^2}\hat{R}_t$$

at time t. Then, for i in $\{1, 2, \dots\}$, set $X_i = \min\{t : A_t = i\}$. For i in $\{1, 2, \dots\}$, define

$$G_i = \inf\left\{t \geq 0 : \hat{B}_t = \inf\{\hat{B}_s : s \leq X_i\}\right\} \text{ and}$$
$$\Sigma_i = \inf\left\{t \geq 0 : \inf\{\hat{B}_s : s \leq G_i + t\} < \inf\{\hat{B}_s : s \leq X_i\}\right\},$$

so that for each i in $\{1, 2, \dots\}$, $\left(\frac{2}{\sigma_+}\hat{R}_t, t \in [G_i, G_i + \Sigma_i]\right)$ encodes the $\mathbb{R}$-tree in the out-$\mathbb{R}$-forest that contains X_i. For each element of $\{(G_i, \Sigma_i) : i = 1, 2, \dots\}$ we will sample the candidates in the $\mathbb{R}$-tree. Fix i, and set $(g, \sigma) = (G_i, \Sigma_i)$. Let $V_1 = \inf\{s > 0 : A(s) = A(g) + 1\}$, so that

$g \leq V_1 \leq g+\sigma$ by definition of (g,σ). Let $\mathcal{T}_g$ be the $\mathbb{R}$-tree encoded by $\left(\frac{2}{\sigma_+}\hat{R}_t, t \in [g, g+\sigma]\right)$ and let $p_g : [g, g+\sigma] \to \mathcal{T}_l$ be the projection onto $\mathcal{T}_g$ given by the encoding. Set

$$||\mathcal{T}_g|| = \sup\left\{\frac{2}{\sigma_+}\hat{R}_t, t \in [g, g+\sigma]\right\},$$

the *height* of $\mathcal{T}_g$.

Suppose we have found candidates $\{V_1, \ldots, V_m\}$. For $V_m \leq s \leq g+\sigma$, let T_s^{mk} be the subtree of $\mathcal{T}_g$ spanned by $p_g(\{g, V_1, \ldots, V_m, s\})$, and let $|T_s^{\mathrm{mk}}|$ be its total length. Then, let V_{m+1} be the first arrival time of a Poisson process on $[V_m, g+\sigma]$ of intensity

$$\frac{\sigma_{-+} + \nu_-}{\mu^2}|T_s^{\mathrm{mk}}|ds.$$

If the process does not contain a point, let $\{V_1, \ldots, V_m\}$ be the candidates of $\mathcal{T}_g$, and set $N = m$. Otherwise, we repeat the inductive step for $\{V_1, \ldots, V_{m+1}\}$. If the induction does not terminate, we set $N = \infty$.

We show that $\mathbb{P}(N = \infty) = 0$, by adapting the argument in Subsection 3.2.2 of [59] to our set-up. Indeed, note that $V_m \leq s \leq V_{m+1}$ implies that $|T_s^{\mathrm{mk}}| < (m+1)||\mathcal{T}_g||$. Therefore,

$$\mathbb{P}\left(N \geq g+1, V_{m+1} - V_m < t \middle| N \geq g\right) \leq \mathbb{P}(E_{m+1} < t),$$

for $(E_k, k \geq 1)$ a sequence of exponential random variables with respective rates

$$\frac{\sigma_{-+} + \nu_-}{\mu^2}k||\mathcal{T}_g||.$$

Then,

$$\mathbb{P}(N = \infty) = \mathbb{P}(N = \infty \text{ and } \sup\{V_i : i \in \mathbb{N}\} < g+\sigma) \leq \mathbb{P}\left(\sum_{i=2}^{\infty} E_k \leq g+\sigma - V_1\right).$$

However, $\sum_{i=2}^{\infty} E_k = \infty$ a.s., because the harmonic series diverges, so, indeed, $\mathbb{P}(N < \infty) = 1$.

Finally, for $1 \leq i \leq N$, let the head corresponding to V_i, which we call W_i, be a uniform pick from the length measure on $T_{V_i}^{\mathrm{mk}}$.

Let $T^{\mathrm{mk}}(g)$ be the subtree of $\mathcal{T}^g$ spanned by $\{g, V_1, \ldots, V_N\}$. Then quotient $T^{\mathrm{mk}}(g)$ by the equivalence relation $\sim$ which identifies V_i and W_i for each $1 \leq i \leq N$ to obtain a rooted metric space

$$\mathcal{M}_g := T^{\mathrm{mk}}(g)/\sim .$$

View $\mathcal{M}_g$ as an element of $\vec{\mathcal{G}}$ in the natural way. To be precise, let the vertex set of $\mathcal{M}_l$ consist of g, W_i for $i \leq N$, and the branch points $V_i \wedge V_j$ for $i \neq j \leq N$. The directions are inherited from $\mathcal{T}^l$, by considering all edges directed away from the root. Remove all edges that do not lie in an SCC of $\mathcal{M}_g$ and delete any isolated vertices that are thus created. Then, apply the smoothing operation as defined in Subsection 3.1.4. This creates a collection $\mathcal{C}_g$ of strongly connected MDMs. Doing this for each $(g, \sigma) \in \{[G_i, \Sigma_i]\}$ yields the collection of strongly connected MDMs $\mathcal{C}$ that has the law of the limit in Theorem 3.1.1.

3.2.2.3 Properties of the limit object

We note that the limit object is encoded by 3 parameters: the out-$\mathbb{R}$-forest is encoded by a Brownian motion with variance σ_+^2 and parabolic drift with coefficient $-(\sigma_{-+} + \nu_-)/(2\mu)$, and the identifications are a Cox process with intensity $(\sigma_{-+} + \nu_-)/\mu^2$ on the length measure of the subtree spanned by the previously found candidates and the currently explored vertex as described in Subsection 3.2.2.2. The limit object that is studied in [59] corresponding to $\lambda = 0$ (i.e. at criticality) is equal to our limit object in the case $\sigma_+^2 = 1$, $-(\sigma_{-+} + \nu_-)/(2\mu) = -1/2$, and $(\sigma_{-+} + \nu_-)/(\mu^2) = 1$. Note that these three conditions are satisfied if we let D^- and D^- be independent Poisson(1) random variables. In [59], some properties of the limit object corresponding to these specific parameters are shown. A quick check shows that the proofs do not depend on the values of the parameters, so we deduce that the same properties also hold for our limit object. Let $\mathcal{M} := \bigcup_{G_i} \mathcal{M}_{G_i}$.

Proposition 3.2.9. *1. The number of complex connected components of $\mathcal{M}$ has finite expectation.*

2. The number of loops of $\mathcal{M}$ is a.s. infinite.

Proposition 3.2.10. *The SCCs of $\mathcal{M}$ all have different lengths almost surely.*

Write $\mathcal{C}$ for the SCCs of $\mathcal{M}$ and $\mathbf{C}_l$ for those of $\mathcal{M}_l$, in decreasing order of length, with $\mathcal{M}_l$ as defined in Subsection 3.2.2.2. Write $\mathcal{C}_{\mathrm{cplx}}$ for the list of complex components of $\mathcal{C}$ in decreasing order of length. For sequences $(K_1, \ldots, K_j)$ and $(J_1, \ldots, J_k)$ of directed multigraphs, write $(J_1, \ldots, J_k) \equiv (K_1, \ldots, K_j)$ if $j = k$ and J_i is isomorphic to K_i for each $i \leq j$. Extend this notation naturally to the case where one or both of the sequences has edge lengths by ignoring the edge lengths.

Theorem 3.2.11. *Let $K_1, \dots, K_j$ be a finite sequence consisting of 3-regular strongly connected directed multigraphs or loops. We have*

$$\mathbb{P}\left[\mathcal{C}_l \equiv (K_1, \dots, K_j)\right] > 0.$$

Assuming that $K_1, \dots, K_j$ are all complex, we also have that

$$\mathbb{P}\left[\mathcal{C}_{cplx} \equiv (K_1, \dots, K_j)\right] > 0.$$

Let $(e_i, 1 \leq i \leq M)$ be an arbitrary ordering of the edges of $K_1, \dots, K_j$. Then, conditionally on $\mathcal{C}_l \equiv (K_1, \dots, K_j)$, (resp. $\mathcal{C}_{cplx} \equiv (K_1, \dots, K_j)$), $\mathcal{C}_l$ (resp. $\mathcal{C}_{cplx}$) gives lengths $(\ell(e_i), 1 \leq i \leq M)$ to these edges, and their joint distribution has full support in

$$\left\{ \mathbf{x} = (x_1, \dots, x_M) \in \mathbb{R}_+^M : \forall 1 \leq i \leq j-1, \sum_{k : e_k \in E(K_i)} x_k \geq \sum_{k : e_k \in E(K_{i+1})} x_k \right\}.$$

3.3 Analysis of the measure change

Recall $(\hat{\mathbf{D}}_{n,1}, \hat{\mathbf{D}}_{n,2}, \dots, \hat{\mathbf{D}}_{n,n})$ are are the degree pairs of the vertices in order of discovery, and that R_n is the number of vertices with positive in-degree. The behaviour of the $\hat{\mathbf{D}}_{m,n}$ for $m \leq R_n$ and $m > R_n$ is rather different. Before R_n, new vertices are discovered with probability proportional to their in-degree. After R_n, all vertices with positive in-degree have already been discovered and we choose to explore the remaining vertices in some uniform order.

Later in Section 3.5, we show that we only need to consider timescales of the order of $m = \Theta(n^{2/3})$. Let $p = \mathbb{P}(D^- > 0)$ such that R_n is distributed as a Binomial(n, p) random variable. We show now that the probability that $m \leq R_n$ will converge exponentially to 1.

Lemma 3.3.1. *If $m = \Theta(n^{2/3})$ then there exists $c > 0$ such that $\mathbb{P}(R_n < m) < e^{-cn}$.*

Proof. If $m = \Theta(n^{2/3})$ then $\mathbb{E}[R_n] - m = pn - m = \Theta(n)$. Thus by Hoeffding's inequality

$$\mathbb{P}(R_n < m) = \mathbb{P}\left(\mathbb{E}[R_n] - R_n > \mathbb{E}[R_n] - m\right) \leq e^{-\frac{2}{n}(\mathbb{E}[R_n]-m)^2} < e^{-cn}$$

for some $c > 0$. □

Hence it is sensible to prove results on the event that $m \leq R_n$.

When discussing the criticality condition, we gave heuristics showing that the limiting distribution of $\hat{D}_{n,1}$ is given by $\mathbf{Z}$ where

$$\mathbb{P}(Z^- = k^-, Z^+ = k^+) = \frac{k^-}{\mu}\mathbb{P}(D^- = k^-, D^+ = k^+).$$

Similarly, $\hat{D}_{n,2}$ is also approximately distributed like $\mathbf{Z}$ for large n, and so on. In this section we in fact prove a precise relation between $\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,m}$ and the sequence $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ of i.i.d. copies of $\mathbf{Z}$.

The results proved in this section do not actually require the criticality condition, so let us define notation for the mean of the $Z_i^\pm$ and the two corresponding centered random walks. Let

$$\lambda_\pm = \mathbb{E}[Z_1^\pm] \quad \text{and} \quad V^\pm(n) = \textstyle\sum_{i=1}^n (Z_i^\pm - \lambda_\pm).$$

The criticality condition is then equivalent to assuming $\lambda_+ = 1$. We also define the notation

$$\Xi_{n-m}^\pm = \sum_{i=m+1}^n D_i^\pm \quad \text{and} \quad \Delta_n = \Xi_n^- - \Xi_n^+$$

such that $\{\Delta_n = 0\}$ is the event that the total out-degree is equal to the total in-degree.

The following lemma asserts the existence of the measure change ϕ_m^n, and its joint scaling limit with the random walks V^- and V^+ when $m = \lfloor n^{2/3} T \rfloor$ for some $T > 0$.

Proposition 3.3.2. *For all positive integers n and m such that $m \leq n$, there exists a function $\phi_m^n : (\mathbb{N} \times \mathbb{N})^n \to [0, \infty)$ such that*

$$\mathbb{E}\left[u(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,m}) \mathbb{1}\{R_n \geq m\} \middle| \Delta_n = 0\right] = \mathbb{E}[u(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)\phi_m^n(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)]$$

for all bounded test functions $u : (\mathbb{N} \times \mathbb{N})^m \to \mathbb{R}$. Define

$$\Phi(n, m) = \phi_m^n(\mathbf{Z}_1, \ldots, \mathbf{Z}_m).$$

Further, let (W^-, W^+) be a pair of correlated standard Brownian motions with correlation

$\mathrm{Corr}(Z_1^-, Z_1^+)$ *and, for* $T > 0$, *define*

$$\Phi(T) = \exp\left(-\frac{\sigma_-}{\mu}\int_0^T s\,\mathrm{d}W_s^- - \frac{\sigma_-^2}{6\mu^2}T^3\right).$$

Then for all $T > 0$,

$$\left(\Phi(n, \lfloor n^{2/3}T \rfloor), \left(n^{-1/3}V^-\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}V^+\left(\lfloor n^{2/3}t\rfloor\right)\right)_{t\in[0,T]}\right)$$
$$\xrightarrow{\text{(d)}} \left(\Phi(T), (\sigma_- W_t^-, \sigma_+ W_t^+)_{t\in[0,T]}\right)$$

in $\mathbb{R} \times \mathbb{D}([0,T], \mathbb{R}^2)$ *as* $n \to \infty$, *even in the absence of the criticality condition.*

The rest of this section is dedicated to proving this proposition.

3.3.1 Exact form of the measure change

To determine the exact form of the measure change, we first need to know the law of the ordering of the first R_n vertices. Let $\mathcal{I}_n = \{i \in [n] : D_i^- > 0\}$. The first R_n vertices we explore in Algorithm 1 will have positive in-degree, thus there exists a random bijection $\Sigma_n : [R_n] \to \mathcal{I}_n$ such that $\hat{\mathbf{D}}_{n,i} = \mathbf{D}_{\Sigma_n(i)}$ for $i = 1, \ldots, R_n$.

Lemma 3.3.3. Σ_n *has law*

$$\mathbb{P}(\Sigma_n = \sigma \mid \mathbf{D}_1, \ldots, \mathbf{D}_n) = \prod_{i=1}^{R_n} \frac{D^-_{\sigma(i)}}{\sum_{j=i}^{R_n} D^-_{\sigma(j)}}.$$

for all bijections $\sigma : [R_n] \to \mathcal{I}_n$.

Proof. In Algorithm 1, a vertex first becomes explored in two ways. Either it is at the start of an out-component or it is discovered when an out-half-edge is paired to one its in-half-edges.

Suppose we have explored m vertices and $m < R_n$. If the next vertex is explored by pairing one of its in-half-edges, then we have chosen it with probability proportional to its in-degree since in- and out-half-edges are paired uniformly at random. Otherwise, it is at the start of a new out-component, and since $m < R_n$, there are still vertices of positive in-degree. Thus we still pick a new vertex with probability proportional to its in-degree.

Therefore in all cases,

$$\begin{aligned}&\mathbb{P}(\Sigma_n(m+1)=\sigma(m+1)\mid\Sigma_n(1)=\sigma(1),\ldots,\Sigma_n(m)=\sigma(m),\mathbf{D}_1,\ldots,\mathbf{D}_n)\\=&\frac{D^-_{\sigma(m+1)}}{\sum_{i\in\mathcal{I}_n}D^-_i-\sum_{j=1}^m D^-_{\sigma(j)}}=\frac{D^-_{\sigma(m+1)}}{\sum_{j=m+1}^{R_n}D^-_{\sigma(j)}}.\end{aligned}$$

From this, repeated applications of the definition of conditional probability yields the desired result. □

Next we establish the form of the measure change when we condition on the exact value of R_n but not $\Delta_n=0$.

Lemma 3.3.4. *For all integers* $0\le r\le n$ *and test functions* $u:(\mathbb{N}\times\mathbb{N})^r\times\mathbb{N}\to\mathbb{R}$,

$$\mathbb{E}\left[u\left(\hat{\mathbf{D}}_{n,1},\ldots,\hat{\mathbf{D}}_{n,r},\textstyle\sum_{i\in\mathcal{I}_n^c}D^+_i\right)\,\middle|\,R_n=r\right]=\mathbb{E}\left[u\left(\mathbf{Z}_1,\ldots,\mathbf{Z}_r,\textstyle\sum_{i=1}^{n-r}E^+_i\right)\psi_r(\mathbf{Z}_1,\ldots,\mathbf{Z}_r)\right]$$

where

$$\psi_r(\mathbf{k}_1,\ldots,\mathbf{k}_r)=\frac{1}{p^r}\prod_{i=1}^r\frac{(r-i+1)\mu}{\sum_{j=i}^r k^-_i}.$$

and $E^+_1,E^+_2,\ldots$ *are i.i.d. random variables such that* E^+_i *has the same distribution as* D^+ *conditioned on* $D^-=0$. *We take the sequences* $(E^+_i)_{i\ge1}$ *and* $(\mathbf{Z}_i)_{i\ge1}$ *to be independent.*

Proof. For any $\mathbf{k}_1,\ldots,\mathbf{k}_m\in\mathbb{N}^+\times\mathbb{N}$ for all i and $s\in\mathbb{N}$.

$$\begin{aligned}&\mathbb{P}\left(\hat{\mathbf{D}}_{n,1}=\mathbf{k}_1,\ldots,\hat{\mathbf{D}}_{n,r}=\mathbf{k}_r,\textstyle\sum_{i\in\mathcal{I}_n^c}D^+_i=s,R_n=r\right)\\=&\sum_{\substack{I\subseteq[n]\\|I|=r}}\sum_{\sigma:[r]\to I}\mathbb{P}\left(\mathbf{D}_{\Sigma_n(1)}=\mathbf{k}_1,\ldots,\mathbf{D}_{\Sigma_n(r)}=\mathbf{k}_r,\textstyle\sum_{i\in\mathcal{I}_n^c}D^+_i=s,\mathcal{I}_n=I,\Sigma_n=\sigma\right)\end{aligned}$$

where the second summation is taken over all bijections $\sigma:[r]\to I$. We examine a single summand.

$$\begin{aligned}&\mathbb{P}\left(\mathbf{D}_{\Sigma_n(1)}=\mathbf{k}_1,\ldots,\mathbf{D}_{\Sigma_n(r)}=\mathbf{k}_r,\textstyle\sum_{i\in\mathcal{I}_n^c}D^+_i=s,\mathcal{I}_n=I,\Sigma_n=\sigma\right)\\=&\mathbb{P}\left(\mathbf{D}_{\sigma(j)}=\mathbf{k}_j\text{ for }j=1,\ldots,r,\textstyle\sum_{i\in I^c}D^+_i=s,D^-_i=0\text{ for }i\in I^c,\Sigma_n=\sigma\right)\\=&\prod_{i=1}^r\frac{k^-_i}{\sum_{j=i}^r k^-_j}\times\prod_{i=1}^r\lambda_{\mathbf{k}_i}\times\mathbb{P}\left(\textstyle\sum_{i\in I^c}D^+_i=s,D^-_i=0\text{ for }i\in I^c\right).\end{aligned}$$

where $\lambda_{\mathbf{k}} = \mathbb{P}(\mathbf{D}_1 = \mathbf{k})$. We have

$$\mathbb{P}\left(\sum_{i\in I^c} D_i^+ = s, D_i^- = 0 \text{ for } i \in I^c\right) = (1-p)^{n-r}\mathbb{P}\left(\sum_{i=1}^{n-r} E_i^+ = s\right).$$

Also

$$\begin{aligned}\prod_{i=1}^r \frac{k_i^-}{\sum_{j=i}^r k_j^-} \times \prod_{i=1}^r \lambda_{\mathbf{k}_i} &= \prod_{i=1}^r \frac{k_i^-}{\mu}\lambda_{\mathbf{k}_i} \times \prod_{i=1}^r \frac{\mu}{\sum_{j=i}^r k_j^-} \\ &= \mathbb{P}(\mathbf{Z}_1 = \mathbf{k}_1, \dots, \mathbf{Z}_r = \mathbf{k}_r) \times \prod_{i=1}^r \frac{\mu}{\sum_{j=i}^r k_j^-}.\end{aligned}$$

Therefore

$$\begin{aligned}&\mathbb{P}\left(\hat{\mathbf{D}}_{n,1} = \mathbf{k}_1, \dots, \hat{\mathbf{D}}_{n,r} = \mathbf{k}_r, \sum_{i\in\mathcal{I}_n^c} D_i^+ = s, R_n = r\right) \\ =&\binom{n}{r} \times r! \times \prod_{i=1}^r \frac{\mu}{\sum_{j=i}^r k_j^-} \times (1-p)^{n-r} \times \mathbb{P}\left(\mathbf{Z}_1 = \mathbf{k}_1, \dots, \mathbf{Z}_r = \mathbf{k}_r, \sum_{i=1}^{n-r} E_i^+ = s\right) \\ =&\binom{n}{r} p^r(1-p)^{n-r} \times \frac{1}{p^r}\prod_{i=1}^r \frac{(r-i+1)\mu}{\sum_{j=i}^r k_i^-} \times \mathbb{P}\left(\mathbf{Z}_1 = \mathbf{k}_1, \dots, \mathbf{Z}_r = \mathbf{k}_r, \sum_{i=1}^{n-r} E_i^+ = s\right).\end{aligned}$$

Finally dividing by $\mathbb{P}(R_n = r) = \binom{n}{r}p^r(1-p)^{n-r}$ gives the desired measure change. □

Using the previous lemma we can prove existence and give the exact form of the desired measure change ϕ_m^n.

Lemma 3.3.5. *For all $m \le n$ and test functions $u : (\mathbb{N}\times\mathbb{N})^m \to \mathbb{R}$,*

$$\mathbb{E}\left[u\left(\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,m}\right) \mathbb{1}\{R_n \ge m\} \,\middle|\, \Delta_n = 0\right] = \mathbb{E}\left[u\left(\mathbf{Z}_1, \dots, \mathbf{Z}_m\right)\phi_m^n(\mathbf{Z}_1, \dots, \mathbf{Z}_m)\right],$$

where

$$\phi_m^n(\mathbf{k}_1, \dots, \mathbf{k}_m) = \frac{1}{\mathbb{P}(\Delta_n = 0)}\mathbb{E}\left[\mathbb{1}\left\{\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-)\right\}\prod_{i=1}^m \frac{(n-i+1)\mu}{\sum_{j=1}^m k_j^- + \Xi_{n-m}^-}\right].$$

Proof. By Lemma 3.3.4, for all $r \ge m$

$$\begin{aligned}&\mathbb{E}\left[u\left(\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,m}\right)\mathbb{1}\{\Delta_n = 0\} \,\middle|\, R_n = r\right] \\ =&\mathbb{E}\left[u\left(\mathbf{Z}_1, \dots, \mathbf{Z}_m\right)\mathbb{1}\left\{\sum_{i=1}^r (Z_i^- - Z_i^+) - \sum_{i=1}^{n-r} E_i^+ = 0\right\}\frac{1}{p^r}\prod_{i=1}^r \frac{(r-i+1)\mu}{\sum_{j=i}^r Z_j^-}\right]\end{aligned}$$

$$=\mathbb{E}\left[u\left(\mathbf{Z}_1,\ldots,\mathbf{Z}_m\right)\mathbb{E}\left[\mathbb{1}\left\{\sum_{i=1}^{r}(Z_i^- - Z_i^+) - \sum_{i=1}^{n-r}E_i^+ = 0\right\}\frac{1}{p^r}\prod_{i=1}^{r}\frac{(r-i+1)\mu}{\sum_{j=i}^{r}Z_j^-}\,\middle|\,\mathbf{Z}_1,\ldots,\mathbf{Z}_m\right]\right]$$

$$=\mathbb{E}\left[u\left(\mathbf{Z}_1,\ldots,\mathbf{Z}_m\right)\tilde{\gamma}_r^{n,m}(\mathbf{Z}_1,\ldots,\mathbf{Z}_m)\right],$$

where

$$\tilde{\gamma}_r^{n,m}(\mathbf{k}_1,\ldots,\mathbf{k}_m) = \mathbb{E}\Bigg[\mathbb{1}\left\{\sum_{i=m+1}^{r}(Z_i^- - Z_i^+) - \sum_{i=1}^{n-r}E_i^+ = \sum_{i=1}^{m}(k_i^+ - k_i^-)\right\}\times$$
$$\frac{1}{p^m}\prod_{i=1}^{m}\frac{(r-i+1)\mu}{\sum_{j=i}^{m}k_j^- + \sum_{j=m+1}^{r}Z_j^-}\frac{1}{p^{r-m}}\prod_{i=m+1}^{r}\frac{(r-i+1)\mu}{\sum_{j=i}^{r}Z_j^-}\Bigg]$$
$$= \mathbb{E}\Bigg[\mathbb{1}\left\{\sum_{i=1}^{r-m}(Z_i^- - Z_i^+) - \sum_{i=1}^{n-r}E_i^+ = \sum_{i=1}^{m}(k_i^+ - k_i^-)\right\}\times$$
$$\frac{1}{p^m}\prod_{i=1}^{m}\frac{(r-i+1)\mu}{\sum_{j=i}^{m}k_j^- + \sum_{j=1}^{r-m}Z_j^-}\frac{1}{p^{r-m}}\prod_{i=1}^{r-m}\frac{(r-m-i+1)\mu}{\sum_{j=i}^{r-m}Z_j^-}\Bigg],$$

since $(\mathbf{Z}_i)_{i=m+1}^{r}$ has the same law as $(\mathbf{Z}_i)_{i=1}^{r-m}$. Then applying Lemma 3.3.4 again shows that

$$\tilde{\gamma}_r^{n,m}(\mathbf{k}_1,\ldots,\mathbf{k}_m) = \mathbb{E}\Bigg[\mathbb{1}\left\{\sum_{i=1}^{r-m}(\hat{D}_{n-m,i}^- - \hat{D}_{n-m,i}^+) - \sum_{i\in\mathcal{I}_{n-m}^c}D_i^+ = \sum_{i=1}^{m}(k_i^+ - k_i^-)\right\}\times$$
$$\frac{1}{p^m}\prod_{i=1}^{m}\frac{(r-i+1)\mu}{\sum_{j=i}^{m}k_j^- + \sum_{j=1}^{r-m}\hat{D}_{n-m,j}^-}\,\Bigg|\,R_{n-m} = r-m\Bigg].$$

Conditional on $R_{n-m} = r-m$, we have

$$\sum_{j=1}^{r-m}(\hat{D}_{n-m,j}^- - \hat{D}_{n-m,j}^+) - \sum_{i\in\mathcal{I}_{n-m}^c}D_i^+ = \Delta_{n-m} \quad\text{and}\quad \sum_{j=1}^{r-m}\hat{D}_{n-m,j}^- = \Xi_{n-m}^-.$$

Therefore,

$$\tilde{\gamma}_r^{n,m}(\mathbf{k}_1,\ldots,\mathbf{k}_m) = \mathbb{E}\left[\frac{1}{p^m}\prod_{i=1}^{m}\frac{(r-i+1)\mu}{\sum_{j=i}^{m}k_j^- + \Xi_{n-m}^-}\mathbb{1}_{A_n}\,\middle|\,R_{n-m} = r-m\right],$$

where

$$A_n(\mathbf{k}_1,\ldots,\mathbf{k}_m) = \left\{\Delta_{n-m} = \sum_{i=1}^{m}(k_i^+ - k_i^-)\right\}.$$

Hence,

$$\mathbb{E}\left[u\left(\hat{\mathbf{D}}_{n,1},\ldots,\hat{\mathbf{D}}_{n,m}\right)\mathbb{1}\{R_n \geq m, \Delta_n = 0\}\right] = \mathbb{E}\left[u\left(\mathbf{Z}_1,\ldots,\mathbf{Z}_m\right)\tilde{\phi}_m^n(\mathbf{Z}_1,\ldots,\mathbf{Z}_m)\right],$$

where

$$\begin{aligned}
&\tilde{\phi}^n_m(\mathbf{k}_1, \dots, \mathbf{k}_m) \\
&= \sum_{r=m}^{n} \binom{n}{r} p^r (1-p)^{n-r} \mathbb{E}\left[\frac{1}{p^m} \prod_{i=1}^{m} \frac{(r-i+1)\mu}{\sum_{j=i}^m k_j^- + \Xi^-_{n-m}} \mathbb{1}_{A_n} \,\middle|\, R_{n-m} = r-m \right] \\
&= \sum_{l=0}^{n-m} \binom{n}{l+m} p^{l+m} (1-p)^{n-m-l} \mathbb{E}\left[\frac{1}{p^m} \prod_{i=1}^{m} \frac{(l+m-i+1)\mu}{\sum_{j=i}^m k_j^- + \Xi^-_{n-m}} \mathbb{1}_{A_n} \,\middle|\, R_{n-m} = l \right].
\end{aligned}$$

We wish to view the sum as an expectation over R_{n-m}. In order to do this, we rewrite the expression so that we are taking a sum over the probabilities of a Binomial$(n-m, p)$ distribution. We can calculate

$$\frac{\binom{n}{l+m} p^{l+m}(1-p)^{n-m-l}}{\binom{n-m}{l} p^l (1-p)^{n-m-l}} = p^m \prod_{i=1}^{m} \frac{(n-i+1)}{(l+m-i+1)}.$$

Therefore,

$$\begin{aligned}
\tilde{\phi}^n_m(\mathbf{k}_1, \dots, \mathbf{k}_m) &= \sum_{l=1}^{n-m} \binom{n-m}{l} p^l (1-p)^{n-m-l} \mathbb{E}\left[\prod_{i=1}^{m} \frac{(n-i+1)\mu}{\sum_{j=i}^m k_j^- + \Xi^-_{n-m}} \mathbb{1}_{A_n} \,\middle|\, R_{n-m} = l \right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{m} \frac{(n-i+1)\mu}{\sum_{j=i}^m k_j^- + \Xi^-_{n-m}} \mathbb{1}_{A_n} \,\middle|\, R_{n-m} \right]\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{m} \frac{(n-i+1)\mu}{\sum_{j=i}^m k_j^- + \Xi^-_{n-m}} \mathbb{1}_{A_n} \right].
\end{aligned}$$

Finally, dividing by $\mathbb{P}(\Delta_n = 0)$ yields the desired form of ϕ^n_m. □

3.3.2 Asymptotic lower bound on the measure change

Recall that our goal in Proposition 3.3.2 is to determine the limiting distribution of

$$\Phi(n, m) = \phi^n_m(\mathbf{Z}_1, \dots, \mathbf{Z}_m),$$

as $n \to \infty$, in the regime where $m = \Theta(n^{2/3})$. When dealing with convergence in distribution, it is suffcient and necessary to work on a sequence of events occuring with high probability. In particular, for the proof of Proposition 3.3.2, we work on the event $\mathcal{E}_m$ where

$$\begin{aligned}
\mathcal{E}_m = \Bigg\{ &\max_{i=1,\dots,m} \left| \textstyle\sum_{j=1}^i (Z_j^- - \lambda_-) \right| \le m^{1/2} \log(m) \\
&\qquad\qquad \text{and} \quad \max_{i=1,\dots,m} \left| \textstyle\sum_{j=1}^i (Z_j^+ - \lambda_+) \right| \le m^{1/2} \log(m) \Bigg\}.
\end{aligned}$$

This says that the centered random walks corresponding to Z_i^+ and Z_i^- both do not deviate by more than $m^{1/2}\log(m)$ in the first m steps. The conditions in Section 3.1.3 ensure each Z_i^+ and Z_i^- has finite variance, thus this event will occur with high probability.

The following lemma is an analogue of Conchon-Kerjan and Goldschmidt [37, Lemma 6.7]. In it we prove a deterministic lower bound on $\phi_m^n(\mathbf{k}_1,\ldots,\mathbf{k}_m)$, for all $\mathbf{k}_1,\ldots,\mathbf{k}_m$ corresponding to the event $\mathcal{E}_m$, up to an error which vanishes as $n\to\infty$.

Proposition 3.3.6. *Define*

$$s^\pm(i) = \textstyle\sum_{j=1}^{i}(k_i^\pm - \lambda_\pm).$$

Suppose that $\mathbf{k}_1,\ldots,\mathbf{k}_m$ *are such that*

$$\max_{i=1,\ldots,m}|s^-(i)| \le m^{\frac{1}{2}}\log(m) \quad \textit{and} \quad \max_{i=1,\ldots,m}|s^+(i)| \le m^{\frac{1}{2}}\log(m) \tag{3.1}$$

Then in the regime $m=\Theta(n^{2/3})$, *as* $n\to\infty$,

$$\phi_m^n(\mathbf{k}_1,\ldots,\mathbf{k}_m) \ge \exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}(s^-(i)-s^-(m)) - \frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right) + o(1),$$

where the $o(1)$ *error term is independent of* $\mathbf{k}_1,\ldots,\mathbf{k}_m$ *satisfying the assumption in Eq.* (3.1).

The fact that we only prove a lower bound may seem strange at first. To understand why this is sufficient, first note that all measure changes are non-negative random variables and have expectation 1. Hence if the sequence of lower bounds on the measure changes converge to a limit that also has expectation 1, then we have not have lost a significant amount of probability mass. It follows that the measure changes converge to the same limit as the lower bounds. This is made formal by Conchon-Kerjan and Goldschmidt [37, Lemma 4.8]. In Proposition 3.3.2 we are considering the joint convergence of the measure change with two other random walks, and thus we adapt [37, Lemma 4.8] to allow for an additional coordinate that is converging jointly with the first coordinate.

Lemma 3.3.7. *Let* $(X_n,Y_n,Z_n)_{n\ge1}$ *be a sequence of* $[0,\infty)\times[0,\infty)\times S$*-valued random variables where* S *is a metric space. Suppose there exists a* $[0,\infty)\times S$*-valued random variable* (Y,Z) *such that the following holds:*

1. $(Y_n,Z_n)\xrightarrow{\text{(d)}}(Y,Z)$ *as* $n\to\infty$.
2. $X_n\ge Y_n$ *almost surely for all* n.

3. $\mathbb{E}[X_n] = 1$ *for all* n *and* $\mathbb{E}[Y] = 1$.

Then $(X_n, Z_n) \xrightarrow{\text{(d)}} (Y, Z)$ *also. Moreover* $(X_n)_{n\geq 1}$ *is a sequence of uniformly integrable random variables.*

The proof of this lemma is obtained by simply adding the corresponding Z_n or Z coordinate to quantities in the proof of [37, Lemma 4.8] and so we will not repeat it here.

3.3.2.1 Discrete local limit theorem

To prove Proposition 3.3.6, we first need to understand the denominator of ϕ_m^n, which, as given by Lemma 3.3.5, is $\mathbb{P}(\Delta_n = 0)$. The random variable Δ_n is a sum of independent integer-valued random variables and the asymptotic behaviour of such a sum being equal to some value is described by the discrete local limit theorem. Such a theorem was first proven by Gnedenko [57]. Here we borrow the presentation from Durrett [52, Section 3.5].

Let $X_1, X_2, \ldots$ be i.i.d. integer-valued random variables with mean μ and finite variance σ^2. Then, by the central limit theorem,

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{(d)}} N(0,1)$$

as $n \to \infty$. Thus we expect $\sum_{i=1}^n X_i$ to be distributed like a $N(n\mu, n\sigma^2)$ random variable for large values of n. Therefore the probability mass function of $\sum_{i=1}^n X_i$ should be well approximated by the probability density function of a $N(n\mu, n\sigma^2)$ distribution, i.e.

$$\mathbb{P}\big(\textstyle\sum_{i=1}^n X_i = s\big) \approx \tfrac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(\tfrac{-(s-n\mu)^2}{2n\sigma^2}\right)$$

for all integers s. Specifically we hope that

$$\sup_{s\in\mathbb{Z}} \left| \mathbb{P}\big(\textstyle\sum_{i=1}^n X_i = s\big) - \tfrac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(\tfrac{-(s-n\mu)^2}{2n\sigma^2}\right) \right| = o(n^{-1/2}). \tag{3.2}$$

This, however, is not always the case. Suppose, for example, that each X_i is almost surely even such that $\mathbb{P}\left(\sum_{i=1}^n X_i = s\right) = 0$ for all odd s. Let s_n be the closest odd integer to $n\mu$. Then

$$\sup_{s\in\mathbb{Z}} \left| \mathbb{P}\big(\textstyle\sum_{i=1}^n X_i = s\big) - \tfrac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(\tfrac{-(s-n\mu)^2}{2n\sigma^2}\right) \right| \geq \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(\frac{-(s_n - n\mu)^2}{2n\sigma^2}\right) = \Theta(n^{-1/2}).$$

Fortunately this kind of periodic behaviour can be mitigated by normalizing the random variables. A one-dimensional random variable X is *lattice* if it is not almost surely constant, and there exists $h > 0$ and $c \in \mathbb{R}$ such that $X \in c + h\mathbb{Z}$ almost surely. The largest such h is called the *span* of X. For example, if X is almost surely even then X has span at least 2. If X is lattice with span h and c is in the support of X, then the affine transform $\frac{1}{h}(X - c)$ is an integer-valued random variable with span 1, for which it can be shown that the approximation in Eq. (3.2) does hold. This gives us the discrete local limit theorem:

Theorem 3.3.8 (Discrete local limit theorem)**.** *Let $X_1, X_2, \ldots$ be i.i.d. $\mathbb{R}$-valued lattice random variables with span h and fix arbitrary $c \in \operatorname{supp}(X_1)$. Then*

$$\sup_{s \in nc + h\mathbb{Z}} \left| \mathbb{P}\big(\textstyle\sum_{i=1}^n X_i = s \big) - \frac{h}{\sqrt{2\pi n \sigma^2}} \exp\left(\frac{-(s-n\mu)^2}{2n\sigma^2} \right) \right| = o(n^{-1/2}).$$

Remark 3.3.9. *For each sequence of integers $(s_n)_{n \geq 1}$ such that $|s_n - n\mu| = \omega(n^{1/2})$, we have that*

$$\frac{1}{\sqrt{2\pi n \sigma^2}} \exp\left(\frac{-(s_n - n\mu)^2}{2n\sigma^2} \right) = o(n^{-1/2}).$$

Hence the discrete local limit theorem (Theorem 3.3.8) tells you only that $\mathbb{P}\left(\sum_{i=1}^n X_i = s_n\right) = o(n^{-1/2})$; it gives no precise characterization of the leading order term.

While this remark will be important later, here Δ_n is centered and we are interested in the probability $\mathbb{P}(\Delta_n = 0)$. In addition, the strong aperiodicity condition in Section 3.1.3 tells us exactly that the $D^- - D^+$ is lattice with span 1. Thus the following is a direct corollary of the discrete local limit theorem (Theorem 3.3.8).

Corollary 3.3.10. *We have*

$$\mathbb{P}(\Delta_n = 0) = \frac{1}{\sqrt{2\pi\sigma^2 n}} + o(n^{-1/2})$$

as $n \to \infty$, where σ is the variance of $D^- - D^+$.

Remark 3.3.11. *The exact value of σ^2 is not important for the asymptotic behaviour of ϕ_m^n because we show later that it will cancel with a term in the numerator of ϕ_m^n.*

3.3.2.2 Exponential Tilting

Next we turn to the numerator of ϕ^n_m. By Lemma 3.3.5, this is given by

$$\mathbb{E}\left[\mathbb{1}\left\{\Delta_{n-m}=\sum_{i=1}^{m}(k_i^+-k_i^-)\right\}\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^-+\Xi^-_{n-m}}\right]. \tag{3.3}$$

For convenience, let $\mathcal{A}_n$ denote the event in the indicator function, i.e.

$$\mathcal{A}_n=\left\{\Delta_{n-m}=\sum_{i=1}^{m}(k_i^+-k_i^-)\right\}.$$

When the $\mathbf{k}_1,\ldots,\mathbf{k}_m$ satisfy the condition in Eq. (3.1), we face two problems in evaluating the expectation in Eq. (3.3).

The first problem concerns the event $\mathcal{A}_n$. To evaluate the expectation we need to understand the asymptotic probability of this event. Unfortunately a naïve application of the discrete local limit theorem will not work in this case, as we now explain. Firstly, note that

$$\sum_{i=1}^{m}(k_i^--k_i^+)=s^-(m)-s^+(m)+(\lambda_+-\lambda_-)m.$$

We have that

$$\lambda_+-\lambda_-=\mathbb{E}[Z^--Z^+]=\tfrac{1}{\mu}\mathbb{E}[D^-D^+-(D^-)^2]$$

which is, in general, non-zero. Then $m=\Theta(n^{2/3})$ whereas, if $\mathbf{k}_1,\ldots,\mathbf{k}_n$ satisfy Eq. (3.1), $s^-(m)$ and $s^+(m)$ are both of order $O(n^{1/3}\log n)$. Therefore

$$\sum_{i=1}^{m}(k_i^--k_i^+)=\Theta(n^{2/3}).$$

In contrast, Δ_{n-m} is centered, so $\mathcal{A}_n$ is looking at the event that Δ_{n-m} takes a value at distance $\Theta(n^{2/3})$ away from its mean. As stated in Remark 3.3.9, the discrete local limit theorem provides no useful information in this regime.

The second problem is that even in absence of the indicator function, the expectation being evaluated in Eq. (3.3) is not dictated by the typical fluctuations of the random variables Ξ^-_{n-m}. In other words, it is not the case that

$$\mathbb{E}\left[\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^-+\Xi^-_{n-m}}\right]\not\approx\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^-+\mathbb{E}[\Xi^-_{n-m}]} \tag{3.4}$$

It turns out that both of these issues can be addressed by introducing a sequence of exponentially tilted measures. The first effect of the exponentially tilted measures will be to shift the mean of Δ_{n-m} in such a way that, after the tilting, the event $\mathcal{A}_n$ concerns only a typical deviation of Δ_{n-m} which can be addressed by a local limit theorem. The second effect is that the expectation being evaluated in Eq. (3.3) will be dictated by the typical fluctations of Ξ^-_{n-m} under the tilted measure.

The next result defines this tilt and then gives asymptotic expansions for cumulant generating function of D^-, the mean of D^- and the mean of D^+ under this tilting.

Lemma 3.3.12. *Define an measure $\mathbb{P}_\theta$, for $\theta \geq 0$, by its Radon–Nikodym derivative*

$$\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}} = \exp\left(-\theta D^- - \alpha(\theta)\right) \quad \textit{where} \quad \alpha(\theta) = \log \mathbb{E}\left[e^{-\theta D^-}\right].$$

Then as $\theta \downarrow 0$ we have

$$\begin{aligned} \alpha(\theta) &= -\mu\theta + \tfrac{1}{2}\operatorname{Var}(D^-)\theta^2 - \tfrac{1}{6}\mathbb{E}\left[(D^- - \mu)^3\right]\theta^3 + o(\theta^3), \\ \mathbb{E}_\theta[D^-] &= \mu - \operatorname{Var}(D^-)\theta + O(\theta^2), \\ \textit{and} \quad \mathbb{E}_\theta[D^+] &= \mu - \operatorname{Cov}(D^-, D^+)\theta + O(\theta^2). \end{aligned}$$

Proof. Since $\mathbb{E}\left[|D^-|^3\right] < \infty$ and D^- is non-negative, by the dominated convergence theorem

$$\mathbb{E}\left[(D^-)^3 \exp(-\theta D^-)\right] = \mathbb{E}\left[(D^-)^3\right] + o(1) \tag{3.5}$$

as $\theta \downarrow 0$. Integrating Eq. (3.5) with respect to θ and applying Fubini's theorem to exchange the order of the expectation and integral gives

$$\mathbb{E}\left[\int_0^\theta (D^-)^3 e^{-\theta' D^-}\,\mathrm{d}\theta'\right] = \mathbb{E}\left[\int_0^\theta \left\{(D^-)^3 + o(1)\right\}\mathrm{d}\theta'\right] = \mathbb{E}\left[(D^-)^3\right]\theta + o(\theta).$$

Evaluating the integral with respect to θ' on the left hand side and rearranging gives that

$$\mathbb{E}\left[(D^-)^2 e^{-\theta D^-}\right] = \mathbb{E}\left[(D^-)^2\right] - \mathbb{E}\left[(D^-)^3\right]\theta + o(\theta).$$

Repeating this method yields

$$\mathbb{E}\left[D^- e^{-\theta D^-}\right] = \mu - \mathbb{E}\left[(D^-)^2\right]\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^3\right]\theta^2 + o(\theta^2), \tag{3.6}$$

$$\text{and}\quad \mathbb{E}\left[e^{-\theta D^-}\right] = 1 - \mu\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^2\right]\theta^2 - \tfrac{1}{6}\mathbb{E}\left[(D^-)^3\right]\theta^3 + o(\theta^3). \tag{3.7}$$

Similarly integrating the equation

$$\mathbb{E}\left[(D^-)^2 D^+ \exp(-\theta D^-)\right] = \mathbb{E}\left[(D^-)^2 D^+\right] + o(1)$$

twice gives

$$\mathbb{E}\left[D^+ e^{-\theta D^-}\right] = \mu\theta - \mathbb{E}\left[D^- D^+\right]\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^2 D^+\right]\theta^2 + o(\theta^2). \tag{3.8}$$

Eq. (3.7) gives the small-θ expansion of the normalising constant of the measure change. Combining this with Eq. (3.6) and Eq. (3.8) yields the expansions for $\mathbb{E}_\theta[D^-]$ and $\mathbb{E}_\theta[D^+]$ respectively. Taking the logarithm of Eq. (3.7) gives the expansion of the cumulant generating function $\alpha(\theta)$. □

To achieve the recentering of Δ_{n-m} we desire, let us define a sequence of tilted measures $\mathbb{P}_n$ defined by their Radon–Nikodym derivative

$$\frac{\mathrm{d}\mathbb{P}_n}{\mathrm{d}\mathbb{P}} = \exp\left(-\theta_n \Xi^-_{n-m} - (n-m)\alpha(\theta_n)\right), \tag{3.9}$$

where $\theta_n = \frac{m}{\mu n}$. This factorises and so $\mathbf{D}_1, \ldots, \mathbf{D}_n$ remain i.i.d. under this tilting, each having the law of $\mathbf{D}$ under $\mathbb{P}_{\theta_n}$. Applying Lemma 3.3.12, we can compute that

$$\mathbb{E}_n[\Delta_{n-m}] = m(\lambda_+ - \lambda_-) + O(n^{1/3}).$$

Hence,

$$\begin{aligned}\textstyle\sum_{i=1}^m (k_i^- - k_i^+) - \mathbb{E}_n[\Delta_{n-m}] &= s^-(m) - s^+(m) + \Big[m(\lambda_+ - \lambda_-) - \mathbb{E}_n[\Delta_{n-m}]\Big] \\ &= O(n^{1/3}\log n),\end{aligned}$$

which is within the $O(n^{1/2})$ range from the mean required for a typical deviation. This justifies our choice of $\theta_n = \frac{m}{\mu n}$.

3.3.2.3 Expansion of the numerator

Remarkably the same tilting to apply the local limit theorem also correctly recenters Ξ^-_{n-m} such that the expectation in Eq. (3.3) is dominated by the typical behaviour of Ξ^-_{n-m} under $\mathbb{P}_n$. Using Lemma 3.3.12, we have that

$$\mathbb{E}_n[\Xi^-_{n-m}] = \mu n - \lambda_- m + O(n^{1/3})$$

under the tilting. Thus we will expand the numerator under the event

$$\mathcal{B}_n = \left\{ |\Xi^-_{n-m} - \mu n + \lambda_- m| \le n^{1/2}\log(n) \right\}.$$

This event is saying that Ξ^-_{n-m} is at 'typical fluctations' from its tilted mean. The next lemma then expands the numerator of ϕ^n_m on the event $\mathcal{B}_n$.

Lemma 3.3.13. *We have that*

$$\mathbb{E}\left[\mathbb{1}_{\mathcal{A}_n\cap\mathcal{B}_n}\prod_{i=1}^m \frac{(n-i+1)\mu}{\sum_{j=1}^m k_j^- + \Xi^-_{n-m}}\right] = \left\{\exp\left(\frac{1}{\mu n}\sum_{i=0}^m (s^-(i) - s^-(m)) - \frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right) + o(1)\right\}$$
$$\times \mathbb{P}_n(\mathcal{A}_n \cap \mathcal{B}_n)$$

where the $o(1)$ term is bounded independently of $\mathbf{k}_1, \dots, \mathbf{k}_m$ satisfying the assumption in Eq. (3.1) on page 123.

Proof. Firstly,

$$\prod_{i=1}^m \frac{(n-i+1)\mu}{\sum_{j=i}^m k_j^- + \Xi^-_{n-m}} = \exp(X_n - Y_n),$$

where

$$X_n = \sum_{i=1}^m \log\left(1 - \frac{i-1}{n}\right) \quad \text{and} \quad Y_n = \sum_{i=1}^m \log\left(\frac{\sum_{j=i}^m k_j^- + \Xi^-_{n-m}}{\mu n}\right).$$

Note that

$$\sum_{j=i}^m k_j^- = s^-(m) - s^-(i-1) + (m-i+1)\lambda_-.$$

For convenience, define

$$\Omega^-_n = \Xi^-_{n-m} - \mu n + \lambda_- m$$

such that $\mathcal{B} = \{|\Omega_n^-| < n^{1/2}\log n\}$. Then we have

$$\begin{aligned} Y_n &= \sum_{i=1}^{m} \log\left(\frac{s^-(m) - s^-(i-1) + (m-i+1)\lambda_- + \Omega_n^- + \mu n - \lambda_- m}{\mu n}\right) \\ &= \sum_{i=1}^{m} \log\left(1 + A_{i,n} + B_{i,n}\right) \end{aligned}$$

where

$$A_{i,n} = \frac{1}{\mu n}\left\{\Omega_n^- - \big[s^-(i-1) - s^-(m)\big]\right\}, \quad B_{i,n} = -\frac{\lambda_-}{\mu n}(i-1).$$

Then on the event $\mathcal{B}_n$,

$$\max_{i=1,\ldots,m} |A_{i,n}| = O(n^{-1/2}\log n) \quad \text{and} \quad \max_{i=1,\ldots,m} |B_{i,n}| = O(n^{-1/3}).$$

where the O bounds are uniform for $\mathbf{k}_1, \ldots, \mathbf{k}_m$ satifying Eq. (3.1). There are $m = \Theta(n^{2/3})$ terms in the summation. Thus to keep all terms of order $\Omega(1)$, we keep terms of order $\Omega(n^{-2/3})$, uniformly in i, when expanding $\log(1 + A_{i,n} + B_{i,n})$. The only such terms are $A_{i,n}, B_{i,n}$ and $B_{i,n}^2$. Moreover,

$$\sum_{i=1}^{m} B_{i,n} = -\frac{\lambda_-}{2\mu}\frac{m^2}{n} + o(1) \quad \text{and} \quad \sum_{i=1}^{m} (B_{i,n})^2 = \frac{\lambda_-^2}{3\mu^2}\frac{m^3}{n^2} + o(1).$$

Therefore,

$$\begin{aligned} Y_n &= \sum_{i=1}^{m} (A_{i,n} + B_{i,n} - \tfrac{1}{2}B_{i,n}^2) + o(1) \\ &= -\frac{1}{\mu n}\sum_{i=0}^{m}\big(s^-(i) - s^-(m)\big) + \frac{m}{\mu n}\Omega_n^- - \frac{\lambda_-}{2\mu}\frac{m^2}{n} - \frac{\lambda_-^2}{6\mu^2}\frac{m^3}{n^2} + o(1), \end{aligned}$$

where we use that $\sum_{i=1}^{m}(s^-(i-1) - s^-(m)) = \sum_{i=0}^{m}(s^-(i) - s^-(m))$.

Similarly we can expand X_n as

$$X_n = -\frac{m^2}{2n} - \frac{m^3}{6n^2} + o(1).$$

Thus,

$$\begin{aligned}
&\mathbb{1}_{\mathcal{A}_n\cap\mathcal{B}_n}\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^-+\Xi_{n-m}^-}\\
=&\exp\left(\frac{1}{\mu n}\sum_{i=1}^{m}(s^-(i)-s^-(m))-\frac{m}{\mu n}\Omega_n^-+\frac{(\lambda_--\mu)}{2\mu}\frac{m^2}{n}+\frac{(\lambda_-^2-\mu^2)}{6\mu^2}\frac{m^3}{n^2}+o(1)\right)\mathbb{1}_{\mathcal{A}_n\cap\mathcal{B}_n}.
\end{aligned}$$

In addition, using Lemma 3.3.12, the measure change can be expanded as

$$\frac{\mathrm{d}\mathbb{P}_n}{\mathrm{d}\mathbb{P}}=\exp\left(-\frac{m}{\mu n}\Omega_n^-+\frac{(\lambda_--\mu)}{2\mu}\frac{m^2}{n}+\frac{(\lambda_-^2-\mu^2)}{6\mu^2}\frac{m^3}{n^2}+\frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}+o(1)\right).$$

Hence,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\mathcal{A}_n\cap\mathcal{B}_n}\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^-+\Xi_{n-m}^-}\right]\\
=&\mathbb{E}_n\left[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}_n}\mathbb{1}_{\mathcal{A}_n\cap\mathcal{B}_n}\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^-+\Xi_{n-m}^-}\right]\\
=&\mathbb{E}_n\left[\mathbb{1}_{\mathcal{A}_n\cap\mathcal{B}_n}\exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}(s^-(i)-s^-(m))-\frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}+o(1)\right)\right]\\
=&\left\{\exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}(s^-(i)-s^-(m))-\frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right)+o(1)\right\}\mathbb{P}_n(\mathcal{A}_n\cap\mathcal{B}_n)
\end{aligned}$$

as required. □

3.3.2.4 Multivariate Local Limit Theorem

To complete the proof of Proposition 3.3.6 we need to understand the asymptotic behaviour of $\mathbb{P}_n(\mathcal{A}_n\cap\mathcal{B}_n)$. Recall an effect of the tilting was to center Δ_{n-m} in such a way that the probability of the event

$$\mathcal{A}_n=\{\Delta_{n-m}=\textstyle\sum_{i=1}^{m}(k_i^+-k_i^-)\}$$

can be addressed by the local limit theorem. However, due to the tilting, $\mathbb{P}_n$ changes with n. In effect, Δ_n under $\mathbb{P}_n$ has the same distribution as $\sum_{i=1}^{n-m}X_{n,i}$ where $(X_{n,i})_{i=1}^n$ has the same joint distribution as $(D_i^--D_i^+)_{i=1}^n$ under $\mathbb{P}_n$. Then $X_{n,1},\dots,X_{n,n}$ are i.i.d. but the distribution of $X_{n,1}$ can change with n. A collection of random variables $(X_{n,1},\dots,X_{n,n})_{n=1}^\infty$ satisfying this property is a *row-wise i.i.d. triangular array*. Thus we require a generalisation of the discrete local limit theorem which can deal with such arrays. In addition, to deal with the event

$$\mathcal{B}_n = \left\{ \left| \Xi^-_{n-m} - \mu n + \lambda_- m \right| \leq n^{1/2} \log n \right\},$$

we will prove a multivariate local limit theorem applicable to $(\Delta_{n-m}, \Xi^-_{n-m})$ under $\mathbb{P}_n$ and then sum over the possible values of Ξ^-_n.

Before we state the result we use, we first define some terminology regarding lattices in $\mathbb{R}^d$. A set of points in $\mathbb{R}^d$ is a *lattice* if there exists a basis $\mathbf{a}_1, \ldots, \mathbf{a}_d$ of $\mathbb{R}^d$ such that

$$\Lambda = \left\{ \textstyle\sum_{i=1}^d n_i \mathbf{a}_i : n_i \in \mathbb{Z} \text{ for } i = 1, \ldots, d \right\}.$$

We say Λ is generated by $\mathbf{a}_1, \ldots, \mathbf{a}_d$. We can summarise the basis by a $n \times n$ matrix A whose columns are $\mathbf{a}_1, \ldots, \mathbf{a}_n$. In other words $A_{ij} = \mathbf{a}^{(i)}_j$. The choice of basis generating a lattice is not unique, and the following lemma adapted from [95, Corollary 4.3a] characterises when two basis generate the same lattice.

Lemma 3.3.14. *Let A and B be $n \times n$ matrices of full rank. Then the columns of A and B generate the same lattice if and only if there exists a matrix U such that U has integer entries,* $\det(U) = \pm 1$ *and $A = UB$.*

Therefore we can define $\det(\Lambda)$ to be $|\det(A)|$ for any matrix A whose columns generate Λ, and this definition is independent of the choice of A.

For integer lattices, we can obtain a canonical choice of the basis generating the lattice. We say a $d \times d$ matrix A is in *Hermite normal form* if A is lower triangular with entries

$$A = \begin{pmatrix} a_{1,1} & & 0 \\ \vdots & \ddots & \\ a_{d,1} & \cdots & a_{d,d} \end{pmatrix}$$

satisfying

1. $a_{i,j}$ is a non-negative integer for all $i = 1, \ldots, d$ and $j \geq i$,

2. $a_{i,i} > 0$ for all $i = 1, \ldots, d$, and

3. $a_{j,i} < a_{i,i}$ for all $j > i$.

Then the following lemma, adapted from [34, Corollary 1], gives existence of a canonical choice of basis generating an integer lattice.

Lemma 3.3.15. *Suppose $\Lambda \subseteq \mathbb{Z}^d$ is an integer lattice. Then there exists a $d \times d$ matrix A in Hermite normal form such that the columns of A form a basis which generates Λ.*

An $\mathbb{R}^d$-valued random variable $\mathbf{X}$ is *non-degenerate* if it is not supported on an affine hyperplane of $\mathbb{R}^d$. $\mathbf{X}$ is *lattice* if it is non-degenerate and supported on a translation of a lattice. To avoid dealing with translations, it is convenient to work with the *symmetrisation* of $\mathbf{X}$. This is the random variable $\mathbf{X}^* = \mathbf{X}_1 - \mathbf{X}_2$ where $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent copies of $\mathbf{X}$. For each lattice Λ, $\mathbf{X}$ is supported on a translation of Λ if and only if $\mathbf{X}^*$ is supported on Λ without translation.

If $\mathbf{X}$ is lattice, the *main lattice* $\Lambda(\mathbf{X})$ of $\mathbf{X}$ is the intersection of all lattices containing the support of $\mathbf{X}^*$. This is in itself a lattice, and is explicitly given by

$$\Lambda(\mathbf{X}) = \bigcup_{k=1}^{\infty} \left\{ \textstyle\sum_{i=1}^{k} n_i \mathbf{x}_i^* : n_i \in \mathbb{Z} \text{ and } \mathbf{x}_i^* \in \operatorname{supp}(\mathbf{X}^*) \text{ for } i = 1, \dots, k \right\}.$$

It will turn out that if $\mathbf{X}$ is an $\mathbb{R}^d$-valued lattice random variable with main lattice Λ, then $\det(\Lambda)$ can be seen as a generalisation of the span of an $\mathbb{R}$-valued random variable.

To deal with the triangular array, we recall the exponential tilt is given by

$$\frac{\mathrm{d}\mathbb{P}_n}{\mathrm{d}\mathbb{P}} = \exp(-\theta_n \Xi^-_{n-m} - (n-m)\alpha(\theta_n))$$

where $\theta_n = \frac{m}{\mu n}$. Since $\theta_n \to 0$, the distribution of $\mathbf{D}_i$ under $\mathbb{P}_n$ is converging to that of $\mathbf{D}_i$ under $\mathbb{P}$ as $n \to \infty$. This allows us to ignore the tilting in the limit.

Theorem 3.3.16. *For each $n \geq 1$ let $\mathbf{X}_n$ be an $\mathbb{R}^d$ valued random variable and*

$$\mathbf{X}_{n,1}, \mathbf{X}_{n,2}, \dots, \mathbf{X}_{n,n}$$

be i.i.d. copies of $\mathbf{X}_n$. Assume that the following holds:

1. *There exists a random variable $\mathbf{X}$ such that $\mathbf{X}_n \xrightarrow{(d)} \mathbf{X}$ as $n \to \infty$.*
2. *$(\|\mathbf{X}_n\|^2)_{n\geq 1}$ is a uniformly integrable sequence of random variables. Explicitly*

$$\lim_{L\to\infty} \sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\left\{\|\mathbf{X}_n\|^2 > L\right\}\right] = 0. \tag{3.10}$$

3. For all n, $\mathbf{X}_n$ and $\mathbf{X}$ are lattice with common main lattice Λ.

Then $\mathbf{X}$ has finite second moment. Further, for each n let $\mathbf{c}_n$ be an arbitrary element in the support of $\sum_{i=1}^n \mathbf{X}_{n,i}$. Then uniformly for $\mathbf{y} \in \mathbf{c}_n + \Lambda$,

$$\mathbb{P}\Big(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y} \Big) = n^{-d/2} \det(\Lambda) f\left(\mathbf{x}_n(\mathbf{y})\right) + o\left(n^{-d/2}\right) \quad \text{where} \quad \mathbf{x}_n(\mathbf{y}) = \tfrac{\mathbf{y} - n\mathbb{E}[\mathbf{X}_n]}{\sqrt{n}}$$

and f is the density of a $N(0, \mathrm{Cov}(\mathbf{X}))$ distribution. This means that

$$\lim_{n\to\infty} \sup_{\mathbf{y} \in \mathbf{c}_n + \Lambda} \left| n^{d/2} \mathbb{P}\Big(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y} \Big) - \det(\Lambda) f(\mathbf{x}_n(\mathbf{y})) \right| = 0.$$

We defer the proof of this to Section 3.7 in the appendix, and instead make a few remarks. Firstly $\mathbf{X}$ is assumed to be lattice and thus non-degenerate. Hence $\mathrm{Cov}(\mathbf{X})$ is invertible, ensuring $N(0, \mathrm{Cov}(\mathbf{X}))$ has a valid density f, which is explicitly given by

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\mathrm{Cov}(\mathbf{X}))}} \exp\left(-\frac{1}{2} \mathbf{x} \cdot \mathrm{Cov}(\mathbf{X})^{-1} \mathbf{x} \right).$$

Secondly, since the $\mathbf{X}_1, \mathbf{X}_2, \ldots$ do not necessarily live in the same probability space we should not technically refer to the sequence $(\|\mathbf{X}_n\|^2)_{n\geq 1}$ as uniformly integrable. However the condition in Eq. (3.10) is still well defined.

We apply Theorem 3.3.16 to $(\Xi^-_{n-m}, \Delta_{n-m})$. Suppose $(D^- - D^+, D^-)$ is non-degenerate and let Λ be its main lattice. By Lemma 3.3.15, Λ is generated by the columns of a matrix A in Hermite normal formal. Since $D^- - D^+$ has span 1, it must be the case that $A_{1,1} = 1$. Thus there is some positive integer q such that

$$A = \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix}.$$

Finally let Σ be the covariance matrix of $(D^- - D^+, D^-)$. With this notation, the following lemma holds:

Lemma 3.3.17. *Suppose $(D^- - D^+, D^-)$ is non-degenerate. For each n, let $\mathbf{c}_n$ be in the support* *of* $(\Delta_{n-m}, \Xi^-_{n-m})$. *Then uniformly for* $(x,y) \in \mathbf{c}_n + \Lambda$,

$$\mathbb{P}_n\left(\Delta_{n-m} = \mathbb{E}\left[\Delta_{n-m}\right] + x,\ \Xi^-_{n-m} = \mathbb{E}\left[\Xi^-_{n-m}\right] + y\right) \\ = \frac{q}{2\pi \det(\Sigma)^{1/2}\, n} \exp\left(\frac{-1}{2n}\begin{pmatrix} x & y \end{pmatrix} \Sigma \begin{pmatrix} x \\ y \end{pmatrix}\right) + o(n^{-1})$$

as $n \to \infty$.

Proof. Let $\mathbf{X} = (D^- - D^+, D^-)$. For each n, let $\mathbf{X}_n$ be distributed as $(D^- - D^+, D^-)$ under $\mathbb{P}_{\theta_n}$. Then

$$\begin{pmatrix} D_1^- - D_1^+ \\ D_1^- \end{pmatrix}, \dots, \begin{pmatrix} D_n^- - D_n^+ \\ D_n^- \end{pmatrix}$$

under $\mathbb{P}_n$ can be seen as n i.i.d. copies of $\mathbf{X}_n$. Since $\theta_n \to 0$, we have that $\mathbf{X}_n \xrightarrow{\text{(d)}} \mathbf{X}$ as $n \to \infty$.

For any $L > 0$,

$$\sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\{\|\mathbf{X}_n\|^2 > L\}\right] = \sup_n \mathbb{E}\left[e^{-\theta_n D^- - \alpha(\theta_n)} \|\mathbf{X}\|^2 \mathbb{1}\{\|\mathbf{X}\|^2 > L\}\right] \\ \leq \left(\sup_n e^{-\alpha(\theta_n)}\right) \mathbb{E}\left[\|\mathbf{X}\|^2 \mathbb{1}\{\|\mathbf{X}\|^2 > L\}\right]$$

since θ_n and D_n^- are non-negative. Since θ_n is convergent,

$$\sup_n e^{-\alpha(\theta_n)} < \infty.$$

Moreover $\mathbb{E}\left[\|\mathbf{X}\|^2 \mathbb{1}\{\|\mathbf{X}\|^2 > L\}\right] \to 0$ as $L \to \infty$ as $\mathbf{X}$ has finite second moment. Thus $(\|\mathbf{X}_n\|^2)_{n \geq 1}$ satisfies the uniform integrability condition in Eq. (3.10).

Finally the exponential tilt does not change the support of the random variables. Thus $\mathbf{X}$ and $\mathbf{X}_n$ share a common main lattice Λ. In addition, $\det(\Lambda) = q$.

Hence the result follows by Theorem 3.3.16. There is a small change in that we are considering a sum of $n - m$ random variables rather than n. However since $m = o(n)$, the same asymptotic result holds. □

Now we show $\mathbb{P}(\mathcal{A}_n, \mathcal{B}_n)$ has the same asymptotic behaviour as $\mathbb{P}(\Delta_n = 0)$. We only prove a lower bound, but this is sufficient for proving Proposition 3.3.6.

Lemma 3.3.18. *Under the assumptions of Proposition 3.3.6,*

$$\mathbb{P}_n\left(\Delta_{n-m}=\sum_{i=1}^{m}(k_i^+-k_i^-),\ |\Xi_{n-m}^- - \mathbb{E}_n[\Xi_{n-m}^-]|\le n^{\frac{1}{2}}\log n\right)\ge \frac{1}{\sqrt{2\pi\sigma^2 n}}(1+o(1)).$$

Proof. For convenience let

$$P_n=\mathbb{P}_n\left(\Delta_{n-m}=\sum_{i=1}^{m}(k_i^+-k_i^-),\ |\Xi_{n-m}^- - \mathbb{E}_n[\Xi_{n-m}^-]|\le n^{\frac{1}{2}}\log n\right).$$

Firstly, suppose $(D^- - D^+, D^-)$ is degenerate. Then since we assume that $D^- - D^+$ is non-constant, it must be the case that either D^- or D^+ is constant. Either way, it becomes the case that

$$\left\{\Delta_{n-m}=\sum_{i=1}^{m}(k_i^+-k_i^-),\ |\Xi_{n-m}^- - \mathbb{E}_n[\Xi_{n-m}^-]|\le n^{\frac{1}{2}}\log n\right\}=\left\{\Delta_{n-m}=\sum_{i=1}^{m}(k_i^+-k_i^-)\right\}.$$

Then applying Theorem 3.3.16, as we did in the proof of Lemma 3.3.17, shows that

$$\mathbb{P}\left(\Delta_{n-m}=\textstyle\sum_{i=1}^{m}(k_i^+-k_i^-)\right)=\frac{1}{\sqrt{2\pi\sigma^2 n}}(1+o(1)).$$

Otherwise assume that $(D^- - D^+, D^-)$ is non-degenerate. Define

$$a_n=\sum_{i=1}^{m}(k_i^+-k_i^-)-\mathbb{E}_n[\Delta_{n-m}].$$

Also let

$$L_n=\left\{y:\left(\textstyle\sum_{i=1}^{m}(k_i^+-k_i^-),y\right)\in \mathbf{c}_n+\Lambda\right\}.$$

L_n has a simpler representation. Fix any $y_0\in L_n$. Then if Λ is generated by the columns of

$$\begin{pmatrix}1 & 0\\ 0 & q\end{pmatrix}$$

we must have $L_n=y_0+q\mathbb{Z}$. Fix an arbitrary $M>0$. Then

$$P_n=\sum_{\substack{y\in L_n\\ |y|\le n^{1/2}\log n}}\mathbb{P}_n\left(\Delta_{n-m}=\mathbb{E}_n[\Delta_{n-m}]+a_n,\ \Xi_{n-m}^-=\mathbb{E}_n[\Xi_{n-m}^-]+y\right)$$

$$\geq \sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \mathbb{P}_n\left(\Delta_{n-m} = \mathbb{E}_n[\Delta_{n-m}] + a_n,\ \Xi^-_{n-m} = \mathbb{E}_n[\Xi^-_{n-m}] + y\right)$$

for all n sufficiently large. By Lemma 3.3.17, using that the error is uniform, we have that

$$P_n \geq \sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \frac{q}{2\pi \det(\Sigma)^{1/2}\, n} \exp\left(\frac{-1}{2n}\begin{pmatrix} a_n \\ y \end{pmatrix} \cdot \Sigma^{-1}\begin{pmatrix} a_n \\ y \end{pmatrix}\right) + o(n^{-1/2}).$$

We wish to factorise the summand. To this end, we make a change of variables. There exists $c \in \mathbb{R}$ such that

$$\mathrm{Cov}(D^- - c(D^- - D^+), D^- - D^+) = 0.$$

Let τ^2 be the variance of $D^- - c(D^- - D^+)$. Then

$$\begin{aligned}\frac{q}{2\pi \det(\Sigma)^{1/2}\, n} &\exp\left(\frac{1}{2n}\begin{pmatrix} a_n \\ y \end{pmatrix} \cdot \Sigma^{-1}\begin{pmatrix} a_n \\ y \end{pmatrix}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2 n}} \exp\left(-\frac{1}{2\sigma^2}\frac{a_n^2}{n}\right) \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right).\end{aligned}$$

We now examine the asymptotic behaviour of a_n. By Lemma 3.3.12,

$$\begin{aligned}\mathbb{E}_n[\Delta_{n-m}] &= (n-m)\mathbb{E}_{\theta_n}[D^- - D^+] \\ &= -(\lambda_- - \lambda_+)m + O(n^{1/3}).\end{aligned}$$

Therefore

$$a_n = s_+(m) - s_-(m) + O(n^{1/3}) = O(n^{1/3}\log n),$$

by the assumption in Eq. (3.1). Thus

$$P_n \geq \frac{1}{\sqrt{2\pi\sigma^2 n}}(1+o(1)) \sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right) + o(n^{-1/2}).$$

Note that

$$\sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right) = \sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \frac{q}{\sqrt{n}}\, g\left(\frac{y-ca_n}{\sqrt{n}}\right)$$

where

$$g(z) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(\frac{-z^2}{2\tau^2}\right).$$

Since $a_n = O(n^{1/3+\epsilon})$, for n sufficiently large

$$\sum_{\substack{y\in L_n \\ |y|\le Mn^{1/2}}} \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right) \ge \sum_{\substack{z\in L_n - ca_n \\ |z|\le \frac{1}{2}Mn^{1/2}}} \frac{q}{\sqrt{n}}\, g\left(\frac{z}{\sqrt{n}}\right) \tag{3.11}$$

$$= \sum_{\substack{z\in \tilde{L}_n \\ |z|\le \frac{1}{2}M}} \frac{q}{\sqrt{n}}\, g(z) \tag{3.12}$$

where

$$\tilde{L}_n = \frac{L_n - ca_n}{\sqrt{n}}.$$

Then $\tilde{L}_n \cap [-\frac{1}{2}M, \frac{1}{2}M]$ is a partition of $[-\frac{1}{2}M, \frac{1}{2}M]$ where adjacent points are distance $q/\sqrt{n}$ apart from each other. Thus Eq. (3.12) is a Riemann sum approximation of an integral. Hence

$$\sum_{\substack{y\in L_n \\ |y|\le Mn^{1/2}}} \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right) \ge (1+o(1)) \int_{-\frac{1}{2}M}^{\frac{1}{2}M} g(z)\,\mathrm{d}z.$$

Thus

$$P_n \ge \frac{1}{\sqrt{2\pi\sigma^2 n}}(1+o(1)) \int_{-\frac{1}{2}M}^{\frac{1}{2}M} g(z)\,\mathrm{d}z.$$

This holds for all $M > 0$, and $\int_{-\infty}^{\infty} g(z)\,\mathrm{d}z = 1$. Therefore,

$$P_n \ge \frac{1}{\sqrt{2\pi\sigma^2 n}}\,(1+o(1))\,,$$

as required. □

3.3.3 Proof of lower bound

Now we are ready to prove Proposition 3.3.6.

Proof of Proposition 3.3.6. By Lemma 3.3.5 and Lemma 3.3.13 we have that

$$\phi(\mathbf{k}_1,\dots,\mathbf{k}_m) \ge \left\{\exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}(s^-(i)-s^-(m)) - \frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right) + o(1)\right\} \frac{\mathbb{P}_n(\mathcal{A}_n\cap\mathcal{B}_n)}{\mathbb{P}(\Delta_n = 0)}$$

where the $o(1)$ term is independent of $\mathbf{k}_1,\dots,\mathbf{k}_m$ satisfying our assumptions. Then by Lemma 3.3.18

and Corollary 3.3.10 we have that

$$\frac{\mathbb{P}_n(\mathcal{A}_n \cap \mathcal{B}_n)}{\mathbb{P}(\Delta_n = 0)} \geq 1 + o(1)$$

where the $o(1)$ term is independent of $\mathbf{k}_1, \ldots, \mathbf{k}_m$ satisfying our assumptions. Thus

$$\phi(\mathbf{k}_1, \ldots, \mathbf{k}_m) \geq \exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}(s^-(i) - s^-(m)) - \frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right) + o(1)$$

as required. ☐

3.3.4 Convergence of the measure change

We are now ready to prove the main result of this section.

Proof of Proposition 3.3.2. The existence of the measure change is covered by Lemma 3.3.5. Define

$$\Gamma(n, m) = \exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}\big(V^-(i) - V^-(m)\big) - \frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right).$$

Then by Donsker's invariance principle,

$$\Big(n^{-1/3}V^-(\lfloor tn^{2/3}\rfloor), n^{-1/3}V^+(\lfloor tn^{2/3}\rfloor)\Big)_{t\geq 0} \xrightarrow{\text{(d)}} \big(\sigma_- W_t^-, \sigma_+ W_t^+\big)_{t\geq 0}$$

in $\mathbb{D}\left([0,\infty), \mathbb{R}^2\right)$, where $(W_t^-, W_t^+)_{t\geq 0}$ are a pair of correlated standard Brownian motions with correlation $\mathrm{Corr}(Z_1^-, Z_1^+)$. We can write

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{\lfloor Tn^{2/3}\rfloor} V^-(i) &= n^{-2/3}\int_0^{\lfloor Tn^{2/3}\rfloor + 1} n^{-1/3}V^-(\lfloor u\rfloor)\,\mathrm{d}u \\ &= \int_0^{n^{-2/3}(\lfloor Tn^{2/3}+1\rfloor)} n^{-1/3}V^-(\lfloor sn^{2/3}\rfloor)\,\mathrm{d}s.\end{aligned}$$

Thus, by the continuous mapping theorem,

$$\frac{1}{n}\sum_{i=0}^{\lfloor Tn^{2/3}\rfloor}\big(V^-(i) - V^-(m)\big) \xrightarrow{\text{(d)}} \int_0^T \sigma_-\left(W_s^- - W_T^-\right)\mathrm{d}s = -\int_0^T \sigma_- s\,\mathrm{d}W_s^-.$$

Hence,

$$\left(\Gamma(n,\lfloor Tn^{2/3}\rfloor),\left(n^{-1/3}V^-(\lfloor tn^{2/3}\rfloor),n^{-1/3}V^+(\lfloor tn^{2/3}\rfloor)\right)_{t\in[0,T]}\right) \xrightarrow{\text{(d)}} (\Phi(T),(\sigma_-W_t^-,\sigma_+W_t^+)_{t\in[0,T]}.)$$

in $\mathbb{R}\times\mathbb{D}([0,T],\mathbb{R})$, as $n\to\infty$. Recall the event

$$\mathcal{E}_m=\left\{\max_{i=1,\dots,m}|V^-(i)|\le m^{1/2}\log m \quad\text{and}\quad \max_{i=1,\dots,m}|V^+(i)|\le m^{1/2}\log m\right\}$$

By Proposition 3.3.6, it is the case that

$$\Phi(n,m)\ge(\Gamma(n,m)+o(1))\mathbb{1}_{\mathcal{E}_m}.$$

The processes $(V^\pm(n))_{n\ge0}$ are discrete martingales. Therefore, by Doob's maximal inequality,

$$\mathbb{P}\left(\max_{i=1,\dots,m}|V^\pm(i)|>m^{1/2}\log(m)\right)\le\frac{\mathbb{E}[(V^\pm(m))^2]}{m(\log m)^2}=\frac{\sigma_\pm^2}{(\log m)^2}\to0$$

as $m\to\infty$. Thus $\mathbb{P}(\mathcal{E}_m)\to1$ as $m\to\infty$. Hence, we still have that

$$\left((\Gamma(n,\lfloor Tn^{2/3}\rfloor)+o(1))\mathbb{1}_{\mathcal{E}_{\lfloor Tn^{2/3}\rfloor}},\left(n^{-1/3}V^-(\lfloor tn^{2/3}\rfloor),n^{-1/3}V^+(\lfloor tn^{2/3}\rfloor)\right)_{t\in[0,T]}\right) \xrightarrow{\text{(d)}} \left(\Phi(T),(\sigma_-W_t^-,\sigma_+W_t^+)_{t\in[0,T]}\right).$$

We have $\mathbb{E}[\Phi(T)]=1$ by a standard stochastic calculus calculation. Therefore, by Lemma 3.3.7, we get the desired result that

$$\left(\Phi(n,\lfloor Tn^{2/3}\rfloor),\left(n^{-1/3}V^-(\lfloor tn^{2/3}\rfloor),n^{-1/3}V^+(\lfloor tn^{2/3}\rfloor)\right)_{t\in[0,T]}\right) \xrightarrow{\text{(d)}} \left(\Phi(T),(\sigma_-W_t^-,\sigma_+W_t^+)_{t\in[0,T]}\right),$$

and that $(\Phi(n,\lfloor Tn^{2/3}\rfloor))_{n\ge1}$ is a uniformly integrable sequence. □

3.4 Convergence of the out-forest

Fix $T>0$. In this section we will show that the Łukasiewicz path and height process corresponding to the out-forest converge under rescaling up to time $\lfloor Tn^{2/3}\rfloor$. Note that the out-forest will contain at least n vertices, so for n large enough, $\lfloor Tn^{2/3} \rfloor \leq n$ and the encoding processes are well-defined up to time $\lfloor Tn^{2/3} \rfloor$.

We will show that the convergence under rescaling of the Łukasiewicz path and height process $(\hat{S}_n^+(k), \hat{H}_n(k), k \leq \lfloor Tn^{2/3} \rfloor)$ occurs jointly with convergence in distribution under rescaling of $(\hat{S}_n^-(k), \hat{P}_n(k), k \leq \lfloor Tn^{2/3} \rfloor)$, for $\hat{S}_n^-(k)$ the number of unpaired in-half-edges of vertices that have been discovered at time k, and $\hat{P}_n(k)$ the number of dummy leaves added in the first k time-steps.

We let $(B_t)_{t \geq 0}$ be a Brownian motion, and define

$$(\hat{B}_t, t \geq 0) := \left(B_t - \frac{\sigma_{-+} + \nu_-}{2\sigma_+ \mu} t^2, t \geq 0 \right).$$

We define the reflected process

$$(\hat{R}_t, t \geq 0) = \left(\hat{B}_t - \inf \left\{ \hat{B}_s : s \leq t \right\}, t \geq 0 \right).$$

The main result of this section is as follows.

Proposition 3.4.1. *It holds that*

$$\left(n^{-1/3} \hat{S}_n^+ \left(\lfloor n^{2/3} t \rfloor \right), n^{-1/3} \hat{H}_n \left(\lfloor n^{2/3} t \rfloor \right), t \leq T \right) \xrightarrow{(d)} \left(\sigma_+ \hat{B}_t, \frac{2}{\sigma_+} \hat{R}_t, t \leq T \right)$$

in $\mathbb{D}([0,T], \mathbb{R})^2$, *and*

$$\left(n^{-2/3} \hat{S}_n^- \left(\lfloor n^{2/3} t \rfloor \right), n^{-1/3} \hat{P}_n \left(\lfloor n^{2/3} t \rfloor \right), t \leq T \right) \xrightarrow{(p)} \left(\nu_- t, \frac{\nu_-}{2\mu} t^2, t \leq T \right)$$

in $\mathbb{D}([0,T], \mathbb{R})^2$ *as* $n \to \infty$.

We prove Proposition 3.4.1 by studying two other forests that are related to the out-forest via a change of measure.

The proof is structured as follows.

1. Recall that $(\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,n})$ are the degree pairs of the vertices in order of discovery. Also recall $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ in an i.i.d. sequence of $\mathbb{N} \times \mathbb{N}$-valued random variables, $\mathbf{Z}_i := (Z_i^-, Z_i^+)$, such that
$$\mathbb{P}(Z_i^- = k^-, Z_i^+ = k^+) = \frac{k^- \mathbb{P}(D^- = k^-, D^+ = k^+)}{\mu}.$$

In Section 3.3, we showed that the law of $(\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,m})$ conditional on $\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$ and $m \le R_n$ is absolutely continuous with respect to that of $(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$, and we showed the convergence under rescaling of the Radon-Nikodym derivative ϕ_m^n for $m = \lfloor Tn^{2/3} \rfloor$.

2. Point 1 motivates us to study a Bienaymé forest with offspring distributed as Z_1^+. The convergence of the Łukasiewicz path of this forest under rescaling follows from Donsker's theorem.

3. In Subsection 3.4.2, we modify the Bienaymé forest in order to include dummy leaves. We add extra randomness, approximating the procedure described in Proposition 3.2.4, in such a way that at some time-steps, a dummy leaf is added. We call the resulting forest *the forest with dummy leaves*. We respect the order of the degrees in the Bienaymé forest, in the sense that for any k, the kth true vertex in the forest with dummy leaves has the same number of children as the kth vertex in the Bienaymé forest. The law of the forest with dummy leaves depends on n, because the probability of finding a dummy leaf depends on n. We then show that the Łukasiewicz path and height process of the forest with dummy leaves converge under rescaling, jointly with the convergence of the Łukasiewicz path and height process of the Bienaymé forest under rescaling up to time $\lfloor Tn^{2/3} \rfloor$.

4. We show convergence under rescaling of the out-forest up to time $\lfloor Tn^{2/3} \rfloor$ by applying the measure change to the forest with dummy leaves and showing that the resulting forest is a good approximation of the out-forest.

3.4.1 Convergence before adding the dummy leaves

We define the two processes

$$\hat{Y}^{\pm}(k) = \sum_{i=1}^{k} (\hat{D}_{n,i}^{\pm} - 1),$$

for $1 \le k \le n$, which encode the degrees in order of discovery.

We will study these processes via the measure change that we defined in Section 3.3. Let

$$Y^{\pm}(k) = \sum_{i=1}^{k} (Z_i^{\pm} - 1)$$

be the corresponding walks for $(\mathbf{Z}_i)_{i=1}^{\infty}$. Then, in the critical case, these are related to the centered random walks $V^{\pm}$ by

$$Y^+(k) = V^+(k) \quad \text{and} \quad Y^-(k) = V^-(k) - (\lambda_- - 1)k = V^-(k) - \nu_- k.$$

Therefore, we obtain the following corollary of Proposition 3.3.2.

Corollary 3.4.2. *Suppose we are in the setting of Proposition 3.3.2 and that the criticality condition holds. Then for all* $T > 0$*,*

$$\left(\Phi(n, \lfloor n^{2/3}T \rfloor), \left(n^{-1/3}V^-\left(\lfloor n^{2/3}t \rfloor\right), n^{-1/3}V^+\left(\lfloor n^{2/3}t \rfloor\right)\right)_{t\in[0,T]}\right) \\ \xrightarrow{\text{(d)}} \left(\Phi(T), (\sigma_- W_t^+, \sigma_+ W_t^+)_{t\in[0,T]}\right)$$

in $\mathbb{R} \times \mathbb{D}([0,T], \mathbb{R}^2)$ *as* $n \to \infty$ *and* $\left(\Phi(n, \lfloor n^{2/3}T \rfloor)\right)_{n\geq 1}$ *is uniformly integrable.*

We will first show that the law of $(\hat{B}_t, t \geq 0)$ is locally absolutely continuous to a Brownian motion and we characterise the Radon–Nikodym derivative. This is the content of the next proposition.

Proposition 3.4.3. *It holds that for* F *a continuous bounded function, and for* $(B_t)_{t\geq 0}$ *a standard Brownian motion,*

$$\begin{aligned} &\mathbb{E}\left[F(\sigma_+ \hat{B}_t, 0 \leq t \leq T)\right] \\ &= \mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_+\mu}\int_0^T s dB_s - \frac{\sigma_{-+}^2 T^3}{6\sigma_+^2\mu^2}\right) F(\sigma_+ B_t, 0 \leq t \leq T)\right]. \end{aligned}$$

Proof. Firstly, we have that for any $t \in [0,T]$ and $\theta > 0$,

$$\begin{aligned} \mathbb{E}\left[\exp\left(-\theta\left(\sigma_+ B_t - \frac{\sigma_{-+}}{2\mu}t^2\right)\right)\right] &= \exp\left(\frac{\sigma_+^2 t}{2}\theta^2 + \frac{\sigma_{-+}t^2}{2\mu}\theta\right) \\ &= \exp\left(-\frac{\sigma_{-+}^2}{2\sigma_+^2\mu^2}\int_0^t\left(s + \frac{\sigma_+^2\theta\mu}{\sigma_{-+}}\right)^2 ds - \frac{\sigma_{-+}^2 t^3}{6\sigma_+^2\mu^2}\right) \\ &= \mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_+\mu}\int_0^t\left(s + \frac{\sigma_+^2\theta\mu}{\sigma_{-+}}\right) dB_s - \frac{\sigma_{-+}^2 t^3}{6\sigma_+^2\mu^2}\right)\right] \\ &= \mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_+\mu}\int_0^t s dB_s - \frac{\sigma_{-+}^2 t^3}{6\sigma_+^2\mu^2}\right)\exp\left(-\theta\sigma_+ B_t\right)\right] \end{aligned}$$

Then, more generally, for $m > 0$, $0 = t_0 \le t_1 \le \dots \le t_m = t$, and $\theta_1, \dots, \theta_m \in \mathbb{R}_+$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\sum_{i=1}^{m}\theta_i\left(\sigma_+(B_t - B_{t_i}) - \frac{\sigma_{-+}}{2\mu}(t_i^2 - t_{i-1}^2)\right)\right)\right] \\
&= \prod_{i=1}^{m}\exp\left(\frac{\sigma_+^2(t_i - t_{i-1})}{2}\theta_i^2 + \frac{\sigma_{-+}(t_i^2 - t_{i-1}^2)}{2\mu}\theta_i\right) \\
&= \prod_{i=1}^{m}\exp\left(-\frac{\sigma_{-+}^2}{2\sigma_+^2\mu^2}\int_{t_{i-1}}^{t_i}\left(s + \frac{\sigma_+^2\theta_i\mu}{\sigma_{-+}}\right)^2 ds - \frac{\sigma_{-+}^2(t_i^3 - t_{i-1}^3)}{6\sigma_+^2\mu^2}\right) \\
&= \prod_{i=1}^{m}\mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_+\mu}\int_{t_{i-1}}^{t_i} s dB_s - \frac{\sigma_{-+}^2(t_i^3 - t_{i-1}^3)}{6\sigma_+^2\mu^2} - \theta_i\sigma_+(B_{t_i} - B_{t_{i-1}})\right)\right] \\
&= \mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_+\mu}\int_0^t s dB_s - \frac{\sigma_{-+}^2 t^3}{6\sigma_+^2\mu^2}\right)\exp\left(-\sum_{i=1}^{m}\theta_i(\sigma_+B_{t_i} - \sigma_+B_{t_{i-1}})\right)\right],
\end{aligned}$$

which proves the result. ☐

Proposition 3.4.4. *We have that*

$$\left(n^{-2/3}\hat{Y}^-\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}\hat{Y}^+\left(\lfloor n^{2/3}t\rfloor\right), 0 \le t \le T\right) \xrightarrow{\text{(d)}} \left(\nu_- t, \sigma_+\hat{B}_t, 0 \le t \le T\right)$$

in the Skorokhod topology as $n \to \infty$.

Proof. We recall from the statement of Corollary 3.4.2 that (W^-, W^+) is a pair of correlated standard Brownian motions with correlation $\mathrm{Corr}(Z_1^-, Z_1^+)$. Let $(B_t^1, t \ge 0)$ and $(B_t^2, t \ge 0)$ be two independent Brownian motions, so that we may define

$$(\sigma_- W_t^-, \sigma_+ W_t^+, t \ge 0) = \left(\frac{\sigma_{-+}}{\sigma_+}B_t^1 + \left(\sigma_-^2 - \frac{\sigma_{-+}^2}{\sigma_+^2}\right)^{1/2} B_t^2, \sigma_+ B_t^1, t \ge 0\right).$$

Then, Corollary 3.4.2 implies that for F a continuous, bounded test function,

$$\begin{aligned}
&\mathbb{E}\left[F\left(n^{-1/3}\hat{Y}^+\left(\lfloor n^{2/3}t\rfloor\right), 0 \le t \le T\right)\right] \\
&= \mathbb{E}\left[F\left(n^{-1/3}\hat{Y}^+\left(\lfloor n^{2/3}t\rfloor\right), 0 \le t \le T\right)\mathbb{1}_{\lfloor Tn^{2/3}\rfloor \le R_n}\right] + o(1) \\
&= \mathbb{E}\left[\Phi(n, \lfloor n^{2/3}T\rfloor)F\left(n^{-1/3}V^+\left(\lfloor n^{2/3}t\rfloor\right), 0 \le t \le T\right)\right] + o(1).
\end{aligned}$$

By the proof of Proposition 3.3.2, we see that for

$$\Gamma(n, m) = \exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}(V^-(i) - V^-(m)) - \frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right),$$

we have that

$$\mathbb{E}\left[\left|\Phi(n,\lfloor n^{2/3}T\rfloor)-\Gamma(n,\lfloor n^{2/3}T\rfloor)\right|\right]\to 0$$

as $n\to\infty$, so it sufficient to show that

$$\mathbb{E}\left[\Gamma(n,\lfloor n^{2/3}T\rfloor)F\left(n^{-1/3}V^{+}\left(\lfloor n^{2/3}t\rfloor\right),0\le t\le T\right)\right]\to\mathbb{E}\left[F\left(\sigma_{+}\hat{B}_t,0\le t\le T\right)\right].$$

Write $V^{+}_{(n)}(t)=n^{-1/3}V^{+}\left(\lfloor n^{2/3}t\rfloor\right)$ and $V^{-}_{(n)}(t)=n^{-1/3}V^{-}\left(\lfloor n^{2/3}t\rfloor\right)$. Then we observe that

$$\Gamma(n,\lfloor n^{2/3}T\rfloor)=\exp\left(\frac{1}{\mu}\int_0^T\left(V^{-}_{(n)}(t)-V^{-}_{(n)}(T)\right)dt-\frac{\sigma_{-}}{6\mu^2}\frac{\lfloor Tn^{2/3}\rfloor^3}{n^2}\right).$$

For a path $x\in\mathbb{D}([0,T],\mathbb{R})$, let

$$\Theta(x,T)=\exp\left(\frac{1}{\mu}\int_0^T(x(t)-x(T))\,dt-\frac{\sigma_{-}}{6\mu^2}T^3\right)$$

so that Θ is a continuous functional of its first argument and

$$\mathbb{E}\left[\left|\Gamma(n,\lfloor n^{2/3}T\rfloor)-\Theta(V^{-}_{(n)},T)\right|\right]\to 0$$

as $n\to\infty$. This implies that it suffices to show that

$$\mathbb{E}\left[\Theta(V^{-}_{(n)},T)F\left(V^{+}_{(n)}(t),0\le t\le T\right)\right]\to\mathbb{E}\left[F\left(\sigma_{+}\hat{B}_t,0\le t\le T\right)\right].$$

But, by the continuity of Θ and Corollary 3.4.2, we get that

$$\begin{aligned}\mathbb{E}\left[\Theta(V^{-}_{(n)},T)F\left(V^{+}_{(n)}(t),0\le t\le T\right)\right]&\to\mathbb{E}\left[\Theta(\sigma_{-}W^{-}_t,T)F\left(\sigma_{+}W^{+}_t,0\le t\le T\right)\right]\\&=\mathbb{E}\left[\exp\left(-\frac{1}{\mu}\int_0^T sd\left(\frac{\sigma_{-+}}{\sigma_{+}}B^1_s+\left(\sigma_{-}^2-\frac{\sigma_{-+}^2}{\sigma_{+}^2}\right)^{1/2}B^2_s\right)-\frac{T^3\sigma_{-}^2}{6\mu^2}\right)F\left(\sigma_{+}B^1_t,0\le t\le T\right)\right]\\&=\mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_{+}\mu}\int_0^T sdB^1_s-\frac{\sigma_{-+}^2T^3}{6\sigma_{+}^2\mu^2}\right)F(\sigma_{+}B^1_t,0\le t\le T)\right].\end{aligned}$$

Then, the fact that $(Y^{-}(k),k\ge 1)$ is a random walk with steps of mean ν_{-} implies that

$$\left(n^{-2/3}Y^{-}\left(\lfloor n^{2/3}t\rfloor\right),t\ge 0\right)\xrightarrow{(\mathrm{p})}(\nu_{-}t,t\ge 0),$$

and then, by repeating the argument above, noting that the change of measure does not affect the deterministic process $(\nu_- t, t \geq 0)$, also

$$\left(n^{-2/3}\hat{Y}^-\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{p})} (\nu_- t, t \geq 0),$$

which proves the statement. □

3.4.2 Adding dummy leaves to a Bienaymé forest

We would like to add dummy leaves to the forest encoded by $(Y^+(l), 1 \leq l \leq k)$. However, in the absence of a true stack of in-edges, we need to approximate the probability of adding a dummy leaf. We do this by approximating the stack size by its mean μn. We use this idea to define the forest with dummy leaves and its Łukasiewicz path $(S_n^+(k), k \geq 1)$ as a function of $(Y^-(k), Y^+(k), k \geq 1)$ and some extra randomness to decide at which time-steps we add a dummy leaf.

1. Set $P_n(1) = 0$, $S_n^+(1) = Z_1^+ - 1$, $S_n^-(1) = Z_1^-$.
2. Suppose we are given $(P_n(l), S_n^+(l), S_n^-(l), 1 \leq l \leq k)$. Define $I^+(k) = \min\{S_n^+(l), l \leq k\}$. Then, with probability

$$p_{k+1} := \frac{S_n^-(k)}{\mu n - k - I^+(k) + 1}\mathbb{1}_{\{I^+(k)=I^+(k-1)\}},$$

   independent from everything else, set $P_n(k+1) = P_n(k) + 1$. Otherwise, set $P_n(k+1) = P_n(k)$.
3. Set

$$S_n^+(k+1) = Y^+(k+1-P_n(k+1)) - P_n(k+1),$$

   and

$$S_n^-(k+1) = Y^-(k+1-P_n(k+1)) - P_n(k+1) - I^+(k) + 1.$$

Let the forest with dummy leaves be the forest with Łukasiewicz path $(S_n^+(k), k \geq 1)$ in which the kth vertex is a dummy leaf if and only if $P^n(k) - P^n(k-1) = 1$.

3.4.2.1 Convergence of the Łukasiewicz path

To show the convergence of the Łukasiewicz path corresponding to the forest with dummy leaves, we will first examine the limit of $(P_n(k), k \geq 1)$ under rescaling. We will first prove tightness, after which we will show convergence.

Lemma 3.4.5. *We have that,*

$$\left(n^{-1/3}P_n\left(\lfloor n^{2/3}t\rfloor\right)\right)_{n\geq 1}$$

is tight for all $t > 0$*.*

Proof. Set $m = \lfloor n^{2/3}t \rfloor$ and fix $\epsilon > 0$. It is trivial that for any $k \leq m$,

$$S^{-}(k) \leq \sum_{i=1}^{k} Z_i^{-} = Y^{-}(k) + k.$$

Moreover, $\mu n - k - I^{+}(l) + 1 > \mu n - k$. Therefore,

$$p_{k+1} \leq \frac{Y^{-}(k) + k}{\mu n - k}.$$

This upper bound is increasing in k. Consequently, conditional on $(Y^{+}(j), Y^{-}(j), j \geq 1)$, we have that $P_n(m)$ is stochastically dominated by a binomial random variable with parameters m and

$$\frac{Y^{-}(m) + m}{\mu n - m} \wedge 1.$$

Since $(Y^{-}(k) + k, k \geq 1)$ is a random walk with steps of finite mean, $\left(n^{-2/3}(Y^{-}(m) + m)\right)_{n\geq 1}$ is tight. Therefore,

$$\left(n^{1/3}\frac{Y^{-}(m) + m}{\mu n - m}\right)_{n\geq 1}$$

is tight, which implies that a binomial random variable with parameters m and

$$\frac{Y^{-}(m) + m}{\mu n - m} \wedge 1$$

multiplied by $n^{-1/3}$ is tight. The statement follows. □

Lemma 3.4.6. *We have*

$$\left(n^{-1/3}P_n(\lfloor n^{2/3}t\rfloor), t \geq 0\right) \xrightarrow{\text{(p)}} \left(\frac{\nu_-}{2\mu}t^2, t \geq 0\right)$$

in $D(\mathbb{R}_+, \mathbb{R})$ *as* $n \to \infty$*.*

Proof. Recall that

$$p_{k+1} = \frac{S_n^-(k)}{\mu n - k - I^+(k) + 1} \mathbb{1}_{\{I^+(k)=I^+(k-1)\}}.$$

Define $M^+(k) = \min\{Y^+(l) : l \leq k\}$ so that $0 \geq I^+(k) \geq M^+(k) - P_n(k)$. Then, by Lemma 3.4.5, the convergence under rescaling of Y^+ shown in Corollary 3.4.2, and the continuous mapping theorem, $\left(n^{-1/3} I^+(\lfloor n^{2/3} t \rfloor)\right)_{n \geq 1}$ is tight for all $t \geq 0$. We will now argue that the indicator, which ensures that the roots are never dummy leaves, does not have an effect on $(P_n(k), k \leq m)$ on the scale of interest. Let $m = \lfloor n^{2/3} t \rfloor$. Define

$$\begin{aligned} E^p(m) &:= \sum_{k=0}^{m-1} \frac{S_n^-(k)}{\mu n - k - I^+(k) + 1} \mathbb{1}_{\{I^+(k) \neq I^+(k-1)\}} \\ &\leq -I^+(m) \frac{Y^-(m) + m}{\mu n - m}, \end{aligned}$$

so since $I^+(m)$ is of order $n^{1/3}$ and $\frac{Y^-(m)+m}{\mu n - m}$ is of order $n^{-1/3}$, $(E^p(m))_{n \geq 1}$ is tight. This means that if we allow the roots to be dummy leaves, with high probability, we would only sample $O(1)$ roots that are dummy leaves up to time $O(n^{2/3})$. This does not affect $(P_n(k), k \leq m)$ on the scale of interest.

Then, the convergence under rescaling of Y^- and Y^+ shown in Corollary 3.4.2, the tightness of $\left(n^{-1/3} I^+(\lfloor n^{2/3} t \rfloor)\right)_{n \geq 1}$ and Lemma 3.4.5 imply that

$$\begin{aligned} &\left(n^{1/3} \frac{S_n^-\left(\lfloor n^{2/3} t \rfloor\right)}{\mu n - \lfloor n^{2/3} t \rfloor - I^{p,+}\left(\lfloor n^{2/3} t \rfloor\right) + 1}, t \geq 0 \right) \\ &= \left(n^{1/3} \frac{Y^-\left(\lfloor n^{2/3} t \rfloor - P_n\left(\lfloor n^{2/3} t \rfloor\right)\right) - P_n\left(\lfloor n^{2/3} t \rfloor\right) - I^+\left(\lfloor n^{2/3} t \rfloor\right) + 1}{\mu n - \lfloor n^{2/3} t \rfloor - I^+\left(\lfloor n^{2/3} t \rfloor\right) + 1}, t \geq 0 \right) \\ &\xrightarrow{(p)} \left(\frac{\nu_-}{\mu} t, t \geq 0 \right) \end{aligned} \tag{3.13}$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$. Then, by the continuous mapping theorem and the tightness of $(E^p(m))_{n \geq 1}$,

$$\left(n^{-1/3} \sum_{k=0}^{\lfloor n^{2/3} t \rfloor} p_k, t \geq 0 \right) \xrightarrow{(p)} \left(\frac{\nu_-}{2\mu} t^2, t \geq 0 \right)$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$.

Let $\mathcal{G} = (\mathcal{G}_k, k \geq 1)$ denote the filtration such that $\mathcal{G}_k$ contains the information on the shape of the forest until time k, including which of the first k vertices are dummy vertices. Then,

$$M_n(k) := \sum_{i=1}^{k} (\mathbb{1}_{\{P_n(i)-P_n(i-1)=1\}} - p_i)$$

is a $\mathcal{G}$-martingale. We claim that $(n^{-1/3}M_n(\lfloor n^{2/3}t \rfloor), t \geq 0)$ converges to 0 in probability in $D(\mathbb{R}_+, \mathbb{R})$. Indeed, for any $t \geq 0$,

$$\begin{aligned} \mathbb{E}[n^{-2/3}M_n(\lfloor n^{2/3}t \rfloor)^2] &= n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3}t \rfloor} \mathbb{E}[\mathbb{E}[(\mathbb{1}_{\{P_n(i)-P_n(i-1)=1\}} - p_i)^2 | \mathcal{G}_{i-1}]] \\ &= n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3}t \rfloor} \mathbb{E}[p_i - p_i^2] \to 0. \end{aligned}$$

Hence, since for all $t \geq 0$,

$$\begin{aligned} n^{-1/3}P_n(\lfloor n^{2/3}t \rfloor) &= n^{-1/3} \sum_{i=1}^{\lfloor n^{2/3}t \rfloor} \mathbb{1}_{\{P_n(i)-P_n(i-1)=1\}} \\ &= n^{-1/3} \sum_{i=0}^{\lfloor n^{2/3}t \rfloor} p_k + n^{-1/3}M_n\left(\lfloor n^{2/3}t \rfloor\right), \end{aligned}$$

we have

$$\left(n^{-1/3}P_n(\lfloor n^{2/3}t \rfloor), t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\frac{\nu_-}{2\mu}t^2, t \geq 0\right),$$

which proves the statement.

□

The convergence of P_n under rescaling implies the convergence of S_n^+ and S_n^- under rescaling, which is the content of the following lemma. Let $(B_t, t \geq 0)$ be a Brownian motion, and define

$$(B_t^{\mathrm{d}}, t \geq 0) = \left(B_t - \frac{\nu_-}{2\mu\sigma_+}t^2, t \geq 0\right).$$

Lemma 3.4.7. *We have*

$$\left(n^{-1/3}Y^+\left(\lfloor n^{2/3}t \rfloor\right), n^{-1/3}S_n^+\left(\lfloor n^{2/3}t \rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\sigma_+B_t, \sigma_+B_t^{\mathrm{d}}, t \geq 0\right)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^2$ *and*

$$\left(n^{-2/3}S_n^-\left(\lfloor n^{2/3}t \rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{p})} (\nu_- t, t \geq 0)$$

in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ *as* $n\to\infty$.

Proof. This follows from the convergence under rescaling of Y^+ and Y^- shown in Corollary 3.4.2 and Lemma 3.4.6, and the expressions

$$S_n^+(k+1) = Y^+(k+1-P_n(k+1)) - P_n(k+1),$$

and

$$S_n^-(k+1) = Y^-(k+1-P_n(k+1)) - P_n(k+1) - I^+(k) + 1.$$

□

3.4.2.2 Convergence of the height process

In this subsection, we will extend Lemma 3.4.7. We will show that, under rescaling, the height process of the forest with dummy leaves converges jointly with the other encoding processes of the forest with dummy leaves. Let $(H_n^+(k), k\geq 1)$ be the height process corresponding to the forest with dummy leaves. Set

$$(R_t^{\mathrm{d}}, t\geq 0) = \left(B_t^{\mathrm{d}} - \inf\left\{B_s^{\mathrm{d}} : s\leq t\right\}, t\geq 0\right).$$

Proposition 3.4.8. *We have that*

$$\begin{aligned}&\left(n^{-1/3}Y^+\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}S_n^+\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}H_n^+\left(\lfloor n^{2/3}t\rfloor\right), t\geq 0\right)\\ &\quad\xrightarrow{(\mathrm{d})}\left(\sigma_+B_t, \sigma_+B_t^{\mathrm{d}}, \frac{2}{\sigma_+}R_t^{\mathrm{d}}, t\geq 0\right)\end{aligned}$$

in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})^3$, *and*

$$\left(n^{-2/3}S_n^-\left(\lfloor n^{2/3}t\rfloor\right), t\geq 0\right)\xrightarrow{(\mathrm{p})}(\nu_- t, t\geq 0)$$

in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ *as* $n\to\infty$.

The difficulty in proving this proposition is the fact that the forest with dummy leaves is not a Bienaymé forest, because the probability of sampling a dummy leaf changes as the exploration is performed. The theory of convergence of height processes under rescaling is well-developed for Bienaymé processes (see e.g. Duquesne and Le Gall [50]), but this is not the case for more

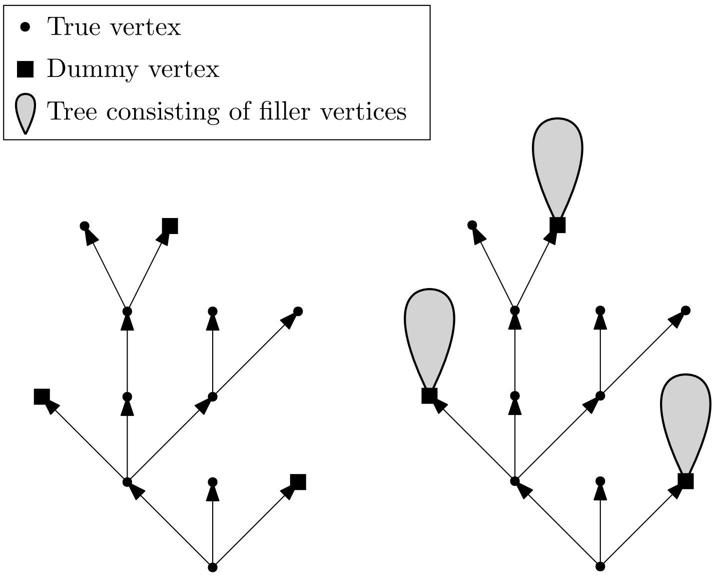


Fig. 3.9: Given a component of the forest with dummy vertices (left), we modify it by sampling independent Bienaymé trees with offspring distributed as Z^+ consisting of filler vertices and identifying each dummy leaf with a root of such a tree. The resulting tree (right) is a Bienaymé tree, and the resulting forest is a Bienaymé forest.

general processes. We will adapt a technique that Broutin, Duquesne and Wang developed in [29] to show the convergence of the height process of an inhomogeneous random graph under rescaling. The key idea is that the forest with dummy leaves itself is not a Bienaymé forest, but we can embed it in a Bienaymé forest that does not depend on n. We call the extra vertices *filler vertices* and call the resulting forest *the forest with dummy and filler vertices.* We then show convergence under rescaling of the height process corresponding to the forest with dummy and filler vertices, and use this to obtain height process convergence for the forest with dummy leaves.

We start by defining the forest with dummy and filler vertices. Informally, we obtain it by modifying the forest with dummy leaves in such a way that a sub-tree consisting of the descendants of a dummy vertex has the same law as a sub-tree consisting of the descendants of a true vertex. We do this by sampling extra Bienaymé trees with offspring distributed as Z^+, whose vertices are all filler vertices, and then identifying their roots with the dummy leaves. The resulting forest is a Bienaymé forest containing true, dummy and filler vertices, in which the forest with true vertices and dummy leaves is embedded. This is illustrated in Figure 3.9.

The formal procedure is as follows. Suppose we are given $(Y^+(k), S_n^+(k), P_n(k), k \geq 1)$, which encodes the forest with dummy leaves.

1. Let $(Y^{\mathrm{f}}(k), k \geq 1)$ be an independent copy of $(Y^+(k), k \geq 1)$, which will encode the

pendant subtrees that consist of filler vertices.

2. Define $\theta_n(k) = k - P_n(k-1) + \min\{j : Y^{\mathrm{f}}(j) = -P_n(k-1)\}$.

3. Set $\Lambda_n(k) = \max\{j : \theta_n(j) \leq k\} - P_n(\max\{j : \theta_n(j) \leq k\})$.

4. We now define

$$(Y^{\mathrm{df}}(k), k \geq 1) = (Y^{+}(\Lambda_n(k)) + Y^{\mathrm{f}}(k - \Lambda_n(k)), k \geq 1) \tag{3.14}$$

   and we let *the forest with dummy and filler vertices* be the forest with Łukasiewicz path $(Y^{\mathrm{df}}(k), k \geq 1)$, in which $P_n(\max\{j : \theta_n(j) \leq k\})$ of the first k vertices are dummy vertices, $\Lambda_n(k)$ of the first k vertices are true vertices, and the rest are filler vertices. We let $(H^{\mathrm{df}}(k), k \geq 1)$ be the height process corresponding to the forest with dummy and filler vertices.

By removing the filler vertices from the forest with dummy and filler vertices, we obtain the original forest with dummy leaves. We make the following observations.

1. We claim that $\theta_n(k)$ is equal to index in depth first order of the kth true or dummy vertex in the forest with dummy and filler vertices. Indeed, note that $\min\{j : Y^{\mathrm{f}}(j) = -P_n(k-1)\}$ is equal to the number of vertices in the first $P_n(k-1)$ trees in the forest encoded by Y^{f}, so that

$$\min\{j : Y^{\mathrm{f}}(j) = -P_n(k-1)\} - P_n(k-1)$$

   is equal to the number of filler vertices in depth-first order until the kth true or dummy vertex.

2. Note that $\Lambda_n(k)$ is the number of true vertices amongst the first k vertices. This follows from the fact that $\max\{j : \theta_n(j) \leq k\}$ is the number of true or dummy vertices amongst the first k vertices.

3. By the previous remark, $(\Lambda_n(k), k \geq 1)$ only takes steps of size 0 or 1. Both $(Y^{+}(k), k \geq 1)$ and $(Y^{\mathrm{f}}(k), k \geq 1)$ are random walks with steps distributed as $Z^{+}-1$, so, by construction, $(Y^{\mathrm{df}}(k), k \geq 1)$ is a random walk with steps distributed as $Z^{+} - 1$, so the forest with dummy and filler vertices is a Bienaymé forest with offspring distributed as Z^{+}.

4. By construction, $(H^{\mathrm{df}}(\theta_n(k)), k \geq 1)$ is the height process corresponding to the forest with dummy vertices. Moreover,

$$(S_n^+(k), k \geq 1) = (Y^{\mathrm{df}}(\theta_n(k)) - E(\theta_n(k)), k \geq 1), \tag{3.15}$$

where $E(k)$ counts the number of children of the kth vertex in the forest with dummy and filler vertices that are filler vertices.

In order to prove Proposition 3.4.8, considering the construction above and Lemma 3.4.7, it is sufficient to prove the following lemma.

Lemma 3.4.9. *There exists a process $(D_t, t \geq 0)$ such that*

$$\left(n^{-1/3}\left[Y^{\mathrm{df}}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right) - E\left(\lfloor n^{2/3}t\rfloor\right)\right], n^{-1/3}H^{\mathrm{df}}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t \geq 0\right)$$
$$\xrightarrow{(\mathrm{d})} \left(\sigma_+ D_t, \frac{2}{\sigma_+}\left(D_t - \inf\left\{D_s, s \leq t\right\}\right), t \geq 0\right)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^2$ as $n \to \infty$ and $\left(\frac{2}{\sigma_+}\left(D_t - \inf\left\{D_s, s \leq t\right\}\right), t \geq 0\right)$ is the height process corresponding to $(\sigma_+ D_t, t \geq 0)$.

The next lemma show that the pathwise construction of $(Y^{\mathrm{df}}(k), H^{\mathrm{df}}(k), k \geq 1)$ converges to its continuous counterpart.

Let $(B_t, t \geq 0)$ and $(B_t^{\mathrm{f}}, t \geq 0)$ be two independent Brownian motions and let

$$\theta(t) := t + \inf\left\{s \geq 0 : \sigma_+ B_s^{\mathrm{f}} < -\frac{\nu_-}{2\mu}t^2\right\},$$

and $\Lambda(t) = \inf\{s \geq 0 : \theta(s) > t\}$. Define

$$\left(B_t^{\mathrm{df}}, t \geq 0\right) := \left(B_{\Lambda(t)} + B_{t-\Lambda(t)}^{\mathrm{f}}, t \geq 0\right) \tag{3.16}$$

and set

$$(R_t^{\mathrm{df}}, t \geq 0) := \left(B_t^{\mathrm{df}} - \inf\{B_s^{\mathrm{df}}, s \leq t\}, t \geq 0\right).$$

Lemma 3.4.10. *We have that $\left((2/\sigma_+)R_t^{\mathrm{df}}, t \geq 0\right)$ is the height process corresponding to $\left(\sigma_+ B_t^{\mathrm{df}}, t \geq 0\right)$. Moreover,*

$$\left(n^{-1/3}Y^{\mathrm{df}}\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}H^{\mathrm{df}}\left(\lfloor n^{2/3}t\rfloor\right), t\geq 0\right) \xrightarrow{\text{(d)}} \left(\sigma_+ B^{\mathrm{df}}_t, \frac{2}{\sigma_+}R^{\mathrm{df}}_t, t\geq 0\right) \tag{3.17}$$

in $D(\mathbb{R}_+,\mathbb{R})^2$*, jointly with*

$$\left(n^{-1/3}Y^{+}\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}Y^{\mathrm{f}}\left(\lfloor n^{2/3}t\rfloor\right), t\geq 0\right) \xrightarrow{\text{(d)}} \left(\sigma_+ B_t, \sigma_+ B^{\mathrm{f}}_t, t\geq 0\right)$$

in $D(\mathbb{R}_+,\mathbb{R})^2$ *and*

$$\left(n^{-2/3}\Lambda_n\left(\lfloor n^{2/3}t\rfloor\right), n^{-2/3}\theta_n\left(\lfloor n^{2/3}t\rfloor\right), t\geq 0\right) \xrightarrow{\text{(d)}} \left(\Lambda(t), \theta(t), t\geq 0\right)$$

in $D(\mathbb{R}_+,\mathbb{R})^2$ *as* $n\to\infty$*. Moreover,*

$$\left(n^{-1/3}Y^{\mathrm{df}}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), n^{-1/3}H^{\mathrm{df}}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t\geq 0\right) \xrightarrow{\text{(d)}} \left(\sigma_+ B^{\mathrm{df}}_{\theta(t)}, \frac{2}{\sigma_+}R^{\mathrm{df}}_{\theta(t)}, t\geq 0\right) \tag{3.18}$$

in $D(\mathbb{R}_+,\mathbb{R})^2$ *as* $n\to\infty$ *jointly with the other convergences.*

In the proof of Lemma 3.4.10 we use the following straightforward technical result that follows immediately from the characterization of convergence in the Skorokhod topology given in Ethier and Kurtz [53, Proposition 3.6.5], .

Lemma 3.4.11. *Suppose* $h_n\to h$ *in* $\mathbb{D}(\mathbb{R}_+,\mathbb{R}_+)$ *and* $f_n\to f$ *in* $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ *as* $n\to\infty$*. Then, if* h_n *and* h *are non-decreasing and* h *is continuous, or if* f *is continuous, then*

$$f_n\circ h_n\to f\circ h$$

in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ *as* $n\to\infty$*.*

We also use the following technical result, that is proved in Appendix 3.8.

Lemma 3.4.12. *If* $f_n\to f$ *in* $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ *as* $n\to\infty$*, and* f *is a continuous function that is not bounded from above, with* $f(0)=0$ *and with unique local maxima, then*

$$(\inf\{t: f_n(t)>s\}, s>0)\to(\inf\{t: f(t)>s\}, s>0)$$

in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ *as* $n\to\infty$*.*

Proof of Lemma 3.4.10. Firstly, note that since $(Y^{\mathrm{df}}(k), k \geq 1)$ encodes a critical Bienaymé forest with offspring variance σ_+^2, the proof of Theorem 1.8 in Le Gall [76] gives us that for $(B_s^*, s \geq 0)$ a Brownian motion,

$$\begin{aligned}&\left(n^{-1/3}Y^{\mathrm{df}}\left(\lfloor n^{2/3}s\rfloor\right), n^{-1/3}H^{\mathrm{df}}\left(\lfloor n^{2/3}s\rfloor\right), s \geq 0\right)\\ &\qquad\xrightarrow{\text{(d)}} \left(\sigma_+ B_s^*, \frac{2}{\sigma_+}\left(B_s^* - \inf\{B_u^* : u \leq s\}\right), s \geq 0\right)\end{aligned} \tag{3.19}$$

in $D(\mathbb{R}_+, \mathbb{R})^2$ as $n \to \infty$, and that $\left(\frac{2}{\sigma_+}(B_s^* - \inf\{B_u^*, u \leq s\}), s \geq 0\right)$ is the height process corresponding to $(\sigma_+ B_s^*, s \geq 0)$. Then, we note that since $(Y^+(k), k \geq 1) \overset{d}{=} (Y^{\mathrm{df}}(k), k \geq 1)$, so that also

$$\left(n^{-1/3}Y^+\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{\text{(d)}} (\sigma_+ B_t, t \geq 0)$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$. Then, since also $(Y^+(k), k \geq 1) \overset{d}{=} (Y^{\mathrm{f}}(k), k \geq 1)$ and by Lemma 3.4.12 and the almost sure uniqueness of the local minima of Brownian, we get that

$$\begin{aligned}&\left(n^{-1/3}Y^{\mathrm{f}}\left(\lfloor n^{2/3}s\rfloor\right), n^{-2/3}\inf\left\{k : n^{-1/3}Y^{\mathrm{f}}(k) \leq -x\right\}, s \geq 0, x \geq 0\right)\\ &\qquad\xrightarrow{\text{(d)}} \left(\sigma_+ B_s^{\mathrm{f}}, \inf\left\{u : \sigma_+ B_u^{\mathrm{f}} < -x\right\}, s \geq 0, x \geq 0\right)\end{aligned} \tag{3.20}$$

in $D(\mathbb{R}_+, \mathbb{R})^2$ as $n \to \infty$.

Since $(P_n(k), k \geq 1)$ is non-decreasing, applying Lemma 3.4.11, and combining the convergence in Eq. (3.20) with Lemma 3.4.6 gives that also

$$\left(n^{-2/3}\inf\left\{k : Y^{\mathrm{f}}(k) \leq -P_n\left(\lfloor n^{2/3}t\rfloor - 1\right)\right\}, t \geq 0\right) \xrightarrow{\text{(d)}} \left(\inf\left\{u : \sigma_+ B_u^{\mathrm{f}} < -\frac{\nu_-}{2\mu}t^2\right\}, t \geq 0\right)$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in Eq. (3.20). Therefore,

$$\left(n^{-2/3}\theta_n\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{\text{(d)}} (\theta(t), t \geq 0) \tag{3.21}$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in Eq. (3.20). Recall that

$$\Lambda_n(k) = \max\{j : \theta_n(j) \leq k\} - P_n(\max\{j : \theta_n(j) \leq k\}).$$

By definition, for all n, $(\theta_n(k), k \geq 1)$ and $(\theta(t), t \geq 0)$ are strictly increasing, so

$$\left(n^{-2/3} \max\{j : \theta_n(j) \leq \lfloor n^{2/3}t \rfloor\}), t \geq 0\right) \xrightarrow{\text{(d)}} (\Lambda(t), t \geq 0)$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in Eq. (3.20) and Eq. (3.21). Since $\max\{j : \theta_n(j) \leq \lfloor n^{2/3}t \rfloor\}$ is of order $n^{2/3}$, and, by Lemma 3.4.6, $P_n(\lfloor n^{2/3}t \rfloor)$ is of order $n^{1/3}$, we get that

$$\left(n^{-2/3}\Lambda_n\left(\lfloor n^{2/3}t \rfloor\right), t \geq 0\right) \xrightarrow{\text{(d)}} (\Lambda(t), t \geq 0)$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in Eq. (3.20) and Eq. (3.21).

To finish the proof, we examine the construction of $(Y^{\text{df}}(k), k \geq 1)$ in Eq. (3.14) and the construction of $(B^{\text{df}}_s, s \geq 0)$ in Eq. (3.16). Again, by Lemma 3.4.11, this implies that

$$\left(n^{-1/3}Y^{\text{df}}\left(\lfloor n^{2/3}t \rfloor\right), t \geq 0\right) \xrightarrow{\text{(d)}} \left(B^{\text{df}}_t, t \geq 0\right)$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with all earlier mentioned convergences. Combining this with the convergence in Eq. (3.19) proves Eq. (3.17). The fact that $(B^{\text{df}}_t, t \geq 0)$ is continuous almost surely and Lemma 3.4.11 then imply Eq. (3.18). □

Lemma 3.4.13. *We have that*

$$\begin{aligned}&\left(n^{-1/3}S^+\left(\lfloor n^{2/3}t \rfloor\right)), n^{-1/3}H^+\left(\lfloor n^{2/3}t \rfloor\right), t \geq 0\right)\\ &\qquad\xrightarrow{\text{(d)}} \left(\sigma_+ B^{\text{df}}_{\theta(t)}, \frac{2}{\sigma_+}\left(B^{\text{df}}_{\theta(t)} - \inf\{B^{\text{df}}_s : s \leq \theta(t)\}\right), t \geq 0\right)\end{aligned}$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^2$ *as* $n \to \infty$.

Proof. By Eq. (3.15), and by Lemma 3.4.10, it is sufficient to show that for any $t > 0$,

$$n^{-1/3} \max_{k \leq \lfloor n^{2/3}t \rfloor} E(k) \xrightarrow{\text{(p)}} 0.$$

We remind the reader that $E(k)$ counts the number children of the kth vertex in the forest with dummy and filler vertices that are filler vertices, so

$$n^{-1/3} \max_{k \leq \lfloor n^{2/3}t \rfloor} E(k) \leq n^{-1/3} \max_{k \leq \theta_n(\lfloor n^{2/3}t \rfloor)} (Y^{\text{f}}(k) - Y^{\text{f}}(k-1) + 1),$$

which converges to 0 by tightness of $\left(n^{-2/3}\theta^n(\lfloor n^{2/3}t\rfloor)\right)_{n\geq 1}$ and the fact that

$$\left(n^{-1/3}\mathrm{Y}^{\mathrm{f}}\left(\lfloor n^{2/3}s\rfloor\right), s\geq 0\right)$$

converges in distribution to a continuous process in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$. □

The following lemma is the last ingredient in the proof of Lemma 3.4.9.

Lemma 3.4.14. *We have that with probability* 1,

$$\left(\frac{2}{\sigma_+}\left(B^{\mathrm{df}}_{\theta(t)}-\inf\{B^{\mathrm{df}}_s : s\leq\theta(t)\}\right), t\leq T\right)=\left(\frac{2}{\sigma_+}\left(B^{\mathrm{df}}_{\theta(t)}-\inf\{B^{\mathrm{df}}_{\theta(s)} : s\leq t\}\right), t\leq T\right),$$

which is continuous, and it is the height process corresponding to $\left(\sigma_+B^{\mathrm{df}}_{\theta(t)}, t\leq T\right)$.

Proof. From [76], we know that $\left(\frac{2}{\sigma_+}R^{\mathrm{df}}_t, t\geq 0\right)$ is the height process corresponding to $\left(\sigma_+B^{\mathrm{df}}_t, t\geq 0\right)$. By definition of the height process, it is sufficient to show that, firstly, with probability 1, $(B^{\mathrm{df}}_{\theta(t)}, t\geq 0)$ is continuous, and, secondly, for all $t\geq 0$, and all s such that $\theta(t-)<s<\theta(t)$, we have $B^{\mathrm{df}}_s>B^{\mathrm{df}}_{\theta(t)}$.

Recall that $(B_t, t\geq 0)$ and $(B^{\mathrm{f}}_t, t\geq 0)$ are two independent Brownian motions,

$$\theta(t)=t+\inf\left\{s\geq 0:\sigma_+B^{\mathrm{f}}_s<-\frac{\nu_-}{2\mu}t^2\right\},$$

we have $\Lambda(t)=\inf\{s\geq 0:\theta(s)>t\}$, and

$$\left(B^{\mathrm{df}}_t, t\geq 0\right):=\left(B_{\Lambda(t)}+B^{\mathrm{f}}_{t-\Lambda(t)}, t\geq 0\right).$$

Firstly, note that the jumps of θ correspond to excursions above the infimum of B^{f}. With probability 1, for each of these excursions, the minimum on the excursion is only attained at the endpoints. This can be seen by the almost sure uniqueness of local minima of Brownian motion. We will work on this event of probability 1.

Now fix t such that $\theta(t-)\neq\theta(t)$ and let $s\in(\theta(t-),\theta(t))$. Observe that Λ is equal to t on $[\theta(t-),\theta(t)]$. For $[\theta(t-),\theta(t))$ this follows by definition of Λ, and for $\theta(t)$ it follows since $(\theta(u):u\geq 0)$ is strictly increasing. This implies that

$$s-\Lambda(s)<\theta(t)-\Lambda(\theta(t))=\inf\left\{u\geq 0:\sigma_+B^{\mathrm{f}}_u<-\frac{\nu_-}{2\mu}t^2\right\}.$$

By our assumption on the minima of the excursions above the infimum of B^{f}, this implies that

$$B^{\mathrm{f}}_{s-\Lambda(s)} > -\frac{\nu_-}{2\mu}t^2 = B^{\mathrm{f}}_{\theta(t)-\Lambda(\theta(t))}$$

where the last equality follows from continuity of B^{f}. Combining this with $\Lambda(s) = \Lambda(\theta(t))$ implies that $B^{\mathrm{df}}_s > B^{\mathrm{df}}_{\theta(t)}$.

Finally,

$$B^{\mathrm{df}}_{\theta(t-)} = B_{\Lambda(\theta(t-))} + B^{\mathrm{f}}_{\theta(t-)-\Lambda(\theta(t-))} = B_t + B^{\mathrm{f}}_{\theta(t-)-t}$$

and by continuity of $(B^{\mathrm{f}}_s, s \geq 0)$,

$$\begin{aligned}
B^{\mathrm{f}}_{\theta(t-)-t} &= B^{\mathrm{f}}\left(\lim_{s\uparrow t}\inf\{u : B^{\mathrm{f}}_u < -\frac{\nu_-}{2\mu}s^2\}\right) \\
&= \lim_{s\uparrow t} B^{\mathrm{f}}\left(\inf\left\{u : B^{\mathrm{f}}_u < -\frac{\nu_-}{2\mu}s^2\right\}\right) \\
&= -\frac{\nu_-}{2\mu^2}t^2 \\
&= B^{\mathrm{f}}_{\theta(t)-t},
\end{aligned}$$

so $B^{\mathrm{df}}_{\theta(t-)} = B^{\mathrm{df}}_{\theta(t)}$. □

3.4.3 Proof of Proposition 3.4.1

We will now combine the convergence of the measure change under rescaling, which is the content of Corollary 3.4.2, and the convergence of the encoding processes of the forest with dummy leaves, which is the content of Proposition 3.4.8, in order to prove Proposition 3.4.1.

Proof of Proposition 3.4.1. Recall that $\hat{P}_n(k)$ denotes the number of dummy leaves amongst the first k vertices in the forest with dummy leaves. Then, as shown in Proposition 3.2.4, the probability that the $(k+1)$th vertex in the out-forest is purple, given the degrees in order of discovery and the dummy leaves amongst the first k vertices is equal to

$$q_{k+1} := \frac{\hat{S}^-_n(k)}{\sum_{i=1}^n D^-_i - k - \hat{I}^+_n(k)} \mathbb{1}_{\{\hat{I}^+_n(k-1)=\hat{I}^+_n(k)\}},$$

where $\hat{I}^+_n(k) = \min\{\hat{S}^+_n(l) : l \leq k\}$. In order to use the results on the forest with dummy leaves, we need to replace the term $\sum_{i=1}^n D^-_i$ in the denominator by μn. Therefore, define a new forest, *the approximate out-forest*, in which the degrees in order of discovery are the same as in the out-forest. However, in this forest, the probability that the $(k+1)$th vertex is a dummy leaf, given the degrees in order of discovery and the dummy leaves amongst the first k vertices, is equal to

$$\tilde{q}_{k+1} := \frac{\tilde{S}_n^-(k)}{\mu n - k - \tilde{I}_n(k)} \mathbb{1}_{\{\tilde{I}_n(k-1)=\tilde{I}_n(k)\}},$$

where $\tilde{S}_n^-(k)$ is the number of unused in-edges of previously discovered vertices in the approximate out-forest up to time k and $-\tilde{I}_n^+(k)$ is the number of components in the approximate out-forest up to time k. We let $\tilde{P}_n(k)$ denote the number of dummy leaves amongst the first k vertices in the approximate out-forest. We claim that there exists a coupling such that

$$\sum_{i=1}^{\lfloor n^{2/3}T \rfloor} |q_i - \tilde{q}_i| \xrightarrow{(\mathrm{p})} 0$$

as $n \to \infty$. Indeed, by the convergence in Proposition 3.4.4,

$$\left(n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3}T \rfloor} \hat{D}_i^n \right)_{n>0}$$

is tight. Moreover, with a slight adaptation to the proof of Lemma 3.4.5, we can show that $\left(n^{-1/3}\tilde{P}_n\left(\lfloor n^{2/3}T\rfloor\right)\right)_{n>0}$ is tight. This, combined with the convergence under rescaling of $(\hat{Y}_n^+(k), k \geq 1)$, implies that also $\left(n^{-1/3}\tilde{I}_n^+\left(\lfloor n^{2/3}T\rfloor\right)\right)_{n>0}$ is tight. Since $D_1^-, \dots, D_n^-$ are i.i.d. random variables with mean μ and finite variance, $\left(n^{-1/2}\left(\sum_{i=1}^n D_i^- - \mu n\right)\right)_{n>0}$ is tight. By using the trivial identity $a/b - c/d = (b(a-c) - c(d-b))/bd$, this implies that $\left(n^{2/3} \max_{k \leq \lfloor n^{2/3}T \rfloor} |q_k - \tilde{q}_k|\right)_{n>0}$ is tight, which implies that there exists a coupling such that $\left(\max_{k\leq\lfloor n^{2/3}T\rfloor} |\hat{P}_n(k) - \tilde{P}_n(k)|\right)_{n>1}$ and $\left(\max_{k\leq\lfloor n^{2/3}T\rfloor} |\hat{I}_n^+(k) - \tilde{I}_n^+(k)|\right)_{n>1}$ are tight, which implies that, again by $a/b - c/d = (b(a-c) - c(d-b))/bd$, $\left(n^{5/6} \max_{k\leq\lfloor n^{2/3}T\rfloor} |q_k - \tilde{q}_k|\right)_{n>0}$ is tight, which implies that

$$\sum_{i=1}^{\lfloor n^{2/3}T \rfloor} |q_i - \tilde{q}_i| \xrightarrow{(\mathrm{p})} 0$$

as $n \to \infty$. Therefore, under the right coupling,

$$\mathbb{P}\left(\max_{k\leq\lfloor n^{2/3}T\rfloor} |\hat{P}_n(k) - \tilde{P}_n(k)| > 0 \right) \to 0.$$

In other words, we can couple the out-forest and the approximate out-forest in such a way that we do not see any difference on the scale of interest. Therefore, we can show convergence under rescaling of the encoding processes of the approximate out-forest instead. To avoid further complicating notation, we will from now on refer to its encoding processes as

$$(\hat{S}_n^+(k), \hat{H}_n, \hat{S}_n^-(k), \hat{P}_n(k), 1 \leq k \leq \lfloor n^{2/3}T \rfloor).$$

Then, these processes are constructed out of sample paths of $(\hat{Y}^+(k), \hat{Y}^-(k), 1 \leq k \leq \lfloor n^{2/3}T \rfloor)$ and independent randomness in exactly the same way as the sample paths of

$$(S_n^+(k), H_n^+(k), S_n^-(k), P_n(k), 1 \leq k \leq \lfloor n^{2/3}T \rfloor)$$

(corresponding to the forest with dummy vertices) are constructed out of sample paths of $(Y^+(k), Y^-(k), 1 \leq k \leq \lfloor n^{2/3}T \rfloor)$ and independent randomness. We will use the following notation:

$$\begin{aligned}
\hat{Y}_{(n)}^+ &:= \left(n^{-1/3}\hat{Y}^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \leq t \leq T\right) \\
\hat{S}_{(n)}^+ &:= \left(n^{-1/3}\hat{S}_n^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \leq t \leq T\right) \\
\hat{H}_{(n)} &:= \left(n^{-1/3}\hat{H}_n\left(\lfloor n^{2/3}t \rfloor\right), 0 \leq t \leq T\right) \\
Y_{(n)}^+ &:= \left(n^{-1/3}Y^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \leq t \leq T\right) \\
S_{(n)}^+ &:= \left(n^{-1/3}S_n^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \leq t \leq T\right) \\
H_{(n)}^+ &:= \left(n^{-1/3}H_n^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \leq t \leq T\right)
\end{aligned}$$

Let $f : D([0,T], \mathbb{R})^3 \to \mathbb{R}$ be a bounded, continuous test-function. Then, for $m = \lfloor n^{2/3}T \rfloor$

$$\begin{aligned}
\mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\right] &= \mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\mathbb{1}_{R_n \geq m}\right] + o(1) \\
&= \mathbb{E}\left[\mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\middle| \hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,m}\right]\mathbb{1}_{R_n \geq m}\right] + o(1) \\
&= \mathbb{E}\left[\Phi(n,m)\mathbb{E}\left[f\left(Y_{(n)}^+, S_{(n)}^+, H_{(n)}^+\right)\middle| \mathbf{Z}_1, \ldots, \mathbf{Z}_m\right]\right] + o(1) \\
&= \mathbb{E}\left[\Phi(n,m)f\left(Y_{(n)}^+, S_{(n)}^+, H_{(n)}^+\right)\right] + o(1),
\end{aligned}$$

where we use that $\mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\middle| \hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,m}\right]$ and $\mathbb{1}_{R_n \geq m}$ are bounded, adapted functions of $\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,m}$, and that $\Phi(n,m)$ is the measure change from $(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$ to $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,m})$. Then, if we repeat the proof of Proposition 3.4.4, using Proposition 3.4.8 to include the convergence of $S^+_{(n)}$ and $H^+_{(n)}$, we obtain that

$$\mathbb{E}\left[f\left(\hat{Y}^+_{(n)}, \hat{S}^+_{(n)}, \hat{H}_{(n)}\right)\right] \to \mathbb{E}\left[\Phi(T) f\left(\sigma_+ B_t, \sigma_+ B^+_t, \frac{2}{\sigma_+} R^+_t, 0 \le t \le T\right)\right].$$

Since

$$(B^+_t, t \ge 0) = \left(B_t - \frac{\nu_-}{2\sigma_+\mu} t^2, t \ge 0\right),$$

Proposition 3.4.3 implies that the limit object has the right law. By Proposition 3.4.8, S^-_n converges in distribution under rescaling to a deterministic process, which will not be affected by the measure change. This completes the proof. □

3.4.4 Conditioning on simplicity

In this section, we will first show that, with high probability, there exists a simple graph with the degree sequence that we sample and we show that the multigraph resulting from the configuration model is simple with probability asymptotically bounded away from 0. Then, we show that when we sample the configuration model according to Algorithm 1, we do not see any loops or multiple edges far beyond our time scale of interest. We will then use an argument by Joseph [70] to show that this implies that Proposition 3.4.1 holds conditional on the resulting multigraph being simple.

We start by showing that with high probability, there exists a simple graph with the degree sequence that we sample. For this, we need the following lemma.

Lemma 3.4.15. *On the event* $\left\{\sum_{i=1}^n D^-_i = \sum_{i=1}^n D^+_i\right\}$, *for all integers* i *and* j *with* $1 \le i+j \le 3$ *or* $\{i,j\} = \{1,3\}$ *it holds that*

$$\frac{1}{n}\sum_{k=1}^{n} (D^-_k)^i (D^+_k)^j \xrightarrow{\mathbb{P}} \mathbb{E}\left[(D^-)^i (D^+)^j\right].$$

Proof. First, for $l, m \in \mathbb{N}$, define $\rho_m(l) = \mathbb{P}\left(\sum_{k=1}^m (D^-_k - D^+_k) = l\right)$. Then, since the second moment of $D^- - D^+$ is finite, the discrete local limit theorem implies that there exists a $C > 0$ such that $\rho_m(l) < Cm^{-1/2}$ for all l, m. Moreover, again by the discrete local limit theorem and because $D^- - D^+$ is strongly aperiodic, there exists a $c > 0$ such that $\rho_m(0) > cm^{-1/2}$ for all m large enough.

Now, let i, j be as in the statement of the lemma. Fix $\epsilon > 0$. Then,

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n}\left|\sum_{k=1}^{\lfloor n\rfloor}\left((D_k^-)^i(D_k^+)^j - \mathbb{E}\left[(D^-)^i(D^+)^j\right]\right)\right| > \epsilon \;\middle|\; \sum_{k=1}^{n} D_k^- = \sum_{k=1}^{n} D_k^+\right)\\
&\le \mathbb{P}\left(\frac{1}{n}\left|\sum_{k=1}^{\lfloor n/2\rfloor}\left((D_k^-)^i(D_k^+)^j - \mathbb{E}\left[(D^-)^i(D^+)^j\right]\right)\right| > \epsilon/2 \;\middle|\; \sum_{k=1}^{n} D_k^- = \sum_{k=1}^{n} D_k^+\right)\\
&+ \mathbb{P}\left(\frac{1}{n}\left|\sum_{k=\lfloor n/2\rfloor+1}^{n}\left((D_k^-)^i(D_k^+)^j - \mathbb{E}\left[(D^-)^i(D^+)^j\right]\right)\right| > \epsilon/2 \;\middle|\; \sum_{k=1}^{n} D_k^- = \sum_{k=1}^{n} D_k^+\right)
\end{aligned}$$

by the triangle inequality, so by symmetry it suffices to show that the second term goes to 0 as $n \to \infty$. Denote

$$A_n = A_n(\mathbf{D}_{\lfloor n/2\rfloor+1}, \ldots, \mathbf{D}_n) = \left\{\frac{1}{n}\left|\sum_{k=\lfloor n/2\rfloor+1}^{n}\left((D_k^-)^i(D_k^+)^j - \mathbb{E}\left[(D^-)^i(D^+)^j\right]\right)\right| > \epsilon/2\right\}$$

so that $\mathbb{P}(A_n) \to 0$ as $n \to \infty$ by the weak law of large numbers. We note that

$$\begin{aligned}
&\mathbb{P}\left(A_n \;\middle|\; \sum_{k=1}^{n} D_k^- = \sum_{k=1}^{n} D_k^+\right)\\
&= \frac{\mathbb{E}\left[\mathbb{1}_{A_n}\mathbb{P}\left(\sum_{k=1}^{\lfloor n/2\rfloor}(D_k^- - D_k^+) = \sum_{k=\lfloor n/2\rfloor+1}^{n}(D_k^+ - D_k^-)\middle| \mathbf{D}_{\lfloor n/2\rfloor+1}, \ldots, \mathbf{D}_n\right)\right]}{\mathbb{P}\left(\sum_{k=1}^{n}(D_k^- - D_k^+) = 0\right)}\\
&= \mathbb{E}\left[\mathbb{1}_{A_n}\frac{\rho_{\lfloor n/2\rfloor}\left(\sum_{k=\lfloor n/2\rfloor+1}^{n}(D_k^+ - D_k^-)\right)}{\rho_n(0)}\right]
\end{aligned}$$

where we use the definition of conditional probability and the tower property in the second line and the independence between $\{\mathbf{D}_1, \ldots \mathbf{D}_{\lfloor n/2\rfloor}\}$ and $\{\mathbf{D}_{\lfloor n/2\rfloor+1}, \ldots \mathbf{D}_n\}$ in the third line. However, by our observations above, there is a C' such that $\frac{\rho_{\lfloor n/2\rfloor}(k)}{\rho_n(0)} < C'$ for all n large enough and for all k, so

$$\mathbb{P}\left(A_n \;\middle|\; \sum_{k=1}^{n} D_k^- = \sum_{k=1}^{n} D_k^+\right) \le C'\mathbb{P}(A_n)$$

which tends to 0. □

This yields the following proposition.

Proposition 3.4.16. *Let* $(\mathbf{D}_{1,n}, \ldots, \mathbf{D}_{n,n})$ *be a progression of sequences of i.i.d. samples from* ν*, conditional on the event that* $\left\{\sum_{k=1}^{n} D_{k,n}^- = \sum_{k=1}^{n} D_{k,n}^+\right\}$*. Then, the probability that there exists a simple digraph with degree sequence* $(\mathbf{D}_{1,n}, \ldots, \mathbf{D}_{n,n})$ *tends to* 1 *as* $n \to \infty$*. Moreover,* *the probability that the configuration model on* $(\mathbf{D}_{1,n},\dots,\mathbf{D}_{n,n})$ *yields a simple graph tends to*

$$\exp\left(-1-\frac{(\mathbb{E}[(D^-)^2]-\mu)(\mathbb{E}[(D^+)^2]-\mu)}{\mu^2}\right)$$

as $n\to\infty$.

Proof. By Lemma 3.4.15, we may work on a probability space where for all non-negative integers i and j with $1\le i+j\le 3$ or $\{i,j\}=\{1,3\}$ it holds that

$$\frac{1}{n}\sum_{k=1}^{n}(D^-_{k,n})^i(D^+_{k,n})^j\to\mathbb{E}\left[(D^-)^i(D^+)^j\right]$$

almost surely as $n\to\infty$.

Now, let $(\mathbf{d}_{1,n},\dots,\mathbf{d}_{n,n})_{n\ge1}$ be a progression of sequences with $\mathbf{d}_{k,n}=(d^-_{k,n},d^+_{k,n})\in\mathbb{N}\times\mathbb{N}$ for each k,n. Assume that for each n it holds that, firstly, $\sum_{k=1}^n d^-_{k,n}=\sum_{k=1}^n d^+_{k,n}$, secondly, $\max_{k\le n}d^-_{k,n}\vee d^+_{k,n}=o(\sqrt{n})$, and, finally, for all non-negative integers i,j such that $1\le i+j\le 2$ there exist positive $a_{i,j}$ such that

$$\frac{1}{n}\sum_{k=1}^{n}(d^-_{k,n})^i(d^+_{k,n})^j\to a_{i,j}.$$

Then, for S_n the number of self-loops and M_n the number of directed edges with multiplicity exceeding 1 in the directed configuration model on vertex set $[n]$ with degree sequence $(\mathbf{d}_{1,n},\dots,\mathbf{d}_{n,n})$, it holds that (S_n,M_n) converges in distribution to (S,M), for S and M two independent Poisson random variables with means $a_{1,1}/a_{1,0}$ and $(a_{2,0}-a_{1,0})(a_{0,2}-a_{0,1})/a_{1,0}^2$ respectively. This follows from a trivial adaptation of the proof of [101, Proposition 7.13], where an analogous property is shown for the undirected configuration model.

Therefore, on the coupling on degree sequences that we consider above, almost surely, the number of self-loops and multiple edges in the configuration model on degree sequence $(\mathbf{D}_{1,n},\dots,\mathbf{D}_{n,n})$ converges in distribution to (S,M) for S and M two independent Poisson random variables with means $\mathbb{E}[D^-D^+]/\mathbb{E}[D^-]=1$ and $(\mathbb{E}[(D^-)^2]-\mathbb{E}[D^-])(\mathbb{E}[(D^+)^2]-\mathbb{E}[D^+])/\mathbb{E}[D^-]^2$ respectively, and in particular, almost surely, the asymptotic probability of sampling a simple graph is bounded away from 0. Here we use that the almost sure convergence of $\frac{1}{n}\sum_{k=1}^n(D^-_{k,n})^3$ implies that $\max\{D^-_{k,n}\}=o(\sqrt{n})$ almost surely and, similarly, we have that $\max\{D^+_{k,n}\}=o(\sqrt{n})$ almost surely. The result follows. $\square$

We will now show that when we sample the configuration model according to Algorithm 1, we do not see any loops or multiple edges far beyond our time scale of interest. We let $B_n(k)$ be the number of 'bad edges' up to time k; to be precise, it equals be the number of self-loops and edges created parallel to an existing edge in the same direction as that edge, up until discovery of the kth vertex of the out-forest. Following [37], we call these anomalous edges.

Proposition 3.4.17. *Suppose $\beta < 1$. Then we have*

$$\mathbb{P}\left(B_n(\lfloor n^\beta \rfloor) > 0\right) \to 0$$

as $n \to \infty$.

Remark 3.4.18. *We adapt the proof of [70, Lemma 7.1] and of [37, Proposition 5.3] to the directed setting. A significant complication is caused by the conditioning on*

$$\left\{\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+\right\}.$$

We observe that in both papers, the proof of the aforementioned result is not fully correct, because the authors use the wrong expression for the probability of sampling an anomalous edge. However, the argument below can be adapted to the setting of [70] and [37] to yield a correct proof.

Proof. We distinguish between the following types of anomalous edges.

Self-loops occur when the out-half-edge of a vertex is paired to an in-half-edge of the same vertex. Let $B_n^1(k)$ be the number of self-loops that are found up to time k. For v explored up to time $\lfloor n^\beta \rfloor$, a vertex with in-degree d_v^- and out-degree d_v^+, there are $d_v^- d_v^+$ possible combinations of an in-half-edge and an out-half-edge that form a self-loop connected to v. Any of these combinations of half-edges is paired with probability bounded above by

$$\frac{1}{\sum_{i=\lfloor n^\beta \rfloor+1}^n \hat{D}_i^-}.$$

Parallel edges occur when an out-half-edge of a vertex is paired to an in-half-edge of one of its previously explored children. Let $B_n^2(k)$ be the number of parallel edges that are found up to time k. For any vertex v with in-degree d_v^-, and a parent $p(v)$ with out-degree $d_{p(v)}^+$, there are at most $d_v^- d_{p(v)}^+$ possible combinations of an in-half-edge and an out-half-edge that form a parallel edge from $p(v)$ to v. Again, any of these combinations of half-edges is paired with probability bounded above by

$$\frac{1}{\sum_{i=\lfloor n^\beta\rfloor+1}^{n}\hat{D}_i^-}.$$

The last type of anomalous edges is a surplus edge with multiplicity greater than 1. Let $B_n^3(k)$ be the number of surplus edges with multiplicity greater than 1 that are found up to time k. For a vertex w with out-degree d_w^+ and a vertex v with in-degree d_v^-, a multiple surplus edge from w to v can only occur if v is discovered before w. In that case, there are at most $(d_w^+)^2(d_v^-)^2$ possible pairs of combinations of half-edges, and each of these pairs appears with probability bounded above by

$$\left(\frac{1}{\sum_{i=\lfloor n^\beta\rfloor+1}^{n}\hat{D}_i^-}\right)^2.$$

Let $p(i)$ denote the index of the parent of the vertex with index i. Also, denote

$$\mathcal{G}^n=\sigma\left(\hat{D}_1^-,\hat{D}_1^+,\dots,\hat{D}_n^-,\hat{D}_n^+\right).$$

Then, by the conditional version of Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left(B_n^1(\lfloor n^\beta\rfloor)>0\middle|\,\mathcal{G}^n\right)&\le\frac{\sum_{i=1}^{\lfloor n^\beta\rfloor}\hat{D}_i^-\hat{D}_i^+}{\sum_{i=\lfloor n^\beta\rfloor+1}^{n}\hat{D}_i^-}\wedge 1,\\
\mathbb{P}\left(B_n^2(\lfloor n^\beta\rfloor)>0\middle|\,\mathcal{G}^n\right)&\le\frac{\sum_{i=1}^{\lfloor n^\beta\rfloor}\hat{D}_i^-\mathbb{E}\left[\hat{D}_{p(i)}^+\middle|\,\mathcal{G}^n\right]}{\sum_{i=\lfloor n^\beta\rfloor+1}^{n}\hat{D}_i^-}\wedge 1,\\
\mathbb{P}\left(B_n^3(\lfloor n^\beta\rfloor)>0\middle|\,\mathcal{G}^n\right)&\le\frac{\sum_{i=1}^{\lfloor n^\beta\rfloor}\sum_{j<i}(\hat{D}_i^+)^2(\hat{D}_j^-)^2}{\left(\sum_{i=\lfloor n^\beta\rfloor+1}^{n}\hat{D}_i^-\right)^2}\wedge 1,
\end{aligned}$$

where we note that $p(i)$ is not adapted to $\mathcal{G}^n$, because ancestral relations in the tree also depend on the surplus edges. However, we observe that by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{i=1}^{\lfloor n^\beta\rfloor}\hat{D}_i^-\mathbb{E}\left[\hat{D}_{p(i)}^+\middle|\,\mathcal{G}^n\right]&\le\left(\sum_{i=1}^{\lfloor n^\beta\rfloor}(\hat{D}_i^-)^2\right)^{1/2}\left(\sum_{i=1}^{\lfloor n^\beta\rfloor}\mathbb{E}\left[\hat{D}_{p(i)}^+\middle|\,\mathcal{G}^n\right]^2\right)^{1/2}\\
&=\left(\sum_{i=1}^{\lfloor n^\beta\rfloor}(\hat{D}_i^-)^2\right)^{1/2}\left(\sum_{j=1}^{\lfloor n^\beta\rfloor}(\hat{D}_j^+)^2\sum_{i=1}^{\lfloor n^\beta\rfloor}\mathbb{E}\left[\mathbb{1}_{j=p(i)}\middle|\,\mathcal{G}^n\right]\right)^{1/2}
\end{aligned}$$

$$\leq \left(\sum_{i=1}^{\lfloor n^\beta \rfloor} (\hat{D}_i^-)^2\right)^{1/2} \left(\sum_{i=1}^{\lfloor n^\beta \rfloor} (\hat{D}_i^+)^3\right)^{1/2}.$$

We will show that

$$\mathbb{P}\left(B_n^1(\lfloor n^\beta \rfloor) + B_n^2(\lfloor n^\beta \rfloor) + B_n^3(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) \xrightarrow{p} 0 \tag{3.22}$$

as $n \to \infty$. We note that

$$\sum_{i=\lfloor n^\beta \rfloor + 1}^{n} \hat{D}_i^- = \sum_{i=1}^{n} D_i^- - \sum_{i=1}^{\lfloor n^\beta \rfloor - 1} \hat{D}_i^-,$$

and by the weak law of large numbers, $\frac{1}{n}\sum_{i=1}^n D_i^- \xrightarrow{p} \mu n$, so Eq. (3.22) follows if we show that

1. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} \hat{D}_i^- \xrightarrow{p} 0$,
2. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} \hat{D}_i^- \hat{D}_i^+ \xrightarrow{p} 0$,
3. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} (\hat{D}_i^-)^2 \xrightarrow{p} 0$, and
4. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} (\hat{D}_i^+)^3 \xrightarrow{p} 0$

as $n \to \infty$. The proposition will then follow from the bounded convergence theorem.

Note that we can only show the convergence of the Radon-Nikodym derivative $\Phi(n, m)$ under rescaling for $m = O(n^{2/3})$, so it is not straightforward to use the measure change to prove results on the time scale $O(n^\beta)$ for $\beta > 2/3$, such as the convergences above. Therefore, instead, we will use *Poissonization* to sample $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{R_n,n})$. This technique was also used by Joseph in [70].

Let R_n be as before, and, conditional on R_n, let $D_1^{0,+}, \ldots, D_{n-R_n}^{0,+}$ i.i.d. random variables with the law of D^+ conditional on the event $\{D^- = 0\}$, and set $S_n = \sum_{i=1}^{n-R_n} D_i^{0,+}$. Suppose $R_n = r$ and $S_n = s$. Let

$$\pi_0(dt, k_1, k_2) = r\mathbb{P}(D^- = k_1, D^+ = k_2 | D^- > 0) k_1 \exp(-k_1 t) dt$$

be a measure on $\mathbb{R}_+ \times \mathbb{N}^2$, and let Π_0 be a Poisson point process with intensity measure π_0 conditional on $\Pi_0(\mathbb{R}, \mathbb{N}, \mathbb{N}) = r$. We view the first coordinate as the time coordinate, and refer to the second and third coordinate as the *point*. Then, the points in Π_0 ordered by time have the same law as $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{r,n})$ (before conditioning on the event $\{\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+\}$).

The intensity of this process is not constant in t, so we perform a time change. Define

$$\mathcal{L}_{\mathbf{D}}(x,y) = \mathbb{E}\left[\exp(-xD^- - yD^+)\middle|\, D^- > 0\right],$$

and set

$$\psi(t) = \left(1 - \mathcal{L}_{\mathbf{D}}(\cdot,0)\right)^{-1},$$

so that, by a trivial adaptation of [70, Lemma 4.1], for

$$\pi_r(dt,k_1,k_2) := \mathbb{P}(D^- = k_1, D^+ = k_2 | D^- > 0) k_1 \exp\left(-k_1\psi(t/r)\right)\psi'(t/r)dt$$

on $(0,r)\times\mathbb{N}^2$, we have that for $t\in(0,r)$, there exists a probability measure P_t on $\mathbb{N}^2$ such that

$$\pi_r(dt,k_1,k_2) = P_t(D^- = k_1, D^+ = k_2)dt.$$

Let Π^r be a Poisson point process with intensity π_r. Define $N_r = \Pi_r((0,r),\mathbb{N},\mathbb{N})$ and $\Delta_r = \int_{(0,r)\times\mathbb{N}^2}(k_1 - k_2)\Pi^r(dt,k_1,k_2) = s$. Then, let $\Pi^{r,s}$ have the law of Π_r conditional on the events $\{N_r = r\}$ and $\{\Delta_r = s\}$. Then, the points of $\Pi^{r,s}$ ordered by time are distributed as $(\hat{\mathbf{D}}_{n,1},\ldots,\hat{\mathbf{D}}_{n,R_n})$ conditional on the events $\left\{\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+\right\}$, $\{R_n = r\}$ and $\{S_n = s\}$. Let $\lambda_t^{r,s}$ be the marginal density of $\Pi^{r,s}$ in t, so that there exists a probability distribution $P_t^{r,s}(k_1,k_2)$ on $\mathbb{N}^2$ such that for $\pi_t^{r,s}(k_1,k_2)$ the marginal intensity measure on $\mathbb{N}^2$ of $\Pi^{r,s}$ in t,

$$\pi_t^{r,s}(k_1,k_2) = \lambda_t^{r,s}P_t^{r,s}(k_1,k_2)$$

for all $k_1,k_2\in\mathbb{N}$.

For any $L>0$, define

$$\mathcal{E}_L = \left\{|R_n - \mathbb{E}[R_n]| \le Ln^{1/2}, |S_n - \mathbb{E}[S_n]| \le Ln^{1/2}\right\}.$$

Then, note that

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta\rfloor}\hat{D}_i^-\hat{D}_i^+ > \epsilon\right) \le & \mathbb{P}(\mathcal{E}_L^c) + \mathbb{P}\left(\Pi_{R_n,S_n}\left((0,2n^\beta),\mathbb{N}^2\right) < n^\beta \middle| \mathcal{E}_L\right) \\ & + \mathbb{P}\left(\frac{1}{n}\int_{(0,2n^\beta)\times\mathbb{N}^2} k_1k_2\Pi_{R_n,S_n}(dt,k_1,k_2) > \epsilon \middle| \mathcal{E}_L\right)\end{aligned}$$

Fix $\epsilon > 0$. By the central limit theorem, we can pick an L such that $\mathbb{P}(\mathcal{E}_L^c) < \epsilon$ for all n. We condition on $\mathcal{E}_L$. Suppose $R_n = r$ and $S_n = s$. Then, for P a Poisson random variable with rate $2n^\beta$,

$$\mathbb{P}\left(\Pi_{r,s}\left((0,2n^\beta),\mathbb{N}^2\right) < n^\beta\right) \leq \frac{\mathbb{P}\left(P < n^\beta\right)}{\mathbb{P}(\Delta_r = s, N_r = r)}$$

We note that the numerator is the probability of a large-deviation event and decreases exponentially fast in n^β, while the local limit theorem yields that the denominator is of order $n^{-1/2}$ uniformly in all r and s that we consider on $\mathcal{E}_L$. This implies that

$$\mathbb{P}\left(\Pi_{R_n,S_n}\left((0,2n^\beta),\mathbb{N}^2\right) < n^\beta \middle| \mathcal{E}_L\right) \to 0$$

as $n \to \infty$. Now, note that for $E_t^{r,s}$ denoting the expectation with respect to $P_t^{r,s}$,

$$\mathbb{E}\left[\frac{1}{n}\int_{(0,2n^\beta)\times\mathbb{N}^2} k_1 k_2 \Pi_{r,s}(dt,k_1,k_2)\right] = \frac{1}{n}\int_{(0,2n^\beta)} \lambda_t^{r,s} E_t^{r,s}[D^- D^+]dt,$$

so we start by bounding $E_t^{r,s}[D^-D^+]$. We note that

$$E_t^{r,s}\left[D^-D^+\right] = E_t^r\left[D^-D^+\middle|\,\Delta_r = s, N_r = r\right] = E_t^r\left[D^-D^+\frac{\mathbb{P}\left[\Delta_n = s, N_r = r\middle|\,\Pi_r(t,D^-,D^+) = 1\right]}{\mathbb{P}\left[\Delta_r = s, N_r = r\right]}\right].$$

By the fact that Π_r is a point process, we have that for k_1, k_2 in $\mathbb{N}$,

$$\mathbb{P}\left[\Delta_r = s, N_r = r\,|\Pi_r(t,k_1,k_2) = 1\right] = \mathbb{P}\left[\Delta_r = s + k_2 - k_1, N_r = r-1\right],$$

so that, since $N_r \sim \text{Poisson}(r)$, and since on the event $\{N_r = r-1\}$ (resp. $\{N_r = r\}$), $\Delta_r - s$ is the sum of $r-1$ (resp. r) i.i.d. random variables with finite variance and mean at most $O(n^{-1/2})$, we observe that, by the local limit theorem,

$$\mathbb{P}\left[\Delta_r = s, N_r = r\,\middle|\hat{D}_t^- = k_1, \hat{D}_t^+ = k_2\right] = O(n^{-1/2}), \text{ and}$$
$$\mathbb{P}\left[\Delta_r = s, N_r = r\right] = \Theta(n^{-1/2})$$

for any k_1 and k_2, and any r and s that we consider on $\mathcal{E}_L$. Therefore, there exists a c_1 such that

$$\frac{\mathbb{P}\left[\Delta_r = s, N_r = r\,\middle|\hat{D}_t^- = k_1, \hat{D}_t^+ = k_2\right]}{\mathbb{P}\left[\Delta_r = s, N_r = r\right]} < c_1$$

for any k_1, k_2, t and n, and any r and s that we consider on $\mathcal{E}_L$. If we show that for some c_2

$$E_t^r\left[\hat{D}^-\hat{D}^+\right] < c_2$$

for all r in the interval that we consider and all $t < 2n^\beta$, it follows that there is a c_3 such that

$$E_t^{r,s}\left[\hat{D}^-\hat{D}^+\right] < c_3$$

for any k_1, k_2, t and n, and any r and s that we consider on $\mathcal{E}_L$. We note that by definition of $\pi_r(dt, k_1, k_1)$,

$$E_t^r\left[\hat{D}^-\hat{D}^+\right] = \frac{\frac{d^3}{dx^2dy}\mathcal{L}_{\mathbf{D}}(x,y)|_{(\psi(t/r),0)}}{\frac{d}{dx}\mathcal{L}_{\mathbf{D}}(x,y)|_{(\psi(t/r),0)}}.$$

Careful analysis of $\mathcal{L}_{\mathbf{D}}(x,y)$ and $\psi(s)$ implies that this quantity is bounded uniformly for all n, all r in the interval that we consider and all $t \in (0, 2n^\beta)$. We refer the reader to the proof of [70, Lemma A.1] for the details of a similar argument in the undirected setting. This implies that

$$\mathbb{E}\left[\frac{1}{n}\int_{(0,2n^\beta)\times\mathbb{N}^2} k_1k_2\Pi_{r,s}(dt,k_1,k_2)\right] \le \frac{C}{n}\mathbb{E}\left[\Pi_{r,s}\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right)\right].$$

Then, we note that for any $x > 0$, for P a Poisson random variable with rate $2n^\beta$,

$$\mathbb{P}\left(\Pi_{r,s}\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right) > (x+1)2n^\beta\right) \le \frac{\mathbb{P}\left[P > (x+1)2n^\beta\right]}{\mathbb{P}\left[\Delta_r = s, N_r = r\right]}.$$

Then, by the local limit theorem and the exponential tail of the Poisson distribution, we obtain that there exist $c_4, c_5 > 0$ such that for all n, all r and s in the interval of interest and all $x > 1$,

$$\mathbb{P}\left(\Pi_{r,s}\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right) > (x+1)2n^\beta\right) \le c_4\exp(-c_5xn^\beta).$$

This implies that there is a constant c_6 such that

$$\mathbb{E}\left[\Pi_{r,s}\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right)\right] \le c_6n^\beta$$

for all n and all r and s that we consider under $\mathcal{E}_L$. It then follows that

$$\mathbb{E}\left[\frac{1}{n}\int_{(0,2n^\beta)\times\mathbb{N}^2} k_1k_2\Pi_{r,s}(dt,k_1,k_2)\right] \to 0$$

as $n \to \infty$ uniformly in all r and s of interest, so for n large enough,

$$\mathbb{P}\left(\frac{1}{n} \int_{(0,2n^\beta)\times\mathbb{N}^2} k_1 k_2 \Pi_{R_n,S_n}(dt, k_1, k_2) > \epsilon \middle| \mathcal{E}_L \right) < \epsilon.$$

This implies that

$$\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} \hat{D}_i^- \hat{D}_i^+ \xrightarrow{p} 0.$$

The other convergences are proved similarly, and the result follows. □

Proposition 3.4.19. *Proposition 3.4.1 holds conditionally on the resulting multigraph being simple.*

Proof. Let $\rho(n) = \inf\{k \geq 1 : B_n(k) > 0\}$, and note that the event that the multigraph formed by the configuration model on n vertices is simple is equal to $\{\rho(n) = \infty\}$. Proposition 3.4.17 shows that we do not observe any anomalous edges far beyond the timescale in which we explore the largest components of the out-forest. This allows us to conclude that all of the results we prove using the exploration up to time $O(n^{2/3})$ are also true conditioned on $\{\rho(n) = \infty\}$. This follows from the proof of Theorem 3.2 in [70]. □

The results that follow are all obtained by studying the exploration up to time $O(n^{2/3})$, so will also be true conditional on the resulting directed multigraph being simple.

3.5 Convergence of the SCCs under rescaling

In this section, we will use the convergence of the out-forest that we obtained in Section 3.4 to show that the SCCs ordered by decreasing number of edges converge under rescaling in the $d_{\vec{\mathcal{G}}}$-product topology.

3.5.1 Convergence of the out-components that contain an ancestral surplus edge

In this subsection, we will prove that the out-components that are explored up to time $O(n^{2/3})$ that contain an ancestral surplus edge converge under rescaling. Recall the definition of $(A_n(k), k \geq 1)$ from Subsection 3.2.1.3, and recall that the out-components that contain a non-trivial SCC are the out-components on which $(A_n(k), k \geq 1)$ increases. Moreover, if $(A_n(k), k \geq 1)$ increases on a component, the law of the first increase time corresponds to the position of the tail of the first ancestral surplus edge in the component.

We first study the convergence of $(\hat{H}_n^\ell(k), k \geq 1)$ under rescaling. This is an extension of Proposition 3.4.1. Recall that for $(B_t, t \geq 0)$ a standard Brownian motion, we defined

$$(\hat{B}_t, t \geq 0) = \left(B_t - \frac{\sigma_{-+} + \nu_-}{2\sigma_+\mu} t^2, t \geq 0\right),$$

and its reflected process

$$(\hat{R}_t, t \geq 0) = \left(\hat{B}_t - \inf\left\{\hat{B}_s : s \leq t\right\}, t \geq 0\right).$$

Proposition 3.5.1. *We have that*

$$\left(n^{-1/3}\hat{S}_n^+\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}\hat{H}_n\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}\hat{H}_n^\ell\left(\lfloor n^{2/3}t\rfloor\right), t \leq T\right) \xrightarrow{(d)} \left(\sigma_+\hat{B}_t, \frac{2}{\sigma_+}\hat{R}_t, \frac{2(\sigma_{-+}+\nu_-)}{\sigma_+\mu}\hat{R}_t, t \leq T\right)$$

in $\mathbb{D}([0,T],\mathbb{R})^3$, jointly with

$$\left(n^{-2/3}\hat{S}_n^-\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}\hat{P}_n\left(\lfloor n^{2/3}t\rfloor\right), t \leq T\right) \xrightarrow{p} \left(\nu_- t, \frac{\nu_-}{2\mu}t^2, t \leq T\right)$$

in $\mathbb{D}([0,T],\mathbb{R})^2$ as $n \to \infty$.

Proof. We use de Raphélis [43, Theorem 1], which states the convergence of the height process of a Bienaymé forest with edge-lengths under a few conditions on the degree and edge length distribution. We will apply this result to the Bienaymé forest with dummy and filler vertices, as defined in Subsection 3.4.2.2.

We equip this forest with edge lengths similarly to how we equipped the out-forest with edge-lengths when we described how to sample the candidates in Subsection 3.2.1.3. We do this as follows. For a dummy or filler vertex with out degree d^+, sample its in-degree with law Z^- for $\mathbf{Z}$ conditional on the event $\{Z^+ = d^+\}$. The in-degree of the true vertices is encoded by $(Y^-(k), k \geq 1)$. Then, for a vertex with in-degree d^-, let the edges connecting it to its children have length $d^- - 1$ (unless it is the root of the component, then let the edges connecting it to its children have length d^-). Let $(H^{\mathrm{df},\ell}(k), k \geq 1)$ be the height process of the resulting forest. We will translate the conditions of Theorem 1 in [43] to our setting and check them. The conditions are as follows.

1. $\mathbb{E}[Z^+] = 1$
2. $1 < \mathbb{E}[(Z^+)^2] < \infty$
3. $\mathbb{E}\left[Z^+ \mathbb{1}_{\{Z^- > x\}}\right] = o(x^{-2})$ as $x \to \infty$.

Under these conditions, using the notation from Subsection 3.4.2.2,

$$\left(n^{-1/3} Y^{\mathrm{df}}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{\mathrm{df}}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{\mathrm{df},\ell}\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right) \\ \xrightarrow{(\mathrm{d})} \left(\sigma_+ B_s, \frac{2}{\sigma_+} R_s, \frac{2(\sigma_{+-} + \nu_-)}{\mu \sigma_+} R_s, t \geq 0\right) \tag{3.23}$$

in $D(\mathbb{R}_+, \mathbb{R})^3$ as $n \to \infty$. Then, we observe that the rest of the argument in Subsections 3.4.2.2 and 3.4.3 can be extended to include the height process with edge lengths. This yields the result.

Therefore, to finish the proof, we need the conditions of Theorem 1 in [43] to hold. The conditions are equivalent to

1. $\mathbb{E}[D^+ D^-] = \mathbb{E}[D^-]$
2. $1 < \frac{\mathbb{E}[(D^+)^2 D^-]}{\mathbb{E}[D^-]} < \infty$
3. $\mathbb{E}\left[D^+ D^- \mathbb{1}_{D^- > x}\right] = o(x^{-2})$ as $x \to \infty$.

Note that the first and second conditions follow directly from the assumptions, and the third condition is implied by $\mathbb{E}[D^+ (D^-)^3] < \infty$. □

Proposition 3.5.2. *We have, jointly with the convergence in Proposition 3.5.1,*

$$\left(A_n\left(\lfloor tn^{2/3} \rfloor\right), t \leq T\right) \xrightarrow{(\mathrm{d})} \left(A_t, t \leq T\right),$$

as $n \to \infty$, where $(A_t, t \geq 0)$ is a Cox process of intensity

$$\frac{2(\sigma_{-+} + \nu_-)}{\sigma_+ \mu^2} \hat{R}_t$$

at time t. The convergence is in $D([0,T], \mathbb{R})$.

Proof. By definition, $(A_n(k), k \geq 1)$ is a counting process with compensator

$$\begin{aligned} A_n^{comp}(k) &= \sum_{i=1}^{k} \frac{\hat{H}^\ell(i)}{\hat{S}_n^-(i)} \mathbb{1}_{\{\hat{P}_n(i)-\hat{P}_n(i-1)=1\}} \\ &= \sum_{j=1}^{\hat{P}_n(k)} \frac{\hat{H}^\ell(\min\{l : \hat{P}_n(l) \geq k\})}{\hat{S}_n^-(\min\{l : \hat{P}_n(l) \geq k\})}. \end{aligned}$$

By Daley and Vere-Jones [41, Theorem 14.2.VIII], the claimed convergence under rescaling of $(A_n(k), k \geq 1)$ follows if we show that

$$\left(A_n^{comp}\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right) \xrightarrow{(d)} \left(\frac{2(\sigma_{-+}+\nu_-)}{\sigma_+\mu^2} \int_0^t \hat{R}_v dv, t \geq 0\right) \tag{3.24}$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in Proposition 3.5.1. Therefore, we will now prove that Eq. (3.24) holds. Since

$$\left(n^{-1/3}\hat{P}_n\left(\lfloor n^{2/3}t \rfloor\right), t \geq 0\right) \xrightarrow{p} \left(\frac{\nu_-}{2\mu}t^2, t \geq 0\right)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$, we get that

$$\begin{aligned} \left(n^{-2/3} \min\{l \geq 1 : n^{-1/3}\hat{P}_n(l) \geq t\}, t \geq 0\right) &\xrightarrow{p} \left(\min\left\{s > 0 : \frac{\nu_-}{2\mu}s^2 > t\right\}, t \geq 0\right) \\ &=: (\tau(t), t \geq 0) \end{aligned}$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$, because $\left(\frac{\nu_-}{2\mu}t^2, t \geq 0\right)$ is strictly increasing. Then, Proposition 3.5.1, Lemma 3.4.11, Slutsky's lemma and the continuous mapping theorem imply that

$$\left(\sum_{j=1}^{\lfloor n^{1/3}t \rfloor} \frac{\hat{H}^\ell(\min\{l : \hat{P}_n(l) \geq k\})}{\hat{S}_n^-(\min\{l : \hat{P}_n(l) \geq k\})}, t \geq 0\right) \xrightarrow{(d)} \left(\frac{2(\sigma_{-+}+\nu_-)}{\sigma_+\mu} \int_0^t \frac{\hat{R}_{\tau(s)}}{\nu_-\tau(s)} ds, t \geq 0\right)$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$. If we combine this with the convergence under rescaling of $(P_n(k), k \geq 1)$ from Lemma 3.4.6 and apply Lemma 3.4.11, some simple analysis then yields Eq. (3.24), which proves the statement. □

3.5.2 Finding the important components in the out-forest

In this subsection, we will show that, conditional on the convergence under rescaling in Proposition 3.5.2, the sequence of intervals that encode the trees with ancestral surplus edges sampled up to time $\lfloor Tn^{2/3}/2 \rfloor$ converges as well under rescaling. We want all of the trees that contain such an ancestral surplus edge to be fully explored by time $\lfloor Tn^{2/3} \rfloor$, so we let T be large enough so that this is likely. To be precise, fix $\epsilon > 0$ and, from now on, let T be large enough such that $\inf\{\hat{B}_t, t \leq T\} < \inf\{\hat{B}_t, t \leq T/2\}$ with probability at least $1 - \epsilon$.

Lemma 3.5.3 is a statement about extracting excursion intervals from deterministic functions with marks, which we will apply to the sample paths of $(\hat{S}_n^+(k), k \geq 1)$ with the increase times of $(A_n(k), k \geq 1)$ playing the rôle of the marks. The lemma tells us that if the sample paths and increase times converge under rescaling, then the beginnings and endpoints of the excursions above the running infimum that contain the increase times converge as well.

Let $(f_n(t), t \leq T)$ for $n \geq 1$, and $(f(t), t \leq T)$ be functions in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$, such that

$$(f_n(t), t \leq T) \to (f(t), t \leq T)$$

in $\mathbb{D}([0, T], \mathbb{R})$ as $n \to \infty$. Assume that $(f(t), t \leq T)$ is continuous and that the local minima of $(f(t), t \geq 0)$ are unique. Moreover, let $(x_i^n)_{1 \leq i \leq m}$, for $n \geq 1$, and $(x_i)_{1 \leq i \leq m}$ be elements of $[0, T]^m$ such that for all $i \in [m]$, $x_i^n \to x_i$ in $[0, T]$ as $n \to \infty$, and such that $f(x_i) - \inf\{f(s) : s \leq x_i\} > 0$ for all $i \in [m]$. Moreover, assume that $\inf\{f(t) : t \leq T\} < \inf\{f(t) : t \leq x_m\}$ and that $\inf\{f_n(t) : t \leq T\} < \inf\{f_n(t) : t \leq x_m^n\}$. For $i \in [m], n \geq 1$, let g_i^n be the left endpoint of the excursion of f_n above its running infimum that contains x_i^n, and let σ_i^n be the length of this excursion, i.e.

$$\begin{aligned} g_i^n &= \inf\{t \geq 0 : f_n(t) = \inf\{f_n(s) : s \leq x_i^n\}\}, \\ \sigma_i^n &= \inf\{t \geq 0 : \inf\{f_n(s) : s \leq g_i^n + t\} < \inf\{f_n(s) : s \leq x_i^n\}\}. \end{aligned}$$

Similarly, let g_i be the left endpoint of the excursion of f above its running infimum that contains x_i, and let σ_i be the length of this excursion, i.e.

$$\begin{aligned} g_i &= \inf\{t \geq 0 : f(t) = \inf\{f(s) : s \leq x_i\}\}, \\ \sigma_i &= \inf\{t \geq 0 : \inf\{f(s) : s \leq g_i + t\} < \inf\{f(s) : s \leq x_i\}\}. \end{aligned}$$

For $S = \{(a_i, b_i), i \in [m]\}$, let $\text{ord}(S)$ be a sequence consisting of the elements of S put in decreasing order of a_i, with ties broken arbitrarily, and concatenated with $(0, 0)_{i \geq 1}$ so that $\text{ord}(S) \in (\mathbb{R}^2)^\infty$.

Lemma 3.5.3. *We have that*

$$\operatorname{ord}\left(\{(g_i^n, \sigma_i^n) : 1 \le i \le m\}\right) \to \operatorname{ord}\left(\{(g_i, \sigma_i) : 1 \le i \le m\}\right)$$

in $(\mathbb{R}^2)^\infty$ *equipped with the product topology as* $n \to \infty$*.*

Note that if a given excursion of f above its running infimum contains multiple marks, only one instance of its left endpoint and excursion length will appear in $\operatorname{ord}\left(\{(g_i^n, \sigma_i^n) : 1 \le i \le m\}\right)$. Therefore, the number of non-zero entries of $\operatorname{ord}\left(\{(g_i^n, \sigma_i^n) : 1 \le i \le m\}\right)$ can vary as n varies, which is why we work in $(\mathbb{R}^2)^\infty$. This lemma is proved in Appendix 3.8.

We now apply this result to our process to extract the excursion intervals that contain the marks representing ancestral backedges that are sampled up to time $\lfloor Tn^{2/3}/2 \rfloor$. We recall the following definitions from Subsection 3.2.1.3. We have that G_i^n is the left endpoint of the excursion of $\hat{S}_n^+$ above its running infimum that encodes the out-component that contains the ith ancestral surplus edge, and Σ_i^n is the length of this excursion. Moreover, G_i and Σ_i are their continuous counterparts. Formally, for $i \in \left\{1, \ldots, A_n\left(\lfloor Tn^{2/3}/2 \rfloor\right)\right\}$,

$$\begin{aligned} G_i^n &= \min\left\{k \ge 1 : \hat{S}_n^+(k) = \min\{\hat{S}_n^{p,+}(l) : l \le X_i^n\}\right\} \text{ and} \\ \Sigma_i^n &= \min\left\{k \ge 1 : \min\left\{\hat{S}_n^{p,+}(l) : l \le G_i^n + k\right\} < \min\left\{\hat{S}_n^{p,+}(l) : l \le X_i^n\right\}\right\}, \end{aligned}$$

and for $i \in \{1, \ldots, A\left(T/2\right)\}$,

$$\begin{aligned} G_i &= \inf\left\{t \ge 0 : \sigma_+ \hat{B}_t = \inf\{\sigma_+ \hat{B}_s : s \le X_i\}\right\} \text{ and} \\ \Sigma_i &= \inf\left\{t \ge 0 : \inf\{\sigma_+ \hat{B}_s : s \le G_i + t\} < \inf\{\sigma_+ \hat{B}_s : s \le X_i\}\right\}. \end{aligned}$$

We recall that the function ord sorts a set of elements by decreasing second coordinate and appends an infinite sequence of zeroes; the formal definition was given before the statement of Lemma 3.5.3.

Proposition 3.5.4. *It holds that*

$$\operatorname{ord}\left(\left\{\left(n^{-2/3} G_i^n, n^{-2/3} \Sigma_i^n\right) : 1 \le i \le A_n\left(\lfloor Tn^{2/3}/2 \rfloor\right)\right\}\right) \xrightarrow{(\mathrm{d})} \operatorname{ord}\left(\{(G_i, \Sigma_i) : 1 \le i \le A(T/2)\}\right)$$

in the product topology on $(\mathbb{R}^2)^\infty$ *as* $n \to \infty$*, jointly with the convergence in Proposition 3.5.2.*

Proof. By Skorokhod's representation theorem, we may work on a probability space where the convergence in Proposition 3.5.2 holds almost surely. We only consider the event on which the convergence holds and $\inf\{\hat{B}_t, t \leq T\} < \inf\{\hat{B}_t, t \leq T/2\}$ holds and claim that we can apply Lemma 3.5.3 to the sample paths of $\left(n^{-1/3}\hat{S}_n^+\left(\lfloor n^{2/3}t\rfloor\right), t \leq T\right)$ with marks

$$\left(n^{-2/3}X_n^i\right)_{1\leq i\leq A_n\left(\lfloor Tn^{2/3}/2\rfloor\right)},$$

where we observe that by the convergence, for n large enough, also

$$\inf\left\{\hat{S}_n^+\left(\lfloor n^{2/3}t\rfloor\right), t \leq T\right\} < \inf\left\{\hat{S}_n^+\left(\lfloor n^{2/3}t\rfloor\right), t \leq T/2\right\}$$

holds. We check the conditions. Firstly, note that by $A_n\left(\lfloor Tn^{2/3}/2\rfloor\right) \to A\left(T/2\right)$ as $n \to \infty$, we can pick n large enough such that $A_n\left(\lfloor Tn^{2/3}/2\rfloor\right) = A\left(T/2\right)$. By the local absolute continuity of $(\hat{B}_t, t \geq 0)$ to a Brownian motion, its local minima are almost surely unique. Since

$$\left(A_n\left(\lfloor tn^{2/3}\rfloor\right), t \leq T/2\right) \overset{a.s.}{\to} \left(A\left(t\right), t \leq T/2\right)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$, we observe that for all $i \in \{1, \ldots, A(T/2)\}$, $n^{-2/3}X_i^n \to X_i$ almost surely in $\mathbb{R}$ as $n \to \infty$. The intensity of the Cox process $(A_t, t \geq 0)$ at time t is proportional to $\hat{R}_t$, so $\hat{R}_{X_i} > 0$ for all i almost surely. This allows us to apply Lemma 3.5.3, and the convergence follows. □

3.5.3 Convergence of the set of candidates

By Lemma 3.5.3, we know that the intervals that encode the out-components that contain an ancestral surplus edge converge under rescaling. This convergence holds jointly with the convergence under rescaling of the first time-step at which an ancestral surplus edge is found in each of these components. We will show that the positions of the other candidates in a component converge as well under rescaling. Recall the procedure to sample candidates that is described in Subsection 3.2.1.3.

We will now show convergence under rescaling of the sequence of candidates in a particular component of $(\hat{F}_n(k), k \geq 1)$.

By Skorokhod's representation theorem, we may work on a probability space where the convergence in Propositions 3.5.2 and 3.5.4 holds almost surely. Let $(g, \sigma) \in \{(G_i, \Sigma_i) : i \leq A(T/2)\}$, so that, for each n large enough, we can find $(g_n, \sigma_n) \in \left\{(G_i^n, \Sigma_i^n) : i \leq A_n\left(\lfloor Tn^{2/3}/2 \rfloor\right)\right\}$ such that $(g_n, \sigma_n) \to (g, \sigma)$. Set $V_1 = \inf\{t \in [g, g+\sigma] : A(t) = A(g) + 1\}$, and similarly, set $V_1^n = \min\{g_n < k \leq g_n + \sigma_n : A_n(k) = A_n(g_n) + 1\}$, which are well-defined by definition of g, σ, g_n and σ_n. By construction, $\{g_n + 1, \ldots, g_n + \sigma_n\}$ encodes an out-component. Call this component $T_{g_n}^n$. We apply the procedure defined in Proposition 3.2.6 to find the candidates in $T_{g_n}^n$. Let $\mathbf{V}_n(g_n)$ denote the sequence of candidates in $T_{g_n}^n$. Similarly, $[g, g+\sigma]$ encodes a component of the out-$\mathbb{R}$-forest. Call this component $\mathcal{T}_g$, and apply the procedure in Subsection 3.2.2.2 to find the candidates in $\mathcal{T}_g$. Denote the sequence of candidates by $\mathbf{V}(g)$.

Proposition 3.5.5. *Jointly with the convergence in Proposition 3.5.4,*

$$n^{-2/3}\mathbf{V}_n(g_n) \xrightarrow{(\mathrm{d})} \mathbf{V}(g)$$

in the product topology.

Proof. We will find a coupling such that $n^{-2/3}\mathbf{V}_n(g_n) \overset{a.s.}{\to} \mathbf{V}(g)$. By the convergence in Propositions 3.5.2 and 3.5.4, $n^{-2/3}V_1^n \overset{a.s.}{\to} V_1$. In general, let V_m^n denote the m^{th} candidate that is found in $T_{g_n}^n$, and let V_m denote the m^{th} candidate that is found in $\mathcal{T}_g$. Suppose that, for some m, we have found a coupling such that

$$n^{-2/3}(V_1^n, \ldots, V_m^n) \overset{a.s.}{\to} (V_1, \ldots, V_m). \tag{3.25}$$

Then, V_{m+1}^n is distributed as the position of the first jump of a counting process $K_{m+1}^n(k)$ on $[0, \infty)$ with compensator

$$K_{comp,m+1}^n(k) = \sum_{i=V_m^n+1}^{k} \frac{\ell\left(T_i^{n,\mathrm{mk}}\right) - m}{\hat{S}^-(i)} \mathbb{1}\{P_n(i) = P_n(i-1) + 1\}$$

for $k \in [V_m^n + 1, g_n + \sigma_n]$ and 0 otherwise, where $T_i^{n,\mathrm{mk}}$ is the subtree of $T_{g_n}^n$ spanned by $\{g_n + 1, V_1^n, \ldots, V_m^n, i\}$. Moreover, for T_s the subtree of $\mathcal{T}_g$ spanned by $\{g, V_1, \ldots, V_m, s\}$, and $|T_s|$ its length as encoded by $\left(\frac{2}{\sigma_+}\hat{R}_t, t \geq 0\right)$, V_{m+1} is the first jump in a counting process $K_{m+1}(t)$ on $[0, \infty)$ with compensator

$$K_{comp,m+1}(t) = \int_{V_m}^{t} \frac{\sigma_{-+} + \nu_-}{\mu^2} |T_s| ds$$

for $t \in [V_m, g+\sigma]$ and 0 otherwise. By the convergence under rescaling of $(\hat{H}^{\ell}_n(k), k \geq 1)$ in Proposition 3.5.1, and by Proposition 3.5.4, we get that the metric structure of $T^n_{g_n}$ with distances defined by $(\hat{H}^{\ell}_n(k), k \geq 1)$, and its projection onto $[n^{-2/3}(g_n+1), n^{-2/3}(g_n+\sigma_n)]$, converge under rescaling to the metric structure of $\mathcal{T}_g$ with distances defined by

$$\left(\frac{2(\sigma_{-+}+\nu_-)}{\sigma_+\mu}\hat{R}_t, t \geq 0\right)$$

and its projection onto $[g, g+\sigma]$. This, combined with Eq. (3.25) implies that

$$\left(n^{-1/3}\ell\left(T^{n,\mathrm{mk}}_{\lfloor tn^{2/3}\rfloor}\right), V_m \leq t \leq g+\sigma\right) \overset{a.s.}{\to} \left(\frac{\sigma_{-+}+\nu_-}{\mu^2}|T^{\mathrm{mk}}_t|, V_m \leq t \leq l+\sigma\right)$$

in $\mathbb{D}([V_m, g+\sigma], \mathbb{R}_+)$ as $n \to \infty$. Then, a similar argument to that used in the proof of Proposition 3.5.2 implies that

$$\left(K^n_{comp,m+1}\left(\lfloor tn^{2/3}\rfloor\right), V_m \leq t \leq g+\sigma\right) \overset{a.s.}{\to} \left(K_{comp,m+1}(t), V_m \leq t \leq g+\sigma\right),$$

$\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ as $n \to \infty$. This implies that

$$(K^n_{m+1}(\lfloor tn^{2/3}\rfloor), t \geq 0) \xrightarrow{(\mathrm{d})} (K_{m+1}(t), t \geq 0)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ as $n \to \infty$ and, in particular, we can find a coupling such that $K_m(\infty) > 0$ if and only if $K^n_m(\infty) > 0$ for all n large enough, and such that on this event,

$$n^{-2/3}V^n_{m+1} \overset{a.s.}{\to} V_{m+1}.$$

If $K_m(\infty) = 0$, set $\mathbf{V}(g) = (V_1, \dots, V_m)$, $\mathbf{V}_n(g_n) = (V^n_1, \dots, V^n_m)$, and the statement follows. If $K_m(\infty) > 0$, apply the induction step to $(V_1, \dots, V_{m+1})$ and $(V^n_1, \dots, V^n_{m+1})$. The fact that $|\mathbf{V}(g)| < \infty$ almost surely, as shown in Subsection 3.2.2.2, implies that the induction terminates. □

The following proposition shows that also the law of the heads of the surplus edges corresponding to a candidate converges under rescaling. Moreover, we show convergence under rescaling in the pointed Gromov-Hausdorff topology of an out-component with the location of the candidates and the heads of their corresponding surplus edges.

Proposition 3.5.6. *Suppose the convergence in Propositions 3.5.2, 3.5.4 and 3.5.5 holds almost surely. Then, for* $\mathbf{V}_n(g_n) = (V_1^n, \ldots, V_{N_n}^n)$*,* $\mathbf{V}(g) = (V_1, \ldots, V_N)$*, let* W_i^n *be the index of the vertex that the surplus edge corresponding to* V_i^n *connects to. Similarly, let* W_i *be the index of the vertex that the surplus edge corresponding to* V_i *connects to. Then,*

$$\begin{aligned}&\left(n^{-1/3}T_{g_n}^n, n^{-2/3}(g_n+1), \left(n^{-2/3}V_1^n, n^{-2/3}W_1^n\right) \ldots, \left(n^{-2/3}V_{N_n}^n, n^{-2/3}W_{N_n}^n\right)\right)\\ &\xrightarrow{\text{(d)}} (\mathcal{T}_g, l, (V_1, W_1), \ldots, (V_N, W_N))\end{aligned}$$

in the $(2N+1)$*-pointed Gromov-Hausdorff topology.*

Proof. For S a subset of the vertices of $T_{g_n}^n$, let $T_{g_n}^n(S)$ denote the subtree of $T_{g_n}^n$ spanned by S. By definition, for $m \leq N_n$, W_m^n is the vertex corresponding to a uniform unpaired in-half-edge of the vertices in $T_{g_n}^n(\{g_n+1, V_1^n, \ldots, V_m^n\})$. By Proposition 3.5.1 and Slutsky's lemma,

$$\left(\frac{\hat{H}_n^\ell\left(\lfloor tn^{2/3}\rfloor\right)}{\hat{H}_n\left(\lfloor tn^{2/3}\rfloor\right)}, t \geq 0\right) \overset{a.s.}{\to} \left(\frac{\sigma_{-+}+\nu_-}{2\mu}, t \geq 0\right)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$, which implies that the law of W_m^n converges to the law of a uniform vertex in $T_{g_n}^n(\{g_n+1, V_1^n, \ldots, V_m^n\})$. Note that, by Theorem 3.4.1, Propositions 3.5.4 and 3.5.5, we know that the height process of $T_{g_n}^n$ converges under rescaling to the height process of $\mathcal{T}_g$, jointly with the convergence under rescaling of the positions of the candidates. By the proof of Proposition 5.4 in [59], this implies that

$$\left(n^{-1/3}T_{g_n}^n, n^{-2/3}g_n+1, n^{-2/3}V_1^n, \ldots, n^{-2/3}V_m^n\right) \overset{a.s.}{\to} (\mathcal{T}_g, g, V_1, \ldots, V_m)$$

in the $(m+1)$-pointed Gromov-Hausdorff topology. Since the relation

$$\left|T_{g_n}^n\left(\{g_n+1, V_1^n, \ldots, V_m^n\}\right)\right| = \left|T_{g_n}^n\left(\{g_n+1, V_1^n, \ldots, V_m^n, W_m^n\}\right)\right|$$

passes to the limit, with $|\cdot|$ denoting the length in the tree as encoded by $(\hat{H}_n(k), k \geq 1)$, the limit in distribution of $n^{-2/3}W_m^n$ is a uniform point on the subtree of $\mathcal{T}_g$ spanned by $(g, V_1, \ldots, V_m)$, which is equal to the law of W_m. This proves the statement. □

The proofs of Propositions 3.5.5 and 3.5.6 imply the following proposition.

Proposition 3.5.7. *By Skorokhod's representation theorem, we may work on a probability space* *where the convergence in Propositions 3.5.5 and 3.5.6 holds almost surely. Let* $T^{n,mk}(g_n)$ *be the subtree of* $T^n_{g_n}$ *spanned by* $\{g_n+1, V^n_1, \dots, V^n_{N_n}\}$*, and similarly, let* $T^{mk}(g)$ *be the subtree of* $\mathcal{T}_g$ *spanned by* $\{g, V_1, \dots, V_N\}$*. Then, also*

$$\begin{aligned}&\left(n^{-1/3}T^{n,mk}(g_n), n^{-2/3}(g_n+1), \left(n^{-2/3}V^n_1, n^{-2/3}W^n_1\right)\dots, \left(n^{-2/3}V^n_{N_n}, n^{-2/3}W^n_{N_n}\right)\right)\\ &\to \left(T^{mk}(g), g, (V_1, W_1), \dots, (V_N, W_N)\right)\end{aligned}$$

almost surely in the $(2N+1)$*-pointed Gromov-Hausdorff topology as* $n\to\infty$*. Also the total length in the trees converges, i.e.*

$$n^{-1/3}\left|T^{n,mk}(g_n)\right| \to \left|T^{mk}(g)\right|$$

almost surely as $n\to\infty$*.*

We now identify the candidates, as described in Subsection 3.2.1.3. In $T^{n,\mathrm{mk}}(g_n)$, set $V^n_i \sim W^n_i$ for each $1 \le i \le N_n$, and set $M^n_{g_n} := T^{n,\mathrm{mk}}(g_n)/\sim$. Moreover, in $T^{\mathrm{mk}}(g)$, set $V_i \sim W_i$ for each $1 \le i \le N$, and set $\mathcal{M}_g := T^{\mathrm{mk}}(g)/\sim$. View both as elements of $\vec{\mathcal{G}}$ in the natural way. To be precise, in $M^n_{g_n}$, let the vertex set consist of g_n+1, W^n_i for $i \le N_n$, and the branch points $V^n_i \wedge V^n_j$ for $i \ne j \le N_n$. Similarly, in $\mathcal{M}_g$, let the vertex set consist of g, W_i for $i \le N$, and the branch points $V_i \wedge V_j$ for $i \ne j \le N$. Then we have the following proposition.

Proposition 3.5.8. *On the probability space where the convergence in Propositions 3.5.5 and 3.5.6 holds almost surely,* $n^{-1/3}M^n_{g_n} \xrightarrow{a.s.} \mathcal{M}_g$ *in* $\vec{\mathcal{G}}$*.*

Proof. The proof is analogous to the proof of Proposition 5.6 in [59]. □

Proposition 3.5.9. *On the probability space where the convergence in Propositions 3.5.5 and 3.5.6 holds almost surely, the SCCs in* $n^{-1/3}M^n_{g_n}$*, listed in decreasing order of length, converge to the SCCs in* $\mathcal{M}_g$*, listed in decreasing order of length, in* $\vec{\mathcal{G}}$ *almost surely as* $n\to\infty$*.*

Proof. This follows from Proposition 5.3 in [59]. This proposition requires that the lengths of the SCCs in $\mathcal{M}_g$ have different lengths almost surely, which is the content of Proposition 3.2.10. □

Proposition 3.5.10. *Let* $T>0$*, and let* $(C^T_i(n), i\ge 1)$ *be the kernels of the SCCs that contain a candidate with label at most* $\lfloor Tn^{2/3}/2\rfloor$*, ordered by length. Similarly, let* $(\mathcal{C}^T_i, i \ge 1)$ *be the* *kernels of the SCCs obtained from the out-$\mathbb{R}$-forest with a candidate before time $T/2$, ordered by length. Then,*

$$\left(n^{-1/3}C_i^T(n), i \geq 1\right) \xrightarrow{(\mathrm{d})} (\mathcal{C}_i^T, i \geq 1)$$

in the $\vec{\mathcal{G}}$-product topology, as $n \to \infty$.

Proof. This follows from Proposition 3.5.4, Proposition 3.5.9, and the fact that all SCCs in the limit object have a different length by Proposition 3.2.10. □

Finally, we claim that we can choose T large enough such that the SCCs with the highest number of edges are explored before time $\lfloor Tn^{2/3} \rfloor$. This is the content of the following lemma. The proof is in the same spirit as Aldous [12, Lemma 9].

Lemma 3.5.11. *For $\delta > 0$ and I an interval, let $SCC(n, I, \delta)$ denote the number of SCCs whose vertices have at total of at least $\delta n^{1/3}$ in-edges (including those which are not part of the SCC) and whose time of first discovery is in $n^{2/3}I$. Then,*

$$\lim_{s\to\infty} \limsup_{n} \mathbb{P}\left(SCC(n, (s, \infty), \delta) \geq 1\right) = 0 \textit{ for all } \delta > 0.$$

Proof. Fix $\delta > 0$. Suppose there is an SCC C with $vn^{1/3}$ total in-edges. Conditionally on this fact, the in-edges that are paired up until the time the first in-edge of C is paired are uniform picks (without replacement) from the total set of in-edges. We use Ξ_n to denote the time of discovery of the first in-edge of C multiplied by $n^{-2/3}$. Then, $\Xi_n \xrightarrow{(\mathrm{d})} \mathrm{Exp}(v)$. Fix $\epsilon > 0$. We have that, by the memoryless property at time s,

$$\mathbb{P}\left(SCC\left(n, (s, 2s), \delta\right) = 0 | SCC\left(n, (s, \infty), \delta\right) \geq 1\right)$$

is asymptotically bounded from above by $\exp(-s\delta)$ by the memoryless property at time s. So that we can find an $s > 0$ such that for all n large enough,

$$\mathbb{P}\left(SCC\left(n, (s, \infty), \delta\right) \geq 1 \text{ and } SCC\left(n, (s, 2s), \delta\right) = 0\right) < \epsilon.$$

We claim that, by possibly increasing s and n, we also get that

$$\mathbb{P}\left(SCC\left(n, (s, 2s), \delta\right) = 0\right) > 1 - \epsilon,$$

which proves the statement. Firstly, we observe that the ratio of the length of an SCC and its total in-degree are asymptotically equal to $\frac{\sigma_{-+}+\nu_{-}}{2\mu}$ by the proof of Proposition 3.5.6. Then, note that it is clear from the description of the limit process that, for s large enough, with probability at most $\epsilon/2$, an SCC with total length at least $\frac{\mu}{\sigma_{-+}+\nu_{-}}\delta$ is discovered after time s. By the convergence of the exploration process on compact time intervals, by choosing n large enough, we can then ensure that

$$\mathbb{P}\left(SCC\left(n,(s,2s),\delta\right)=0\right)>1-\epsilon.$$

We conclude that

$$\mathbb{P}(SCC\left(n,(s,\infty),\delta\right)\geq 1)\leq 2\epsilon. \qquad \square$$

Note that the number of edges in an SCC is bounded from below by the total number of in-edges of vertices in the SCC.

We now show that for any j and any $\epsilon>0$, we can pick T large enough such the j largest components in $(\mathcal{C}_i, i\geq 1)$ are contained in $(\mathcal{C}_i^T, i\geq 1)$ with probability at least $1-\epsilon$.

Lemma 3.5.12. *For all j holds that*

$$\lim_{T\to\infty}\mathbb{P}\left(\forall i\leq j, \mathcal{C}_i\in(\mathcal{C}_i^T, i\geq 1)\right)=1.$$

Proof. Fix $\epsilon>0$. By [59, Proposition 5.10] adapted to our limit object, for k large enough, with probability $1-\epsilon/2$, the j largest components of $(\mathcal{C}_i, i\geq 1)$ are contained in the k largest components of the out-forest with identifications. Moreover, for T large enough, with probability $1-\epsilon/2$, the k largest excursions above the infimum of a Brownian motion with negative parabolic drift occur before time T (see [15, Section 3]). This implies the statement. $\square$

Theorem 3.1.1 then follows from Proposition 3.5.10, Lemma 3.5.11 and Lemma 3.5.12.

3.6 Open problems

Our work contains the first quantitative results on the directed configuration model at criticality, and is the second metric space convergence result for a directed graph model (after the directed Erdős-Rényi graph was studied in [59]), and many interesting unresolved questions remain.

1. The law of our limit object is defined by three parameters that are functions of the (mixed) moments of the degree distribution. Does a different choice of parameters always give a different limit distribution? If so, are the laws absolutely continuous to one another?

2. Our methods show that the diameter of the configuration model at criticality is $\Omega(n^{1/3})$ in probability, which is in contrast with the off-critical cases (for deterministic degrees), in which the diameter is $\Theta(\log(n))$ in probability [30]. We conjecture that the diameter is in fact $\Theta(n^{1/3})$ in probability. Goldschmidt and Maazoun are working on this question for the directed Erdős-Rényi graph at criticality.

3. In [59], the authors show convergence of the sequence of SCCs in the ℓ_1-sense, which is stronger than the product topology as considered by us. This for example implies that for the directed Erdős-Rényi graph, under rescaling, the total length in the SCCs converges in distribution to some finite random variable. Also for undirected configuration models, there are no results that show metric space convergence in a topology on the sequence of components that is stronger than the product topology [24, 37, 23].

4. We conjecture that, just like the directed Erdős-Rényi graph [59], the directed configuration model gives rise to a critical window, that in some sense interpolates between subcritical and supercritical models. It would be interesting to adapt our methods to the critical window.

5. In future work, we plan to extend our understanding of the SCCs by studying the directed graphs in which they are embedded. A first step would be to study all vertices that can be reached from the non-trivial strongly components. This would illuminate connections between the SCCs and expose the fractal structure of the directed graph, which is not observed when only studying the SCCs themselves.

6. Another natural next step is to study the model under weaker moment conditions. The first condition to eliminate would be $\mathbb{E}\left[(D^-)^i(D^+)^j\right] < \infty$ for $(i,j) = (1,3)$ and $(i,j) = (3,1)$. Removing the former condition would in some sense make the identifications less uniform on the ancestral lines. To be precise, $(\hat{H}^\ell(k)/\hat{H}(k), k \geq 0)$ will not necessarily converge to a constant process under rescaling of time, which means that the in-edges that can be used to form surplus edges are spread out less uniformly on the out-components. We have reason to believe that this would place the model in a different universality class,

but further research is needed to confirm this. Removing the latter condition requires an adaptation of the proof of Proposition 3.4.17 that does not use the Cauchy-Schwarz inequality. Also the heavy-tailed case is not well-understood, but given our results, it is natural to expect that a potential limit object would be embedded in a tilted stable tree as defined in [37]. Moreover, one could define hybrid models by letting the tail-behaviour of the in- and out-degrees be different.

7. We conjecture that the rank-1 inhomogeneous directed random graph model under suitable conditions is part of the same universality class as the directed Erdős-Rényi graph [59] and the model we consider in this work. We believe that our methods and the methods of [59] can be adapted to obtain a metric space scaling limit for the inhomogeneous directed random graph model, and we intend to pursue this in future work.

3.7 Multivariate triangular local limit theorem

The goal of this section is to prove Theorem 3.3.16. This can be deduced from Mukhin [87, Corollary 1]. However, Mukhin's result is more general than is needed to prove Theorem 3.3.16. As a result, the conditions which we need to check in order to apply Mukhin's result are rather complicated. Instead, we offer here an elementary proof.

First, we recall some definitions. An $\mathbb{R}^d$-valued random variable $\mathbf{X}$ is lattice if it is non-degenerate and is supported on the translation of some lattice. The symmetrisation of $\mathbf{X}$ is given by $\mathbf{X}^* = \mathbf{X}_1 - \mathbf{X}_2$ where $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent copies of $\mathbf{X}$. If $\mathbf{X}$ is lattice, the main lattice of $\mathbf{X}$ is given by

$$\Lambda = \bigcup_{m=1}^{\infty} \left\{\textstyle\sum_{i=1}^m n_i \mathbf{x}_i^* : n_i \in \mathbb{Z} \text{ and } \mathbf{x}_i^* \in \operatorname{supp}(\mathbf{X}^*) \text{ for all } i = 1, \ldots, m\right\}.$$

Now we restate Theorem 3.3.16.

Theorem 3.3.16. *For each $n \geq 1$ let $\mathbf{X}_n$ be an $\mathbb{R}^d$ valued random variable and*

$$\mathbf{X}_{n,1}, \mathbf{X}_{n,2}, \ldots, \mathbf{X}_{n,n}$$

be i.i.d. copies of $\mathbf{X}_n$. Assume that the following holds:

1. *There exists a random variable $\mathbf{X}$ such that $\mathbf{X}_n \xrightarrow{\text{(d)}} \mathbf{X}$ as $n \to \infty$.*
2. *$(\|\mathbf{X}_n\|^2)_{n\geq 1}$ is a uniformly integrable sequence of random variables. Explicitly*

$$\lim_{L\to\infty} \sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\left\{\|\mathbf{X}_n\|^2 > L\right\}\right] = 0. \tag{3.10}$$

3. *For all n, $\mathbf{X}_n$ and $\mathbf{X}$ are lattice with common main lattice Λ.*

Then $\mathbf{X}$ has finite second moment. Further, for each n let $\mathbf{c}_n$ be an arbitrary element in the support of $\sum_{i=1}^n \mathbf{X}_{n,i}$. Then uniformly for $\mathbf{y} \in \mathbf{c}_n + \Lambda$,

$$\mathbb{P}\Big(\textstyle\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\Big) = n^{-d/2} \det(\Lambda) f\left(\mathbf{x}_n(\mathbf{y})\right) + o\left(n^{-d/2}\right) \quad \textit{where} \quad \mathbf{x}_n(\mathbf{y}) = \tfrac{\mathbf{y} - n\mathbb{E}[\mathbf{X}_n]}{\sqrt{n}}$$

and f is the density of a $N(0, \operatorname{Cov}(\mathbf{X}))$ distribution. This means that

$$\lim_{n\to\infty} \sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda} \left| n^{d/2}\mathbb{P}\Big(\textstyle\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\Big) - \det(\Lambda) f(\mathbf{x}_n(\mathbf{y}))\right| = 0.$$

Before we prove Theorem 3.3.16, we first prove a sequence of lemmas. Our proof of the local limit theorem will use characteristic functions. Let $\mathbf{X}$ be $\mathbb{R}^d$-valued. We use the convention that the characteristic function of $\mathbf{X}$ is given by

$$\phi(\mathbf{u}) = \mathbb{E}\left[e^{i\mathbf{u}\cdot\mathbf{X}}\right].$$

The following lemma shows the points at which the characteristic function of a lattice random variables attains 1 in absolute value can be precisely characterised when the main lattice is known. This is an adaptation of [96, P.67, T1].

Lemma 3.7.1. *Suppose* $\mathbf{X}$ *is lattice with main lattice* $\mathbb{Z}^d$ *and characteristic function* ϕ*. Then* $|\phi(\mathbf{u})| = 1$ *if and only if* $\mathbf{u} \in (2\pi\mathbb{Z})^d$*.*

Proof. If every coordinate of $\mathbf{u}$ is a multiple of 2π, then $\mathbf{u}\cdot\mathbf{X}$ has support in $t + 2\pi\mathbb{Z}$ for some $t \in \mathbb{R}$. Therefore $e^{i\mathbf{u}\cdot\mathbf{X}}$ is constant and hence $|\phi(\mathbf{u})| = 1$.

For the converse, note the characteristic function of the symmetrisation $\mathbf{X}^*$ satisfies

$$\mathbb{E}\left[e^{i\mathbf{u}\cdot\mathbf{X}^*}\right] = \mathbb{E}\left[e^{i\mathbf{u}\mathbf{X}_1}\right]\mathbb{E}\left[e^{-i\mathbf{u}\mathbf{X}_2}\right] = \left|\mathbb{E}\left[e^{i\mathbf{u}\mathbf{X}}\right]\right|^2 = 1.$$

Thus $e^{i\mathbf{u}\cdot\mathbf{x}^*} \in 2\pi\mathbb{Z}$ for all $\mathbf{x}^*$ in the support of $\mathbf{X}^*$. Since the fundamental lattice of $\mathbf{X}$ is $\mathbb{Z}^d$, there exists $\mathbf{x}_1^*, \dots, \mathbf{x}_m^*$ in the support of $\mathbf{X}^*$ and $k_1, \dots, k_m \in \mathbb{Z}$ such that

$$\sum_{i=1}^{m} k_i \mathbf{x}_i^* = (1, 0, \dots, 0).$$

Therefore,

$$u^{(1)} = \sum_{i=1}^{m} k_i \mathbf{u}\cdot\mathbf{x}_i^* \in 2\pi\mathbb{Z}.$$

Repeating this argument for the other coordinates of $\mathbf{u}$ shows all coordinates of $\mathbf{u}$ are multiples of 2π. □

The next lemma shows convergence of the means and covariance of $\mathbf{X}_n$ to that of $\mathbf{X}$, and moreover shows the uniform integrability condition still holds after centering the random variables.

Lemma 3.7.2. *Suppose conditions (1) and (2) of Theorem 3.3.16 hold. Then, as* $n \to \infty$*,*

$$\mathbb{E}[\mathbf{X}_n] \to \mathbb{E}[\mathbf{X}] \quad \textit{and} \quad \mathrm{Cov}(\mathbf{X}_n) \to \mathrm{Cov}(\mathbf{X}).$$

Further for each n, let $\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbb{E}[\mathbf{X}_n]$, and $\hat{\mathbf{X}} = \mathbf{X} - \mathbb{E}[\mathbf{X}]$. Then the uniform integrability condition in Eq. (3.10) holds for the centered random variables $(\hat{\mathbf{X}}_n)_{n\geq 1}$. This means that

$$\lim_{L\to\infty} \sup_n \mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2 \mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2 > L\right\}\right] = 0.$$

Proof. By Skorokhod's representation theorem, we can assume without loss of generality that $(\mathbf{X}_n)_{n\geq 1}$ and $\mathbf{X}$ are in the same probability space and $\mathbf{X}_n \to \mathbf{X}$ almost surely as $n \to \infty$. Then, the condition in Eq. (3.10) gives uniform integrability of $(\|\mathbf{X}_n\|_2^2)_{n\geq 1}$. Thus, by Vitali's convergence theorem, $\mathbf{X}_n \to \mathbf{X}$ in L^2 as $n \to \infty$. Therefore, $\mathbf{X}$ has finite second moment and the mean and covariance of $\mathbf{X}_n$ converge to that of $\mathbf{X}$.

Since the means converge, the centerings $\hat{\mathbf{X}}_n \to \hat{\mathbf{X}}$ in L^2 as $n \to \infty$ also. Thus, $(\|\hat{\mathbf{X}}_n\|_2^2)_{n\geq 1}$ is uniformly integrable by the converse statement in Vitali's theorem, as required. □

The following lemma shows that we have a normal central limit theorem.

Lemma 3.7.3. *Suppose we are in the setting of Theorem 3.3.16. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}_{n,i} - \mathbb{E}[\mathbf{X}_n]) \xrightarrow{(\mathrm{d})} N(0, \Sigma)$$

as $n \to \infty$.

Proof. We use the Lindeberg-Feller central limit theorem. We will use the notation $\Sigma = \mathrm{Cov}(\mathbf{X})$, $\Sigma_n = \mathrm{Cov}(\mathbf{X}_n)$, $\hat{\mathbf{X}}_{n,i} = \mathbf{X}_{n,i} - \mathbb{E}[\mathbf{X}_n]$ and $\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbb{E}[\mathbf{X}_n]$. We will reduce the problem to the one-dimensional case. By the Cramér–Wold device it is sufficient to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{u} \cdot \hat{\mathbf{X}}_{n,i} \xrightarrow{(\mathrm{d})} N(0, \mathbf{u} \cdot \Sigma \mathbf{u})$$

for all $\mathbf{u} \in \mathbb{R}^d$. Define

$$A_{n,i} = \frac{1}{\sqrt{n}} \mathbf{u} \cdot \hat{\mathbf{X}}_{n,i}.$$

Then by the version of the Lindeberg–Feller central limit theorem stated by Durrett in [52, P.128-129, Theorem 3.4.10], to complete the proof it suffices to check that

1. $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[A_{n,i}^2] = \mathbf{u} \cdot \Sigma \mathbf{u}$.

2. For all $\epsilon > 0$, $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left[A_{n,i}^2 \mathbb{1}\left\{|A_{n,i}| > \epsilon\right\}\right] = 0$.

To check condition (1),

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}[A_{n,i}^2]=\lim_{n\to\infty}\mathbb{E}\Big[(\mathbf{u}\cdot\hat{\mathbf{X}}_n)^2\Big]=\lim_{n\to\infty}\mathbf{u}\cdot\Sigma_n\mathbf{u}=\mathbf{u}\cdot\Sigma\mathbf{u}$$

by Lemma 3.7.2. To check condition (2), for all $\epsilon>0$

$$\begin{aligned}\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\Big[A_{n,i}^2\mathbb{1}\left\{|A_{n,i}|>\epsilon\right\}\Big]&=\lim_{n\to\infty}\mathbb{E}\Big[(\mathbf{u}\cdot\hat{\mathbf{X}}_n)^2\mathbb{1}\Big\{(\mathbf{u}\cdot\hat{\mathbf{X}}_n)^2>\epsilon^2 n\Big\}\Big]\\&\le\|\mathbf{u}\|^2\lim_{n\to\infty}\mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2\mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2>\frac{\epsilon^2}{\|\mathbf{u}\|^2}n\right\}\right]\\&\le\|\mathbf{u}\|^2\lim_{n\to\infty}\sup_k\mathbb{E}\left[\|\hat{\mathbf{X}}_k\|^2\mathbb{1}\left\{\|\hat{\mathbf{X}}_k\|^2>\frac{\epsilon^2}{\|\mathbf{u}\|^2}n\right\}\right]\\&=0\end{aligned}$$

by Lemma 3.7.2. □

The last lemma we prove provides bounds on the absolute value of the characteristic functions of $\mathbf{X}_n$. This will be used to apply the dominated convergence theorem in the main proof.

Lemma 3.7.4. *Suppose we are in the setting of Theorem 3.3.16. Moreover assume that the common main lattice* Λ *is* $\mathbb{Z}^d$*. Let* $\phi_n(\mathbf{u})$ *be the characteristic function of* $\hat{\mathbf{X}_n}=\mathbf{X}_n-\mathbb{E}[\mathbf{X}_n]$*. Then there exist* $\delta,c>0$*,* $\rho\in(0,1)$ *and* N *such that for all* $n\ge N$

1. $|\phi_n(\mathbf{u})|\le 1-c\|\mathbf{u}\|^2$ *for all* $\mathbf{u}\in S(\delta)$*, and*

2. $|\phi_n(\mathbf{u})|\le\rho$ *for all* $\mathbf{u}\in S(\pi)\setminus S(\delta)$

where, for all $r>0$*,* $S(r)=[-r,r]^d$*.*

Proof. Firstly we use a analytical lemma stated by Durrett in [52, P.116, Lemma 3.3.19]. By that lemma, there exists a constant $A>0$ such that

$$\left|e^{ix}-\left(1+ix-\tfrac{1}{2}x^2\right)\right|\le A\min\{|x|,1\}x^2$$

for all $x\in\mathbb{R}$. Then applying this with $x=\mathbf{u}\cdot(\mathbf{X}_n-\mathbb{E}[\mathbf{X}_n])$

$$|\phi_n(\mathbf{u})|\le\left|1-\tfrac{1}{2}\mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u}\right|+R_n(\mathbf{u})$$

where

$$R_n(\mathbf{u}) \leq A\mathbb{E}\left[\min\{|\mathbf{u}\cdot\hat{\mathbf{X}}_n|, 1\}(\mathbf{u}\cdot\hat{\mathbf{X}}_n)^2\right].$$

We provide bounds on R_n and $|1-\frac{1}{2}\mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u}|$, starting with $|1-\frac{1}{2}\mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u}|$.

Let $\lambda_{\mathrm{min}n}$ and $\lambda_{\mathrm{max}n}$ be the minimum and maximum eigenvalues of $\mathrm{Cov}(\mathbf{X}_n)$ respectively. Then, by standard theory for quadratic forms,

$$\lambda_{\mathrm{min}n}\|\mathbf{u}\|^2 \leq \mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u} \leq \lambda_{\mathrm{max}n}\|\mathbf{u}\|^2.$$

Moreover, let λ_{min} and λ_{max} be the minimum and maximum eigenvalues of $\mathrm{Cov}(\mathbf{X})$ respectively. The eigenvalues of a matrix are continuous in its entries and $\mathrm{Cov}(\mathbf{X}_n)\to\mathrm{Cov}(\mathbf{X})$ by Lemma 3.7.2. Therefore $\lambda_{\mathrm{min}n}\to\lambda_{\mathrm{min}}$ and $\lambda_{\mathrm{max}n}\to\lambda_{\mathrm{max}}$ as $n\to\infty$.

We have assumed that $\mathrm{Cov}(\mathbf{X})$ is non-degenerate thus $\lambda_{\mathrm{min}}>0$. Hence, there exists N such that for all $n\geq N$,

$$\frac{1}{2}\lambda_{\mathrm{min}} \leq \lambda_{\mathrm{min}n} \leq \lambda_{\mathrm{max}n} \leq 2\lambda_{\mathrm{max}}.$$

There also exists $\delta_1>0$ sufficiently small that $\lambda_{\mathrm{max}}\|\mathbf{u}\|^2<1$ for all $\mathbf{u}\in S(\delta_1)$. Then for all $n\geq N$ and $\mathbf{u}\in S(\delta_1)$,

$$\left|1-\tfrac{1}{2}\mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u}\right| = 1-\tfrac{1}{2}\mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u} \leq 1-\tfrac{1}{4}\lambda_{\mathrm{min}}\|\mathbf{u}\|^2. \tag{3.26}$$

To bound R_n, by the Cauchy-Schwarz inequality

$$R_n(\mathbf{u}) \leq AE_n(\mathbf{u})\|\mathbf{u}\|^2 \quad \text{where} \quad E_n(\mathbf{u}) = \mathbb{E}[\min\{\|\mathbf{u}\|\|\hat{\mathbf{X}}_n\|, 1\}\|\hat{\mathbf{X}}_n\|^2].$$

Then for all $L>0$, splitting the expectation into the case where $\|\hat{\mathbf{X}}_n\|^2\leq L^2$ and the case when $\|\hat{\mathbf{X}}_n\|^2>L^2$,

$$\begin{aligned}\sup_n E_n(\mathbf{u}) &\leq L^2\min\{L\|\mathbf{u}\|, 1\} + \sup_n \mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2 \mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2 > L^2\right\}\right] \\ &\to \sup_n \mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2 \mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2 > L^2\right\}\right]\end{aligned}$$

as $\mathbf{u}\to 0$. This holds for all $L>0$, hence taking the limit $L\to\infty$ and using Lemma 3.7.2 we obtain that $\lim_{\mathbf{u}\to 0}\sup_n E_n(\mathbf{u}) = 0$. Thus, there exists δ_2 such that for all $\mathbf{u} \in S(\delta_2)$

$$R_n(\mathbf{u}) \leq \frac{1}{8}\lambda_{\min}\|\mathbf{u}\|^2. \tag{3.27}$$

Thus setting $\delta = \min\{\delta_1, \delta_2\}$, for all $n \geq N$ and $\mathbf{u} \in S(\delta)$

$$|\phi_n(\mathbf{u})| \leq 1 - c\|\mathbf{u}\|^2,$$

where $c = \frac{1}{8}\lambda_{\min}$.

We now address the second bound. let ϕ be the characteristic function of $\mathbf{X}$. We assume $\mathbf{X}$ has main lattice $\mathbb{Z}^d$, thus $|\phi(\mathbf{u})| = 1$ if and only if every entry of $\mathbf{u}$ is a multiple of 2π by Lemma 3.7.1. In particular $|\phi(\mathbf{u})| < 1$ for all $\mathbf{u} \in S(\pi) \setminus S(\delta)$. ϕ is continuous and $S(\pi) \setminus S(\delta)$ is compact. Therefore there exists $\epsilon > 0$ such that $\sup_{\mathbf{u}\in S(\pi)\setminus S(\delta)}|\phi(\mathbf{u})| \leq 1 - \epsilon$.

Since $\mathbf{X}_n \xrightarrow{(\mathrm{d})} \mathbf{X}$ as $n \to \infty$, $\phi_n \to \phi$ uniformly on compact sets. Therefore there exists N such that for all $n \geq N$

$$\sup_{\mathbf{u}\in S(\pi)\setminus S(\delta)} |\phi_n(\mathbf{u})| \leq \rho = 1 - \tfrac{1}{2}\epsilon. \qquad \square$$

We are finally ready to prove Theorem 3.3.16

Proof of Theorem 3.3.16. We first address the case where the main lattice of $\mathbf{X}$ and all $\mathbf{X}_n$ is $\mathbb{Z}^d$. The main trick in the proof is to notice that if n is integer valued then

$$\mathbb{1}\{n = 0\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{inu}\,\mathrm{d}u.$$

For all $\mathbf{y} \in \mathbf{c}_n + \mathbb{Z}^d$, $\sum_{i=1}^n \mathbf{X}_{n,i} - \mathbf{y} \in \mathbb{Z}^d$, so

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\right) &= \mathbb{E}\left[\frac{1}{(2\pi)^d}\int_{S(\pi)} e^{i\mathbf{u}\cdot\left(\sum_{i=1}^n \mathbf{X}_{n,i} - \mathbf{y}\right)}\,\mathrm{d}\mathbf{u}\right] \\ &= \frac{1}{(2\pi)^d}\int_{S(\pi)} \phi_n(\mathbf{u})^n e^{-i\mathbf{u}\cdot(\mathbf{y} - n\mathbb{E}[\mathbf{X}_n])}\,\mathrm{d}\mathbf{u},\end{aligned}$$

where $\phi_n(\mathbf{u}) = \mathbb{E}[e^{i\mathbf{u}\cdot(\mathbf{X}_n - \mathbb{E}[\mathbf{X}_n])}]$ and $S(r) = [-r, r]^d$ for all $r > 0$. Recall

$$\mathbf{x}_n = n^{-1/2}(\mathbf{y} - n\mathbb{E}[\mathbf{X}_n]).$$

Then, changing variables with $\mathbf{s} = \sqrt{n}\mathbf{u}$,

$$n^{d/2}\mathbb{P}\left(\sum_{i=1}^{n} \mathbf{X}_{n,i} = \mathbf{y}\right) = \frac{1}{(2\pi)^d}\int_{S(\pi\sqrt{n})} \phi_n(\mathbf{s}/\sqrt{n})^n e^{-i\mathbf{s}\cdot\mathbf{x}_n}\,\mathrm{d}\mathbf{s}.$$

By the Fourier inversion theorem,

$$f(\mathbf{x}) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \psi(\mathbf{s})e^{-i\mathbf{s}\cdot\mathbf{x}}\,\mathrm{d}\mathbf{s}$$

where ψ is the characteristic function of the $N(0, \mathrm{Cov}(\mathbf{X}))$ distribution. Therefore

$$\begin{aligned}
&\sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda}\left|n^{d/2}\mathbb{P}\left(\textstyle\sum_{i=1}^{n} \mathbf{X}_{n,i} = \mathbf{y}\right) - f(\mathbf{x}_n(\mathbf{y}))\right| \\
&\qquad = \sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda}\left|\int_{\mathbb{R}^d}\left(\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi_n(\mathbf{s}/\sqrt{n})^n - \psi(\mathbf{s})\right)e^{-i\mathbf{s}\cdot\mathbf{x}_n(\mathbf{y})}\,\mathrm{d}\mathbf{s}\right| \\
&\qquad \le \int_{\mathbb{R}^d}\left|\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi_n(\mathbf{s}/\sqrt{n})^n - \psi(\mathbf{s})\right|\mathrm{d}\mathbf{s}.
\end{aligned}$$

We apply the dominated convergence theorem. To dominate the integrand, first note that ψ is integrable. Secondly let δ, c, ρ and N be as in Lemma 3.7.4. For all $n \ge N$ and for all $\mathbf{s} \in S(\delta\sqrt{n})$,

$$|\phi_n(\mathbf{s}/\sqrt{n})|^n \le (1 - c\|s\|^2/n)^n \le e^{-c\|s\|^2}.$$

Let $C = -\log(\rho)$. Note if $\mathbf{s} \in S(\pi\sqrt{n})$ then $\|\mathbf{s}\|^2 \le \pi^2 dn$. Thus for all $n \ge N$ and $\mathbf{s} \in S(\pi\sqrt{n}) \setminus S(\delta\sqrt{n})$

$$|\phi_n(\mathbf{s}/\sqrt{n})|^n \le e^{-Cn} \le e^{-\frac{C}{\pi^2 d}\|\mathbf{s}\|^2}.$$

Hence for all $n \ge N$,

$$\left|\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi_n(\mathbf{s}/\sqrt{n})^n - \psi(\mathbf{s})\right| \le e^{-c\|\mathbf{s}\|^2} + e^{-\frac{C}{\pi^2 d}\|\mathbf{s}\|^2} + |\psi(\mathbf{s})|$$

where, in particular, the right hand side is integrable. By Lemma 3.7.3,

$$\phi_n(\mathbf{s}/\sqrt{n})^n \to \psi(\mathbf{s})$$

as $n \to \infty$ for all $\mathbf{s} \in \mathbb{R}^d$. Thus for all $\mathbf{s} \in \mathbb{R}^d$

$$\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi(\mathbf{s}/\sqrt{n})^n \to \psi(\mathbf{s})$$

as $n \to \infty$. Hence by the dominated convergence theorem

$$\lim_{n\to\infty} \sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda} \left| n^{d/2}\mathbb{P}\left(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\right) - f(\mathbf{x}_n) \right| = 0,$$

as required.

Finally we generalise to any main lattice Λ. Suppose that Λ is generated by the columns of the invertible matrix A. Then A, viewed as a linear transform, is a isomorphism mapping $\mathbb{Z}^d$ to Λ. Thus $A^{-1}\mathbf{X}_n$ and $A^{-1}\mathbf{X}$ will have common lattice $\mathbb{Z}^d$ for all n. Moreover we can check the remaining assumptions of Theorem 3.3.16 still hold, thus uniformly for $\mathbf{y}$ in the translation of Λ containing the support of $\sum_{i=1}^n \mathbf{X}_{n,i}$,

$$\mathbb{P}\left(\sum_{i=1}^n A^{-1}\mathbf{X}_{n,i} = A^{-1}\mathbf{y}\right) = \frac{1}{\sqrt{(2\pi n)^d \det\tilde{\Sigma}}} \exp\left(-\frac{1}{2}(A^{-1}\mathbf{x}_n)^T\tilde{\Sigma}^{-1}(A^{-1}\mathbf{x}_n)\right) + o(n^{-d/2}).$$

where $\tilde{\Sigma} = \mathrm{Cov}(A^{-1}\mathbf{X})$. This simplifies to

$$\mathbb{P}\left(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\right) = \frac{1}{\sqrt{(2\pi n)^d \det\tilde{\Sigma}}} \exp\left(-\frac{1}{2}\mathbf{x}_n^T(A\tilde{\Sigma}A^T)^{-1}\mathbf{x}_n\right) + o(n^{-d/2}).$$

We have that

$$\tilde{\Sigma} = \mathrm{Cov}(A^{-1}\mathbf{X}) = A^{-1}\,\mathrm{Cov}(\mathbf{X})(A^{-1})^T.$$

Therefore

$$\det(\tilde{\Sigma}) = \det(A)^{-2}\det(\mathrm{Cov}(\mathbf{X})) = \det(\Lambda)^{-2}\det(\mathrm{Cov}(\mathbf{X}))$$

and so

$$\mathbb{P}\left(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\right) = \frac{\det(\Lambda)}{\sqrt{(2\pi n)^d \det(\mathrm{Cov}\,\mathbf{X})}} \exp\left(-\frac{1}{2}\mathbf{x}_n^T\,\mathrm{Cov}(\mathbf{X})^{-1}\mathbf{x}_n\right) + o(n^{-d/2}),$$

as required. □

3.8 Proof of technical lemmas

Proof of Lemma 3.4.12. Denote $g_n(s) = \inf\{t : f_n(t) > s\}$ and $g(s) = \inf\{t : f(t) > s\}$. By Proposition 3.6.5 in the book by Ethier and Kurtz [53], it is sufficient to show that for any $s > 0$, for any $s_n \to s$,

1. $\max\{|g_n(s_n) - g(s)|, |g_n(s_n) - g(s-)|\} \to 0$;
2. If $u_n \leq s_n$ for all n, $s_n \to s$, $u_n \to s$ and $g_n(s_n) \to g(s-)$, then $g_n(u_n) \to g(s-)$;
3. If $u_n \geq s_n$ for all n, $s_n \to s$, $u_n \to s$ and $g_n(s_n) \to g(s)$, then $g_n(u_n) \to g(s)$.

Fix $s > 0$. If $g(s-) = g(s)$, the result is straightforward, so we focus on $g(s-) < g(s)$.

We start by proving the first property. Fix $\epsilon > 0$ and suppose $s_n \to s$. We observe that $g(s-) < g(s)$ implies that f has a local maximum at $g(s-)$ and that $f(g(s-)) = f(g(s)) = s$. By the uniqueness of local maxima of f and the definition of g, there exists a $\delta_1 > 0$ such that for all $t < g(s-) - \epsilon$, we have that $f(t) < s - \delta_1$. Similarly, there exists a $\delta_2 > 0$ such that for all $g(s-) + \epsilon < t < g(s) - \epsilon$, we have that $f(t) < s - \delta_2$. Moreover, define

$$\delta_3 = \sup\{f(t) : g(s) < t < g(s) + \epsilon\} - s,$$

so that, by definition of g, we have that $\delta_3 > 0$. Define $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Now, let n be large enough such that $\sup_{t\in[0,g(s)+\epsilon]} |f_n(s) - f(s)| < \delta/2$ and $|s_n - s| < \delta/2$. Then, it holds that

1. $f_n(t) < s - \delta/2 < s_n$ for all $t < g(s-) - \epsilon$;
2. $f_n(t) < s - \delta < s_n$ for all $g(s-) + \epsilon < t < g(s) - \epsilon$;
3. There is a $g(s) < t < g(s) + \epsilon$ such that $f_n(t) > s + \delta/2 > s_n$.

These tree facts imply that $g_n(s_n) \subseteq [g(s-) - \epsilon, g(s-) + \epsilon] \cup [g(s) - \epsilon, g(s) + \epsilon]$, which proves the first of the three conditions.

Then, the second and third property follow immediately from the first property and the monotonicity of g_n and g. □

Proof of Lemma 3.5.3. First, note that g_i^n, σ_i^n, g_i, and σ_i are well-defined for all $i \in [m]$, $n \geq 1$ by $\inf\{f(t) : t \leq T\} < \inf\{f(t) : t \leq x_m\}$ and $\inf\{f_n(t) : t \leq T\} < \inf\{f_n(t) : t \leq x_m^n\}$.

Fix i. We will first show that $g_i^n \to g_i$ and $\sigma_i^n \to \sigma_i$ as $n \to \infty$. Firstly, note that by the assumption that $f(x_i) - \inf\{f(s) : s \leq x_i\} > 0$ and the continuity of f, $g_i < x_i < g_i + \sigma_i$. Fix $0 < \epsilon < \min\{x_i - g_i, g_i + \sigma_i - x_i\}/2$. We claim that the following conditions are sufficient for $g_i^n \to g_i$ and $\sigma_i^n \to \sigma_i$ as $n \to \infty$. For all n large enough,

1. $g_i + \epsilon < x_i^n < g_i + \sigma_i - \epsilon$
2. $\inf\{f_n(s) : s \in (g_i - \epsilon, g_i + \epsilon)\} < \inf\{f_n(s) : s \in [g_i + \epsilon, g_i + \sigma_i - \epsilon]\}$,

3. $\inf\{f_n(s) : s \in (g_i - \epsilon, g_i + \epsilon)\} < \inf\{f_n(s) : s \in [0, g_i - \epsilon]\}$,

4. $\inf\{f_n(s) : s \in (g_i + \sigma_i - \epsilon, g_i + \sigma_i + \epsilon)\} < \inf\{f_n(s) : s \in [0, g_i + \sigma_i - \epsilon]\}$

Indeed, conditions 1, 2 and 3 imply $|g_i^n - g_i| < \epsilon$, while conditions 1, 2 and 4 imply $|(g_i^n + \sigma_i^n) - (g_i + \sigma_i)| < \epsilon$. Note that condition 1 holds for n large enough by definition of ϵ and the convergence of x_i^n to x_i. To show the other conditions, define

$$\begin{aligned}
\delta_1 &= \inf\{f(s) : s \in [g_i + \epsilon, g_i + \sigma_i - \epsilon]\} - \inf\{f(s) : s \in (g_i - \epsilon, g_i + \epsilon)\} \\
\delta_2 &= \inf\{f(s) : s \in [0, g_i - \epsilon]\} - \inf\{f(s) : s \in (g_i - \epsilon, g_i + \epsilon)\} \\
\delta_3 &= \inf\{f(s) : s \in [0, g_i + \sigma_i - \epsilon]\} - \inf\{f(s) : s \in (g_i + \sigma_i - \epsilon, g_i + \sigma_i + \epsilon)\}.
\end{aligned}$$

By uniqueness of local minima and the definition of g_i and σ_i, we have $\delta := \min\{\delta_1, \delta_2, \delta_3\}/3 > 0$. Then, note that for n large enough, $\sup\{|f_n(s) - f(s)| : s \le g_i + \epsilon\} < \delta$, which implies conditions 2, 3, and 4 for such n.

Since i was arbitrary, and m is finite, we find that

$$(g_i^n, \sigma_i^n)_{1 \le i \le m} \to (g_i, \sigma_i)_{1 \le i \le m}$$

in $\mathbb{R}^{2m}$ as $n \to \infty$.

We now claim that $g_i^n \to g_i$ and $g_j^n \to g_i$ implies that $g_i^n = g_j^n$ for n large enough. Indeed, by definition of g_i^n, g_j^n and σ_i^n, we have that $g_i^n < g_j^n$ implies that $g_j^n - g_i^n \ge \sigma_i^n$, and by the argument above, $\sigma_i^n \to \sigma_i > 0$, so $g_i^n - g_j^n \to 0$ can only hold if $g_i^n = g_j^n$ for n large enough. This implies that

$$\#\{(g_i^n, \sigma_i^n) : 1 \le i \le m\} \to \#\{(g_i, \sigma_i) : 1 \le i \le m\}.$$

Then, the result follows. □

Acknowledgements

The authors would like to thank Christina Goldschmidt and Robin Stephenson for many fruitful discussions, and for kindly allowing us to use some of their figures, and Igor Kortchemski for his advice on local limit theorems.

Chapter 4

Conclusion

In this section, I will discuss other research that I have undertaken during my doctorate but which is not included in this thesis, as well as future research directions.

4.1 Random trees

In Chapter 2, we used the bijection between $\mathcal{T}(n)$ and $[n]^{n-1}$ that was introduced by Addario-Berry, the author, Maazoun and Martin in [11], restricted to trees with a particular degree sequence, to study the height of uniformly random trees with a given degree sequence. We discuss some extensions and further applications of the bijection.

Firstly, as discussed in [11], for any n and k the bijection can be extended to unrooted trees with vertex set $[n]$, forests with with vertex set $[n]$ and k trees, and rooted trees with vertex set $[n]$ and k marks on vertices, which yields a straightforward procedure to sample uniformly random elements in these classes. This has some powerful applications. For example, when we use the bijection to sample a uniform rooted tree with vertex set $[n]$ and 1 mark, the path between the root and the mark is encoded by the elements of the sequence up to the first repeat. If we then consider the bijection between rooted trees with vertex set $[n]$ and 1 mark and maps $f : [n] \to [n]$, the path between the root and the mark encodes the cyclic structure of the map, and we get immediate access to the law of the cyclic structure in a uniform map by studying the law of the random sequence up to the first repeat. By further exploiting this encoding, we can obtain easy new proofs for the results by Aldous and Pitman on the asymptotics of uniformly random mappings in [17].

Furthermore, in [11], we define a growth procedure for random trees with a given degree sequence. To be precise, given a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$, a natural number $d' > 0$ and a $\mathrm{T}_{\mathrm{d}} \in_u \mathcal{T}_{\mathrm{d}}$, we define a procedure to obtain a rooted tree with the law of a uniformly random element of $\mathcal{T}_{\mathrm{d}'}$ with degree sequence $\mathrm{d}' = (d_1, \ldots, d_n, d', 0, \ldots, 0)$ by a random local regrafting of T_{d}. This in particular gives a new straightforward method to grow uniformly random d-ary trees with n vertices. The existence of a 'nice' growth procedure for uniformly random objects in a class $\mathcal{C}_n$ from objects in a class $\mathcal{C}_{n-1}$ is often an ingredient in showing that a Markov chain defined on $\mathcal{C}_n$ with the uniform distribution on $\mathcal{C}_n$ as its stationary distribution mixes in polynomial time, so the new growth procedure for trees with a given degree sequence may have applications in this direction that the author and Caraceni will explore in future work.

A further application of the bijection, that is also discussed in Chapter 2, is to tree-rooted graphs. Consider a multiset $\mathcal{M}$ of elements of $[n]$ with size (counted with multiplicity) at least $n-1$. Then construct a sequence of length $n-1$ by uniform sampling without replacement from $\mathcal{M}$, and consider the tree T on $[n]$ encoded by this sequence via the bijection. (If $\mathcal{M}$ itself has $n-1$ elements then T has the law of a uniform tree with a given degree sequence, where for $i \in [n]$, the multiplicity of i in $\mathcal{M}$ is the degree of i in T. More generally, the number of copies of i in the subsample is the degree of i in T, and T is a uniform tree with this degree sequence.) It turns out that T has the law of a uniform spanning tree of a random *tree-rooted graph* with degree sequence defined by $\mathcal{M}$. In [5], a different sampling method is used to show that in the finite variance regime, the spanning trees of large random tree-rooted graphs converge after rescaling to the Brownian continuum random tree. In future work, Addario-Berry and the author plan to build on the current work to study distances in and convergence of random tree-rooted graphs for other degree regimes.

Other future directions of this line of research are discussed in Section 2.5.

4.2 The configuration model

In [46], we show joint convergence under rescaling of the height process and Łukasiewicz path of forests of supercritical Bienaymé trees for which the Łukasiewicz path under rescaling lies in the domain of attraction of a supercritical Lévy process. This extends the results by Duquesne and Le Gall [50] that only cover forests for which the Łukasiewicz path under rescaling is in the domain of attraction of a critical or subcritical Lévy process. Moreover, the limit object of the corresponding height process is the height process of a supercritical continuous state branching process as constructed by Lambert [75], illustrating that his definition is consistent with the discrete height process.

Then, in the second part of the paper, we use the result to obtain the metric space scaling limit of a uniformly random graph with an i.i.d. degree sequence in the α-stable regime in the critical window. This extends the result by Conchon-Kerjan and Goldschmidt [37], who show the result at one point in the critical window. The limit object that they obtain is called the *stable graph*. We show that the critical window that we consider corresponds to bond percolation on a uniformly random graph with a supercritical i.i.d. degree sequence with power-law tail behaviour.

In an upcoming work [47], we discuss a new critical window for the configuration model that is defined via inhomogeneous site-percolation where a vertex with degree d is closed at a rate proportional to its degree. We will now give a more formal definition of the model that we consider. Let $\alpha \in (1,2)$. We consider n vertices and let their degrees be i.i.d. samples $(D_1, \dots, D_n)$. We require, firstly, that $\mathbb{P}(D_1 = k) \sim ck^{-(\alpha+2)}$ for some $c > 0$ so that we are in the α-stable regime and, secondly, that $\mathbb{E}[D_1^2] = 2\mathbb{E}[D_1] =: 2\mu$ so that we are at the critical point. Then, we sample a uniform graph with this degree sequence, fix a $\lambda \in 0$ and then close a vertex of degree d with probability $1 - e^{-\frac{\lambda}{\mu} n^{-1/(\alpha+1)} d}$. We then show Gromov–Hausdorff–Prokhorov convergence of the largest components under rescaling. Notably, under this definition of the critical window, the components of the limit object are independent conditional on their sizes and are small copies of the components of the stable graph. A similar result cannot be expected to hold when the critical window is defined with bond percolation. Namely, in the stable graph, branchpoints are hubs with infinite degree with different local times. These hubs are not affected by cutting according to the length-measure or bond percolation, so performing such a procedure creates small components with atypically large hubs, which cannot have the same law as small components of the stable graph. There are striking analogies between the limit object as obtained by us in [47] and the limit under rescaling of the Erdős–Rényi graph in the critical window, which suggests that the procedure as considered by us is the natural definition of the critical window for the heavy-tailed configuration model.

For the model that we consider in [47], there is a natural coupling between the model at different values of λ that can be considered as a process in time, analogous to the coupling of the critical window in the Erdős–Rényi graph. In future work, we want to show joint convergence under rescaling in the critical window as a whole. As a first step, we want to show convergence of the component sizes as a process in time and we later want to extend this to GHP-convergence of the components themselves as a process in time, echoing the results of Rossignol on the Erdős–Rényi graph [93].

Also many natural questions remain for the directed configuration model, as discussed in Section 3.6. In particular, we are interested in studying the model under different moment conditions, which will likely reveal as yet undiscovered limit objects and universality classes. Furthermore, we will extend our understanding of the strongly connected components by studying the directed graphs in which they are embedded. A first step will be to study all vertices that are accessible from the non-trivial strongly connected components by a single directed path. This will in particular reveal the oriented paths between the strongly connected components and expose the fractal structure of the directed graph, which is not observed when only studying the strongly connected components. These questions will be the content of future work of the author, Goldschmidt and Xie.

Bibliography

[1] Romain Abraham, Jean-François Delmas, and Patrick Hoscheit. A note on the Gromov–Hausdorff–Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.*, 18, 2013. ISSN 1083-6489. doi: 10.1214/EJP.v18-2116. URL http://doi.org/10.1214/EJP.v18-2116.

[2] L. Addario-Berry. Tail bounds for the height and width of a random tree with a given degree sequence. *Random Struct. Algorithms*, 41(2):253–261, 2012. ISSN 1042-9832. doi: 10.1002/rsa.20438. URL https://doi.org/10.1002/rsa.20438.

[3] L. Addario-Berry, N. Broutin, and C. Goldschmidt. The continuum limit of critical random graphs. *Probab. Theory Related Fields*, 152(3-4):367–406, 2012. ISSN 0178-8051. doi: 10.1007/s00440-010-0325-4. URL https://doi.org/10.1007/s00440-010-0325-4.

[4] Louigi Addario-Berry. Most trees are short and fat. *Probab. Theory Related Fields*, 173(1-2):1–26, 2019. ISSN 0178-8051. doi: 10.1007/s00440-018-0829-x. URL https://doi.org/10.1007/s00440-018-0829-x.

[5] Louigi Addario-Berry and Jordan Barrett. Random tree-weighted graphs. *Stochastic Analysis, Random Fields and Integrable Probability, Adv. Studies in Pure Math., MSJ*, 87:1–59, 2021. doi: 10.2969/aspm/08710001. URL https://doi.org/10.2969/aspm/08710001.

[6] Louigi Addario-Berry and Serte Donderwinkel. Random trees have height $O(\sqrt{n})$. Preprint arXiv:2201.11773, 2022. URL https://arxiv.org/abs/2201.11773.

[7] Louigi Addario-Berry, Nicolas Broutin, and Christina Goldschmidt. Critical random graphs: Limiting constructions and distributional properties. *Electron. J. Probab.*, 15:741–775, 2010. ISSN 1083-6489. doi: 10.1214/EJP.v15-772. URL https://doi.org/10.1214/EJP.v15-772.

[8] 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. ISSN 0091-1798. doi: 10.1214/12-AOP758. URL https://doi.org/10.1214/12-AOP758.

[9] Louigi Addario-Berry, Nicolas Broutin, Christina Goldschmidt, and Grégory Miermont. The scaling limit of the minimum spanning tree of the complete graph. *Ann. Probab.*, 45(5):3075–3144, 2017. ISSN 00911798. doi: 10.1214/16-AOP1132. URL https://doi.org/10.1214/16-AOP1132.

[10] Louigi Addario-Berry, Anna Brandenberger, Jad Hamdan, and Céline Kerriou. Universal height and width bounds for random trees. Preprint arXiv:2105.03195, 2021. URL https://arxiv.org/abs/2105.03195.

[11] Louigi Addario-Berry, Serte Donderwinkel, Mickaël Maazoun, and James B. Martin. A new proof of Cayley's formula. Preprint arXiv:2107.09726, 2021. URL https://arxiv.org/abs/2107.09726.

[12] D. Aldous. The Continuum Random Tree II: An overview. In M. T. Barlow and N. H. Bingham, editors, *Stochastic Analysis: Proceedings of the Durham Symposium on Stochastic Analysis, 1990*, London Mathematical Society Lecture Note Series, pages 23–70. Cambridge University Press, Cambridge, 1991. doi: 10.1017/CBO9780511662980.003. URL https://doi.org/10.1017/CBO9780511662980.003.

[13] David Aldous. The Continuum Random Tree I. *Ann. Probab.*, 19(1):1–28, 1991. ISSN 0091-1798, 2168-894X. doi: 10.1214/aop/1176990534. URL https://doi.org/10.1214/aop/1176990534.

[14] David Aldous. The Continuum Random Tree III. *Ann. Probab.*, 21(1):248–289, 1993. ISSN 0091-1798. doi: 10.1214/aop/1176989404. URL http://doi.org/10.1214/aop/1176989404.

[15] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997. ISSN 0091-1798, 2168-894X. doi: 10.1214/aop/1024404421. URL https://doi.org/10.1214/aop/1024404421.

[16] David J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985. doi: 10.1007/BFb0099421. URL `https://doi.org/10.1007/BFb0099421`.

[17] David J. Aldous and Jim Pitman. Brownian bridge asymptotics for random mappings. *Random Struct. Algorithms*, 5(4):487–512, 1994. ISSN 1042-9832. doi: 10.1007/3-540-34266-4_10. URL `https://doi.org/10.1007/3-540-34266-4_10`.

[18] Eleanor Archer, Asaf Nachmias, and Matan Shalev. The GHP scaling limit of uniform spanning trees in high dimensions. Preprint arXiv:2112.0120, 2021. URL `https://arxiv.org/abs/2112.01203`.

[19] Krishna B. Athreya and Peter Ney. *Branching processes.* Grundlehren der mathematischen Wissenschaften ; Bd.196. Springer-Verlag, Berlin, 1972. ISBN 9780387057903. doi: 10.1007/978-3-642-65371-1. URL `https://doi.org/10.1007/978-3-642-65371-1`.

[20] Olivier Bernardi. Bijective counting of tree-rooted maps and shuffles of parenthesis systems. *Electron. J. Combin.*, 14(1):Research Paper 9, 36, 2007. doi: 10.37236/928. URL `https://doi.org/10.37236/928`.

[21] S.N. Bernstein. *Theory of Probability (Russian).* Moscow, 1927.

[22] S. Bhamidi, S. Sen, and X. Wang. Continuum limit of critical inhomogeneous random graphs. *Probab. Theory Related Fields*, 169(1-2):565–641, 2017. ISSN 0178-8051. doi: 10.1007/s00440-016-0737-x. URL `https://doi.org/10.1007/s00440-016-0737-x`.

[23] Shankar Bhamidi, Souvik Dhara, Remco van der Hofstad, and Sanchayan Sen. Global lower mass-bound for critical configuration models in the heavy-tailed regime. Preprint arXiv:2005.02566, 2020. URL `https://arxiv.org/abs/2005.02566`.

[24] Shankar Bhamidi, Souvik Dhara, Remco van der Hofstad, and Sanchayan Sen. Universality for critical heavy-tailed network models: Metric structure of maximal components. *Electron. J. Probab.*, 25:1–57, 2020. ISSN 1083-6489, 1083-6489. doi: 10.1214/19-EJP408. URL `https://doi.org/10.1214/19-EJP408`.

[25] Arthur Blanc-Renaudie. Limit of trees with fixed degree sequence. Preprint arXiv:2110.03378, 2021. URL `https://arxiv.org/abs/2110.03378`.

[26] Mindaugas Bloznelis, Friedrich Götze, and Jerzy Jaworski. Birth of a strongly connected giant in an inhomogeneous random digraph. *J. Appl. Probab.*, 49(3):601–611, 2012. ISSN 0021-9002, 1475-6072. doi: 10.1239/jap/1346955320. URL `https://doi.org/10.1239/jap/1346955320`.

[27] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980. ISSN 0195-6698. doi: 10.1016/S0195-6698(80)80030-8. URL `https://doi.org/10.1016/S0195-6698(80)80030-8`.

[28] Mireille Bousquet-Mélou and Julien Courtiel. Spanning forests in regular planar maps. *J. Combin. Theory Ser. A*, 135:1–59, 2015. ISSN 0097-3165. doi: 10.1016/j.jcta.2015.04.002. URL `https://doi.org/10.1016/j.jcta.2015.04.002`.

[29] Nicolas Broutin, Thomas Duquesne, and Minmin Wang. Limits of multiplicative inhomogeneous random graphs and Lévy trees: Limit theorems. *Probab. Theory Related Fields*, July 2021. ISSN 1432-2064. doi: 10.1007/s00440-021-01075-z. URL `https://doi.org/10.1007/s00440-021-01075-z`.

[30] Xing Shi Cai and Guillem Perarnau. The diameter of the directed configuration model. Preprint arXiv:2003.04965, 2020. URL `https://arxiv.org/abs/2003.04965`.

[31] Xing Shi Cai and Guillem Perarnau. The giant component of the directed configuration model revisited. *Lat. Am. J. Probab. Math. Stat.*, 18:1517, 2021. doi: 10.30757/ALEA.v18-55. URL `https://doi.org/10.30757/ALEA.v18-55`.

[32] Xing Shi Cai and Guillem Perarnau. Minimum stationary values of sparse random directed graphs. Preprint arXiv:2010.07246, 2021. URL `https://arxiv.org/abs/2010.07246`.

[33] Junyu Cao and Mariana Olvera-Cravioto. Connectivity of a general class of inhomogeneous random digraphs. *Random Struct. Algorithms*, 56(3):722–774, 2019. ISSN 1042-9832. doi: 10.1002/rsa.20892. URL `https://doi.org/10.1002/rsa.20892`.

[34] J. W. S. Cassels. *An Introduction to the Geometry of Numbers*. Classics in Mathematics. Springer-Verlag, Berlin Heidelberg, 1997. ISBN 978-3-540-61788-4. doi: 10.1007/978-3-642-62035-5. URL `https://doi.org/10.1007/978-3-642-62035-5`.

[35] P. Chassaing, J. F. Marckert, and M. Yor. The height and width of simple trees. In *Mathematics and computer science (Versailles, 2000)*, Trends Math., pages 17–30.

Birkhäuser, Basel, 2000. doi: 10.1007/978-3-0348-8405-1_2. URL https://doi.org/10.1007/978-3-0348-8405-1_2.

[36] Ningyuan Chen and Mariana Olvera-Cravioto. Directed random graphs with given degree distributions. *Stoch. Syst.*, 3(1):147–186, 2013. ISSN 1946-5238. doi: 10.1214/12-SSY076. URL https://doi.org/10.1214/12-SSY076.

[37] Guillaume Conchon-Kerjan and Christina Goldschmidt. The stable graph: the metric space scaling limit of a critical random graph with i.i.d. power-law degrees. Preprint arXiv:2002.04954v3, 2020. URL https://arxiv.org/abs/2002.04954v3.

[38] Colin Cooper and Alan Frieze. The Size of the Largest Strongly Connected Component of a Random Digraph with a Given Degree Sequence. *Combin. Probab. Comput.*, 13(3): 319–337, 2004. ISSN 1469-2163, 0963-5483. doi: 10.1017/S096354830400611X. URL https://doi.org/10.1017/S096354830400611X.

[39] Cecil C. Craig. On the Tchebychef inequality of Bernstein. *Ann. Math. Statist.*, 4(2): 94–102, 1933. ISSN 0003-4851. doi: 10.1214/aoms/1177732803. URL https://www.doi.org/10.1214/aoms/1177732803.

[40] Nicolas Curien, Bénédicte Haas, and Igor Kortchemski. The CRT is the scaling limit of random dissections. *Random Struct. Algorithms*, 47(2):304–327, 2015. ISSN 1042-9832. doi: 10.1002/rsa.20554. URL https://doi.org/10.1002/rsa.20554.

[41] D. J. Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and Its Applications, An Introduction to the Theory of Point Processes. Springer-Verlag, New York, second edition, 2008. ISBN 978-0-387-21337-8. doi: 10.1007/978-0-387-49835-5. URL https://doi.org/10.1007/978-0-387-49835-5.

[42] N. G. de Bruijn, D. E. Knuth, and S. O. Rice. The average height of planted plane trees. In *Graph theory and computing*, pages 15–22. 1972. doi: 10.1016/B978-1-4832-3187-7.50007-6. URL https://doi.org/10.1016/B978-1-4832-3187-7.50007-6.

[43] Loïc de Raphélis. Scaling limit of multitype Galton–Watson trees with infinitely many types. *Ann. Inst. H. Poincaré Probab. Statist.*, 53(1):200–225, February 2017. ISSN 0246-0203. doi: 10.1214/15-AIHP713. URL https://doi.org/10.1214/15-AIHP713.

[44] Souvik Dhara, Remco van der Hofstad, Johan S. H. van Leeuwaarden, and Sanchayan Sen. Critical window for the configuration model: Finite third moment degrees. *Electron. J. Probab.*, 22:1–33, 2017. ISSN 1083-6489, 1083-6489. doi: 10.1214/17-EJP29. URL `https://doi.org/10.1214/17-EJP29`.

[45] Souvik Dhara, Remco van der Hofstad, Johan S. H. van Leeuwaarden, and Sanchayan Sen. Heavy-tailed configuration models at criticality. *Ann. Inst. H. Poincaré Probab. Statist.*, 56(3):1515–1558, 2020. ISSN 0246-0203. doi: 10.1214/19-AIHP980. URL `https://doi.org/10.1214/19-AIHP980`.

[46] Serte Donderwinkel. Height process convergence of supercritical Galton–Watson forests, with an application to the configuration model in the critical window. Preprint arXiv:2105.12109, 2021. URL `https://arxiv.org/abs/2105.12109`.

[47] Serte Donderwinkel. Self-similarity of the stable graph: a new critical window for the heavy-tailed configuration model, *Preprint in preparation.* 2022.

[48] Serte Donderwinkel and Zheneng Xie. Universality for the directed configuration model with random degrees: metric space convergence of the strongly connected components at criticality. Preprint arXiv:2105.11434, 2021. URL `https://arxiv.org/abs/2105.11434`.

[49] Thomas Duquesne. A limit theorem for the contour process of conditioned Galton–Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003. ISSN 0091-1798. doi: 10.1214/aop/1048516543. URL `https://doi.org/10.1214/aop/1048516543`.

[50] Thomas Duquesne and Jean-François Le Gall. *Random Trees, Lévy Processes and Spatial Branching Processes.* Number 281 in Astérisque. Société mathématique de France, 2002. URL `http://numdam.org/item/AST_2002__281__R1_0/`.

[51] Thomas Duquesne and Jean-François Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005. ISSN 0178-8051. doi: 10.1007/s00440-004-0385-4. URL `https://doi.org/10.1007/s00440-004-0385-4`.

[52] Rick Durrett. *Probability: Theory and Examples.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fifth edition, 2019. ISBN 978-1-108-47368-2. doi: 10.1017/9781108591034. URL `https://doi.org/10.1017/9781108591034`.

[53] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 1986. ISBN 978-0-471-08186-9.

[54] Philippe Flajolet and Andrew Odlyzko. The average height of binary trees and other simple trees. *J. Comput. System Sci.*, 25(2):171–213, 1982. ISSN 0022-0000. doi: 10.1016/0022-0000(82)90004-6. URL `https://doi.org/10.1016/0022-0000(82)90004-6`.

[55] Philippe Flajolet, Zhicheng Gao, Andrew Odlyzko, and Bruce Richmond. The distribution of heights of binary trees and other simple trees. *Combin. Probab. Comput.*, 2(2):145–156, 1993. ISSN 0963-5483. doi: 10.1017/S0963548300000560. URL `https://doi.org/10.1017/S0963548300000560`.

[56] Jean-François Le Gall. Random real trees. *Ann. Fac. Sci. Toulouse Math.*, Ser. 6, 15(1): 35–62, 2006. doi: 10.5802/afst.1112. URL `https://doi.org/10.5802/afst.1112`.

[57] B. V. Gnedenko. On a local limit theorem of the theory of probability. *Uspekhi Mat. Nauk*, 3(3(25)):187–194, 1948.

[58] Christina Goldschmidt and Bénédicte Haas. A line-breaking construction of the stable trees. *Electron. J. Probab.*, 20, 2015. ISSN 1083-6489. doi: 10.1214/EJP.v20-3690. URL `https://doi.org/10.1214/EJP.v20-3690`.

[59] Christina Goldschmidt and Robin Stephenson. The scaling limit of a critical random directed graph. Preprint arXiv:1905.05397, 2021. URL `https://arxiv.org/abs/1905.05397`.

[60] A. Graf. On the strongly connected components of random directed graphs with given degree sequences. Thesis, 2016. URL `http://hdl.handle.net/10012/10681`.

[61] Ewain Gwynne, Adrien Kassel, Jason Miller, and David B. Wilson. Active spanning trees with bending energy on planar maps and SLE-decorated Liouville quantum gravity for $\kappa > 8$. *Comm. Math. Phys.*, 358(3):1065–1115, 2018. ISSN 0010-3616. doi: 10.1007/s00220-018-3104-1. URL `https://doi.org/10.1007/s00220-018-3104-1`.

[62] Ewain Gwynne, Nina Holden, and Xin Sun. A mating-of-trees approach for graph distances in random planar maps. *Probab. Theory Related Fields*, 177(3-4):1043–1102, 2020. ISSN 0178-8051. doi: 10.1007/s00440-020-00969-8. URL `https://doi.org/10.1007/s00440-020-00969-8`.

[63] Ewain Gwynne, Jason Miller, and Scott Sheffield. The Tutte embedding of the mated-CRT map converges to Liouville quantum gravity. *Ann. Probab.*, 49(4):1677 – 1717, 2021. doi: 10.1214/20-AOP1487. URL https://doi.org/10.1214/20-AOP1487.

[64] Bénédicte Haas and Grégory Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.*, 9:57 – 97, 2004. doi: 10.1214/EJP.v9-187. URL https://doi.org/10.1214/EJP.v9-187.

[65] 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. ISSN 0091-1798. doi: 10.1214/11-AOP686. URL https://doi.org/10.1214/11-AOP686.

[66] Frank Harary and Geert Prins. The number of homeomorphically irreducible trees, and other species. *Acta Math.*, 101:141–162, 1959. ISSN 0001-5962. doi: 10.1007/BF02559543. URL https://doi.org/10.1007/BF02559543.

[67] Nina Holden and Xin Sun. SLE as a mating of trees in Euclidean geometry. *Comm. Math. Phys.*, 364(1):171–201, 2018. ISSN 0010-3616. doi: 10.1007/s00220-018-3149-1. URL https://doi.org/10.1007/s00220-018-3149-1.

[68] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012. doi: 10.1214/11-PS188. URL https://doi.org/10.1214/11-PS188.

[69] Svante Janson and Malwina J. Luczak. A new approach to the giant component problem. *Random Struct. Algorithms*, 34(2):197–216, 2009. ISSN 1098-2418. doi: 10.1002/rsa.20231. URL https://doi.org/10.1002/rsa.20231.

[70] Adrien Joseph. The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.*, 24(6):2560–2594, 2014. ISSN 1050-5164, 2168-8737. doi: 10.1214/13-AAP985. URL https://doi.org/10.1214/13-AAP985.

[71] Valentin F. Kolchin. *Random mappings*. Translation Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York, 1986. ISBN 0-911575-16-2. Translated from the Russian, With a foreword by S. R. S. Varadhan.

[72] Igor Kortchemski. Sub-exponential tail bounds for conditioned stable Bienaymé–Galton–Watson trees. *Probab. Theory Related Fields*, 168(1-2):1–40, 2017. ISSN 0178-8051. doi: 10.1007/s00440-016-0704-6. URL `https://doi.org/10.1007/s00440-016-0704-6`.

[73] Ivan Kryven. Emergence of the giant weak component in directed random graphs with arbitrary degree distributions. *Phys. Rev. E*, 94(1):012315, July 2016. doi: 10.1103/PhysRevE.94.012315. URL `https://doi.org/10.1103/PhysRevE.94.012315`.

[74] Ivan Kryven. Finite connected components in infinite directed and multiplex networks with arbitrary degree distributions. *Phys. Rev. E*, 96(5):052304, November 2017. doi: 10.1103/PhysRevE.96.052304. URL `https://doi.org/10.1103/PhysRevE.96.052304`.

[75] Amaury Lambert. The genealogy of continuous-state branching processes with immigration. *Probab. Theory Related Fields*, 122(1):42–70, 2002. ISSN 0178-8051. doi: 10.1007/s004400100155. URL `https://doi.org/10.1007/s004400100155`.

[76] Jean-François Le Gall. Random trees and applications. *Probab. Surveys*, 2:245–311, 2005. ISSN 1549-5787. doi: 10.1214/154957805100000140. URL `https://doi.org/10.1214/154957805100000140`.

[77] Jean-Francois Le Gall and Yves Le Jan. Branching processes in Lévy processes: the exploration process. *Ann. Probab.*, 26(1):213–252, 1998. ISSN 0091-1798. doi: 10.1214/aop/1022855417. URL `https://doi.org/10.1214/aop/1022855417`.

[78] Yiting Li, Xin Sun, and Samuel S. Watson. Schnyder woods, SLE(16), and Liouville quantum gravity. Preprint arXiv:1705.03573, May 2017. URL `https://arxiv.org/abs/1705.03573`.

[79] Tomasz Łuczak. The number of trees with large diameter. *J. Aust. Math. Soc.*, 58(3):298–311, 1995. ISSN 0263-6115. doi: 10.1017/S1446788700038337. URL `https://doi.org/10.1017/S1446788700038337`.

[80] Jean-Francois Marckert and Abdelkader Mokkadem. The depth first processes of Galton–Watson trees converge to the same Brownian excursion. *Ann. Probab.*, 31(3):1655–1678, 2003. ISSN 0091-1798. doi: 10.1214/aop/1055425793. URL `https://doi.org/10.1214/aop/1055425793`.

[81] Colin McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 195–248, New York, 1998. Springer Verlag. doi: 10.1007/978-3-662-12788-9_6. URL https://doi.org/10.1007/978-3-662-12788-9_6.

[82] A. Meir and J. W. Moon. The distance between points in random trees. *J. Combinatorial Theory*, 8:99–103, 1970. ISSN 0021-9800. doi: 10.1016/S0021-9800(70)80012-6. URL https://doi.org/10.1016/S0021-9800(70)80012-6.

[83] A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian J. Math.*, 30(5):997–1015, 1978. ISSN 0008-414X. doi: 10.4153/CJM-1978-085-0. URL https://doi.org/10.4153/CJM-1978-085-0.

[84] Jason Miller and Scott Sheffield. Liouville quantum gravity spheres as matings of finite-diameter trees. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(3):1712–1750, 2019. ISSN 0246-0203. doi: 10.1214/18-AIHP932. URL https://doi.org/10.1214/18-AIHP932.

[85] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms*, 6(2-3):161–180, 1995. ISSN 1098-2418. doi: 10.1002/rsa.3240060204. URL https://doi.org/10.1002/rsa.3240060204.

[86] Michael Molloy and Bruce Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, 7(3):295–305, 1998. ISSN 1469-2163, 0963-5483. doi: 10.1017/S0963548398003526. URL https://doi.org/10.1017/S0963548398003526.

[87] A. B. Mukhin. Local limit theorems for lattice random variables. *Theory Probab. Appl.*, 36(4):698–713, 1992. ISSN 0040-585X. doi: 10.1137/1136086. URL https://doi.org/10.1137/1136086.

[88] R. C. Mullin. On the enumeration of tree-rooted maps. *Canadian J. Math.*, 19:174–183, 1967. ISSN 0008-414X. doi: 10.4153/CJM-1967-010-x. URL https://doi.org/10.4153/CJM-1967-010-x.

[89] Konstantinos Panagiotou, Benedikt Stufler, and Kerstin Weller. Scaling limits of random graphs from subcritical classes. *Ann. Probab.*, 44(5):3291–3334, 2016. ISSN 0091-1798. doi: 10.1214/15-AOP1048. URL https://doi.org/10.1214/15-AOP1048.

[90] A. Rényi and G. Szekeres. On the height of trees. *J. Austral. Math. Soc.*, 7:497–507, 1967. ISSN 0263-6115. doi: 10.1017/S1446788700004432. URL `https://doi.org/10.1017/S1446788700004432`.

[91] J. Riordan. The enumeration of trees by height and diameter. *IBM J. Res. Develop.*, 4: 473–478, 1960. ISSN 0018-8646. doi: 10.1147/rd.45.0473. URL `https://doi.org/10.1147/rd.45.0473`.

[92] Oliver Riordan. The phase transition in the configuration model. *Combin. Probab. Comput.*, 21(1-2):265–299, 2012. ISSN 1469-2163, 0963-5483. doi: 10.1017/S0963548311000666. URL `https://doi.org/10.1017/S0963548311000666`.

[93] Raphaël Rossignol. Scaling limit of dynamical percolation on critical Erdős–Rényi random graphs. *Ann. Probab.*, 49(1):322–399, 2021. ISSN 0091-1798. doi: 10.1214/20-AOP1472. URL `https://doi.org/10.1214/20-AOP1472`.

[94] Gennady Samorodnitsky, Sidney Resnick, Don Towsley, Richard Davis, Amy Willis, and Phyllis Wan. Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *J. Appl. Probab.*, 53(1):146–161, 2016. ISSN 0021-9002, 1475-6072. doi: 10.1017/jpr.2015.15. URL `https://doi.org/10.1017/jpr.2015.15`.

[95] Alexander Schrijver. *Theory of Linear and Integer Programming.* John Wiley & Sons, June 1998. ISBN 978-0-471-98232-6.

[96] Frank Spitzer. *Principles of Random Walk.* Graduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1964. ISBN 978-1-4757-4229-9. doi: 10.1007/978-1-4757-4229-9. URL `https://doi.org/10.1007/978-1-4757-4229-9`.

[97] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1999. ISBN 0-521-56069-1; 0-521-78987-7. doi: 10.1017/CBO9780511609589. URL `https://doi.org/10.1017/CBO9780511609589`. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[98] Benedikt Stufler. Limits of random tree-like discrete structures. *Probab. Surv.*, 17:318–477, 2020. doi: 10.1214/19-PS338. URL `https://doi.org/10.1214/19-PS338`.

[99] 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. doi: 10.1007/BFb0071532. URL https://doi.org/10.1007/BFb0071532.

[100] Giovanni Torriano. *Piazza Universale Di Proverbi Italiani: Or, A Common Place Of Italian Proverbs And Proverbial Phrases. Digested in Alphabetical Order by way of Dictionary. Interpreted, and occasionally Illustrated with Notes. Together with A Supplement of Italian Dialogves.* F. and T.W. for the Author, 1666. URL https://leme.library.utoronto.ca/lexicons/1115/details.

[101] Remco van der Hofstad. *Random Graphs and Complex Networks*. Cambridge University Press, Cambridge, 2017. ISBN 9781316779422. doi: 10.1017/9781316779422. URL http://ebooks.cambridge.org/ref/id/CBO9781316779422.

[102] T. Walsh and A. B. Lehman. Counting rooted maps by genus. II. *J. Combinatorial Theory Ser. B*, 13:122–141, 1972. ISSN 0095-8956. doi: 10.1016/0095-8956(72)90049-4. URL https://www.sciencedirect.com/science/article/pii/0095895672900494.