Scaling limits of critical directed graphs

Zheneng Xie

Oriel College

University of Oxford

A thesis submitted for the degree of

Doctor of Statistics

2023

Acknowledgements

I would like to thank my supervisor Prof. Christina Goldschmidt whose assistance, support and guidance has been invaluable and incalculable. I'm grateful for my colleague Serte Donderwinkel without whom the work on the directed configuration model would not be possible. I'd also like to thank the EPSRC and the Random Systems CDT for providing me with a scholarship funding my studies. Finally I would like to thank my family for their love and support.

Abstract

The scaling limit of connected components in critical undirected random graphs has been well studied and shown to be universal between different critical models of undirected random graphs. Work by Goldschmidt and Stephenson has established the scaling limit of a the strongly connected components of critical directed Erdős-Rényi random graphs. We investigate whether the same scaling limits arise in different critical models of random directed graphs.

We study two models in particular: the directed configuration model and inhomogeneous directed random graphs. Our first main result is a scaling limit for the strongly connected components in a critical directed configuration model. Our proofs rely on two main ingredients: a depth first exploration of the directed configuration model to construct an out-forest for the model and also the identification of a distinguished subset of edges not in the out-forest, called candidates, which are the only edges not in the out-forest which can be part of a strongly connected component. The number of such candidate edges remains bounded asymptotically and will converge to point identifications on the real trees that arise as the scaling limits of the out-forests. We can then extract a directed multigraph with edge lengths by taking the root of each real tree and the point identifications, then orienting each edge away from the root. From these multigraphs, we can extract the continuum strongly connected components. We also present work in progress on establishing a similar scaling limit for a family of rank-1 inhomogeneous random directed Poissonian graphs. We use similar techniques to the directed configuration model but use a coupling with a queueing process instead to construct an out-forest.

Contents

Chapter 1

Introduction

1.1 Undirected graphs

1.1.1 The Erdős–Rényi model

One of the first models of random graphs studied was the Erdős–Rényi random graph model. There are two variants of this model – the $G(n, m)$ model introduced by Erdős and Rényi [29] and the $G(n, p)$ model introduced by Gilbert [32]. Both have been studied extensively by Erdős and Rényi. Due to its nicer independence properties, we will be looking at the $G(n, p)$ model.

The $G(n, p)$ model is defined for $n \in \mathbb{N}_+$ and $p \in [0, 1]$. It is the random graph on n vertices where each of the $\binom{n}{2}$ possible edges are included in the graph with probability p independently of the other edges. More formally, $G \sim G(n, p)$ if G is a random graph on $[n]$ with law given by

$$\mathbb{P}(G = H) = p^{e(H)}(1 - p)^{\binom{n}{2} - e(H)}$$

where $e(H)$ is the number of edges in H.

The connectivity of $G(n, p)$ is a well studied area, tracing back to the very first papers on the model. What is of most relevance to us is the phase transition in the size of the largest component of $G(n, p)$. This transition was first observed by Erdős and Rényi [30]. The following version of the result is taken from [38, Theorem 5.4].

Theorem 1.1. *Consider $G(n, p)$ where $p = c/n$. Let $C_1^n, C_2^n, \dots$ be the components of G in decreasing order of size.*

1. *If $c > 1$ then $|C_1^n| = \Theta_p(n)$ and $|C_2^n| = O_p(\log n)$.*
2. *If $c < 1$ then $|C_1^n| = O_p(\log n)$.*

Much more can be said about the critical phase when $c = 1$. In the highly influential work [4], Aldous encodes an exploration of the graph by a random process. Using this technique, it is proved that not only do the components in the critical window scale as $n^{2/3}$, but in fact there is an intimate link between the sizes of the components and excursions of a Brownian motion with quadratic drift above its running minimum. Aldous not only proves a limit for the component sizes but also for the surpluses of the components. The *surplus* of a connected graph G is given by

$$e(G) - |G| + 1.$$

A connected graph has surplus 0 if and only if it is a tree.

Theorem 1.2. *Consider $G \sim G(n, n^{-1}+\lambda n^{-4/3})$ for $\lambda \in \mathbb{R}$. In decreasing order of size, let $C_1^n, C_2^n, \ldots$ be the components of G. In decreasing order of length, let $\gamma_1, \gamma_2, \ldots$ be the excursions of*

$$W_t^\lambda - \min_{s\in[0,t]} W_s^\lambda \quad \textit{where} \quad W_t^\lambda = W_t + \lambda t - \tfrac{1}{2}t^2$$

and $(W_t)_{t\geq 0}$ is a Brownian motion. Let $|\gamma_i|$ be the length of γ_i. Then

$$n^{-2/3}(|C_1^n|, |C_2^n|, \ldots) \xrightarrow{\text{(d)}} (|\gamma_1|, |\gamma_2|, \ldots) \tag{1.1}$$

as $n \to \infty$ with respect to the ℓ^2 metric on decreasing sequences. Further, let Z_i^n be the surplus of C_i^n. Let $\mu(\gamma_i)$ be the area under the excursion γ_i. Then conditional on $(\gamma_1, \gamma_2, \ldots)$, let $Z_1, Z_2, \ldots$ be independent with distribution $Z_i \sim \mathrm{Po}(\mu(\gamma_i))$. Then, jointly with the convergence in Eq. (1.1),

$$(Z_1^n, Z_2^n, \ldots) \xrightarrow{\text{(d)}} (Z_1, Z_2, \ldots)$$

as $n \to \infty$ with respect to the product topology. This means initial segments of arbitrary length converge.

Note the surpluses of the components require no scaling in order to converge. This is important for later.

Addario-Berry, Broutin, and Goldschmidt [1] showed not only that the component sizes have a scaling limit, but that the rescaled components viewed as a whole can be understood as converging to a sequence of metric spaces. To make this more precise we will first need to recall definitions on distances between compact metric spaces.

Let (M, d) be a metric space. Then given two non-empty compact subsets $C_1 \subseteq M$ and $C_2 \subseteq M$, the *Hausdorff* distance d_H between them is defined to be

$$d_H(C_1, C_2) = \max\left\{ \sup_{x_1\in C_1} d(x_1, C_2), \sup_{x_2\in C_2} d(x_2, C_1), \right\}$$

where $d(x, C) = \inf_{y\in C} d(x, y)$ is the distance between a point x and a set C. Then for two non-empty compact metric spaces (X, d_X) and (Y, d_Y), the *Gromov–Hausdorff* distance d_{GH} between X and Y is defined to be

$$d_{GH}(X, Y) = \sup_M \sup_{i_X} \sup_{i_Y} d_H(i_X(X), i_Y(Y))$$

where the supremum is taken over all metric spaces M and all isometric embeddings $i_X : X \to M$ and $i_Y : Y \to M$.

Note that any finite connected graph G can be interpreted as a compact metric space consisting of the vertex set equipped with the graph distance. We will use the notation αM for the metric space M' where the metric is scaled by α, meaning $d'(x, y) = \alpha d(x, y)$. Then the following is the key result of [1].

Theorem 1.3. *Let* $G \sim G(n, n^{-1} + \lambda n^{-4/3})$*. In decreasing order of size, let* $C_1^n, C_2^n, \ldots$ *be the components of* G*. Then viewing the* C_i^n *as compact metric spaces under graph distance, there exists a sequence of random compact metric spaces* $M_1, M_2, \ldots$ *such that*

$$n^{-1/3}(C_1^n, C_2^n, \ldots) \to (M_1, M_2, \ldots)$$

as $n \to \infty$ *with respect to the* ℓ^4 *metric based on the Gromov-Hausdorff distance.*

$M_1, M_2, \ldots$ are constructed from a one-parameter family of random compact metric spaces $M(\sigma)$. We will not give the details here as they are unnecessary for the sequel. Conditional on the excursions $\gamma_1, \gamma_2, \ldots$ as defined in Theorem 1.2, M_i is distributed as $M(|\gamma_i|)$ independently of the other M_j.

1.1.2 Universality

These limiting objects turn out to be not specific to the critical Erdős–Rényi model. Two other models that exhibit the same scaling limit, under suitable conditions, are the configuration model and inhomogeneous random graphs, as we now explain.

To sample the configuration model, we start with a degree sequence $(d_1, \ldots, d_n)$ of a graph. Then we take n vertices with $d_1, \ldots, d_n$ half-edges respectively. After that we randomly pair the half-edges to sample the configuration model. Further, conditional on the resulting graph being simple we obtain a uniform random graph with the given degree sequence. A phase transition for these random models is proven by Molloy and Reed [46]. Bhamidi and Sen [5] and Conchon–Kerjan and Goldschmidt [19] both address the scaling limits for critical configuration models, but take quite different approaches. Bhamidi and Sen [5] addresses deterministic degree sequences whereas Conchon–Kerjan and Goldschmidt [19] addresses i.i.d. random degree sequences, who both show (with slight caveats) that the objects $M(\sigma)$ show up as the scaling limit, under appropriate assumptions.

An inhomogeneous random graph is similar to an Erdős–Rényi graph, in that different edges are present or absent independently of each other, but the probability of inclusion for each edge may vary. Bollobás, Janson, and Riordan [11] prove a phase transition for these models. Bhamidi, Sen, and Wang [6] and Broutin, Duquesne, and Wang [13] show that the objects $M(\sigma)$ also appear as the scaling limit for critical rank-1 graphs, a type of inhomogeneous random graph, under appropriate assumptions.

The overarching philosophy of this thesis is that this universality should also be present for corresponding models of directed random graphs, and this will be addressed in Section 1.2.4. Before we do so, we should first go over background material, motivating examples and definitions for directed graphs.

1.2 Directed graphs

When working with directed graphs (digraphs) there are two notions of connectivity - strong and weak connectivity. We will be working with the strong notion. Let D be a digraph. Then we say a vertex $w \in D$ is *reachable* from a vertex $v \in D$ if there exists a directed path from v to w. We say v and w are *strongly connected*, written $v \leftrightarrow w$, if both v is reachable from w and w is reachable from v. Then $\leftrightarrow$ is an equivalence relation on the vertex set and the equivalence classes are the *strongly connected components* (SCCs). We will use the term SCC interchangeably for the set of vertices and the digraph they induce. Finally, D is itself strongly connected if it has exactly one SCC, or in other words any pair of vertices in D are strongly connected.

1.2.1 Directed Erdős–Rényi model

The *directed Erdős–Rényi model* $D(n,p)$ is the random graph on n vertices where each of the possible $n(n-1)$ directed edges are present with probability p independently of the other edges. Łuczak [45] and Karp [40] showed this model exhibits a phase transition in the size of the largest SCC similar to that of the largest component of $G(n,p)$. We rephrase the result from Karp [40].

Theorem 1.4. *Let $C_1^n, C_2^n, \ldots$ be the SCCs of $D(n,p)$ in decreasing order of size where $p = \frac{c}{n}$.*

1. *If $c > 1$ then $|C_1^n| = \Theta_p(n)$ and $|C_2^n| = O_p(\log n)$.*
2. *If $c < 1$ then $|C_1^n| = O_p(\log n)$.*

Hence, taking inspiration from the undirected case, it is reasonable to try and find a scaling limit for SCCs in the critical phase.

In defining the scaling limit of components of an undirected graph we used metric spaces. For the scaling limit of SCCs we face two problems. The first is that metric spaces can't capture the directionality of edges. Metrics must be symmetric. The second more subtle but, perhaps, more fundamental problem is that SCCs are not fractal structures. SCCs in digraphs and components in undirected graphs look very different.

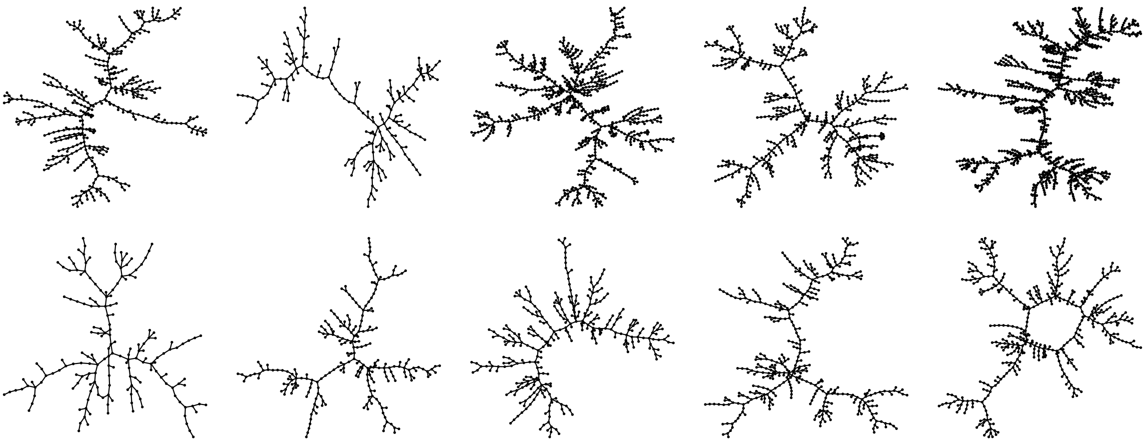
To illustrate this point, Fig. 1.1a shows some samples of the largest critical component in the undirected case. As can be seen, the components of the undirected graph exhibit a complex tree structure. In contrast, Fig. 1.1b shows the SCCs of critical directed graphs consist of finitely many cycles.

A much better point of comparison is the 2-core of the undirected components. The 2-core of a graph G is defined to be the maximal subgraph of G with minimum degree 2. Algorithmically it can be obtained by repeatedly deleting leaves from the graph until no more leaves are left, including new leaves created by previous deletions. A connected graph can be decomposed into its 2-core and then a collection of pendant trees 'hanging' off the 2-core. Thus 2-cores show the cyclical structure of G, much as SCCs show the cyclical structure of digraphs.

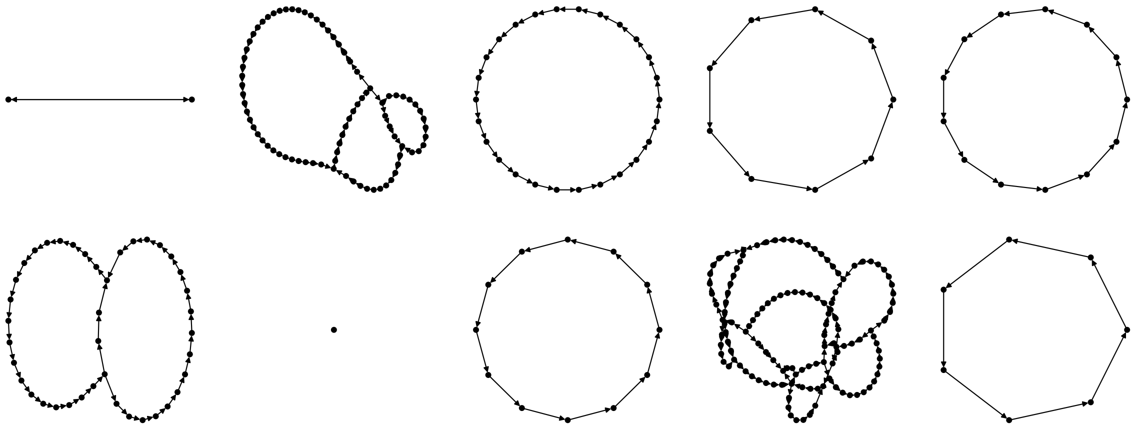
Sample 2-cores for the critical undirected case are plotted in Fig. 1.1c. Since no scaling on the surplus is required in Theorem 1.2, the surpluses of the critical components remain asymptotically bounded. Moreover, for a connected graph G with surplus s, the number of cycles in G can be bounded by 2^s. This can be proved by picking a spanning tree of G and noting that every distinct cycle in G must use a distinct subset of edges not in the spanning tree. In particular even though the lengths of paths in the 2-core are diverging to infinity, the actual number of cycles in the 2-core will remain finite.

Since the 2-cores have finite structure, it is excessive to view them as metric spaces. Instead, to capture their structure, we will replace paths in the graph by single edges weighted by the length of the path they are replacing. This will result in a weighted multigraph. We can then define a distance between weighted multigraphs by taking an ℓ^∞ distance between the weights on the edges if the multigraphs are isomorphic, or setting the distance to be infinite if they are not isomorphic. We can then scale the graphs by scaling the weights on the edges.

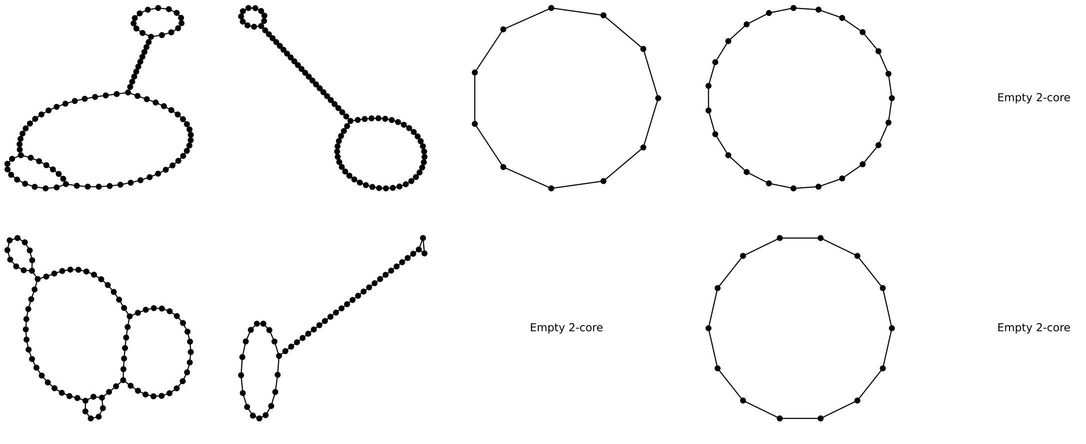
It is reasonable to conjecture that the structure of the SCCs will also remain finite in the limit. Then the above method is easily adapted to the directed case by simply working with weighted directed multigraphs instead. This is the approach taken by Goldschmidt and Stephenson [34].



(a) The largest component from samples of $G(n, 1/n)$



(b) The largest SCC from samples of $D(n, 1/n)$



(c) The 2-core of the largest component from samples of $G(n, 1/n)$

Fig. 1.1: Plots of samples of critical random graphs. In all cases $n = 10000$.

While the 2-cores provide motivation, the formal relationship between the 2-cores of the undirected Erdős-Rényi model and the SCCs of the directed Erdős-Rényi model remains an open question. It remains to be seen if the limiting distribution of the sequence of 2-cores is absolutely continuous with respect to the limiting distribution of the sequence of SCCs (once the direction of edges in the SCC are removed), and, if so, what the density function would be.

We will go through the notation and definitions in the directed case more formally. A *directed multigraph* is a tuple (V, E, r) where

1. V is a finite set of *vertices*,
2. E is a finite set of *edges*, and
3. $r : E \to (V, V)$ maps E to the set of all directed edges in the graph.

Associated with r are two functions $r_1, r_2 : E \to V$ such that $r(e) = (r_1(e), r_2(e))$. Then the *tail* of an edge e is $r_1(e)$ and the *head* of e is $r_2(e)$. It's more appropriate to think about the weights as lengths in our case, so rather than using the term 'weighted directed multigraph' we will use the term *metric directed multigraph* (MDM). An MDM is a tuple (V, E, r, l) where (V, E, r) is a directed multigraph and $l : E \to [0, \infty)$ assigns a *length* to each edge. We use $\mathfrak{L}$ to denote the MDM which consists of a single vertex and a self-loop of length 0.

Consider two MDMs $M = (V, E, r, l)$ and $M' = (V', E', r', l')$. An *isomorphism* (i_V, i_E) between M and M' is a pair of bijections $i_V : V \to V'$ and $i_E : E \to E'$ such that for all $e \in E$

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

If such an isomorphism exists then we say that M and M' are *isomorphic*.

We define a distance $d_{\vec{\mathcal{G}}}$ between M and M' by

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

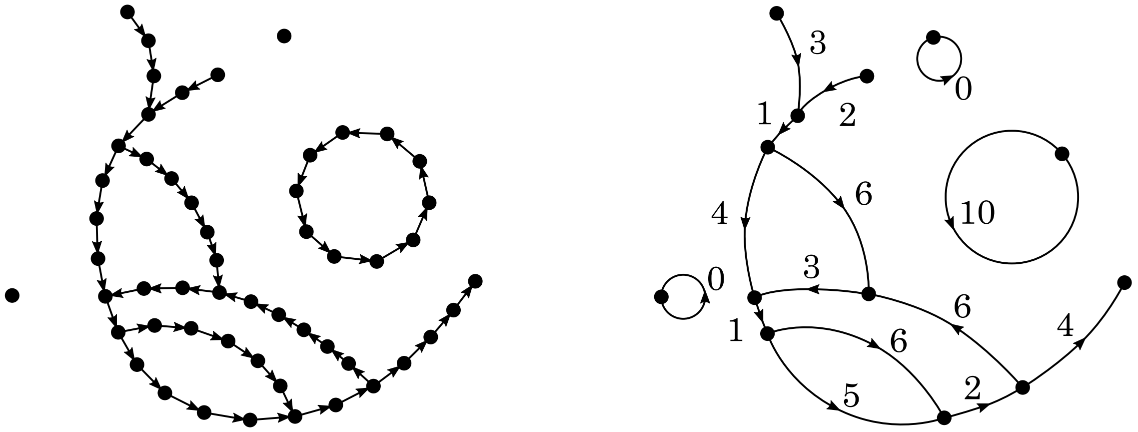
where the infimum is taken over all isomorphisms (i_V, i_E) between M and M'.

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 $r(e_2) = (w, v)$,

(a) The graph before smoothing w

(b) The graph before smoothing w

Fig. 1.2: Smoothing a vertex w



(a) D

(b) $\kappa(D)$

Fig. 1.3: An example of a digraph D and its kernel $\kappa(D)$

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 Fig. 1.2. Then the *kernel* of a digraph D is obtained by doing the following:

1. Assign a 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.
3. Replace all singletons by $\mathfrak{L}$.

Denote the resulting MDM by $\kappa(D)$. An example is shown in Fig. 1.3. We can now state the main theorem in [34].

Theorem 1.5. *Let $C_1^n, C_2^n, \ldots$ be the SCCs of $D(n,p)$ in descending order of size. We pad this sequence by an infinite tail of singletons. Then there exists a sequence of random MDMs $M_1, M_2, \ldots$ such that*

$$n^{-1/3}(\kappa(C_1^n), \kappa(C_2^n), \ldots) \xrightarrow{\text{(d)}} (M_1, M_2, \ldots)$$

as $n \to \infty$ with respect to an ℓ^1 distance using $d_{\vec{\mathcal{G}}}$. The scaling is applied to the lengths of edges in the MDM.

The description of the limiting random MDMs is rather involved and will be explained in the chapter on the directed configuration model. Also, for any fixed i, we expect C_i^n to not be a singleton with high probability as $n \to \infty$. Therefore, M_i will not be a singleton for any i. Hence, we insist that the kernel of a singleton is $\mathfrak{L}$ so that it can be isomorphic to the limiting object.

1.2.2 Directed configuration model

Description of the model

We adapt here the introduction from [26]. 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 [20]. 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 [20, 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.

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 $\mathbf{D}_1, \ldots, \mathbf{D}_n$ 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 $n \to \infty$.

Criticality

Let us examine when such a random graph is critical. 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 a 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[1] with offspring distribution Z^+. It is well known that such trees exhibit critical behaviour when $\mathbb{E}[Z^+] = 1$. This is equivalent to assuming $\mathbb{E}[D^-D^+] = E[D^-]$.

Cooper and Frieze [20] studied this phase transition for a deterministic degree sequence $\mathbf{d}_1, \ldots, \mathbf{d}_n$. 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

---

[1] 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. [2].

on whether d is strictly greater than or less than 1. In Chapter 2 we show corresponding condition, $\mathbb{E}[Z^-] = 1$, is also the correct criticality condition to take for i.i.d. random degree sequences.

1.2.3 Inhomogeneous random directed graphs

Description of the model

Inhomogeneous directed random graphs are a natural generalisation of directed Erdős-Rényi random graphs where we allow the connection probability between two vertices i and j to depend on i and j. We do this by assigning two weights to each vertex which determine how likely an edge is to exist coming into and going out of the vertex respectively. This is formally described below.

Our work will draw heavily on Broutin et al. [13] and [12]. We will consider a similar model of an inhomogeneous random graph but modified for the digraph setting. Consider a *bi-weight sequence* $\mathbb{W}_n = (\mathbf{w}_1^n, \dots, \mathbf{w}_n^n)$, where each $\mathbf{w}_i^n \in (0, \infty) \times (0, \infty)$ is a *bi-weight*: $w_i^{n,-}$ is the *in-weight* and $w_i^{n,+}$ is the *out-weight* of vertex i. Further let $l_n^\pm = \sum_{i=1}^n w_i^{n,\pm}$ and $l_n = l_n^- + l_n^+$. We say D is distributed as a *inhomogeneous random directed graph* (IRDG) on n vertices with bi-weight sequence $\mathbb{W}_n$, written $D \sim D(\mathbb{W}_n)$, if D is a random digraph on $[n]$ where each directed edge (i, j) is present independently with probability

$$\mathbb{P}((i,j) \in D) = 1 - \exp\left(-\frac{w_i^{n,+} w_j^{n,-}}{l_n}\right).$$

This is the directed analogue to the undirected random Poissonian graphs, also known as the Norros-Reittu model. For the largest connected components in the undirected case, a metric space scaling limit is established by Broutin et al. [13].

A more general model of directed inhomogeneous random graphs using a kernel function is studied by Cao and Olvera-Cravioto [17]. While we would also like to study more general graphs, even in the undirected case the metric space scaling limit of the connected components is only well understood for rank-1 graphs. Therefore, for now, we are restricting our attention to the Poissonian case.

We see that for any pair of distinct vertices i and j, a higher in-weight $w_i^{n,-}$ will lead to a higher probability of an edge (j, i) leading into i existing, whereas a higher out-weight $w_i^{n,+}$ will lead to a higher probability an edge (i, j) leading out of i existing. We will be working in regimes where

$w_i^{n,+}w_i^{n,-}/l_n$ will be small, thus the approximation

$$\mathbb{P}((i,j) \in D) \approx \frac{w_i^{n,+}w_j^{n,-}}{l_n}$$

will be valid.

The directed Erdős-Rényi random $D(n,p)$ is a special case of the IRDG $D(\mathbb{W}_n)$ where all weights are equal. Specifically, setting $\mathbf{w}_i = (c,c)$ for all $i = 1, \ldots, n$ for $c \in \mathbb{R}$ corresponds to a $D(n, 1-e^{-c/2n})$ random graph. For large n this is approximately a $D(n, \frac{c}{2n})$ random graph.

Criticality

Cao and Olvera-Cravioto [17] studied whether or not the largest SCC will take up an asymptotically positive proportion of the vertices for general directed inhomogeneous random graphs. We will rephrase their results here for IRDGs. Let μ_n be the measure on $\mathbb{R}^2$ which is the empirical measure of the in- and out-weights, meaning

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(w_i^{n,-}, w_i^{n,+})}.$$

Then the following is an application of [17, Proposition 3.13].

Theorem 1.6. *Suppose there exists a Borel probability measure supported on* $(0,\infty)\times(0,\infty)$ *such that* $\mu_n \to \mu$ *weakly. We also assume the following:*

1. *There exists a positive constant* θ *such that* $\frac{l_n}{n} \to \theta$ *as* $n \to \infty$*.*
2. *The following two limits hold and are finite:*

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{w_i^{n,+}w_j^{n,-}}{\theta} &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\left(1 - e^{-w_i^{n,+}w_j^{n,-}/l_n}\right) \\ &= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \frac{w_1^+ w_2^-}{\theta}\mu(\mathrm{d}\mathbf{w}_1)\mu(\mathrm{d}\mathbf{w}_2).\end{aligned}$$

Let C_1^n *be the largest SCC of* $D(\mathbb{W}_n)$*. Then there exists a constant* ρ *such that*

$$\frac{|C_1^n|}{n} \xrightarrow{(\mathrm{p})} \rho.$$

Further $\rho > 0$ *if and only if* $\mathbb{E}[W^-W^+] > \theta$ *where* $(W^-, W^+) \sim \mu$*.*

We have that $\mathbb{E}[W^- W^+] \approx \frac{1}{n}\sum_{i=1}^n w_i^{n,-} w_i^{n,+}$ and $\theta \approx \frac{1}{n} l_n$. This suggests we have criticality when

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n w_i^{n,-} w_i^{n,+}}{l_n} = 1.$$

The exact rate of this convergence will determine whether or not we have criticality; specifically, whether or not the largest SCCs are all $\Theta_p(n^{1/3})$ in size.

To conjecture a critical window for this model, recall that we see critical behaviour for the Erdős-Rényi random graph $D(n, p_n)$ when $p_n = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$ for some $\lambda \in \mathbb{R}$. Using the approximation $1 - e^{-x} \approx x$ for small x, this approximately corresponds to a IRDG where $w_i^{n,-} = w_i^{n,+} = 2 + \frac{2\lambda}{n^{4/3}}$. Hence,

$$\frac{\sum_{i=1}^n w_i^{n,-} w_i^{n,+}}{l_n} = 2\lambda n^{-1/3}.$$

Thus, we conjecture we get critical behaviour, meaning the largest SCCs are all of order $\Theta(n^{1/3})$ in size, when there exists a constant $\alpha \in \mathbb{R}$ such that

$$\left(1 - \frac{\sum_{i=1}^n w_i^{n,-} w_i^{n,+}}{l_n}\right) n^{1/3} \to \alpha \tag{1.2}$$

as $n \to \infty$. This will be our conjectured criticality condition. This mirrors the assumption made in [13] to get a metric space scaling limit in the undirected case.

1.2.4 Aim of the thesis

Since the directed configuration model and IRDGs both exhibit a phase transition in the size of the largest SCC, it is reasonable to conjecture that the scaling limits we obtain for the directed Erdős–Rényi model will also appear at criticality under certain moment conditions on the vertex degrees. The aim of this thesis is to demonstrate this universality. In particular, we conjecture the following to be true:

Conjecture 1.7. *Let $C_1^n, C_2^n, \ldots$ be the SCCs of a random graph on n vertices sampled from either*

1. *the directed Erdős–Rényi model, or*
2. *the directed configuration model, or*
3. *the inhomogeneous random directed graph model*

at criticality. Suppose $C_1^n, C_2^n, \ldots$ are listed in decreasing order of size. We pad this sequence by an *infinite tail of singletons. Then there exists a four parameter family of random sequences of MDMs* $\mathcal{M}(a,b,c,d) = (M_1, M_2, \ldots)$ *such that, under appropriate additional assumptions, there is a choice of parameters* a, b, c, d *such that*

$$n^{-1/3}(\kappa(C_1^n), \kappa(C_2^n), \ldots) \xrightarrow{(\mathrm{d})} \mathcal{M}(a,b,c,d)$$

as $n \to \infty$ *with respect to an* ℓ^1 *distance using* $d_{\vec{\mathcal{G}}}$*. The scaling is applied to the lengths of edges in the MDM. Further*

The part of the result on the directed Erdős–Rényi model has been proved by Goldschmidt and Stephenson [34] and is stated here in Theorem 1.5. No additional assumptions are needed. The part of the result for the directed configuration model is proven in joint work with Serte Donderwinkel in [26]. We present this result in Chapter 2, along with a detailed explanation of the additional assumptions. The part of the result on IRDGs remains a conjecture. In Chapter 3, we present a conjecture scaling limit and the additional assumptions we believe would be required. Under these assumptions, we prove a scaling limit for the height process of a subforest of the IRDG. We also summarise what remains to be proved in order to prove the conjecture. We discuss the parameters in more detail in Section 1.3.3 after we introduce background material on discrete and real trees.

Our proof techniques rely on exploring the out-forest of a directed graph and showing a Gromov-Hausdorff convergence of the out-forest to a sequence of real trees. The exploration gives a planar order to the vertices making this is subgraph of the directed graph a sequence of plane trees. These plane trees have the property that for any vertex v, all vertices w that can be reached by v and have not yet appeared in the planar order will be a descendant of v. The interaction between edges in the out-forest and a subset of edges called *candidates* (see Section 2.2.1) will create the SCCs.

In the last section of this chapter, we will cover some background material and terminology on both discrete and real trees which is essential for the rest of the work in this thesis.

1.3 Trees

1.3.1 Plane Trees

Basic definitions of plane trees

A *plane tree* is a rooted finite tree where an order is specified for the children of each vertex. To formalise this it is convenient to use *Ulam-Harris* labelling for the vertices. The ideas is to label each vertex by a finite sequence of positive integers where $(i_1, \ldots, i_{k+1})$ will label the i_{k+1}-th child of the vertex labelled $(i_1, \ldots, i_k)$. Let

$$\mathcal{U} = \bigcup_{n=0}^{\infty} \mathbb{N}^n$$

where $\mathbb{N}^0 = \{\emptyset\}$ is a singleton set containing the empty tuple $\emptyset = ()$. This is the set of all possible labels. Then a *plane tree* T is finite subset of $\mathcal{U}$ satisfying the following properties:

1. $\emptyset \in T$.
2. If $(a_1, \ldots, a_k, a_{k+1}) \in T$, for $k \geq 0$, then $(a_1, \ldots, a_k) \in T$.
3. If $(a_1, \ldots, a_k) \in T$ then $(a_1, \ldots, a_{k-1}, j) \in T$ for $j = 1, \ldots, a_k$.

These labels fully encode the graph structure of the plane tree. Then $\emptyset$ is the root vertex of T. For each $v \in T$, its *height* $h(v)$ is given by its length $|v|$, or equivalently its graph distance from the root.

The *contour walk* of the vertices is a sequence of vertices $w_1, \ldots, w_{2n-1}$ defined as follows. Let w_0 be the root $\emptyset$. Given w_i, let w_{i+1} be the smallest child of w_i that is not included in $\{w_1, \ldots, w_i\}$ if such a child exists. If such a child does not exist, let w_{i+1} be the parent of w_i. The walk will start and end at the root and visit all vertices. The *depth-first order* of the vertices is the same as the lexicographical order. It can also be obtained from the contour walk by keeping only the first occurrence of each vertex.

Coding functions of plane trees

Consider a plane tree with n vertices. Suppose the contour walk is given by $w_1, \ldots, w_{2n-1}$ and suppose the vertices in depth-first order are given by $v_1, \ldots, v_n$.

The *contour process* (referred to as the *search depth* in [3]) is the function $f : [2n-1] \to \mathbb{N}$ where $f(i) = h(w_i)$. This is related to the *Harris path*, which is the function $\tilde{f} : \{0, \ldots, 2n\} \to \mathbb{N}$ where

$\tilde{f}(0) = \tilde{f}(2n) = 0$ and $\tilde{f}(i) = f(i) + 1$ for $i \in [2n-1]$.

The *height process* is the function $h : [n] \to \mathbb{N}$ where $h(i) = h(v_i)$ (this overloads the symbol h, but it should be clear from context which function is being used).

The *Łukasiewicz path* is a function $g : \{0, \ldots, n\} \to \mathbb{N}$ defined inductively as follows:

1. $g(0) = 0$.
2. $g(i+1) = g(i) +$ number of children of $v_{i+1} - 1$.

We can obtain the height function from the Łukasiewicz path as follows:

$$h(k) = \#\left\{ j = 1, \ldots, k : g(j) = \min_{i=\{j,j+1,\ldots,k\}} g(i) \right\}. \tag{1.3}$$

The contour walk is obtained from the DFS order by backtracking. Similarly, we can obtain the contour function from the height process by adding sections to the path. Suppose $h(i+1) \neq h(i) + 1$. Then we decrease in steps of 1 until we reach $h(i+1) - 1$, then go up to $h(i+1)$.

Finally, we can recover the plane tree from the contour process. To formalise this process, define the pseudometric

$$\tilde{d}(i,j) = f(i) + f(j) - 2 \min_{k \in \{i \wedge j, \ldots, i \vee j\}} f(k).$$

Let $\sim$ be the equivalence relation where $i \sim j$ if $\tilde{d}(i,j) = 0$. Let $V = [2n-1]/\sim$ and let $\mathrm{cl}(i)$ denote the equivalence class of i. The function d induces a metric d on V where $d(\mathrm{cl}(i), \mathrm{cl}(j)) = \tilde{d}(i,j)$ which is the graph distance of a tree with vertex set V. We can recover the edges of the tree by adding an edge between any two vertices that are at d-distance 1 from each other. We root the tree at $\mathrm{cl}(1)$. Finally, we order the children of each vertex by the smallest index in their equivalence class in order to recover our plane tree.

Hence, the height process, contour walk and Łukasiewicz path all fully encode the structure of the plane tree.

These definitions can all be extended to a sequence of plane trees by concatenating the functions for each tree side by side in the order they appear in the sequence.

1.3.2 Real trees

The continuum limit of the out-forests we construct will be a sequence of real trees.

Definition of real trees

A *real tree* (or $\mathbb{R}$-*tree*) is a metric space $(\mathcal{T}, d)$ such that for any two points $a, b \in \mathcal{T}$ the following properties hold:

1. There exists a unique isometry $f_{a,b} : [0, d(a, b)] \to \mathcal{T}$ such that $f_{a,b}(0) = a$ and $f_{a,b}(d(a, b)) = b$.
2. If $q : [0, 1] \to \mathcal{T}$ is any injective continuous map such that $q(0) = a$ and $q(1) = b$ then q and $f_{a,b}$ have identical images.

Real trees via continuous excursions

Let $f : [0, l] \to [0, \infty)$ be a continuous function such that $f(0) = f(l) = 0$. We describe now how to obtain a real tree from such a function. For $s, t \in [0, l]$, define the pseudometric

$$d(s, t) = f(s) + f(t) - 2 \min_{r \in [s \wedge t, s \vee t]} f(r).$$

Let $\sim$ be the equivalence relation where $s \sim t$ if $d(s, t) = 0$. Then $\mathcal{T} = [0, l]/\sim$ equipped with the metric induced by d is a real tree and $\mathrm{cl}(s)$ be the equivalence class containing s. The pseudometric d induces a metric $\tilde{d}$ on $\mathcal{T}$ by $\tilde{d}(\mathrm{cl}(s), \mathrm{cl}(t)) = d(s, t)$. Then $(\mathcal{T}, \tilde{d})$ is a real tree encoded by f. This is akin to the how we recover plane trees from a discrete contour function, as described in Section 1.3.1. In fact we can linearly interpolate the discrete contour functions to a continuous function, and the corresponding real tree will be the metric space obtained by adding line segments of length 1 between connected vertices in the discrete tree.

The following theorem, taken from [44, Lemma 2.3], shows the encoding is continuous with respect to the uniform norm on functions and the Gromov-Hausdorff topology on real trees.

Theorem 1.8. *Let f and f' be two continuous functions $[0, l] \to [0, \infty)$ which both start and end at 0. Let $\mathcal{T}$ and $\mathcal{T}'$ be the real trees encoded f and f' respectively. Then*

$$d_{GH}(\mathcal{T}, \mathcal{T}') \leq \sup_{t \in [0, l]} |f(t) - f'(t)|.$$

The Łukasiewicz path in the continuum

The above theorem allows us to prove convergence of random discrete trees to random real trees in distribution with respect to the Gromov-Hausdorff topology if we can prove convergence of the contour processes. However, it is not straightforward to directly prove convergence of the contour process. Instead, it is often easier to first prove convergence of the Łukasiewicz path because the Łukasiewicz path is constructed as a random walk, to which we can apply tools such as Donsker's theorem or martingale convergence theorems.

Duquesne and Le Gall [27] cover in great detail the case where the Łukasiewicz path will converge to a Lévy process X which almost surely does not drift to $+\infty$ and has paths of infinite variation. Following Eq. (1.3), we define the corresponding height process to be the 'measure' of the set

$$\left\{ s \leq t : X_s = \inf_{s \leq r \leq t} X_r \right\}.$$

However, the Lebesgue measure of this set will be 0, and we instead need to consider a local time. Under additional assumptions, they show there is a continuous process H such that for every fixed $t \geq 0$, the following limit holds in probability:

$$H_t = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^t \mathbb{1} \left\{ X_s \leq \inf_{s \leq r \leq t} X_r + \epsilon \right\}.$$

We refer to this as the *continuous height process* associated with X. In particular, in [27, Section 1.2], they showed that if $X_t = -\alpha t + \sigma B_t$ where $\alpha \in \mathbb{R}$, $\sigma > 0$ and B is a standard Brownian motion, then H_t can take the form

$$H_t = \frac{2}{\sigma^2} \left(X_t - \inf_{s \leq t} X_s \right).$$

In the continuum, the difference between the contour process and the height process becomes blurred. In [27, Section 2.4], it is shown for a sequence of critical Galton-Watson trees that is the height processes converge to H when rescaled then the contour process will converge to $(H_{t/2})_{t \geq 0}$ under the same rescaling. Thus the contour process and height process in the continuum are the same up to a 1/2 rescaling in the argument.

1.3.3 The parameters of the scaling limit

Recall the 4 parameter family of random sequences of MMDs $cM(a, b, c, d)$ in Conjecture 1.7. The parameters a, b and c will correspond to the linear term, quadratic term and Brownian motion of the Łukasiewicz path $X(t) = at - \frac{1}{2}bt^2 + cB_t$ that will be used to construct the scaling limit of the out-forest.

The parameter d is used to control the intensity of the Cox processes used to construct point identifications on the real trees, corresponding to the candidate edges. A formal construction of the limiting object is found in Section 2.2.2 The value of these parameters for the directed Erdős-Rényi model is $a = \lambda$ and $b = c = d = 1$. The value of these parameters in the context of the configuration model can be found in Section 2.2.2. For IRDGs, conjectured values of a, b and c are given in Section 3.1.3.

Chapter 2

Directed Configuration Model

2.1 Introduction

This chapter is joint work with Serte Donderwinkel. We have submitted a version [26] which has been published in the Electronic Journal of Probability.

We recall here the definition of the the model $\vec{G}_n(\nu)$ from Section 1.2.2. Let ν be a distribution on $\mathbb{N} \times \mathbb{N}$. We then sample $\mathbf{D}_1, \ldots, \mathbf{D}_n$ as i.i.d. samples from ν conditioned on the event that

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

Then $\vec{G}_n(\nu)$ is a digraph chosen uniformly at random from all digraphs with degree sequence $\mathbf{D}_1, \ldots, \mathbf{D}_n$.

2.1.1 Assumptions

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

1. $\mathbb{E}[(D^-)^i (D^+)^j] < \infty$ for $1 \le i + j \le 3$, $(i,j) = (1,3)$ 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. $\mathbb{E}[D^- D^+] = \mathbb{E}[D^-]$ or $\mathbb{E}[D^- D^+] = \mathbb{E}[D^+]$, where both statements are equivalent supposing the second condition holds.

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.

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 [54, Page 42, P1], the third condition ensures that the event has positive probability for all sufficiently large $n \ge 1$. 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, which is discussed in Section 1.2.2.

Recall that (Z^-, Z^+) is distributed as (D^-, D^+) size-biased by D^-, that is

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

for all $k^-, k^+ \in \mathbb{N}$.

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

2.1.2 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.

Theorem 2.1. *There exists a sequence $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ of random strongly connected MDMs such that*

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

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 2.2.2.

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 [34]. 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.

Theorem 2.2. *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 Theorem 2.2 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 2.1 has the following immediate corollaries, which were previously unknown.

Corollary 2.3. *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 $(\mathbb{R}^2)^{\mathbb{N}}$.

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

Corollary 2.4. *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, $\text{Diam}\left(\vec{G}_n(\nu)\right) = \Omega_p(n^{1/3})$.

This follows because the diameter is at least the length of the shortest path between any two vertices in the largest SCC. We know from the scaling limit that there are points in the SCC where the distance between them scales like $n^{1/3}$, thus the diameter grows at least at quickly as $cn^{1/3}$ for some $c > 0$. The actual scaling rate of the diameter is still an open question.

2.1.3 Previous work

The configuration model was introduced by Bollobás [10] 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 [55, 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 [46, 47, 37]. The law of component sizes at criticality and in the critical window were obtained by Riordan [50] 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 [23] and in the heavy-tailed regime [24]. Bhamidi, Dhara, van der Hofstad and Sen obtained metric space convergence in the critical window in [7], a result that the authors later improved to a stronger topology in [8].

Configuration models with a random degree sequence are considered in [39], [19], and [25]. Joseph [39] 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 [19] show Gromov-Hausdorff-Prokhorov convergence of the rescaled components ordered by decreasing size at criticality in these two regimes. The results in [19] in the heavy-tailed regime are extended to the critical window by the first author in [25]. Our techniques are closely related to the techniques introduced in [19].

Some results have been obtained for other directed graph models. Cao and Olvera-Cravioto [17] 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 [9], in which a smaller class of inhomogeneous directed graphs is considered. Samorodnitsky, Resnick, Towsley, Davis, Willis and Wan [51] studied the tails of the degree distribution in the directed preferential attachment model. As previously mentioned, Goldschmidt and Stephenson [34] 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 [20]. They consider a deterministic degree sequence under a number of conditions. As discussed previously in Section 1.2.2, a phase transition for the SCCs occurs when a parameter d is equal to 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 [35]. These results are in contrast with the critical case, with Corollary 2.3, 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 [16], 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 2.4, which says that in our set-up the diameter is $\Omega(n^{1/3})$ in probability at criticality. Then, in [14], 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 [15], they study the behaviour of a random walk on a directed configuration model.

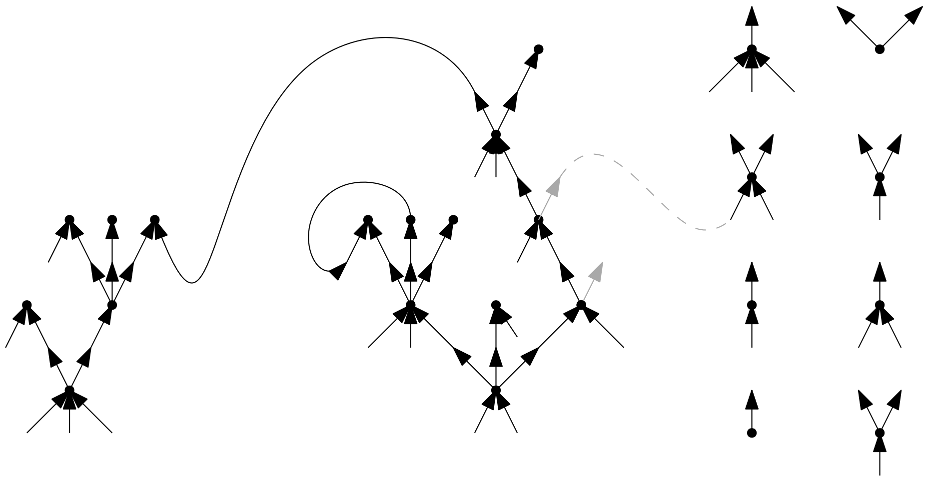
A necessary and sufficient condition for the existence of a giant weakly connected component for the directed configuration model with a deterministic degree sequence is discussed in the physics literature by Kryven [41]. He also studies the distribution of the in- and out-components in [42].

The directed configuration model with random in- and out-degrees is also considered by Chen and Olvera-Cravioto [18] 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.

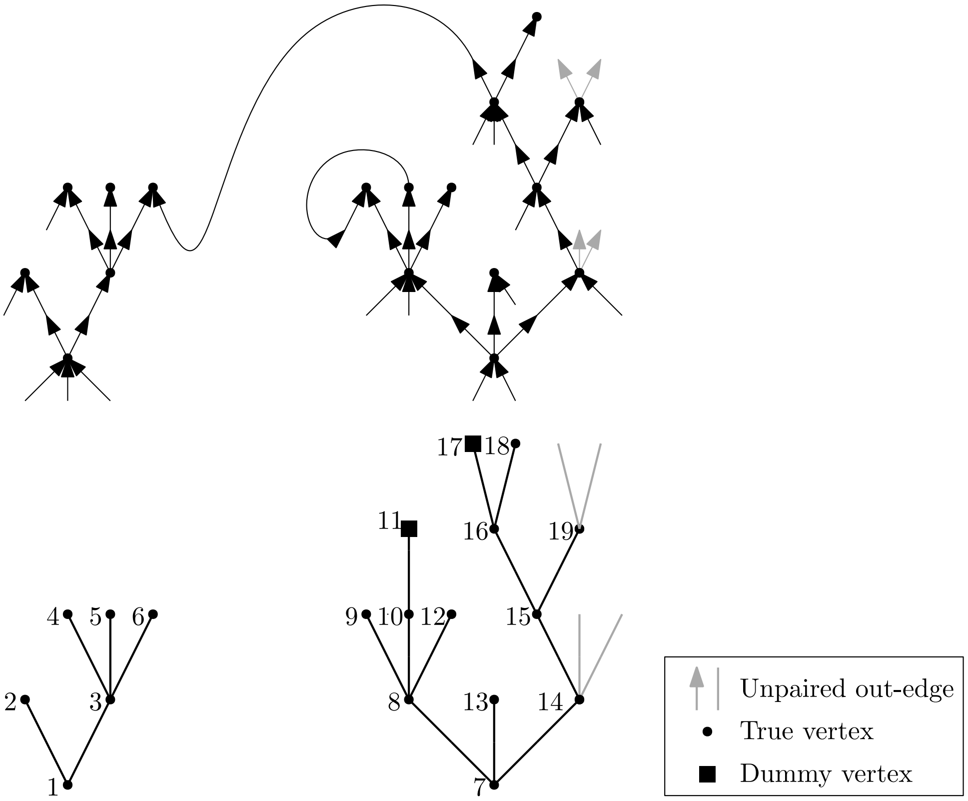
2.1.4 Proof outline

We will study the law of the SCCs of a uniform directed graph with degree sequence $(\mathbf{D}_1, \dots, \mathbf{D}_n)$, 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 2.1a. The definition of the out-forest is illustrated in Figure 2.1b.

The sampling procedure uses a queue of unpaired out-edges (represented by the label of their



(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. 2.1: Partial constructions of the configuration model and out-forest

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, $\hat{s}^+(k)$ is akin to a Łukasiewiscz 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 2.1b.

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 2.5. 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

```
Data: A set of vertices $V = \{v_1, \dots, v_n\}$ with degree pairs
      $(d^-(v_1), d^+(v_1)), \dots, (d^-(v_n), d^+(v_n))$ 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;
$(d^-_{\text{unpaired}}(v_1), \dots, d^-_{\text{unpaired}}(v_n)) \leftarrow (d^-(v_1), \dots, d^-(v_n))$          // 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 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 2.6 and Proposition 2.7), 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.

We note the following key facts. Firstly, the order in which the true vertices are discovered does not depend on the positions of the dummy leaves. Secondly, 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 2.2.1. The analogous sampling procedure for the limit object is described in Subsection 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 2.34. 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 2.3 , 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 2.52, Lemma 2.54, and Proposition 2.55.
3. We show that the positions of the heads of the surplus edges corresponding to the candidates converge, which is the content of Proposition 2.56.
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 [34] to show that the cutting procedure converges. This summarised in Corollary 2.60.

5. 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 2.48.

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 2.61.

2.2 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. [46, 47, 37] 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 2.2.1, 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 2.2.1 and 2.2.1 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 2.2.2, 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 2.2.1 and 2.2.1.

2.2.1 The discrete case

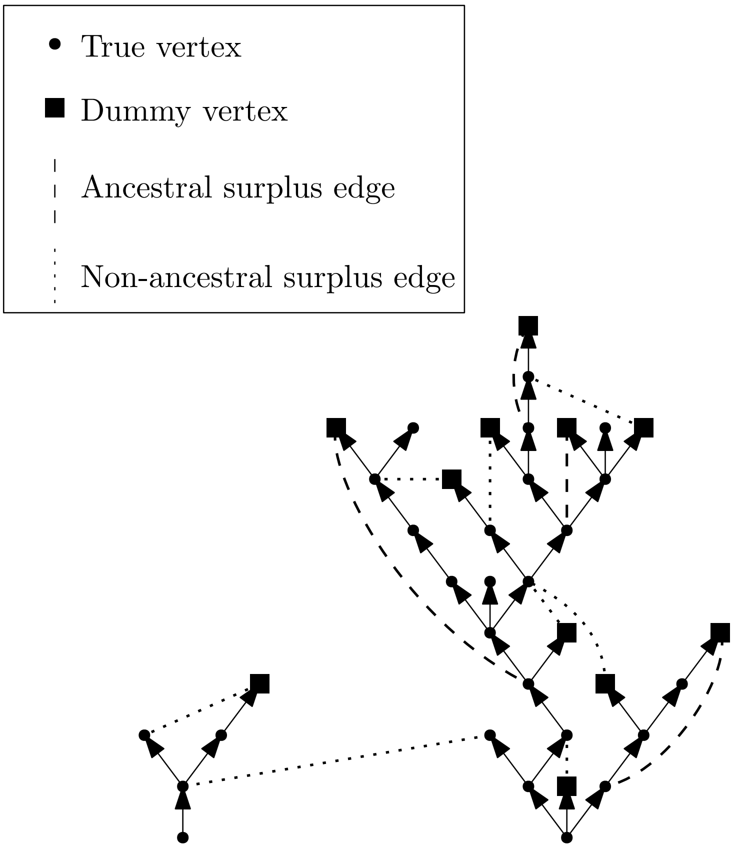
We will discuss the different type of edges that we can encounter in the exploration. Recall from Subsection 2.1.4 that by slight abuse of terminology, we call the dummy leaf that corresponds to a surplus edge its tail.

Necessary conditions for an edge to be part of an SCC

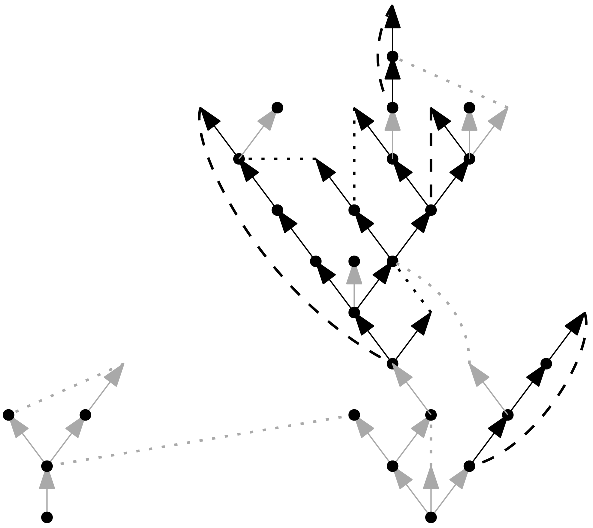
Amongst the surplus edges, *ancestral surplus edges*, which are surplus edges that point from a vertex to one of its ancestors, play a special role. All other surplus edges are called *non-ancestral*. This is illustrated in Figure 2.2a. In Figure 2.2b we show how surplus edges affect the structure of the SCCs. This is the content of the next lemma.

Lemma 2.5. *The following facts hold for SCCs.*

1. *The vertices of an SCC are contained in precisely one of the components of the out-forest.*



(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. 2.2: We illustrate the different types of surplus edges and how they affect the structure of the SCCs.

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 2.5 motivates the following definition.

Definition 2.6. *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 an ancestor of a candidate.*

The following proposition is at the core of our strategy to study the SCCs.

Proposition 2.7. *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 2.6 and Lemma 2.5. □

Proposition 2.7 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.

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 2.3. The out-forest is obtained in the following way. Let $(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,n})$ 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 $(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,n})$. 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 $\widehat{D}^+_{n,m+1}$.
2. Otherwise,
    (a) With probability proportional to the total in-degree of the undiscovered vertices, i.e. proportional to $\sum_{i=m+1}^n \widehat{D}^-_{n,i}$, let the next vertex in depth-first order be a true vertex with out-degree $\widehat{D}^+_{n,m+1}$.
    (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 2.8. *Suppose that the sequence of degrees in order of discovery $(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,n})$ is given. Suppose that after time-step k, there are still unpaired out-half-edges. Suppose that, for $1 \leq l \leq k$, up* *to time l, $\hat{P}_n(l)$ surplus edges have been sampled. Then,*

$$\left(\hat{S}_n^+(l), 1 \le l \le k\right) := \left(\sum_{i=1}^{l-\hat{P}_n(l)} \widehat{D}_{n,i}^+ - l, 1 \le l \le k\right)$$

is the Łukasiewicz path of the out-forest up to time k. Moreover, for

$$\left(\hat{I}_n^+(l), 1 \le l \le k\right) := \left(\min\left\{\hat{S}_n^+(m) : 1 \le m \le l\right\}, 1 \le l \le k\right),$$

define

$$\left(\hat{S}_n^-(l), 1 \le l \le k\right) := \left(\sum_{i=1}^{l-\hat{P}_n(l)} \widehat{D}_{n,i}^- - l - \hat{I}_n^+(l) + 1, 1 \le l \le k\right),$$

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

$$\frac{\hat{S}_n^-(k+1)}{\sum_{i=1}^n D_i^- - k - \hat{I}_n^+(k) + 1} \mathbb{1}_{\left\{\hat{I}_n^+(k) = \hat{I}_n^+(k-1)\right\}}.$$

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 $(\widehat{D}_{n,1}^+, \dots, \widehat{D}_{n,k-\hat{P}_n(k)}^+)$, 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 $\sum_{i=1}^{k-\hat{P}_n(k)} \widehat{D}_{n,i}^-$. 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 $\sum_{i=1}^n D_i^- - k - \hat{I}_n^+(k) + 1$. 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. $\square$

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 2.1.4 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 $(\hat{H}_n^{\ell}(k), k \geq 1)$. Recall from Lemma 2.5 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 2.9. *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*

$$\frac{\hat{H}_n^\ell(k)}{\hat{S}_n^-(k)} \mathbb{1}_{\{\hat{P}_n(k)-\hat{P}_n(k-1)=1\}}.$$

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 $T_{g_n}^n$ 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 $T_{g_n}^n$ corresponds to a dummy leaf $V_1^n \in [g_n + 2, g_n + \sigma_n]$. Let $V_1^n < k \leq g_n + \sigma_n$, and suppose the candidates found up to time k are given by $V_1^n, \dots, V_m^n$. Let $T_k^{n,\mathrm{mk}}$ be the subtree of $T_{g_n}^n$ spanned by $\{g_n + 1, V_1^n, \dots, V_m^n, k\}$, and let $\ell(T_k^{n,\mathrm{mk}})$ be its total length with edge lengths as encoded by $(\hat{H}^\ell(i), i \in [g_n + 1, g_n + \sigma_n])$.

Proposition 2.10. *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 2.6, 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 2.11. *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 2.8, 2.9, 2.10, and 2.11 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 $\{V_2^n, \ldots, V_{N_n}^n\}$ according to the procedure described in the statement of Proposition 2.10.

(b) For the tails $V_1^n, \ldots, V_{N_n}^n$, sample their corresponding heads $W_1^n, \ldots, W_{N_n}^n$ respectively according to the procedure described in the statement of Proposition 2.10.

(c) Let $T^{n,\mathrm{mk}}(g_n)$ be the subtree of $T_{g_n}^n$ spanned by $\{g_n + 1, V_1^n, \ldots, V_{N_n}^n\}$. 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 2.3a, 2.3b and 2.3c for an illustration of this procedure.

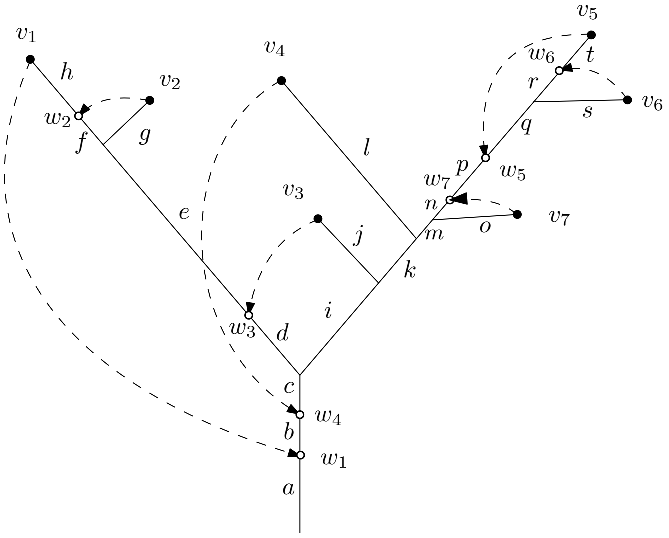
2.2.2 The continuum case

We will define 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 [34].

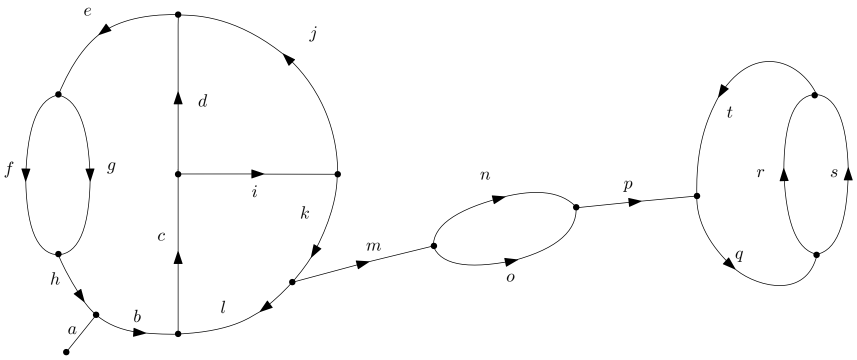
The limit object

Let $B = (B_t, t \geq 0)$ be a standard Brownian motion, and set

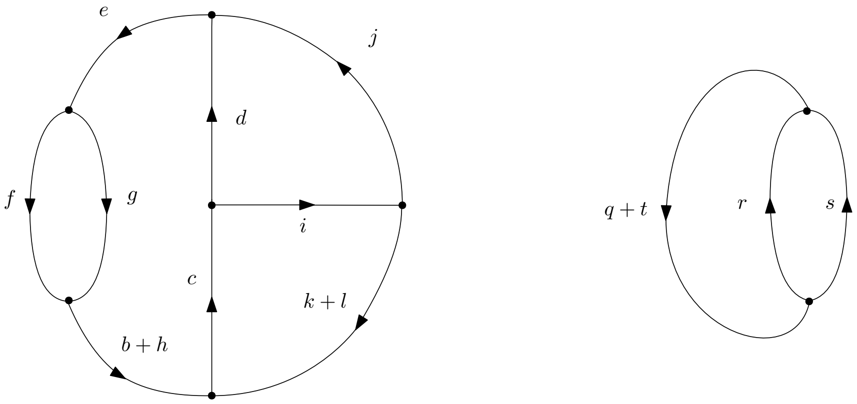
$$\hat{B} = \left(\hat{B}_t, t \geq 0\right) = \left(B_t - \frac{\sigma_{-+} + \nu_{-}}{2\sigma_{+}\mu} t^2, t \geq 0\right).$$



(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. 2.3: We illustrate the procedure to find the SCCs in a component of the out-forest after finding the candidates. Taken from [34] with permission of the authors.

Remark 2.12. *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

$$\hat{R} = (\hat{R}_t, t \geq 0) = \left(\hat{B}_t - \inf\left\{\hat{B}_s : s \leq t\right\}, t \geq 0\right).$$

Then, it follows from the argument in Section 2.4 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 $\{1, 2, \ldots\}$, set $X_i = \min\{t : A_t = i\}$. For i in $\{1, 2, \ldots\}$, 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 $\{1, 2, \ldots\}$, $\left(\frac{2}{\sigma_+}\hat{R}_t, t \in [G_i, G_i + \Sigma_i]\right)$ encodes the $\mathbb{R}$-tree in the out-$\mathbb{R}$-forest that contains X_i. For each element of $\{(G_i, \Sigma_i) : i = 1, 2, \ldots\}$ 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 [34] to our set-up. Indeed, note that $V_m \le s \le V_{m+1}$ implies that $|T_s^{\mathrm{mk}}| < (m+1)||\mathcal{T}_g||$. Therefore,

$$\mathbb{P}\left(N \ge g+1, V_{m+1}-V_m < t \middle| N \ge g\right) \le \mathbb{P}(E_{m+1} < t),$$

for $(E_k, k\ge 1)$ a sequence of exponential random variables with respective rates

$$\frac{\sigma_{-+}+\nu_-}{\mu^2}k||\mathcal{T}_g||.$$

Then,

$$\mathbb{P}\left(N=\infty\right)=\mathbb{P}\left(N=\infty \text{ and } \sup\{V_i : i\in\mathbb{N}\} < g+\sigma\right) \le \mathbb{P}\left(\sum_{i=2}^{\infty} E_k \le g+\sigma-V_1\right).$$

However, $\sum_{i=2}^{\infty} E_k=\infty$ a.s., because the harmonic series diverges, so, indeed, $\mathbb{P}\left(N<\infty\right)=1$.

Finally, for $1\le i\le N$, 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 $1\le i\le N$ 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 $\mathcal{M}_g$ consist of g, W_i for $i\le N$, and the branch points $V_i\wedge V_j$ for $i\ne j\le N$. The directions are inherited from $\mathcal{T}_g$, 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 Section 1.2.1. 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 2.1.

The parameters of the limit object

As mentioned in Section 1.3.3, we can parametrise the limiting object as $\mathcal{M}(a,b,c,d) = (\mathcal{C}_1, \mathcal{C}_2, \ldots)$ where

$$a = 0, \tag{2.1}$$

$$b = \frac{\sigma_{-+} + \nu_-}{\mu}, \tag{2.2}$$

$$c = \sigma_+, \text{ and} \tag{2.3}$$

$$d = \frac{\sigma_{-+} + \nu_-}{\mu^2}. \tag{2.4}$$

Here a, $-\frac{1}{2}b$ and c are the coefficients of the linear term, quadratic term and the Brownian motion in $\sigma_+ \hat{B}$, which is the Łukasiewic path used to construct the real tree limit of the out-forest. The value d is the coefficient the height process used to construct point identifications. We have that $a = 0$ because we are not considering a scaling window here, and are instead assuming exact criticality.

Properties of the limit object

We note that the distribution of 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 2.2.2. The limit object that is studied in [34] 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 [34], 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 2.13. *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 2.14. *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 2.2.2. Write $\mathcal{C}_{\text{cplx}}$ for the list of complex components of $\mathcal{C}$ in decreasing order of length. For sequences $(K_1, \dots, K_j)$ and $(J_1, \dots, J_k)$ of directed multigraphs, write $(J_1, \dots, J_k) \equiv (K_1, \dots, K_j)$ if $j = k$ and J_i is isomorphic to K_i for each $i \le j$. Extend this notation naturally to the case where one or both of the sequences has edge lengths by ignoring the edge lengths.

Theorem 2.15. *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 $(e_i, 1 \le i \le M)$ 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 $(\ell(e_i), 1 \le i \le M)$ to these edges, and their joint distribution has full support in

$$\left\{ \mathbf{x} = (x_1, \dots, x_M) \in \mathbb{R}_+^M : \forall 1 \le i \le j-1, \sum_{k: e_k \in E(K_i)} x_k \ge \sum_{k: e_k \in E(K_{i+1})} x_k \right\}.$$

2.3 Analysis of the measure change

Recall $(\widehat{\mathbf{D}}_{n,1}, \widehat{\mathbf{D}}_{n,2}, \ldots, \widehat{\mathbf{D}}_{n,n})$ are are the degree pairs of the vertices in order of discovery. Let R_n be the number of vertices with positive in-degree. The behaviour of the $\widehat{\mathbf{D}}_{n,m}$ 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 2.5, 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 2.16. *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 $\widehat{D}_{n,1}$ is given by $\mathbf{Z}$ where

$$\mathbb{P}(Z^- = k^-, Z^+ = k^+) = \frac{k^-}{\mu}\mathbb{P}(D^- = k^-, D^+ = k^+).$$

Similarly, $\widehat{D}_{n,2}$ is also approximately distributed like $\mathbf{Z}$ for large n, and so on. In this section we in fact prove a precise relation between $\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,m}$ and a 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

$$\Xi^{\pm}_{n-m} = \sum_{i=m+1}^{n} D_i^{\pm} \quad \text{and} \quad \Delta_n = \Xi^-_n - \Xi^+_n$$

such that $\{\Delta_n = 0\}$ is the event that the total out-degree is equal to the total in-degree.

The following proposition asserts the existence of the measure change ϕ^n_m, 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 2.17. *For all positive integers n and m such that $m \leq n$, there exists a function $\phi^n_m : (\mathbb{N} \times \mathbb{N})^n \to [0, \infty)$ such that*

$$\mathbb{E}\left[u(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,m}) \mathbb{1}\{R_n \geq m\} \middle| \Delta_n = 0 \right] = \mathbb{E}[u(\mathbf{Z}_1, \ldots, \mathbf{Z}_m) \phi^n_m(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)]$$

for all bounded test functions $u : (\mathbb{N} \times \mathbb{N})^m \to \mathbb{R}$. Define

$$\Phi(n, m) = \phi^n_m(\mathbf{Z}_1, \ldots, \mathbf{Z}_m).$$

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 $T > 0$,

$$\left(\Phi(n, \lfloor n^{2/3}T \rfloor), \left(n^{-1/3} V^- \left(\lfloor n^{2/3}t \rfloor \right), n^{-1/3} V^+ \left(\lfloor n^{2/3}t \rfloor \right) \right)_{t \in [0,T]} \right) \xrightarrow{(\mathrm{d})} \left(\Phi(T), (\sigma_- W^-_t, \sigma_+ W^+_t)_{t \in [0,T]} \right)$$

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.

2.3.1 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 $\widehat{\mathbf{D}}_{n,i} = \mathbf{D}_{\Sigma_n(i)}$

for $i = 1, \ldots, R_n$.

Lemma 2.18. *Σ_n has law given by*

$$\mathbb{P}(\Sigma_n = \sigma \mid \mathbf{D}_1, \ldots, \mathbf{D}_n) = \prod_{i=1}^{R_n} \frac{D^-_{\sigma(i)}}{\sum_{j=i}^{R_n} D^-_{\sigma(j)}}.$$

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 of 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 2.19. *For all integers $0 \le r \le n$ and test functions $u : (\mathbb{N}\times\mathbb{N})^r \times \mathbb{N} \to \mathbb{R}$,*

$$\mathbb{E}\left[u\left(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,r}, \textstyle\sum_{i\in\mathcal{I}_n^c} D^+_i\right) \Big| R_n = r\right] = \mathbb{E}\left[u\left(\mathbf{Z}_1, \ldots, \mathbf{Z}_r, \textstyle\sum_{i=1}^{n-r} E^+_i\right)\psi_r(\mathbf{Z}_1, \ldots, \mathbf{Z}_r)\right]$$

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 $s \in \mathbb{N}$.

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\mathbf{D}}_{n,1} = \mathbf{k}_1, \ldots, \widehat{\mathbf{D}}_{n,r} = \mathbf{k}_r, \textstyle\sum_{i\in\mathcal{I}_n^c} D_i^+ = s, R_n = r\right) \\
&= \sum_{\substack{I\subseteq[n]\\|I|=r}} \sum_{\sigma:[r]\to I} \mathbb{P}\left(\mathbf{D}_{\Sigma_n(1)} = \mathbf{k}_1, \ldots, \mathbf{D}_{\Sigma_n(r)} = \mathbf{k}_r, \textstyle\sum_{i\in\mathcal{I}_n^c} D_i^+ = s, \mathcal{I}_n = I, \Sigma_n = \sigma\right)
\end{aligned}$$

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

$$\mathbb{P}\left(\textstyle\sum_{i\in I^c} D_i^+ = s, D_i^- = 0 \text{ for } i \in I^c\right) = (1-p)^{n-r}\mathbb{P}\left(\textstyle\sum_{i=1}^{n-r} E_i^+ = s\right).$$

Also

$$\begin{aligned}
\prod_{i=1}^{r} \frac{k_i^-}{\sum_{j=i}^r k_j^-} \times \prod_{i=1}^{r} \lambda_{\mathbf{k}_i} &= \prod_{i=1}^{r} \frac{k_i^-}{\mu}\lambda_{\mathbf{k}_i} \times \prod_{i=1}^{r} \frac{\mu}{\sum_{j=i}^r k_j^-} \\
&= \mathbb{P}(\mathbf{Z}_1 = \mathbf{k}_1, \ldots, \mathbf{Z}_r = \mathbf{k}_r) \times \prod_{i=1}^{r} \frac{\mu}{\sum_{j=i}^r k_j^-}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\mathbf{D}}_{n,1} = \mathbf{k}_1, \ldots, \widehat{\mathbf{D}}_{n,r} = \mathbf{k}_r, \textstyle\sum_{i\in\mathcal{I}_n^c} D_i^+ = s, R_n = r\right) \\
=&\binom{n}{r} \times r! \times \prod_{i=1}^{r} \frac{\mu}{\sum_{j=i}^r k_j^-} \times (1-p)^{n-r} \times \mathbb{P}\left(\mathbf{Z}_1 = \mathbf{k}_1, \ldots, \mathbf{Z}_r = \mathbf{k}_r, \textstyle\sum_{i=1}^{n-r} E_i^+ = s\right) \\
=&\binom{n}{r} p^r(1-p)^{n-r} \times \frac{1}{p^r} \prod_{i=1}^{r} \frac{(r-i+1)\mu}{\sum_{j=i}^r k_i^-} \times \mathbb{P}\left(\mathbf{Z}_1 = \mathbf{k}_1, \ldots, \mathbf{Z}_r = \mathbf{k}_r, \textstyle\sum_{i=1}^{n-r} E_i^+ = s\right).
\end{aligned}$$

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 2.20. *For all $m \leq n$ and test functions $u : (\mathbb{N} \times \mathbb{N})^m \to \mathbb{R}$,*

$$\mathbb{E}\left[u\left(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,m}\right) \mathbb{1}\{R_n \geq m\} \,\middle|\, \Delta_n = 0\right] = \mathbb{E}\left[u\left(\mathbf{Z}_1, \ldots, \mathbf{Z}_m\right) \phi_m^n(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)\right],$$

where

$$\phi_m^n(\mathbf{k}_1, \ldots, \mathbf{k}_m) = \frac{1}{\mathbb{P}(\Delta_n = 0)} \mathbb{E}\left[\mathbb{1}\left\{\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-)\right\} \prod_{i=1}^m \frac{(n-i+1)\mu}{\sum_{j=1}^m k_j^- + \Xi_{n-m}^-}\right].$$

Proof. By Lemma 2.19, for all $r \geq m$

$$\begin{aligned}
&\mathbb{E}\left[u\left(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,m}\right) \mathbb{1}\{\Delta_n = 0\} \,\middle|\, R_n = r\right] \\
=&\mathbb{E}\left[u\left(\mathbf{Z}_1, \ldots, \mathbf{Z}_m\right) \mathbb{1}\left\{\sum_{i=1}^r (Z_i^- - Z_i^+) - \sum_{i=1}^{n-r} E_i^+ = 0\right\} \frac{1}{p^r} \prod_{i=1}^r \frac{(r-i+1)\mu}{\sum_{j=i}^r Z_j^-}\right] \\
=&\mathbb{E}\left[u\left(\mathbf{Z}_1, \ldots, \mathbf{Z}_m\right) \mathbb{E}\left[\mathbb{1}\left\{\sum_{i=1}^r (Z_i^- - Z_i^+) - \sum_{i=1}^{n-r} E_i^+ = 0\right\} \frac{1}{p^r} \prod_{i=1}^r \frac{(r-i+1)\mu}{\sum_{j=i}^r Z_j^-} \,\middle|\, \mathbf{Z}_1, \ldots, \mathbf{Z}_m\right]\right] \\
=&\mathbb{E}\left[u\left(\mathbf{Z}_1, \ldots, \mathbf{Z}_m\right) \tilde{\gamma}_r^{n,m}(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)\right],
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\gamma}_r^{n,m}(\mathbf{k}_1, \ldots, \mathbf{k}_m) &= \mathbb{E}\Bigg[\mathbb{1}\left\{\sum_{i=1}^r (Z_i^- - Z_i^+) - \sum_{i=1}^{n-r} E_i^+ = 0\right\} \times \\
&\qquad \frac{1}{p^r} \prod_{i=1}^r \frac{(r-i+1)\mu}{\sum_{j=i}^r Z_j^-} \,\Bigg|\, \mathbf{Z}_1 = \mathbf{k}_1, \ldots, \mathbf{Z}_m = \mathbf{k}_m\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{1}\left\{\sum_{i=m+1}^r (Z_i^- - Z_i^+) - \sum_{i=1}^{n-r} E_i^+ = \sum_{i=1}^m (k_i^+ - k_i^-)\right\} \times \\
&\qquad \frac{1}{p^m} \prod_{i=1}^m \frac{(r-i+1)\mu}{\sum_{j=i}^m k_j^- + \sum_{j=m+1}^r Z_j^-} \frac{1}{p^{r-m}} \prod_{i=m+1}^r \frac{(r-i+1)\mu}{\sum_{j=i}^r Z_j^-}\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{1}\left\{\sum_{i=1}^{r-m} (Z_i^- - Z_i^+) - \sum_{i=1}^{n-r} E_i^+ = \sum_{i=1}^m (k_i^+ - k_i^-)\right\} \times \\
&\qquad \frac{1}{p^m} \prod_{i=1}^m \frac{(r-i+1)\mu}{\sum_{j=i}^m k_j^- + \sum_{j=1}^{r-m} Z_j^-} \frac{1}{p^{r-m}} \prod_{i=1}^{r-m} \frac{(r-m-i+1)\mu}{\sum_{j=i}^{r-m} Z_j^-}\Bigg],
\end{aligned}$$

since $(\mathbf{Z}_i)_{i=m+1}^{r}$ has the same law as $(\mathbf{Z}_i)_{i=1}^{r-m}$. Then applying Lemma 2.19 again shows that

$$\tilde{\gamma}_r^{n,m}(\mathbf{k}_1,\ldots,\mathbf{k}_m) = \mathbb{E}\Bigg[\mathbb{1}\left\{\sum_{i=1}^{r-m}(\widehat{D}_{n-m,i}^{-}-\widehat{D}_{n-m,i}^{+}) - \sum_{i\in\mathcal{I}_{n-m}^c} D_i^{+} = \sum_{i=1}^{m}(k_i^{+}-k_i^{-})\right\}\times \\ \frac{1}{p^m}\prod_{i=1}^{m}\frac{(r-i+1)\mu}{\sum_{j=i}^{m}k_j^{-}+\sum_{j=1}^{r-m}\widehat{D}_{n-m,j}^{-}}\ \Bigg|\ R_{n-m}=r-m\Bigg].$$

Conditional on $R_{n-m}=r-m$, we have

$$\sum_{j=1}^{r-m}(\widehat{D}_{n-m,j}^{-}-\widehat{D}_{n-m,j}^{+}) - \sum_{i\in\mathcal{I}_{n-m}^c} D_i^{+} = \Delta_{n-m} \quad\text{and}\quad \sum_{j=1}^{r-m}\widehat{D}_{n-m,j}^{-} = \Xi_{n-m}^{-}.$$

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,

$$\mathbb{E}\left[u\left(\widehat{\mathbf{D}}_{n,1},\ldots,\widehat{\mathbf{D}}_{n,m}\right)\mathbb{1}\{R_n\geq m,\Delta_n=0\}\right] = \mathbb{E}\left[u\left(\mathbf{Z}_1,\ldots,\mathbf{Z}_m\right)\tilde{\phi}_m^{n}(\mathbf{Z}_1,\ldots,\mathbf{Z}_m)\right],$$

where

$$\begin{aligned}
&\tilde{\phi}_m^{n}(\mathbf{k}_1,\ldots,\mathbf{k}_m)\\
&=\sum_{r=m}^{n}\binom{n}{r}p^r(1-p)^{n-r}\mathbb{E}\left[\frac{1}{p^m}\prod_{i=1}^{m}\frac{(r-i+1)\mu}{\sum_{j=i}^{m}k_j^{-}+\Xi_{n-m}^{-}}\mathbb{1}_{A_n}\ \middle|\ R_{n-m}=r-m\right]\\
&=\sum_{l=0}^{n-m}\binom{n}{l+m}p^{l+m}(1-p)^{n-m-l}\mathbb{E}\left[\frac{1}{p^m}\prod_{i=1}^{m}\frac{(l+m-i+1)\mu}{\sum_{j=i}^{m}k_j^{-}+\Xi_{n-m}^{-}}\mathbb{1}_{A_n}\ \middle|\ R_{n-m}=l\right].
\end{aligned}$$

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. □

2.3.2 Asymptotic lower bound on the measure change

Recall that our goal in Proposition 2.17 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 sufficient and necessary to work on a sequence of events occuring with high probability. In particular, for the proof of Proposition 2.17, we work on the event $\mathcal{E}_m$ where

$$\mathcal{E}_m = \Bigg\{\max_{i=1,\dots,m}\left|\textstyle\sum_{j=1}^i(Z_j^- - \lambda_-)\right| \le m^{1/2}\log(m)\\ \qquad\qquad\text{and}\quad \max_{i=1,\dots,m}\left|\textstyle\sum_{j=1}^i(Z_j^+ - \lambda_+)\right| \le m^{1/2}\log(m)\Bigg\}.$$

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 2.1.1 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 [19, Lemma 6.7]. In it we prove a deterministic lower bound on $\phi^n_m(\mathbf{k}_1,\dots,\mathbf{k}_m)$, for all $\mathbf{k}_1,\dots,\mathbf{k}_m$ corresponding to the event $\mathcal{E}_m$, up to an error which vanishes as $n \to \infty$.

Proposition 2.21. *Define*

$$s^{\pm}(i) = \textstyle\sum_{j=1}^i(k_i^{\pm} - \lambda_{\pm}).$$

Suppose that $\mathbf{k}_1, \dots, \mathbf{k}_m$ *are such that*

$$\max_{i=1,\dots,m} |s^-(i)| \le m^{\frac{1}{2}} \log(m) \quad \textit{and} \quad \max_{i=1,\dots,m} |s^+(i)| \le m^{\frac{1}{2}} \log(m) \tag{2.5}$$

Then in the regime $m = \Theta(n^{2/3})$*, as* $n \to \infty$*,*

$$\phi^n_m(\mathbf{k}_1, \dots, \mathbf{k}_m) \ge \exp\left(\frac{1}{\mu n} \sum_{i=0}^{m} (s^-(i) - s^-(m)) - \frac{\sigma_-}{6\mu^2} \frac{m^3}{n^2} \right) + o(1),$$

where the $o(1)$ *error term is independent of* $\mathbf{k}_1, \dots, \mathbf{k}_m$ *satisfying the assumption in Eq. (2.5).*

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 [19, Lemma 4.8]. In Proposition 2.17 we are considering the joint convergence of the measure change with two other random walks, and thus we adapt [19, Lemma 4.8] to allow for an additional coordinate that is converging jointly with the first coordinate.

Lemma 2.22. *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. $(Y_n, Z_n) \xrightarrow{(d)} (Y, Z)$ *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 $(X_n, Z_n) \xrightarrow{(d)} (Y, Z)$ *also. Moreover* $(X_n)_{n\ge1}$ *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 [19, Lemma 4.8] and so we will not repeat it here.

Discrete local limit theorem

To prove Proposition 2.21, we first need to understand the denominator of ϕ_m^n, which, as given by Lemma 2.20, 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 [33]. Here we borrow the presentation from Durrett [28, 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

$$\sup_{s\in\mathbb{Z}}\left|\mathbb{P}\big(\textstyle\sum_{i=1}^n X_i = s\big) - \tfrac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(\tfrac{-(s-n\mu)^2}{2n\sigma^2}\right)\right| = o(n^{-1/2}). \tag{2.6}$$

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. (2.6) does hold. This gives us the discrete local limit theorem:

Theorem 2.23 (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 2.24. *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 2.23) 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 2.1.1 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 2.23).

Corollary 2.25. *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 2.26. *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.*

Exponential Tilting

Next we turn to the numerator of ϕ_m^n. By Lemma 2.20, this is given by

$$\mathbb{E}\left[\mathbb{1}\left\{\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-)\right\} \prod_{i=1}^m \frac{(n-i+1)\mu}{\sum_{j=i}^m k_j^- + \Xi_{n-m}^-}\right]. \tag{2.7}$$

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. (2.5), we face two problems in evaluating the expectation in Eq. (2.7).

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. (2.5), $s^-(m)$ 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 2.24, 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. (2.7) is not dictated by the typical fluctuations of the random variables Ξ_{n-m}^-. In other words, it is not the case that

$$\mathbb{E}\left[\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^- + \Xi_{n-m}^-}\right] \not\approx \prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{j=i}^{m}k_j^- + \mathbb{E}[\Xi_{n-m}^-]} \tag{2.8}$$

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. (2.7) 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 2.27. *Define a 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

$$\mathbb{E}\left[(D^-)^3 \exp(-\theta D^-)\right] = \mathbb{E}\left[(D^-)^3\right] + o(1) \tag{2.9}$$

as $\theta \downarrow 0$. Integrating Eq. (2.9) 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

$$\mathbb{E}\left[D^- e^{-\theta D^-}\right] = \mu - \mathbb{E}\left[(D^-)^2\right]\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^3\right]\theta^2 + o(\theta^2), \tag{2.10}$$

$$\text{and} \quad \mathbb{E}\left[e^{-\theta D^-}\right] = 1 - \mu\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^2\right]\theta^2 - \tfrac{1}{6}\mathbb{E}\left[(D^-)^3\right]\theta^3 + o(\theta^3). \tag{2.11}$$

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

$$\mathbb{E}\left[D^+ e^{-\theta D^-}\right] = \mu\theta - \mathbb{E}\left[D^- D^+\right]\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^2 D^+\right]\theta^2 + o(\theta^2). \tag{2.12}$$

Eq. (2.11) gives the small-θ expansion of the normalising constant of the measure change. Combining this with Eq. (2.10) and Eq. (2.12) yields the expansions for $\mathbb{E}_\theta[D^-]$ and $\mathbb{E}_\theta[D^+]$ respectively. Taking the logarithm of Eq. (2.11) 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

$$\frac{\mathrm{d}\mathbb{P}_n}{\mathrm{d}\mathbb{P}} = \exp\left(-\theta_n \Xi^-_{n-m} - (n-m)\alpha(\theta_n)\right), \tag{2.13}$$

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 2.27, 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 $\theta_n = \frac{m}{\mu n}$.

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. (2.7) is dominated by the typical behaviour of Ξ^-_{n-m} under $\mathbb{P}_n$. Using Lemma 2.27, 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

$$\mathcal{B}_n = \left\{|\Xi^-_{n-m} - \mu n + \lambda_- m| \leq n^{1/2}\log(n)\right\}.$$

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 2.28. *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* $\mathbf{k}_1,\ldots,\mathbf{k}_m$ *satisfying the assumption in Eq.* (2.5) *on page 53.*

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

$$A_{i,n}=\frac{1}{\mu n}\left\{\Omega_n^--\left[s^-(i-1)-s^-(m)\right]\right\},\quad B_{i,n}=-\frac{\lambda_-}{\mu n}(i-1).$$

Then on the event $\mathcal{B}_n$,

$$\max_{i=1,\ldots,m}|A_n^i|=O(n^{-1/2}\log n)\quad\text{and}\quad\max_{i=1,\ldots,m}|B_n^i|=O(n^{-1/3}).$$

where the O bounds are uniform for $\mathbf{k}_1,\ldots,\mathbf{k}_m$ satifying Eq. (2.5). 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,

$$\sum_{i=1}^{m} B_n^i = -\frac{\lambda_-}{2\mu}\frac{m^2}{n} + o(1) \quad \text{and} \quad \sum_{i=1}^{m} (B_n^i)^2 = \frac{\lambda_-^2}{3\mu^2}\frac{m^3}{n^2} + o(1).$$

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 $\sum_{i=1}^{m} \left(s^-(i-1) - s^-(m)\right) = \sum_{i=0}^{m} \left(s^-(i) - s^-(m)\right)$.

Similarly we can expand X_n as

$$X_n = -\frac{m}{2n} - \frac{m^3}{3n^2} + o(1).$$

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 2.27, 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. □

Multivariate Local Limit Theorem

To complete the proof of Proposition 2.21 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

$$\mathcal{A}_n = \left\{\Delta_{n-m} = \sum_{i=1}^{m}(k_i^+ - k_i^-)\right\}$$

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 $X_{n,1}, \ldots, X_{n,n}$ are i.i.d. but the distribution of $X_{n,1}$ can change with n. A collection of random variables $(X_{n,1}, \ldots, X_{n,n})_{n=1}^\infty$ 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

$$\Lambda = \left\{\sum_{i=1}^{d} n_i \mathbf{a}_i : n_i \in \mathbb{Z} \text{ for } i = 1, \ldots, d\right\}.$$

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 $A_{ij} = \mathbf{a}_j^{(i)}$. The choice of basis generating a lattice is not unique, and the following lemma adapted from [52, Corollary 4.3a] characterises when two basis generate the same lattice.

Lemma 2.29. *Let A and B be $n \times n$ matrices of full rank. Then the columns of A and B generate the same matrix if and only if there exists a matrix U such that U has integer entries,* $\det(U) = \pm 1$ *and $A = BU$.*

iherefore 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. $a_{i,j} < a_{i,i}$ for all $j = 1, \ldots, d$, in other words the unique maximal entry in each row is on the diagonal.

Then the following lemma, adapted from [52, Corollary 4.3b], gives existence of a canonical choice of basis generating an integer lattice.

Lemma 2.30. *Suppose $\Lambda \subseteq \mathbb{Z}^d$ is a lattice. Then there exists a unique $d \times d$ 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

$$\Lambda(\mathbf{X}) = \bigcup_{k=1}^{\infty} \left\{ \textstyle\sum_{i=1}^{k} n_i \mathbf{x}_i^* : n_i \in \mathbb{Z} \text{ and } \mathbf{x}_i^* \in \operatorname{supp}(\mathbf{X}^*) \text{ for } i = 1, \ldots, k \right\}.$$

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 2.31. *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 $\mathbf{X}_n \xrightarrow{(\mathrm{d})} \mathbf{X}$ as $n \to \infty$.*
2. *$(\|\mathbf{X}_n\|^2)_{n\geq 1}$ is a uniformly integrable sequence of random variables. Explicitly*
$$\lim_{L\to\infty} \sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\left\{\|\mathbf{X}_n\|^2 > L\right\}\right] = 0. \tag{2.14}$$
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 $\mathbf{y} \in \mathbf{c}_n + \Lambda$,

$$\mathbb{P}\Big(\textstyle\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\Big) = n^{-d/2}\det(\Lambda) f\left(\mathbf{x}_n(\mathbf{y})\right) + o\left(n^{-d/2}\right) \quad \textit{where} \quad \mathbf{x}_n(\mathbf{y}) = \frac{\mathbf{y}-n\mathbb{E}[\mathbf{X}_n]}{\sqrt{n}}$$

and f is the density of a $N(0, \mathrm{Cov}(\mathbf{X}))$ distribution. This means that

$$\lim_{n\to\infty} \sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda} \left| n^{d/2}\mathbb{P}\Big(\textstyle\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\Big) - \det(\Lambda) f(\mathbf{x}_n(\mathbf{y}))\right| = 0.$$

We defer the proof of this to Appendix B 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. (2.14) is still well defined.

We apply Theorem 2.31 to $(\Xi^-_{n-m}, \Delta_{n-m})$. Suppose $(D^- - D^+, D^-)$ is non-degenerate and let Λ be its main lattice. By Lemma 2.30, 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 exists positive integers p and q such that

$$A = \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}.$$

Finally let Σ be the covariance matrix of $(D^- - D^+, D^-)$. With this notation, the following lemma holds:

Lemma 2.32. *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 $(x, y) \in \mathbf{c}_n + \Lambda$,*

$$\begin{aligned} &\mathbb{P}_n\left(\Delta_{n-m} = \mathbb{E}[\Delta_{n-m}] + x,\ \Xi^-_{n-m} = \mathbb{E}[\Xi^-_{n-m}] + y\right) \\ &\qquad = \frac{q}{2\pi \det(\Sigma)^{1/2}\, n} \exp\left(\frac{-1}{2n}\begin{pmatrix} x & y \end{pmatrix} \Sigma \begin{pmatrix} x \\ y \end{pmatrix}\right) + o(n^{-1}) \end{aligned}$$

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

$$\begin{pmatrix} D_1^- - D_1^+ \\ D_1^- \end{pmatrix}, \ldots, \begin{pmatrix} D_n^- - D_n^+ \\ D_n^- \end{pmatrix}$$

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 $\mathbf{X}_n \xrightarrow{(d)} \mathbf{X}$ as $n \to \infty$.

For any $L > 0$,

$$\begin{aligned} \sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\{\|\mathbf{X}_n\|^2 > L\}\right] &= \sup_n \mathbb{E}\left[e^{-\theta_n D^- - \alpha(\theta_n)} \|\mathbf{X}\|^2 \mathbb{1}\{\|\mathbf{X}\|^2 > L\}\right] \\ &\leq \left(\sup_n e^{-\alpha(\theta_n)}\right) \mathbb{E}\left[\|\mathbf{X}\|^2 \mathbb{1}\{\|\mathbf{X}\|^2 > L\}\right] \end{aligned}$$

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. (2.14).

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 2.31. 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 2.21.

Lemma 2.33. *Under the assumptions of Proposition 2.21,*

$$\mathbb{P}_n\left(\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-),\ |\Xi_{n-m}^- - \mathbb{E}_n[\Xi_{n-m}^-]| \leq n^{\frac{1}{2}} \log n\right) \geq \frac{1}{\sqrt{2\pi\sigma^2 n}}(1 + o(1)).$$

Proof. For convenience let

$$P_n = \mathbb{P}_n\left(\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-),\ |\Xi_{n-m}^- - \mathbb{E}_n[\Xi_{n-m}^-]| \leq n^{\frac{1}{2}} \log n\right).$$

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

$$\left\{\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-),\ |\Xi_{n-m}^- - \mathbb{E}_n[\Xi_{n-m}^-]| \leq n^{\frac{1}{2}} \log n\right\} = \left\{\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-)\right\}.$$

Then applying Theorem 2.31, as we did in the proof of Lemma 2.32, 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

$$L_n = \Big\{ y : \Big(\textstyle\sum_{i=1}^m (k_i^+ - k_i^-), y \Big) \in \mathbf{c}_n + \Lambda \Big\}.$$

L_n has a simpler representation. Fix any $y_0 \in L_n$. Then if Λ is generated by the columns of

$$\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$$

we must have $L_n = y_0 + q\mathbb{Z}$. Fix an arbitrary $M > 0$. Then

$$\begin{aligned} P_n &= \sum_{\substack{y \in L_n \\ |y| \le n^{1/2} \log n}} \mathbb{P}_n \left(\Delta_{n-m} = \mathbb{E}_n[\Delta_{n-m}] + a_n,\ \Xi^-_{n-m} = \mathbb{E}_n[\Xi^-_{n-m}] + y \right) \\ &\ge \sum_{\substack{y \in L_n \\ |y| \le M n^{1/2}}} \mathbb{P}_n \left(\Delta_{n-m} = \mathbb{E}_n[\Delta_{n-m}] + a_n,\ \Xi^-_{n-m} = \mathbb{E}_n[\Xi^-_{n-m}] + y \right) \end{aligned}$$

for all n sufficiently large. By Lemma 2.32, using that the error is uniform, we have that

$$P_n \ge \sum_{\substack{y \in L_n \\ |y| \le M n^{1/2}}} \frac{q}{2\pi \det(\Sigma)^{1/2}\, n} \exp\left(\frac{-1}{2n} \begin{pmatrix} a_n \\ y \end{pmatrix} \cdot \Sigma^{-1} \begin{pmatrix} a_n \\ y \end{pmatrix} \right) + o(n^{-1/2})$$

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

$$\begin{aligned} &\frac{q}{2\pi \det(\Sigma)^{1/2}\, n} \exp\left(\frac{1}{2n} \begin{pmatrix} a_n \\ y \end{pmatrix} \cdot \Sigma^{-1} \begin{pmatrix} a_n \\ y \end{pmatrix} \right) \\ &\qquad = \frac{1}{\sqrt{2\pi\sigma^2 n}} \exp\left(-\frac{1}{2\sigma^2} \frac{a_n^2}{n} \right) \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2} \frac{(y - ca_n)^2}{n} \right). \end{aligned}$$

We now examine the asymptotic behaviour of a_n. By Lemma 2.27,

$$\begin{aligned} \mathbb{E}_n[\Delta_{n-m}] &= (n-m)\mathbb{E}_{\theta_n}[D^- - D^+] \\ &= -(\lambda_- - \lambda_+)m + O(n^{1/3}). \end{aligned}$$

Therefore

$$a_n = s_+(m) - s_-(m) + O(n^{1/3}) = O(n^{1/3}\log n),$$

by the assumption in Eq. (2.5). 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

$$\sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right) \geq \sum_{\substack{z\in L_n - ca_n \\ |z|\leq \frac{1}{2}Mn^{1/2}}} \frac{q}{\sqrt{n}}\, g\left(\frac{z}{\sqrt{n}}\right) \tag{2.15}$$

$$= \sum_{\substack{z\in \tilde{L}_n \\ |z|\leq \frac{1}{2}M}} \frac{q}{\sqrt{n}}\, g(z) \tag{2.16}$$

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. (2.16) is a Riemann sum approximation of an integral. Hence

$$\sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right) \geq (1+o(1)) \int_{-\frac{1}{2}M}^{\frac{1}{2}M} g(z)\,\mathrm{d}z.$$

Thus

$$P_n \geq \frac{1}{\sqrt{2\pi\sigma^2 n}}(1+o(1)) \int_{-\frac{1}{2}M}^{\frac{1}{2}M} g(z)\,\mathrm{d}z.$$

This holds for all $M > 0$, and $\int_{-\infty}^{\infty} g(z)\,\mathrm{d}z = 1$. Therefore,

$$P_n \geq \frac{1}{\sqrt{2\pi\sigma^2 n}}\left(1+o(1)\right),$$

as required. ☐

2.3.3 Proof of lower bound

Now we are ready to prove Proposition 2.21.

Proof of Proposition 2.21. By Lemma 2.20 and Lemma 2.28 we have that

$$\phi(\mathbf{k}_1,\ldots,\mathbf{k}_m) \geq \left\{\exp\left(\frac{1}{\mu n}\sum_{i=0}^{m}(s^-(i)-s^-(m)) - \frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right) + o(1)\right\}\frac{\mathbb{P}_n(\mathcal{A}_n\cap\mathcal{B}_n)}{\mathbb{P}(\Delta_n=0)}$$

where the $o(1)$ term is independent of $\mathbf{k}_1,\ldots,\mathbf{k}_m$ satisfying our assumptions. Then by Lemma 2.32 and Corollary 2.25 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. ☐

2.3.4 Convergence of the measure change

We are now ready to prove the main result of this section.

Proof of Proposition 2.17. The existence of the measure change is covered by Lemma 2.20. 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,

$$\left(n^{-1/3}V^-(\lfloor tn^{2/3}\rfloor), n^{-1/3}V^+(\lfloor tn^{2/3}\rfloor)\right)_{t\geq 0} \xrightarrow{(\mathrm{d})} \left(\sigma_- W_t^-, \sigma_+ W_t^+\right)_{t\geq 0}$$

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,

$$\frac{1}{n}\sum_{i=0}^{\lfloor Tn^{2/3}\rfloor}\big(V^-(i)-V^-(m)\big) \xrightarrow{\text{(d)}} \int_0^T \big(W_s^- - W_T^-\big)\,\mathrm{d}s = -\int_0^T s\,\mathrm{d}W_s^-.$$

Hence,

$$\begin{aligned}&\left(\Gamma(n,\lfloor Tn^{2/3}\rfloor), \left(n^{-1/3}V^-(\lfloor tn^{2/3}\rfloor), n^{-1/3}V^+(\lfloor tn^{2/3}\rfloor)\right)_{t\in[0,T]}\right) \\ &\qquad\qquad\xrightarrow{\text{(d)}} \big(\Phi(T), (\sigma_- W_t^-, \sigma_+ W_t^+)_{t\in[0,T]}\big)\end{aligned}$$

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 2.21, it is the case that

$$\Phi(n,m) \ge (\Gamma(n,m)+o(1))\mathbb{1}_{\mathcal{E}_m}.$$

The processes $(V^\pm(n))_{n\ge 0}$ are discrete martingales. Therefore, by Doob's maximal inequality,

$$\mathbb{P}\left(\max_{i=1,\dots,m}|V^\pm(i)| > m^{1/2}\log(m)\right) \le \frac{\mathbb{E}[(V^\pm(m))^2]}{m(\log m)^2} = \frac{\sigma_\pm^2}{(\log m)^2} \to 0$$

as $m\to\infty$. Thus $\mathbb{P}(\mathcal{E}_m)\to 1$ as $m\to\infty$. Hence, we still have that

$$\begin{aligned}&\left((\Gamma(n,\lfloor Tn^{2/3}\rfloor)+o(1))\mathbb{1}_{E_{\lfloor Tn^{2/3}\rfloor}}, \left(n^{-1/3}V^-(\lfloor tn^{2/3}\rfloor), n^{-1/3}V^+(\lfloor tn^{2/3}\rfloor)\right)_{t\in[0,T]}\right) \\ &\qquad\qquad\xrightarrow{\text{(d)}} \big(\Phi(T), (\sigma_- W_t^-, \sigma_+ W_t^+)_{t\in[0,T]}\big).\end{aligned}$$

We have $\mathbb{E}[\Phi(T)] = 1$ by a standard stochastic calculus calculation. Therefore, by Lemma 2.22, 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 $(\Phi(n,\lfloor Tn^{2/3}\rfloor))_{n\geq 1}$ is a uniformly integrable sequence. □

2.4 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 2.34. *It holds that*

$$\left(n^{-1/3} \hat{S}_n^+ \left(\lfloor n^{2/3} t \rfloor \right), n^{-1/3} \hat{H}_n \left(\lfloor n^{2/3} t \rfloor \right), t \leq T \right) \xrightarrow{\text{(d)}} \left(\sigma_+ \hat{B}_t, \frac{2}{\sigma_+} \hat{R}_t, t \leq T \right)$$

in $\mathbb{D}([0,T], \mathbb{R})^2$, *and*

$$\left(n^{-2/3} \hat{S}_n^- \left(\lfloor n^{2/3} t \rfloor \right), n^{-1/3} \hat{P}_n \left(\lfloor n^{2/3} t \rfloor \right), t \leq T \right) \xrightarrow{\text{(p)}} \left(\nu_- t, \frac{\nu_-}{2\mu} t^2, t \leq T \right)$$

in $\mathbb{D}([0,T], \mathbb{R})^2$ *as* $n \to \infty$.

We prove Proposition 2.34 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 $(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{n,n})$ 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 2.3, we showed that the law of $(\widehat{\mathbf{D}}_{n,1}, \dots, \widehat{\mathbf{D}}_{n,m})$ 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 $m = \lfloor T n^{2/3} \rfloor$.

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 2.4.2, we modify the Bienaymé forest in order to include dummy leaves. We add extra randomness, approximating the procedure described in Proposition 2.8, 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 $\lfloor T n^{2/3} \rfloor$.

4. We show convergence under rescaling of the out-forest up to time $\lfloor T n^{2/3} \rfloor$ 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.

2.4.1 Convergence before adding the dummy leaves

We define the two processes

$$\hat{Y}^{\pm}(k) = \sum_{i=1}^{k} (\widehat{D}_{n,i}^{\pm} - 1),$$

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 2.3. 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 2.17.

Corollary 2.35. *Suppose we are in the setting of Proposition 2.17 and that the criticality condition holds. Then for all* $T > 0$*,*

$$\left(\Phi(n, \lfloor n^{2/3}T \rfloor), \left(n^{-1/3}V^-\left(\lfloor n^{2/3}t \rfloor\right), n^{-1/3}V^+\left(\lfloor n^{2/3}t \rfloor\right)\right)_{t\in[0,T]}\right) \xrightarrow{(\mathrm{d})} \left(\Phi(T), (\sigma_- W_t^+, \sigma_+ W_t^+)_{t\in[0,T]}\right)$$

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.*

Let $(\hat{B}_t, t \geq 0)$ be distributed as follows. For F a suitable test function, and for $(B_t)_{t\geq 0}$ a Brownian motion,

$$\begin{aligned}
&\mathbb{E}\left[F(\sigma_+ \hat{B}_t, 0 \leq t \leq T)\right] \\
&= \mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_+ \mu}\int_0^T s dB_s - \frac{\sigma_{-+}^2 T^3}{6\sigma_+^2 \mu^2}\right) F(\sigma_+ B_t, 0 \leq t \leq T)\right].
\end{aligned}$$

Proposition 2.36. *We have that*

$$\left(n^{-2/3}\hat{Y}^-\left(\lfloor n^{2/3}t \rfloor\right), n^{-1/3}\hat{Y}^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \leq t \leq T\right) \xrightarrow{(\mathrm{d})} \left(\nu_- t, \sigma_+ \hat{B}_t, 0 \leq t \leq T\right)$$

in the Skorokhod topology as $n \to \infty$.

Proof. We recall from the statement of Corollary 2.35 that (W^-, W^+) is a pair of correlated standard Brownian motions with correlation $\mathrm{Corr}(Z_1^-, Z_1^+)$. Let $(B_t^1, t \geq 0)$ and $(B_t^2, t \geq 0)$ be two independent Brownian motions, so that we may define

$$(\sigma_- W_t^-, \sigma_+ W_t^+, t \geq 0) = \left(\frac{\sigma_{-+}}{\sigma_+} B_t^1 + \left(\sigma_-^2 - \frac{\sigma_{-+}^2}{\sigma_+^2}\right)^{1/2} B_t^2, \sigma_+ B_t^1, t \geq 0\right).$$

Then, Corollary 2.35 implies that for F a continuous, bounded test function,

$$\begin{aligned}
&\mathbb{E}\left[F\left(n^{-1/3}\hat{Y}^{+}\left(\lfloor n^{2/3}t\rfloor\right),0\leq t\leq T\right)\right]\\
&=\mathbb{E}\left[F\left(n^{-1/3}\hat{Y}^{+}\left(\lfloor n^{2/3}t\rfloor\right),0\leq t\leq T\right)\mathbb{1}_{\lfloor Tn^{2/3}\rfloor\leq R_n}\right]+o(1)\\
&=\mathbb{E}\left[\Phi(n,\lfloor n^{2/3}T\rfloor)F\left(n^{-1/3}V^{+}\left(\lfloor n^{2/3}t\rfloor\right),0\leq t\leq T\right)\right]+o(1).
\end{aligned}$$

By the proof of Proposition 2.17, 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

$$\mathbb{E}\left[\Gamma(n,\lfloor n^{2/3}T\rfloor)F\left(n^{-1/3}V^{+}\left(\lfloor n^{2/3}t\rfloor\right),0\leq t\leq T\right)\right]\to\mathbb{E}\left[F\left(\sigma_{+}\hat{B}_{t},0\leq t\leq T\right)\right].$$

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

$$\mathbb{E}\left[\Theta(V^{-}_{(n)},T)F\left(V^{+}_{(n)}(t),0\leq t\leq T\right)\right]\to\mathbb{E}\left[F\left(\sigma_{+}\hat{B}_{t},0\leq t\leq T\right)\right].$$

But, by the continuity of Θ and Corollary 2.35, we get that

$$\begin{aligned}
&\mathbb{E}\left[\Theta(V_{(n)}^{-},T)F\left(V_{(n)}^{+}(t),0\leq t\leq T\right)\right]\to\mathbb{E}\left[\Theta(\sigma_{-}W_t^{-},T)F\left(\sigma_{+}W_t^{+},0\leq t\leq T\right)\right]\\
&\quad=\mathbb{E}\left[\exp\left(-\frac{1}{\mu}\int_0^T sd\left(\frac{\sigma_{-+}}{\sigma_{+}}B_s^1+\left(\sigma_{-}^2-\frac{\sigma_{-+}^2}{\sigma_{+}^2}\right)^{1/2}B_s^2\right)-\frac{T^3\sigma_{-}^2}{6\mu^2}\right)F\left(\sigma_{+}B_t^1,0\leq t\leq T\right)\right]\\
&\quad=\mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_{+}\mu}\int_0^T sdB_s^1-\frac{\sigma_{-+}^2T^3}{6\sigma_{+}^2\mu^2}\right)F(\sigma_{+}B_t^1,0\leq t\leq T)\right].
\end{aligned}$$

Then, the fact that $(Y(k),k\geq 1)$ is a random walk with steps of mean ν_{-} implies that

$$\left(n^{-2/3}Y^{-}\left(\lfloor n^{2/3}t\rfloor\right),t\geq 0\right)\xrightarrow{(\mathrm{p})}(\nu_{-}t,t\geq 0),$$

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. □

The following proposition characterises the distribution of $(\hat{B}_t,0\leq t\leq T)$.

Proposition 2.37. *We have that*

$$(\sigma_{+}\hat{B}_t,0\leq t\leq T)\overset{d}{=}\left(\sigma_{+}B_t-\frac{\sigma_{-+}}{2\mu}t^2,0\leq t\leq T\right),$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion.

Proof. Firstly, we have that for any $t\in[0,T]$ and $\theta>0$,

$$\begin{aligned}
\mathbb{E}\left[\exp(-\theta\sigma_{+}\hat{B}_t)\right]&=\mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_{+}\mu}\int_0^t sdB_s-\frac{\sigma_{-+}^2t^3}{6\sigma_{+}^2\mu^2}-\theta\sigma_{+}B_t\right)\right]\\
&=\mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_{+}\mu}\int_0^t\left(s+\frac{\sigma_{+}^2\theta\mu}{\sigma_{-+}}\right)dB_s-\frac{\sigma_{-+}^2t^3}{6\sigma_{+}^2\mu^2}\right)\right]\\
&=\exp\left(-\frac{\sigma_{-+}^2}{2\sigma_{+}^2\mu^2}\int_0^t\left(s+\frac{\sigma_{+}^2\theta\mu}{\sigma_{-+}}\right)^2ds-\frac{\sigma_{-+}^2t^3}{6\sigma_{+}^2\mu^2}\right)\\
&=\exp\left(\frac{\sigma_{+}^2t}{2}\theta^2+\frac{\sigma_{-+}t^2}{2\mu}\theta\right)\\
&=\mathbb{E}\left[\exp\left(-\theta\left(\sigma_{+}B_t-\frac{\sigma_{-+}}{2\mu}t^2\right)\right)\right].
\end{aligned}$$

Then, more generally, for $m > 0$, $0 = t_0 \leq t_1 \leq \cdots \leq t_m = T$, and $\theta_1, \ldots, \theta_m \in \mathbb{R}_+$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\sum_{i=1}^{m}\theta_i(\sigma_+\hat{B}_{t_i} - \sigma_+\hat{B}_{t_{i-1}})\right)\right] \\
&= \prod_{i=1}^{m}\mathbb{E}\left[\exp\left(-\frac{\sigma_{-+}}{\sigma_+\mu}\int_{t_{i-1}}^{t_i} s dB_s - \frac{\sigma_{-+}^2(t_i^3 - t_{i-1}^3)}{6\sigma_+^2\mu^2} - \theta_i\sigma_+(B_{t_i} - B_{t_{i-1}})\right)\right] \\
&= \prod_{i=1}^{m}\exp\left(-\frac{\sigma_{-+}^2}{2\sigma_+^2\mu^2}\int_{t_{i-1}}^{t_i}\left(s + \frac{\sigma_+^2\theta_i\mu}{\sigma_{-+}}\right)^2 ds - \frac{\sigma_{-+}^2(t_i^3 - t_{i-1}^3)}{6\sigma_+^2\mu^2}\right) \\
&= \prod_{i=1}^{m}\exp\left(\frac{\sigma_+^2(t_i - t_{i-1})}{2}\theta_i^2 + \frac{\sigma_{-+}(t_i^2 - t_{i-1}^2)}{2\mu}\theta_i\right) \\
&= \mathbb{E}\left[\exp\left(-\sum_{i=1}^{m}\theta_i\left(\sigma_+(B_t - B_{t_i}) - \frac{\sigma_{-+}}{2\mu}(t_i^2 - t_{i-1}^2)\right)\right)\right],
\end{aligned}$$

which proves the result. □

2.4.2 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 $P^n(k) - P^n(k-1) = 1$.

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 2.38. *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 $(Y^+(j), Y^-(j), j \geq 1)$, $n^{-1/3}P_n(m)$ 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

$$\frac{Y^-(m) + m}{\mu n - m} \wedge 1$$

is tight. The statement follows. □

Lemma 2.39. *We have*

$$\left(n^{-1/3}P_n(\lfloor n^{2/3}t\rfloor), t \geq 0\right) \xrightarrow{(\mathrm{p})} \left(\frac{\nu_-}{2\mu}t^2, t \geq 0\right)$$

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 2.38, the convergence under rescaling of Y^+ shown in Corollary 2.35, 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 2.35, the tightness of $\left(n^{-1/3} I^+(\lfloor n^{2/3} t \rfloor)\right)_{n \geq 1}$ and Lemma 2.38 imply that

$$\begin{aligned} &\left(n^{1/3} \frac{S_n^-\left(\lfloor n^{2/3} t \rfloor\right)}{\mu n - \lfloor n^{2/3} t \rfloor - I^{p,+}\left(\lfloor n^{2/3} t \rfloor\right) + 1}, t \geq 0 \right) \\ &= \left(n^{1/3} \frac{Y^-\left(\lfloor n^{2/3} t \rfloor - P_n\left(\lfloor n^{2/3} t \rfloor\right)\right) - P_n\left(\lfloor n^{2/3} t \rfloor\right) - I^+\left(\lfloor n^{2/3} t \rfloor\right) + 1}{\mu n - \lfloor n^{2/3} t \rfloor - I^+\left(\lfloor n^{2/3} t \rfloor\right) + 1}, t \geq 0 \right) \\ &\xrightarrow{(p)} \left(\frac{\nu_-}{\mu} t, t \geq 0 \right) \end{aligned} \tag{2.17}$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$. Then, by the continuous mapping theorem and the tightness of $(E^p(m))_{n \geq 1}$,

$$\left(n^{-1/3} \sum_{i=0}^{\lfloor n^{2/3} t \rfloor} p_k, t \geq 0 \right) \xrightarrow{(p)} \left(\frac{\nu_-}{2\mu} t^2, t \geq 0 \right)$$

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

$$\left(n^{-1/3} P_n(\lfloor n^{2/3} t \rfloor), t \geq 0 \right) \xrightarrow{\text{(d)}} \left(\frac{\nu_-}{2\mu} t^2, t \geq 0 \right),$$

which proves the statement.

$\square$

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

$$(B_t^{\text{d}}, t \geq 0) = \left(B_t - \frac{\nu_-}{2\mu\sigma_+} t^2, t \geq 0 \right).$$

Lemma 2.40. *We have*

$$\left(n^{-1/3} Y^+ \left(\lfloor n^{2/3} t \rfloor \right), n^{-1/3} S_n^+ \left(\lfloor n^{2/3} t \rfloor \right), t \geq 0 \right) \xrightarrow{\text{(d)}} (\sigma_+ B_t, \sigma_+ B_t^{\text{d}}, t \geq 0)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^2$ *and*

$$\left(n^{-2/3} S_n^- \left(\lfloor n^{2/3} t \rfloor \right), t \geq 0 \right) \xrightarrow{\text{(p)}} (\nu_- t, t \geq 0)$$

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 2.35 and Lemma 2.39, and the expressions

$$S_n^+(k+1) = Y^+(k+1-P_n(k+1)) - P_n(k+1),$$

and

$$S_n^-(k+1) = Y^-(k+1-P_n(k+1)) - P_n(k+1) - I^+(k) + 1.$$

□

Convergence of the height process

In this subsection, we will extend Lemma 2.40. 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 2.41. *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 [27]), but this is not the case for more general processes. We will adapt a technique that Broutin, Duquesne and Wang developed in [12] to show

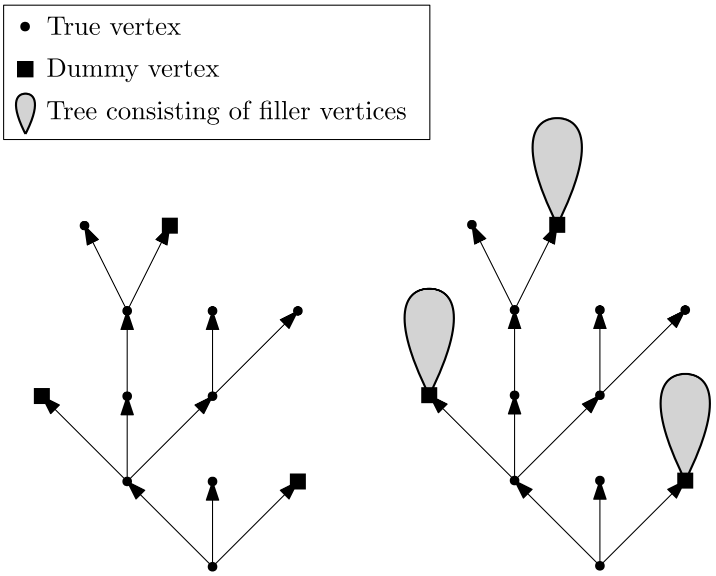


Fig. 2.4: Given a component of the forest with dummy vertices (left), we modify it by sampling independent Bienaymé trees with offspring distributed as Z^+ consisting of filler vertices and identifying each dummy leaf with a root of such a tree. The resulting tree (right) is a Bienaymé tree, and the resulting forest is a Bienaymé forest.

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 2.4.

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

$$(Y^{\mathrm{df}}(k), k \geq 1) = (Y^+(\Lambda_n(k)) + Y^{\mathrm{f}}(k - \Lambda_n(k)), k \geq 1) \tag{2.18}$$

   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,

$$(S_n^+(k), k \geq 1) = (Y^{\mathrm{df}}(\theta_n(k)) - E(\theta_n(k)), k \geq 1), \tag{2.19}$$

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 2.41, considering the construction above and Lemma 2.40, it is sufficient to prove the following lemma.

Lemma 2.42. *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 $(B^{\mathrm{f}}_t, t \geq 0)$ be two independent Brownian motions and let

$$\theta(t) := t + \inf\left\{s \geq 0 : \sigma_+ B^{\mathrm{f}}_s < -\frac{\nu_-}{2\mu}t^2\right\},$$

and $\Lambda(t) = \inf\{s \geq 0 : \theta(s) > t\}$. Define

$$\left(B^{\mathrm{df}}_t, t \geq 0\right) := \left(B_{\Lambda(t)} + B^{\mathrm{f}}_{t-\Lambda(t)}, t \geq 0\right) \tag{2.20}$$

and set

$$\left(R^{\mathrm{df}}_t, t \geq 0\right) := \left(B^{\mathrm{df}}_t - \inf\{B^{\mathrm{df}}_s, s \leq t\}, t \geq 0\right).$$

Lemma 2.43. *We have that* $\left((2/\sigma_+)R^{\mathrm{df}}_t, t \geq 0\right)$ *is the height process corresponding to* $\left(\sigma_+ B^{\mathrm{df}}_t, t \geq 0\right)$. *Moreover,*

$$\left(n^{-1/3}Y^{\mathrm{df}}\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}H^{\mathrm{df}}\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\sigma_+ B^{\mathrm{df}}_t, \frac{2}{\sigma_+}R^{\mathrm{df}}_t, t \geq 0\right) \tag{2.21}$$

in $D(\mathbb{R}_+, \mathbb{R})^2$, *jointly with*

$$\left(n^{-1/3}Y^{+}\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}Y^{\mathrm{f}}\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\sigma_+ B_t, \sigma_+ B^{\mathrm{f}}_t, t \geq 0\right)$$

in $D(\mathbb{R}_+, \mathbb{R})^2$ *and*

$$\left(n^{-2/3}\Lambda_n\left(\lfloor n^{2/3}t\rfloor\right), n^{-2/3}\theta_n\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{d})} (\Lambda(t), \theta(t), t \geq 0)$$

in $D(\mathbb{R}_+, \mathbb{R})^2$ *as* $n \to \infty$. *Moreover,*

$$\left(n^{-1/3}Y^{\mathrm{df}}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), n^{-1/3}H^{\mathrm{df}}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\sigma_+ B^{\mathrm{df}}_{\theta(t)}, \frac{2}{\sigma_+}R^{\mathrm{df}}_{\theta(t)}, t \geq 0\right) \quad (2.22)$$

in $D(\mathbb{R}_+, \mathbb{R})^2$ *as* $n \to \infty$ *jointly with the other convergences.*

In the proof of Lemma 2.43 we use the following straightforward technical result that follows immediately from the characterization of convergence in the Skorokhod topology given in Ethier and Kurtz [31, Proposition 3.6.5], .

Lemma 2.44. *If* $h_n \to h$ *and* $f_n \to f$ *in* $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ *as* $n \to \infty$, *and* h_n *and* h *are monotone non-decreasing and* h *is continuous, then*

$$h_n \circ f_n \to h \circ f$$

and

$$f_n \circ h_n \to f \circ h$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ *as* $n \to \infty$.

We also use the following technical result, that is proved in Appendix C.

Lemma 2.45. *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 2.43. Firstly, note that since $(Y^{\mathrm{df}}(k), k \geq 1)$ encodes a critical Galton-Watson forest with offspring variance σ_+^2, the proof of Theorem 1.8 in Le Gall [43] gives us that for $(B^*_s, s \geq 0)$ a

Brownian motion,

$$\begin{aligned}\left(n^{-1/3}Y^{\mathrm{df}}\left(\lfloor n^{2/3}s\rfloor\right),n^{-1/3}H^{\mathrm{df}}\left(\lfloor n^{2/3}s\rfloor\right),s\geq 0\right)&\\ \xrightarrow{\mathrm{(d)}}\left(\sigma_+B^*_s,\frac{2}{\sigma_+}\left(B^*_s-\inf\{B^*_u:u\leq s\}\right),s\geq 0\right)&\end{aligned}\tag{2.23}$$

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

$$\left(n^{-1/3}Y^+\left(\lfloor n^{2/3}t\rfloor\right),t\geq 0\right)\xrightarrow{\mathrm{(d)}}(\sigma_+B_t,t\geq 0)$$

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 2.45 and the almost sure uniqueness of the local minima of Brownian, we get that

$$\begin{aligned}\left(n^{-1/3}Y^{\mathrm{f}}\left(\lfloor n^{2/3}s\rfloor\right),n^{-2/3}\inf\left\{k:n^{-1/3}Y^{\mathrm{f}}(k)\leq -x\right\},s\geq 0,x\geq 0\right)&\\ \xrightarrow{\mathrm{(d)}}\left(\sigma_+B^{\mathrm{f}}_s,\inf\left\{u:\sigma_+B^{\mathrm{f}}_u<-x\right\},s\geq 0,x\geq 0\right)&\end{aligned}\tag{2.24}$$

in $D(\mathbb{R}_+,\mathbb{R})^2$ as $n\to\infty$.

Since $(P_n(k),k\geq 1)$ is non-decreasing, applying Lemma 2.44, and combining the convergence in Eq. (2.24) with Lemma 2.39 gives that also

$$\left(n^{-2/3}\inf\left\{k:Y^{\mathrm{f}}(k)\leq -P_n\left(\lfloor n^{2/3}t\rfloor-1\right)\right\},t\geq 0\right)\xrightarrow{\mathrm{(d)}}\left(\inf\left\{u:\sigma_+B^{\mathrm{f}}_u<-\frac{\nu_-}{2\mu}t^2\right\},t\geq 0\right)$$

in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ jointly with the convergence in Eq. (2.24). Therefore,

$$\left(n^{-2/3}\theta_n\left(\lfloor n^{2/3}t\rfloor\right),t\geq 0\right)\xrightarrow{\mathrm{(d)}}(\theta(t),t\geq 0)\tag{2.25}$$

in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ jointly with the convergence in Eq. (2.24). 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

$$\left(n^{-2/3}\max\{j:\theta_n(j)\leq\lfloor n^{2/3}t\rfloor\}),t\geq 0\right)\xrightarrow{\mathrm{(d)}}(\Lambda(t),t\geq 0)$$

in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ jointly with the convergence in Eq. (2.24) and Eq. (2.25). Since $\max\{j : \theta_n(j)\le\lfloor n^{2/3}t\rfloor\}$ is of order $n^{2/3}$, and, by Lemma 2.39, $P_n(\lfloor n^{2/3}t\rfloor)$ is of order $n^{1/3}$, we get that

$$\left(n^{-2/3}\Lambda_n\left(\lfloor n^{2/3}t\rfloor\}\right),t\ge 0\right)\xrightarrow{(\mathrm{d})}(\Lambda(t),t\ge 0)$$

in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ jointly with the convergence in Eq. (2.24) and Eq. (2.25).

To finish the proof, we examine the construction of $(Y^{\mathrm{df}}(k),k\ge 1)$ in Eq. (2.18) and the construction of $(B^{\mathrm{df}}_s,s\ge 0)$ in Eq. (2.20). Note that $\Lambda_n(k)$ and $k-\Lambda_n(k)$ are non-decreasing. Again, by Lemma 2.44, this implies that

$$\left(n^{-1/3}Y^{\mathrm{df}}\left(\lfloor n^{2/3}t\rfloor\right),t\ge 0\right)\xrightarrow{(\mathrm{d})}\left(B^{\mathrm{df}}_t,t\ge 0\right)$$

in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ jointly with all earlier mentioned convergences. Combining this with the convergence in Eq. (2.23) proves Eq. (2.21). The fact that $(\theta_n(k),k\ge 1)$ is non-decreasing and Lemma 2.44 then imply Eq. (2.22). □

Lemma 2.46. *We have that*

$$\left(n^{-1/3}S^+\left(\lfloor n^{2/3}t\rfloor\right)),n^{-1/3}H^+\left(\lfloor n^{2/3}t\rfloor\right),t\ge 0\right)$$
$$\xrightarrow{(\mathrm{d})}\left(\sigma_+B^{\mathrm{df}}_{\theta(t)},\frac{2}{\sigma_+}\left(B^{\mathrm{df}}_{\theta(t)}-\inf\{B^{\mathrm{df}}_s:s\le\theta(t)\}\right),t\ge 0\right)$$

in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})^2$ *as* $n\to\infty$.

Proof. By Eq. (2.19), and by Lemma 2.43, it is sufficient to show that for any $t>0$,

$$n^{-1/3}\max_{k\le\lfloor n^{2/3}t\rfloor}E(k)\xrightarrow{(\mathrm{p})}0.$$

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

$$n^{-1/3}\max_{k\le\lfloor n^{2/3}t\rfloor}E(k)\le n^{-1/3}\max_{k\le\theta_n(\lfloor n^{2/3}t\rfloor)}(Y^{\mathrm{f}}(k)-Y^{\mathrm{f}}(k-1)+1),$$

which converges to 0 by tightness of $\left(n^{-2/3}\theta^n(\lfloor n^{2/3}t\rfloor)\right)_{n\ge 1}$ and the fact that

$$\left(n^{-1/3}Y^{\mathrm{f}}\left(\lfloor n^{2/3}s\rfloor\right),s\ge 0\right)$$

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 2.42.

Lemma 2.47. *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 [43], 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 $B^{\mathrm{df}}_{\theta(t-)} = B^{\mathrm{df}}_{\theta(t)}$. □

2.4.3 Proof of Proposition 2.34

We will now combine the convergence of the measure change under rescaling, which is the content of Corollary 2.35, and the convergence of the encoding processes of the forest with dummy leaves, which is the content of Proposition 2.41, in order to prove Proposition 2.34.

Proof of Proposition 2.34. 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 2.8, 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 2.36,

$$\left(n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3}T \rfloor} \widehat{D}_i^n \right)_{n>0}$$

is tight. Moreover, with a slight adaptation to the proof of Lemma 2.38, we can show that the sequence $\left(n^{-1/3}\tilde{P}_n\left(\lfloor n^{2/3}T \rfloor\right)\right)_{n>0}$ 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 $D_1^-, \ldots, D_n^-$ 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

$$(\hat{S}_n^+(k), \hat{H}_n, \hat{S}_n^-(k), \hat{P}_n(k), 1 \le k \le \lfloor n^{2/3}T \rfloor).$$

Then, these processes are constructed out of sample paths of $(\hat{Y}^+(k), \hat{Y}^-(k), 1 \le k \le \lfloor n^{2/3}T \rfloor)$ and independent randomness in exactly the same way as the sample paths of

$$(S_n^+(k), H_n^+(k), S_n^-(k), P_n(k), 1 \le k \le \lfloor n^{2/3}T \rfloor)$$

(corresponding to the forest with dummy vertices) are constructed out of sample paths of $(Y^+(k), Y^-(k), 1 \le k \le \lfloor n^{2/3}T \rfloor)$ and independent randomness. We will use the following notation:

$$\begin{aligned}
\hat{Y}_{(n)}^+ &:= \left(n^{-1/3}\hat{Y}^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \le t \le T\right)\\
\hat{S}_{(n)}^+ &:= \left(n^{-1/3}\hat{S}_n^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \le t \le T\right)\\
\hat{H}_{(n)} &:= \left(n^{-1/3}\hat{H}_n\left(\lfloor n^{2/3}t \rfloor\right), 0 \le t \le T\right)\\
Y_{(n)}^+ &:= \left(n^{-1/3}Y^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \le t \le T\right)\\
S_{(n)}^+ &:= \left(n^{-1/3}S_n^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \le t \le T\right)\\
H_{(n)}^+ &:= \left(n^{-1/3}H_n^+\left(\lfloor n^{2/3}t \rfloor\right), 0 \le t \le T\right)
\end{aligned}$$

Let $f : D([0,T], \mathbb{R})^3 \to \mathbb{R}$ be a bounded, continuous test-function. Then, for $m = \lfloor n^{2/3}T \rfloor$

$$\begin{aligned}
\mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\right] &= \mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\mathbb{1}_{R_n \ge m}\right] + o(1)\\
&= \mathbb{E}\left[\mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\middle| \hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,m}\right]\mathbb{1}_{R_n \ge m}\right] + o(1)\\
&= \mathbb{E}\left[\Phi(n,m)\mathbb{E}\left[f\left(Y_{(n)}^+, S_{(n)}^+, H_{(n)}^+\right)\middle| \mathbf{Z}_1, \dots, \mathbf{Z}_m\right]\right] + o(1)\\
&= \mathbb{E}\left[\Phi(n,m) f\left(Y_{(n)}^+, S_{(n)}^+, H_{(n)}^+\right)\right] + o(1),
\end{aligned}$$

where we use that $\mathbb{E}\left[f\left(\hat{Y}_{(n)}^+, \hat{S}_{(n)}^+, \hat{H}_{(n)}\right)\middle| \hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,m}\right]$ and $\mathbb{1}_{R_n \ge m}$ are bounded, adapted functions of $\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,m}$, and that $\Phi(n,m)$ is the measure change from $(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$ to $(\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,m})$. Then, if we repeat the proof of Proposition 2.36, using Proposition 2.41 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

$$(B_t^+, t \geq 0) = \left(B_t - \frac{\nu_-}{2\sigma_+\mu} t^2, t \geq 0\right),$$

Proposition 2.37 implies that the limit object has the right law. By Proposition 2.41, S_n^- converges in distribution under rescaling to a deterministic process, which will not be affected by the measure change. This completes the proof. □

2.4.4 The convergence of the out-forest holds conditionally on the multigraph being simple

We will now show that the parts of the directed multigraph we observe far beyond the timescale of interest are with high probability simple. We will then use an argument by Joseph [39] to show that this implies that Proposition 2.34 holds conditional on the resulting multigraph being simple. 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 [19], we call these anomalous edges.

Proposition 2.48. *Suppose $\beta < 1$. Then we have*

$$\mathbb{P}\left(B_n(\lfloor n^\beta \rfloor) > 0\right) \to 0$$

as $n \to \infty$.

Remark 2.49. *We adapt the proof of [39, Lemma 7.1] and of [19, 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 [39] and [19] 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

$$\frac{1}{\sum_{i=\lfloor n^\beta \rfloor+1}^n \widehat{D}_i^-}.$$

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

$$\frac{1}{\sum_{i=\lfloor n^\beta \rfloor+1}^n \widehat{D}_i^-}.$$

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

$$\left(\frac{1}{\sum_{i=\lfloor n^\beta \rfloor+1}^n \widehat{D}_i^-}\right)^2.$$

Let $p(i)$ denote the index of the parent of the vertex with index i. Also, denote

$$\mathcal{G}^n = \sigma\left(\widehat{D}_1^-, \widehat{D}_1^+, \ldots, \widehat{D}_n^-, \widehat{D}_n^+\right).$$

Then, by the conditional version of Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left(B_n^1(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) &\le \frac{\sum_{i=1}^{\lfloor n^\beta \rfloor} \widehat{D}_i^- \widehat{D}_i^+}{\sum_{i=\lfloor n^\beta \rfloor+1}^n \widehat{D}_i^-} \wedge 1, \\
\mathbb{P}\left(B_n^2(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) &\le \frac{\sum_{i=1}^{\lfloor n^\beta \rfloor} \widehat{D}_i^- \mathbb{E}\left[\widehat{D}_{p(i)}^+ \middle| \mathcal{G}^n\right]}{\sum_{i=\lfloor n^\beta \rfloor+1}^n \widehat{D}_i^-} \wedge 1, \\
\mathbb{P}\left(B_n^3(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) &\le \frac{\sum_{i=1}^{\lfloor n^\beta \rfloor} \sum_{j<i} (\widehat{D}_i^+)^2 (\widehat{D}_j^-)^2}{\left(\sum_{i=\lfloor n^\beta \rfloor+1}^n \widehat{D}_i^-\right)^2} \wedge 1,
\end{aligned}$$

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,

$$\begin{aligned}
\sum_{i=1}^{\lfloor n^\beta \rfloor} \widehat{D}_i^- \mathbb{E}\left[\widehat{D}_{p(i)}^+ \middle| \mathcal{G}^n \right] &\leq \left(\sum_{i=1}^{\lfloor n^\beta \rfloor} (\widehat{D}_i^-)^2 \right)^{1/2} \left(\sum_{i=1}^{\lfloor n^\beta \rfloor} \mathbb{E}\left[\widehat{D}_{p(i)}^+ \middle| \mathcal{G}^n \right]^2 \right)^{1/2} \\
&= \left(\sum_{i=1}^{\lfloor n^\beta \rfloor} (\widehat{D}_i^-)^2 \right)^{1/2} \left(\sum_{j=1}^{\lfloor n^\beta \rfloor} (\widehat{D}_j^+)^2 \sum_{i=1}^{\lfloor n^\beta \rfloor} \mathbb{E}\left[\mathbb{1}_{j=p(i)} \middle| \mathcal{G}^n \right] \right)^{1/2} \\
&\leq \left(\sum_{i=1}^{\lfloor n^\beta \rfloor} (\widehat{D}_i^-)^2 \right)^{1/2} \left(\sum_{i=1}^{\lfloor n^\beta \rfloor} (\widehat{D}_i^+)^3 \right)^{1/2} .
\end{aligned}$$

We will show that

$$\mathbb{P}\left(B_n^1(\lfloor n^\beta \rfloor) + B_n^2(\lfloor n^\beta \rfloor) + B_n^3(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n \right) \xrightarrow{p} 0 \tag{2.26}$$

as $n \to \infty$. We note that

$$\sum_{i=\lfloor n^\beta \rfloor + 1}^{n} \widehat{D}_i^- = \sum_{i=1}^{n} D_i^- - \sum_{i=1}^{\lfloor n^\beta \rfloor - 1} \widehat{D}_i^-,$$

and by the weak law of large numbers, $\frac{1}{n}\sum_{i=1}^n D_i^- \xrightarrow{p} \mu n$, so Eq. (2.26) follows if we show that

1. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} \widehat{D}_i^- \xrightarrow{p} 0$,
2. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} \widehat{D}_i^- \widehat{D}_i^+ \xrightarrow{p} 0$,
3. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} (\widehat{D}_i^-)^2 \xrightarrow{p} 0$, and
4. $\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} (\widehat{D}_i^+)^3 \xrightarrow{p} 0$

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 $(\widehat{\mathbf{D}}_{n,1}, \ldots, \widehat{\mathbf{D}}_{R_n,n})$. This technique was also used by Joseph in [39].

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 $(\widehat{\mathbf{D}}_{n,1},\dots,\widehat{\mathbf{D}}_{r,n})$ (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 [39, 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 $(\widehat{\mathbf{D}}_{n,1},\dots,\widehat{\mathbf{D}}_{n,R_n})$ 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

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} \widehat{D}_i^- \widehat{D}_i^+ > \epsilon\right) \leq \mathbb{P}(\mathcal{E}_L^c) + \mathbb{P}\left(\Pi_{R_n,S_n}\left((0,2n^\beta),\mathbb{N}^2\right) < n^\beta \middle| \mathcal{E}_L\right)$$
$$+ \mathbb{P}\left(\frac{1}{n}\int_{(0,2n^\beta)\times\mathbb{N}^2} k_1 k_2 \Pi_{R_n,S_n}(dt,k_1,k_2) > \epsilon \middle| \mathcal{E}_L\right)$$

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 $\mathbb{N}$,

$$\mathbb{P}\left[\Delta_r = s, N_r = r \left|\Pi_r(t,k_1,k_2) = 1\right.\right] = \mathbb{P}\left[\Delta_r = s + k_2 - k_1, N_r = r - 1\right],$$

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,

$$\mathbb{P}\left[\Delta_r = s, N_r = r \,\middle|\, \widehat{D}_t^- = k_1, \widehat{D}_t^+ = k_2\right] = O(n^{-1/2}), \text{ and}$$
$$\mathbb{P}\left[\Delta_r = s, N_r = r\right] = \Theta(n^{-1/2})$$

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

$$\frac{\mathbb{P}\left[\Delta_r = s, N_r = r \,\middle|\, \widehat{D}_t^- = k_1, \widehat{D}_t^+ = k_2\right]}{\mathbb{P}\left[\Delta_r = s, N_r = r\right]} < c_1$$

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

$$E_t^r\left[\widehat{D}^- \widehat{D}^+\right] < c_2$$

for all r in the interval that we consider and all $t < 2n^\beta$, it follows that there is a c_3 such that

$$E_t^{r,s}\left[\widehat{D}^- \widehat{D}^+\right] < c_3$$

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 $\pi_r(dt, k_1, k_1)$,

$$E_t^r\left[\widehat{D}^- \widehat{D}^+\right] = \frac{\frac{d^3}{dx^2 dy}\mathcal{L}_{\mathbf{D}}(x,y)|_{(\psi(t/r),0)}}{\frac{d}{dx}\mathcal{L}_{\mathbf{D}}(x,y)|_{(\psi(t/r),0)}}.$$

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 [39, Lemma A.1] for the details of a similar argument in the undirected setting. This implies that

$$\mathbb{E}\left[\frac{1}{n}\int_{(0,2n^\beta)\times\mathbb{N}^2} k_1 k_2 \Pi_{r,s}(dt, k_1, k_2)\right] \le \frac{C}{n}\mathbb{E}\left[\Pi_{r,s}\left((0,2n^\beta), \mathbb{N}, \mathbb{N}\right)\right].$$

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 $x > 1$,

$$\mathbb{P}\left(\Pi_{r,s}\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right) > (x+1)2n^\beta\right) \le c_4 \exp(-c_5 x n^\beta).$$

This implies that there is a constant c_6 such that

$$\mathbb{E}\left[\Pi_{r,s}\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right)\right] \le c_6 n^\beta$$

for all n and all r and s that we consider under $\mathcal{E}_L$. It then follows that

$$\mathbb{E}\left[\frac{1}{n}\int_{(0,2n^\beta)\times\mathbb{N}^2} k_1 k_2 \Pi_{r,s}(dt,k_1,k_2)\right] \to 0$$

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

$$\frac{1}{n}\sum_{i=1}^{\lfloor n^\beta \rfloor} \widehat{D}_i^- \widehat{D}_i^+ \xrightarrow{p} 0.$$

The other convergences are proved similarly, and the result follows. ☐

Proposition 2.50. *Proposition 2.34 holds conditionally on the resulting multigraph being simple.*

Proof. Let $\rho(n) = \inf\{k \ge 1 : B_n(k) > 0\}$, 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 2.48 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 [39]. ☐

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.

2.5 Convergence of the SCCs under rescaling

In this section, we will use the convergence of the out-forest that we obtained in Section 2.4 to show that the SCCs ordered by decreasing number of edges converge under rescaling in the $d_{\vec{\mathcal{G}}}$-product topology.

2.5.1 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 2.2.1, 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 2.34. 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 2.51. *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 [22, 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 2.4.2.

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 2.2.1. 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 [22] 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 2.4.2,

$$\left(n^{-1/3} Y^{\mathrm{df}}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{\mathrm{df}}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{\mathrm{df},\ell}\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\sigma_+ B_s, \frac{2}{\sigma_+} R_s, \frac{2(\sigma_{+-} + \nu_-)}{\mu \sigma_+} R_s, t \geq 0\right) \tag{2.27}$$

in $D(\mathbb{R}_+, \mathbb{R})^3$ as $n \to \infty$. Then, we observe that the rest of the argument in Subsections 2.4.2 and 2.4.3 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 [22] 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 $\mathbb{E}[D^+(D^-)^3] < \infty$. $\square$

Proposition 2.52. *We have, jointly with the convergence in Proposition 2.51,*

$$\left(A_n\left(\lfloor tn^{2/3}\rfloor\right), t \leq T\right) \xrightarrow{(\mathrm{d})} (A_t, t \leq T),$$

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 [21, Theorem 14.2.VIII], the claimed convergence under rescaling of $(A_n(k), k \geq 1)$ follows if we show that

$$\left(A_n^{comp}\left(\lfloor tn^{2/3}\rfloor\right), t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\frac{2(\sigma_{-+} + \nu_-)}{\sigma_+ \mu^2} \int_0^t \hat{R}_v dv, t \geq 0\right) \tag{2.28}$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in Proposition 2.51. Therefore, we will now prove that Eq. (2.28) 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 2.51, Lemma

2.44, Slutsky's lemma and the continuous mapping theorem imply that

$$\left(\sum_{j=1}^{\lfloor n^{1/3}t\rfloor} \frac{\hat{H}^{\ell}(\min\{l : \hat{P}_n(l) \geq k\})}{\hat{S}_n^-(\min\{l : \hat{P}_n(l) \geq k\})}, t \geq 0\right) \xrightarrow{(\mathrm{d})} \left(\frac{2(\sigma_{-+} + \nu_-)}{\sigma_+\mu} \int_0^t \frac{\hat{R}_{\tau(s)}}{\nu_-\tau(s)} ds, t \geq 0\right)$$

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 2.39 and apply Lemma 2.44, some simple analysis then yields Eq. (2.28), which proves the statement. $\square$

2.5.2 Finding the important components in the out-forest

In this subsection, we will show that, conditional on the convergence under rescaling in Proposition 2.52, 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 2.53 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.

$$\begin{aligned}g_i^n &= \inf\left\{t \geq 0 : f_n(t) = \inf\{f_n(s) : s \leq x_i^n\}\right\}, \\ \sigma_i^n &= \inf\left\{t \geq 0 : \inf\{f_n(s) : s \leq g_i^n + t\} < \inf\{f_n(s) : s \leq x_i^n\}\right\}.\end{aligned}$$

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.

$$\begin{aligned}g_i &= \inf\left\{t \geq 0 : f(t) = \inf\{f(s) : s \leq x_i\}\right\}, \\ \sigma_i &= \inf\left\{t \geq 0 : \inf\{f(s) : s \leq g_i + t\} < \inf\{f(s) : s \leq x_i\}\right\}.\end{aligned}$$

For $S = \{(a_i, b_i), i \in [m]\}$, let $\operatorname{ord}(S)$ 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 $\operatorname{ord}(S) \in (\mathbb{R}^2)^\infty$.

Lemma 2.53. *We have that*

$$\operatorname{ord}\left(\{(g_i^n, \sigma_i^n) : 1 \leq i \leq m\}\right) \to \operatorname{ord}\left(\{(g_i, \sigma_i) : 1 \leq i \leq m\}\right)$$

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 $\operatorname{ord}\left(\{(g_i^n, \sigma_i^n) : 1 \leq i \leq m\}\right)$. Therefore, the number of non-zero entries of $\operatorname{ord}\left(\{(g_i^n, \sigma_i^n) : 1 \leq i \leq m\}\right)$ can vary as n varies, which is why we work in $(\mathbb{R}^2)^\infty$. This lemma is proved in Appendix C.

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 2.2.1. 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 $i \in \left\{1, \ldots, A_n\left(\lfloor Tn^{2/3}/2 \rfloor\right)\right\}$,

$$\begin{aligned}G_i^n &= \min\left\{k \geq 1 : \hat{S}_n^+(k) = \min\{\hat{S}_n^{p,+}(l) : l \leq X_i^n\}\right\} \text{ and} \\ \Sigma_i^n &= \min\left\{k \geq 1 : \min\left\{\hat{S}_n^{p,+}(l) : l \leq G_i^n + k\right\} < \min\left\{\hat{S}_n^{p,+}(l) : l \leq X_i^n\right\}\right\},\end{aligned}$$

and for $i \in \{1, \ldots, A(T/2)\}$,

$$G_i = \inf\left\{t \geq 0 : \sigma_+ \hat{B}_t = \inf\{\sigma_+ \hat{B}_s : s \leq X_i\}\right\} \text{ and}$$
$$\Sigma_i = \inf\left\{t \geq 0 : \inf\{\sigma_+ \hat{B}_s : s \leq G_i + t\} < \inf\{\sigma_+ \hat{B}_s : s \leq X_i\}\right\}.$$

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 2.53.

Proposition 2.54. *It holds that*

$$\operatorname{ord}\left(\left\{\left(n^{-2/3}G_i^n, n^{-2/3}\Sigma_i^n\right) : 1 \leq i \leq A_n\left(\lfloor Tn^{2/3}/2\rfloor\right)\right\}\right) \xrightarrow{(\mathrm{d})} \operatorname{ord}\left(\{(G_i, \Sigma_i) : 1 \leq i \leq A(T/2)\}\right)$$

in the product topology on $(\mathbb{R}^2)^\infty$ *as* $n \to \infty$*, jointly with the convergence in Proposition 2.52.*

Proof. By Skorokhod's representation theorem, we may work on a probability space where the convergence in Proposition 2.52 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 2.53 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 $A_n\left(\lfloor Tn^{2/3}/2\rfloor\right) \to A(T/2)$ as $n \to \infty$, we can pick n large enough such that $A_n\left(\lfloor Tn^{2/3}/2\rfloor\right) = A(T/2)$. By the local absolute continuity of $(\hat{B}_t, t \geq 0)$ to a Brownian motion, its local minima are almost surely unique. Since

$$\left(A_n\left(\lfloor tn^{2/3}\rfloor\right), t \leq T/2\right) \overset{a.s.}{\to} \left(A(t), t \leq T/2\right)$$

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 2.53, and the convergence follows. □

2.5.3 Convergence of the set of candidates

By Lemma 2.53, 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 2.2.1.

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 2.52 and 2.54 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 2.10 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 2.2.2 to find the candidates in $\mathcal{T}_g$. Denote the sequence of candidates by $\mathbf{V}(g)$.

Proposition 2.55. *Jointly with the convergence in Proposition 2.54,*

$$n^{-2/3}\mathbf{V}_n(g_n) \xrightarrow{(\mathrm{d})} \mathbf{V}(g)$$

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 2.52 and 2.54, $n^{-2/3}V_1^n \overset{a.s.}{\to} V_1$. 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

$$n^{-2/3}(V_1^n, \ldots, V_m^n) \overset{a.s.}{\to} (V_1, \ldots, V_m). \tag{2.29}$$

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

$$K^n_{comp,m+1}(k) = \sum_{i=V^n_m+1}^{k} \frac{\ell\left(T_i^{n,\mathrm{mk}}\right) - m}{\hat{S}^-(i)} \mathbb{1}\{P_n(i) = P_n(i-1)+1\}$$

for $k \in [V^n_m + 1, g_n + \sigma_n]$ and 0 otherwise, where $T_i^{n,\mathrm{mk}}$ is the subtree of $T^n_{g_n}$ spanned by $\{g_n + 1, V^n_1, \dots, V^n_m, i\}$. Moreover, for T_s the subtree of $\mathcal{T}_g$ spanned by $\{g, V_1, \dots, V_m, s\}$, 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 2.51, and by Proposition 2.54, 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. (2.29) 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 2.52 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 $\mathbf{V}(g) = (V_1, \dots, V_m)$, $\mathbf{V}_n(g_n) = (V_1^n, \dots, V_m^n)$, and the statement follows. If $K_m(\infty) > 0$, apply the induction step to $(V_1, \dots, V_{m+1})$ and $(V_1^n, \dots, V_{m+1}^n)$. The fact that $|\mathbf{V}(g)| < \infty$ almost surely, as shown in Subsection 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 2.56. *Suppose the convergence in Propositions 2.52, 2.54 and 2.55 holds almost surely. Then, for* $\mathbf{V}_n(g_n) = (V_1^n, \dots, V_{N_n}^n)$, $\mathbf{V}(g) = (V_1, \dots, V_N)$, *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,*

$$\begin{aligned}&\left(n^{-1/3}T_{g_n}^n, n^{-2/3}(g_n+1), \left(n^{-2/3}V_1^n, n^{-2/3}W_1^n\right) \dots, \left(n^{-2/3}V_{N_n}^n, n^{-2/3}W_{N_n}^n\right)\right)\\ &\xrightarrow{\text{(d)}} (\mathcal{T}_g, l, (V_1, W_1), \dots, (V_N, W_N))\end{aligned}$$

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 $m \le N_n$, W_m^n is the vertex corresponding to a uniform unpaired in-half-edge of the vertices in $T_{g_n}^n(\{g_n+1, V_1^n, \dots, V_m^n\})$. By Proposition 2.51 and Slutsky's lemma,

$$\left(\frac{\hat{H}_n^\ell\left(\lfloor tn^{2/3}\rfloor\right)}{\hat{H}_n\left(\lfloor tn^{2/3}\rfloor\right)}, t \ge 0\right) \overset{a.s.}{\to} \left(\frac{\sigma_{-+}+\nu_-}{2\mu}, t \ge 0\right)$$

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 $T_{g_n}^n(\{g_n+1, V_1^n, \dots, V_m^n\})$. Note that, by Theorem 2.34, Propositions 2.54 and 2.55, 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 [34], this implies that

$$\left(n^{-1/3}T_{g_n}^n, n^{-2/3}g_n+1, n^{-2/3}V_1^n, \dots, n^{-2/3}V_m^n\right) \overset{a.s.}{\to} (\mathcal{T}_g, g, V_1, \dots, V_m)$$

in the $(m+1)$-pointed Gromov-Hausdorff topology. Since the relation

$$\left|T_{g_n}^n(\{g_n+1, V_1^n, \dots, V_m^n\})\right| = \left|T_{g_n}^n(\{g_n+1, V_1^n, \dots, V_m^n, W_m^n\})\right|$$

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 2.55 and 2.56 imply the following proposition.

Proposition 2.57. *By Skorokhod's representation theorem, we may work on a probability space where the convergence in Propositions 2.55 and 2.56 holds almost surely. Let $T^{n,mk}(g_n)$ be the subtree of $T_{g_n}^n$ spanned by $\{g_n + 1, V_1^n, \ldots, V_{N_n}^n\}$, and similarly, let $T^{mk}(g)$ be the subtree of $\mathcal{T}_g$ spanned by $\{g, V_1, \ldots, V_N\}$. Then, also*

$$\left(n^{-1/3}T^{n,mk}(g_n), n^{-2/3}(g_n+1), \left(n^{-2/3}V_1^n, n^{-2/3}W_1^n\right) \ldots, \left(n^{-2/3}V_{N_n}^n, n^{-2/3}W_{N_n}^n\right)\right)$$
$$\to \left(T^{mk}(g), g, (V_1, W_1), \ldots, (V_N, W_N)\right)$$

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 2.2.1. In $T^{n,\mathrm{mk}}(g_n)$, set $V_i^n \sim W_i^n$ for each $1 \leq i \leq N_n$, and set $M_{g_n}^n := T^{n,\mathrm{mk}}(g_n)/\sim$. Moreover, in $T^{\mathrm{mk}}(g)$, set $V_i \sim W_i$ for each $1 \leq i \leq N$, 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 $M_{g_n}^n$, let the vertex set consist of $g_n + 1$, W_i^n for $i \leq N_n$, and the branch points $V_i^n \wedge V_j^n$ for $i \neq j \leq N_n$. Similarly, in $\mathcal{M}_g$, let the vertex set consist of g, W_i for $i \leq N$, and the branch points $V_i \wedge V_j$ for $i \neq j \leq N$. Then we have the following proposition.

Proposition 2.58. *On the probability space where the convergence in Propositions 2.55 and 2.56 holds almost surely, $n^{-1/3}M_{g_n}^n \overset{a.s.}{\to} \mathcal{M}_g$ in $\vec{\mathcal{G}}$.*

Proof. The proof is analogous to the proof of Proposition 5.6 in [34]. □

Proposition 2.59. *On the probability space where the convergence in Propositions 2.55 and 2.56 holds almost surely, the SCCs in $n^{-1/3}M_{g_n}^n$, 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 [34]. This proposition requires that the lengths of the SCCs in $\mathcal{M}_g$ have different lengths almost surely, which is the content of Proposition 2.14. ☐

Proposition 2.60. *Let $T > 0$, and let $(C_i^T(n), i \geq 1)$ 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 $(\mathcal{C}_i^T, i \geq 1)$ 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 2.54, Proposition 2.59, and the fact that all SCCs in the limit object have a different length by Proposition 2.14. ☐

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 [3, Lemma 9].

Lemma 2.61. *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 2.56. 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

$$\mathbb{P}(SCC\left(n,(s,\infty),\delta\right)\geq 1)\leq 2\epsilon.$$

□

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 2.62. *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 [34, 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 [4, Section 3]). This implies the statement. □

Theorem 2.1 then follows from Proposition 2.60, Lemma 2.61 and Lemma 2.62.

Acknowledgements for this chapter

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. Finally, the authors would like to thank an anonymous referee for very thorough reading of Donderwinkel and Xie [26], on which this chapter is based, and for many useful comments.

Chapter 3

Inhomogeneous Random Directed Graphs

3.1 Introduction

In this section we present work for proving the scaling limit of strongly connected components for a class of inhomogeneous random directed graphs.

We recall here the definition of $D(\mathbb{W}_n)$ from Section 1.2.3. Consider a *bi-weight sequence* $\mathbb{W}_n = (\mathbf{w}_1^n, \dots, \mathbf{w}_n^n)$, where each $\mathbf{w}_i^n \in (0,\infty)\times(0,\infty)$ is a *bi-weight*: $w_i^{n,-}$ is the *in-weight* and $w_i^{n,+}$ is the *out-weight* of vertex i. Further, let $l_n^\pm = \sum_{i=1}^n w_i^{n,\pm}$ and $l_n = l_n^- + l_n^+$. We say D is distributed as an *inhomogeneous random directed graph* (IRDG) on n vertices with bi-weight sequence $\mathbb{W}_n$, written $D \sim D(\mathbb{W}_n)$, if D is a random digraph on $[n]$ where each directed edge (i,j) is present independently with probability

$$\mathbb{P}((i,j)\in D) = 1 - \exp\left(-\frac{w_i^{n,+}w_j^{n,-}}{l_n}\right).$$

3.1.1 Previous work

Söderberg [53] first introduced a model for undirected inhomogeneous random graphs as an extension of the Erdős-Rényi model where each vertex has a different type and the probability of each edge being realized depends on the type of the two vertices in the edge. This model was studied formally by Bollobás et al. [11], where they demonstrate a phase transition in the largest connected component like what is seen in the Erdős–Rényi random graph. Our work on the scaling limits is based heavily on Broutin et al. [13], who determined a metric space scaling limit for a class of rank one inhomogeneous random graphs. In their model, instead of a bi-weight, each vertex has a single weight w_i^n, for $i = 1,\dots,n$, and each undirected edge (i,j) is present with probability

$$1 - \exp\left(-\frac{w_i^n w_j^n}{\sum_{i=1}^n w_i^n}\right).$$

For the directed inhomogeneous model, Cao and Olvera-Cravioto [17] studies phase transition in the size of the largest strongly connected component. A restatement of their results and how it is used to conjecture a critical window for our model is found in Section 1.2.3.

3.1.2 Assumptions

Following the work of Broutin et al. [13], we will not be assuming convergence of the empirical distribution of in- and out-weights, but will be assuming convergence of $\frac{l_n}{n}$. Further, in order to get sensible limits, we will assume the empirical third moments of the weight distribution will also converge to finite constants. This is akin to the finite moment assumptions for the directed configuration model.

In [13] they showed that when the maximum weight of a vertex is $\Omega(n^{1/3})$ in size then the scaling limit of the undirected components will no longer be Brownian but instead related to a Lévy tree. While we are interested in a similar phenomenon for directed graphs, we will first tackle the problem where the conjectured limit is Brownian. Thus, we will also assume the maximum in- and out-weight are $o(n^{1/3})$ in size.

We summarise these assumptions below, along with a reminder of the criticality assumption from Section 1.2.3 for easy reference. We assume there exist finite positive constants κ and $\beta_{a,b}$, for every pair of non-negative integers a and b satisfying $a+b \leq 3$, and a real constant α such that the following limits hold:

$$\frac{l_n}{n} = \frac{\sum_{i=1}^{n}(w_i^{n,-} + w_i^{n,+})}{n} \to \frac{1}{\kappa} \tag{3.1a}$$

$$\frac{\sum_{i=1}^{n}(w_i^{n,-})^a (w_i^{n,+})^b}{l_n} \to \beta_{a,b} \quad \forall a+b \leq 3. \tag{3.1b}$$

$$n^{-1/3} \max\left\{w_{\max}^{n,-}, w_{\max}^{n,+}\right\} \to 0 \quad \text{where } w_{\max}^{n,\pm} = \max_{i=1,\ldots,n} w_i^{n,\pm} \tag{3.1c}$$

$$\left(1 - \frac{\sum_{i=1}^{n} w_i^{n,-} w_i^{n,+}}{l_n}\right) n^{1/3} \to \alpha. \tag{3.1d}$$

It is convenient to work with $\kappa = \lim_{n\to\infty} \frac{n}{l_n}$ rather than $\lim_{n\to\infty} \frac{l_n}{n}$ as we will be dividing by l_n in most of our limits. Similarly to the directed configuration model, a further bound is required on one of the empirical fourth moments when we analyse adding back edges to an out-forest exploration of the IRDG. So we also assume:

$$\sup_n \frac{\sum_{i=1}^{n}(w_i^{n,-})^3 (w_i^{n,+})}{l_n} < \infty. \tag{3.2}$$

3.1.3 Conjectured result

Like for the scaling limits of the connected components of critical undirected random graphs, we believe the scaling limits of the SCCs found for the directed configuration model should exhibit universality. Hence, we conjecture the following:

Conjecture 3.1. *Suppose the $(\mathbb{W}_n)_{n=0}^{\infty}$ is a sequence of bi-weight sequences such that the assumptions in Eqs. (3.1a) to (3.1d) hold. Let $C_1^n, C_2^n, \ldots$ be the SCCs of $D(\mathbb{W}_n)$ in descending order of size. We pad this sequence by an infinite tail of singletons. Then there exists a four parameter family of random*

MDMs $\mathcal{M}(a,b,c,d) = (M_1, M_2, \ldots)$ *such that*

$$n^{-1/3}(\kappa(C_1^n), \kappa(C_2^n), \ldots) \xrightarrow{\text{(d)}} \mathcal{M}(a,b,c,d)$$

as $n \to \infty$ *with respect to an* ℓ^1 *distance using* $d_{\vec{\mathcal{G}}}$*. This is the same 4 parameter family as for the directed configuration model and the directed Erdős-Rényi model. Further*

$$a = -\alpha, \quad b = \kappa\beta_{2,1}, \quad \textit{and} \quad c = \sqrt{\beta_{1,2}}.$$

We do not yet have a conjectured value for the parameter d. Justification for the conjectured values of the parameters a, b and c are found in the next section.

3.1.4 Finite queue coupling

Before we present our main result for this chapter, we present a decomposition of an IRDG. Like for the directed configuration model, the strategy for proving convergence of the SCCs is to prove a metric space convergence of an out-forest for the IRDG, then prove a point process convergence of important additional edges. For the IRDG, there is a nice coupling of the IRDG with a queueing process that naturally yields an out-forest exploration of the IRDG.

Definition of the queue

We can couple an IRDG $D(\mathbb{W}_n)$ with a queueing process satisfying the folowing description:

1. There are customers $1, \ldots, n$ arriving into the queue.
2. The arrival time E_i of customer i is distributed as an $\text{Exp}(w_i^{n,-}/l_n)$ random variable, independently of the other customers.
3. The service time of customer i is $w_i^{n,+}$.
4. There is a single server with a last-in-first-out (LIFO) service policy, meaning that the arrival of a new customer will interrupt any ongoing service, and then immediately enter service.

Let $L^n(t)$ be the total *load* on the server at time t. This is the time required for the server to finish serving all the customers present in the queue at time t. To compute $L^n(t)$, we first define the *algebraic* *load* on the server, which is given by

$$Y^n(t) = -t + \sum_{i=1}^{n} w_i^{n,+} \mathbb{1}\left\{E_i \leq t\right\}.$$

The summation gives the total service time of customers that have arrived into the queue by time t. The $-t$ term accounts for the service that has been completed, but this keeps subtracting service even when the server is idle. Thus the true load on the server is given by the excursions of Y^n above its running infinum, i.e.

$$L^n(t) = Y^n(t) - \inf_{s\in[0,t]} Y^n(s).$$

Let $H^n(t)$ be the number of customers in the queue at time t. We refer to this as the *height process* associated with the algebraic load Y^n. Note we have the formula

$$H^n(t) = \#\left\{ s \in [0,t] : \inf_{r\in[s,t]} Y^n(r) > Y^n(s-) \right\},$$

so that we can obtain H^n from the process Y^n without needing to refer to the queue.

The queueing process gives rise naturally to an associated *queueing forest* F^n which is a sequence of directed trees where i is a child of j if and only if customer i interrupts the service of j. Each tree is rooted at a customer which arrives when the server is idle. This is well defined because there are almost surely no arrivals at the same time. See Fig. 3.1 for an example of the algebraic load annotated with the types of customers arriving in the queue, and the corresponding directed forest. This forest will act as a spanning forest for the IRDG.

Then algebraic load and H^n are continuous-time counterparts of the Łukasiewicz path and height process of this forest, hence why H^n is itself refered to as the height process.

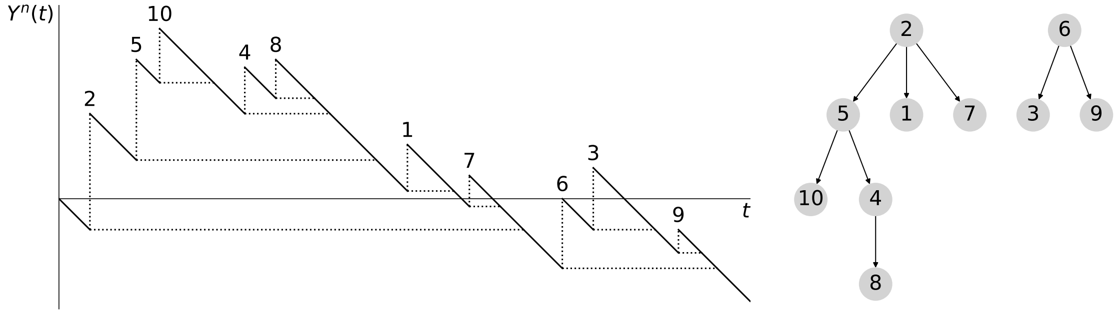


Fig. 3.1: Algebraic load Y^n annotated with types of arriving customers. The associated forest is shown on the right.

Sampling of additional edges

To obtain the IRDG, we need to add edges to the queueing forest. For all (i, j) such that $i \neq j$, let $W_{i,j}^{n,+}$ be the remaining service time of i once j arrives into the queue. We take $W_{i,i}^{n,+} = w_i^{n,+}$ for convenience. First note the following properties of $W_{i,j}^{n,+}$:

1. $W_{i,j}^{n,+} = w_i^{n,+}$ if j arrives before i in the queue.
2. $W_{i,j}^{n,+} = 0$ if j arrives after i leaves the queue, in other words if j is not a direct descendant of i in F^n.
3. $0 < W_{i,j}^{n,+} < w_i^{n+}$ if j arrives while i is in the queue or is being served, in other words if j is a direct descendant of i in F^n.

Then conditionally on the process Y^n (and therefore F^n and the values of $W_{i,j}^{n,+}$), for each $i, j \in [n]$, we add the directed edge (i, j) to F^n with probability

$$1 - \exp\left(-\frac{W_{i,j}^{n,+} w_j^{n,-}}{l_n}\right)$$

independently of the other edges. We ignore any multiple edges created by resampling existing edges of F^n. The resulting graph is distributed as the IRDG $D(\mathbb{W}_n)$, which is proven in Section 3.2.1.

Edges added in this manner from i to j where j arrives before i are called *back edges*. Otherwise they are *surplus edges*. This definition is different from the directed configuration model where all edges that are not part of the out-forest are labelled as surplus, and is more akin to the classification of edges used by Goldschmidt and Stephenson [34].

It is the interaction between the back edges and the edges of the queueing forest that creates the SCCs of the IRDG. Notice that the surplus edges do not affect the directed connectivity structure of the out-forest since they go from i to j where there is an existing directed path from i to j. Thus their removal would not affect which vertices are in which SCC, although surplus edges can still be included in SCCs if they are between two vertices in the same SCC.

3.1.5 Main Result

Our main result is the convergence of the height function of the forest F^n in the Skorokhod topology.

Theorem 3.2. *Then under the assumptions in Eqs.* (3.1a) *to* (3.1d)*,*

$$\left(n^{-1/3}H^n\left(n^{2/3}t\right)\right)_{t\geq 0} \xrightarrow{\text{(d)}} H$$

where B is a standard Brownian motion,

$$Y(t) = -\alpha t - \frac{1}{2}\kappa\beta_{2,1}t^2 + \sqrt{\beta_{1,2}}B_t \quad \text{and} \quad H(t) = \frac{2}{\beta_{1,2}}(Y(t) - \inf_{s\in[0,t]} Y(s)).$$

The form of Y is what motivates our conjectured parameters.

The proof of this is presented in Section 3.2 with the following structure. First in Section 3.2.1, we prove that the coupling presented in Section 3.2.1 is valid. We then study a Markovian modification of the queueing process in Section 3.2.2. The forest corresponding to the Markovian modification of the queueing process is a Bienaymé-Galton-Watson tree, and we show convergence of the corresponding height processes. In Section 3.2.3, we establish and prove the scaling limit of the algebraic load for the finite queue. We then use a method in [12], which we call the red-blue coupling, to derive convergence of the height process H^n, thus proving Theorem 3.10.

3.1.6 Next steps

We describe the steps required to complete this proof. We need to prove convergence of a counting process which counts the number of ancestral back-edges which come out of vertices corresponding to customers we have seen by time t. This would be the counterpart of Proposition 2.52. In Appendix A we present a scaling limit for the height process of F^n where we augment the tree with edge lengths, which would be the starting point of proving such a counterpart. The rate of such a counting process would determine the parameter d, which we have not yet conjectured a value for.

After we do so, we would need to separate out a subset of 'important' *candidate* edges from the back-edges. Following Section 2.5, we would use the limiting height process and point process to prove convergence of these candidate edges jointly with the out-trees containing those candidate edges.

We would also need to handle the surplus edges. While they do not affect the connectivity structure, they will still be included if they appear between two vertices in the same SCC. However, we conjecture this will be an event of vanishing probability for SCCs we observe by time $n^{2/3}T$ for $T > 0$. This would also need to be proven.

We would also need an of analogue of Lemma 2.61 showing for all $\delta > 0$. In particular let $A(n, t, \delta)$ be the event that by time $tn^{2/3}$ in the queueing process, there is still a customer yet to arrive which corresponds to a vertex in a SCC with size greater than $\epsilon n^{1/3}$. We need to show that

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}(A(n, t, \delta)) = 0.$$

3.2 Convergence of the height process

3.2.1 Coupling with IRDGs

Firstly we prove the coupling in Section 3.1.4 is valid, meaning we need to show the graph constructed in the section is distributed as the IRDG $D(\mathbb{W}_n)$. We will prove this inductively. Let $I_1, I_2, \ldots, I_n$ be the customers in order of arrival, and $\mathcal{I}_k = \{I_1, \ldots, I_k\}$ be the set of the first k customers. Let F_k^n be the associated queueing forest for the first k customers. F_k^n is then the subforest of F^n induced by the vertices $I_1, \ldots, I_k$.

Then conditional on the process Y^n up to the arrival of the kth customer, we add edges (i, j) to F_k^n independently with probability

$$\begin{cases} 1 - \exp\left(-\dfrac{W_{i,j}^{n,+} w_j^{n,-}}{l_n}\right) & \text{if } i \in \mathcal{I}_k \text{ and } j \in \mathcal{I}_k, & (3.3a) \\ 1 - \exp\left(-\dfrac{W_{i,I_k}^{n,+} w_j^{n,-}}{l_n}\right) & \text{if } i \in \mathcal{I}_k \text{ and } j \notin \mathcal{I}_k, & (3.3b) \\ 1 - \exp\left(-\dfrac{w_i^{n,+} w_j^{n,-}}{l_n}\right) & \text{if } i \notin \mathcal{I}_k & (3.3c) \end{cases}$$

to obtain the random digraph G_k^n. In particular G_n^n is the construction described previously since $\mathcal{I}_n = [n]$, so all edges are sampled according to case 3.3a.

Proposition 3.3. *Each G_k^n has the same distribution as $D(\mathbb{W}_n)$.*

Proof. We prove this by induction on k. G_0^n is distributed as $D(\mathbb{W}_n)$ since F_0^n is an empty forest and $\mathcal{I}_0$ is empty. Thus, all edges as sampled according to case 3.3c.

Suppose G_k^n is distributed as $D(\mathbb{W}_n)$. Then first consider sampling the edges from I_k to vertices $j \notin \mathcal{I}_k$. This can be done by taking independent exponential random variables $\tilde{E}_j \sim \mathrm{Exp}(w_j^{n,-}/l_n)$ and adding an edge if $\tilde{E}_j < W_{I_k,I_k}^{n,+} = W_{I_k}^{n,+}$. The $\tilde{E}_j$ can be interpreted as the time required for customer j to arrive after customer I_k, conditional on the queue up to the arrival of I_k. These are still independent exponentials with the required parameters due to the memoryless property of exponentials.

Let $\tilde{E} = \min_{j \notin \mathcal{I}_k} \tilde{E}_j = \tilde{E}_{I_{k+1}}$ be the time of the next arrival. If $\tilde{E} < W_{I_k}^{n,+}$, then the edge from I_k to the next vertex I_{k+1} is in included as the new edge in F_{k+1}^n. The remaining edges from I_k to vertices $j \notin \mathcal{I}_k$ should still be included if and only if $\tilde{E}_j < W^{n,+}_{I_k,I_k}$, which is equivalent to

$$\tilde{E}_j - \tilde{E} < W^{n,+}_{I_k,I_k} - \tilde{E}.$$

By properties of competing exponentials (see [49, Theorem 2.3.2]), conditionally on $\tilde{E}$ and I_{k+1}, the $\tilde{E}_j - \tilde{E}$ are still independent and exponentially distributed with parameter $w^{n,-}_j/l_n$. Further $W^{n,+}_{I_k,I_k} - \tilde{E}$ is exactly the service time remaining for customer I_k once I_{k+1} joins the queue. Thus,

$$W^{n,+}_{I_k,I_k} - \tilde{E} = W^{n,+}_{I_k,I_{k+1}}.$$

Hence, conditional on $\tilde{E}$ and I_{k+1}, these edges are now added independently with probability

$$1 - \exp(-W^{n,+}_{I_k,I_{k+1}} w^{n,-}_j/l_n).$$

Further, we still have $W^{n,+}_{i,I_k} = W^{n,+}_{i,I_{k+1}}$ for $i \neq I_k$. Hence, the probabilities used to add the remaining edges for G^n_k match the construction of G^n_{k+1}. Thus, they have the same distribution.

On the other hand, if $\tilde{E} > W^{n,+}_{I_k}$, then this corresponds to customer I_k leaving before I_{k+1} arrives. In that case there are no edges from I_k to $j \notin \mathcal{I}_k$. This matches with the construction of G^n_k since in this case we would have $W^{n,+}_{I_k,I_{k+1}} = 0$. Then by the memoryless property again, the times after the departure of I_k to the arrival of the jth customer, for $j \notin \mathcal{I}_k$, are still distributed as independent exponentials with parameter $w^{n,-}_j/l_n$. So we can repeat the previous argument with I_{k-1} (it is possible to have $W^{n,+}_{I_{k-1},I_k} = 0$ if I_{k-1} left the queue before I_k arrived) until either I_{k+1} interrupts a customer, or all customers depart.

Thus, in all cases, the constructions of G^n_k and G^n_{k+1} coincide. Hence, G^n_{k+1} is also distributed as $D(\mathbb{W}_n)$. The result follows by induction on k. □

3.2.2 The infinite queue

The algebraic load for the finite queue is not Markovian. This is because we need to know which customers have already been served and have left the queue in order to know the future behaviour of the queue, and this cannot be inferred from the current queue length. To mitigate this, we can instead consider a queue where there are infinitely many customers separated into n types.

- There are infinitely many customers of types $1, \ldots, n$ arriving into the queue.

- Customers of type i arrive according to a Poisson process of rate $w_i^{n,-}/l_n$.
- The service time of customers of type i is $w_i^{n,+}$.
- The arrival of customers of different types are independent of each other.
- There is a single server with a LIFO service policy.

As for the finite queue, we can define the algebraic load of the infinite queue as

$$X^n(t) = -t + \sum_{i=1}^{n} w_i^{n,+} N_i^n(t)$$

where $N_i^n(t)$ is the number of customers of type i that have arrived by time t. Hence, $(N_i^n(t))_{t\geq 0}$ are independent Poisson processes where N_i^n has rate $w_i^{n,-}/l_n$ for $1 \leq i \leq n$. We can similarly define an associated height process

$$\mathcal{H}^n(t) = \# \left\{ s \in [0,t] : \inf_{r\in[s,t]} X^n(r) > X^n(s-) \right\}.$$

For future proofs, it is useful to recall some bounds on the Taylor remainder for the exponential function. Let R_k denote the kth Taylor remainder of e^{-x}, meaning

$$R_k(x) = e^{-x} - \sum_{l=0}^{k} \frac{(-1)^l}{l!} x^l.$$

Then for all $x \geq 0$, by the integral form of the Taylor remainder, we have the bound

$$|R_k(x)| = \left| \int_0^x \frac{(-1)^k}{k!} e^{-t}(x-t)^k \, \mathrm{d}t \right| \leq \frac{1}{k!} x^k (1 - e^{-x}) \leq \frac{1}{k!} x^{k+1}. \tag{3.4}$$

Scaling limits for the infinite queue

In contrast to the finite queue, the algebraic load for the infinite queue is Markov. And it turns out to be a Lévy process. Using this, we can derive the scaling limit of X^n when appropriately rescaled. Let B be a standard Brownian motion and let

$$X(t) = -\alpha t + \sqrt{\beta_{1,2}} B_t$$

for $t \geq 0$, where we recall from Eq. (3.1b) that

$$\beta_{1,2} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} w_i^{n,-}(w_i^{n,+})^2}{l_n}.$$

The next lemma shows that X is the desired scaling limit.

Lemma 3.4. *Under the assumptions in Eqs.* (3.1a) *to* (3.1d)*,*

$$\left(n^{-1/3}X^n\left(n^{2/3}t\right)\right)_{t\geq 0} \xrightarrow{\text{(d)}} (X(t))_{t\geq 0}$$

as $n \to \infty$ with respect to the Skorokhod topology on $\mathbb{D}([0,\infty),\mathbb{R})$.

Proof. Let ϕ_n be the Lévy exponent of X^n, meaning that

$$\mathbb{E}\left[e^{-\lambda X^n(t)}\right] = e^{t\phi_n(\lambda)}$$

for all $\lambda \in \mathbb{R}$. As X^n and X are both Lévy processes, a result shown in Jacod and Shiryaev [36, Chapter VII, Corollary 3.6] states that the convergence in this lemma holds if and only if

$$n^{-1/2}X^n(n^{2/3}) \xrightarrow{\text{(d)}} X(1)$$

in $\mathbb{R}$ as $n \to \infty$. Since the moment generating functions of $n^{-1/3}X^n(n^{2/3})$ and $X(1)$ are $n^{2/3}\phi_n(n^{-1/3}\lambda)$ and $\phi(\lambda)$ respectively, both of which are finite for all $\lambda \in \mathbb{R}$, it suffices to show that

$$n^{2/3}\phi_n(n^{-1/3}\lambda) \to \alpha\lambda + \tfrac{1}{2}\beta_{1,2}\lambda^2$$

for all $\lambda \in \mathbb{R}$. We can compute that

$$\begin{aligned}
\phi_n(\lambda) &= \lambda + \sum_{i=1}^{n} \frac{w_i^{n,-}}{l_n}\left(e^{-\lambda w^+_{n,i}} - 1\right) \\
&= \left(1 - \sum_{i=1}^{n} \frac{w_i^{n,-}w_i^{n,+}}{l_n}\right)\lambda + \frac{1}{2}\sum_{i=1}^{n} \frac{(w_i^{n,-})(w_i^{n,+})^2}{l_n}\lambda^2 + \sum_{i=1}^{n} \frac{w_i^{n,-}}{l_n} R_2\left(w_i^{n,+}\lambda\right).
\end{aligned}$$

The Lévy exponent of the rescaled process $\left(n^{-1/3}X^n\left(n^{2/3}t\right)\right)_{t\geq 0}$ is then given by

$$n^{2/3}\phi_n(n^{-1/3}\lambda) = n^{1/3}\left(1-\sum_{i=1}^{n}\frac{w_i^{n,-}w_i^{n,+}}{l_n}\right)\lambda + \frac{1}{2}\sum_{i=1}^{n}\frac{(w_i^{n,-})(w_i^{n,+})^2}{l_n}\lambda^2 \\ + n^{2/3}\sum_{i=1}^{n}\frac{w_i^{n,-}}{l_n}R_2\left(n^{-1/3}w_i^{n,+}\lambda\right).$$

By the assumption in Eq. (3.1d), the first summand will converge to $\alpha\lambda$. By the assumption in Eq. (3.1b), the second summand will converge to $\frac{1}{2}\beta_{1,2}\lambda^2$. Finally, by using the inequality in Eq. (3.4), we can bound the third summand as follows:

$$\left|n^{2/3}\sum_{i=1}^{n}\frac{w_i^{n,-}}{l_n}R_2\left(n^{-1/3}w_i^{n,+}\lambda\right)\right| \leq \frac{1}{2}n^{-1/3}\sum_{i=1}^{n}\frac{(w_i^{n,-})(w_i^{n,+})^3}{l_n}\lambda^3 \\ \leq \frac{1}{2}\frac{w_{\max}^{n,+}}{n^{1/3}}\sum_{i=1}^{n}\frac{(w_i^{n,-})(w_i^{n,+})^2}{l_n}\lambda^3$$

which will converge to 0 by the assumptions in Eqs. (3.1b) and (3.1c). ☐

We now address the scaling limit of the contour process. Recall from [27, Eq. 1.7] that the continuous height process associated with X is given by

$$\mathcal{H}(t) = \frac{2}{\beta_{1,2}}\left(X(t) - \inf_{s\in[0,t]}X(s)\right).$$

Once convergence of $(n^{-1/3}X^n(n^{2/3}t))_{t\geq 0}$ to X is established, a combination of Propositions 2.2 and 2.3 in [12] gives the following result

Proposition 3.5. *Under the assumptions in Eqs.* (3.1a) *to* (3.1d)*, as* $n\to\infty$*,*

$$\left(\left(n^{-1/3}X^n\left(n^{2/3}t\right)\right)_{t\geq 0}, \left(n^{-1/3}\mathcal{H}_n\left(n^{2/3}t\right)\right)_{t\geq 0}\right) \xrightarrow{(d)} (X,\mathcal{H})$$

on $\mathbb{D}([0,\infty),\mathbb{R})\times\mathbb{C}([0,\infty],\mathbb{R})$ *equipped with the product topology of the Skorokhod topology on* $\mathbb{D}([0,\infty),\mathbb{R})$ *and the uniform topology on* $\mathbb{C}([0,\infty],\mathbb{R})$*.*

Proof. Proposition 2.2 in [12] gives the required convergence of the rescaled height process if we can check conditions (C1) to (C4). In our context, (C1) corresponds to our assumption in Eq. (3.1d), (C2) corresponds to our assumption in Eq. (3.1b) and (C3) corresponds to our assumption in Eq. (3.1c). We are considering a scaling regime $(a_n^{-1}X^n_{b_nt})_{t\geq 0}$ for $a_n = n^{1/3}$ and $b_n = n^{2/3}$. Here $\lim_{n\to\infty}\frac{b_n}{a_n^2} = 1$ which is strictly positive, hence by the remark in paragraph after [12, Proposition 2.2], condition (C4) is also satisfied. $\square$

3.2.3 Convergence of processes for the finite queue

Let

$$\begin{aligned}
A(t) &= \frac{1}{2}\kappa\beta_{2,1}t^2, \\
Y(t) &= X(t) + A(t) = -\alpha t - \frac{1}{2}\kappa\beta_{2,1}t^2 + \sqrt{\beta_{1,2}}B_t, \quad \text{and} \\
H(t) &= \frac{2}{\beta_{1,2}}\left(Y(t) - \inf_{s\in[0,t]} Y(s)\right)
\end{aligned}$$

for $t \geq 0$.

Convergence of the algebraic load

We start by proving that Y^n will converge to Y when properly rescaled. To do this, we couple the finite queue with the infinite queue using a first arrival coupling. Since the first arrival time of a Poisson process of rate λ is distributed exponentially with rate λ, we can recover our finite queue by only considering the first customer of each type. The algebraic load of the resulting queue is given by

$$Y^n(t) = -t + \sum_{i=1}^{n} w_i^{n,+}\mathbb{1}\{N_i^n(t) \geq 1\}.$$

We also define the process

$$A^n(t) = \sum_{i=1}^{n} w_i^{n,+}(N_i^n(t) - 1)_+$$

which counts the total queue time contributed by customers from repeated types at time t. We call any jump in X^n that is due to a customer from a repeated type an *excess jump*. Then Y^n is obtained by removing excess jumps from X^n. The processes X^n, Y^n and A^n are related by $X^n = Y^n + A^n$.

This coupling allows us to derive the scaling limit of Y^n using the scaling limits of X^n and A^n. We first prove the scaling limit of A^n.

Lemma 3.6. *Under the assumptions in Eqs. (3.1a) to (3.1d),*

$$\mathbb{E}\left[\sup_{t\leq t_0}\left|n^{-1/3}A^n\left(n^{2/3}t\right) - \tfrac{1}{2}\kappa\beta_{2,1}t^2\right|\right] \to 0 \tag{3.5}$$

as $n \to \infty$ *for all* $t_0 \geq 0$. *In particular* $\left(n^{-1/3}A^n\left(n^{2/3}t\right)\right)_{t\geq 0} \xrightarrow{(d)} \left(\tfrac{1}{2}\kappa\beta_{2,1}t^2\right)_{t\geq 0}$ *with respect to the*

Skorokhod topology on $\mathbb{D}([0,\infty),\mathbb{R})$.

Proof. Suppose N is a Poisson process of rate λ. Let T be the time of the first jump of N, which will be an $\mathrm{Exp}(\lambda)$ random variable. Then $(N(t)-1)_+$ is the process obtained by skipping the first jump of the process. Hence

$$(N(t)-1)_+ = N((t-T)_+ + T) - N(T) = \tilde{N}((t-T)_+)$$

where $\tilde{N}(t) = N(t+T) - N(T)$. By the strong Markov property of Poisson processes, $\tilde{N}$ is another Poisson process that is independent of T. Since $(\tilde{N}(t) - \lambda t)_{t\geq 0}$ is a martingale,

$$(N(t)-1)_+ - \lambda(t-T)_+ = \tilde{N}((t-T)_+) - \lambda(t-T)_+$$

is also a martingale. Therefore, if we define

$$C^n(t) = \sum_{i=1}^{n} \frac{w_i^{n,-} w_i^{n,+}}{l_n}(t - T_i^n)_+,$$

where T_i^n is time N_i^n makes its first jump, then $A^n - C^n$ is a martingale. Hence, by Doob's L^2 inequality we have that

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\leq t_0}\left(A^n(t) - C^n(t)\right)^2\right] &\leq 4\mathbb{E}\left[\left(A^n(t_0) - C^n(t_0)\right)^2\right] \\
&= 4\mathbb{E}\left[\left(\sum_{i=1}^{n} w_i^{n,+}\left\{(N_i^n(t_0)-1)_+ - \tfrac{w_i^{n,-}}{l_n}(t_0 - T_i^n)_+\right\}\right)^2\right] \\
&= 4\sum_{i=1}^{n}(w_i^{n,+})^2\mathbb{E}\left[\left((N_i^n(t_0)-1)_+ - \tfrac{w_i^{n,-}}{l_n}(t_0 - T_i^n)_+\right)^2\right],
\end{aligned}$$

where the last equality follows because the summands are independent with mean 0. We can compute

$$\begin{aligned}
\mathbb{E}\left[\left\{(N(t)-1)_+ - \lambda(t-T)_+\right\}^2\right] &= \int_0^t \mathbb{E}\left[\left\{\tilde{N}(t-s) - \lambda(t-s)\right\}^2\right]\lambda e^{-\lambda s}\,\mathrm{d}s \\
&= R_1(\lambda t) \leq \lambda^2 t^2,
\end{aligned}$$

by the bound in Eq. (3.4). Therefore,

$$\mathbb{E}\left[\sup_{t\leq t_0}\left(A^n(t) - C^n(t)\right)^2\right] \leq 4\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})^2}{l_n^2}t_0^2 \leq 4w_{\max}^{n,+}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})}{l_n^2}t_0^2.$$

Hence, after rescaling we get that

$$\mathbb{E}\left[\sup_{t\le t_0}\left(n^{-1/3}A^n\left(n^{2/3}t\right)-n^{-1/3}C^n\left(n^{2/3}t\right)\right)^2\right]\le 4\frac{w_{\max}^{n,+}}{n^{1/3}}\frac{n}{l_n}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})}{l_n}t_0^2,$$

which converges to 0 by the assumptions in Eqs. (3.1a) to (3.1c). Thus, it suffices to prove that

$$\mathbb{E}\left[\sup_{t\le t_0}\left(n^{-1/3}C^n\left(n^{2/3}t\right)-\tfrac{1}{2}\kappa\beta_{2,1}t^2\right)^2\right]\to 0$$

as $n\to\infty$. To help with this, note that

$$(t-T)_+=\int_0^t \mathbb{1}\left\{N(u)\ge 1\right\}\mathrm{d}u,$$

for all $t\ge 0$. Therefore

$$\begin{aligned}n^{-1/3}C^n\left(n^{2/3}t\right)&=n^{-1/3}\sum_{i=1}^{n}\frac{w_i^{n,-}w_i^{n,+}}{l_n}\int_0^{n^{2/3}t}\mathbb{1}\left\{N_i^n(u)\ge 1\right\}\mathrm{d}u\\&=\int_0^t\left(n^{1/3}\sum_{i=1}^{n}\frac{w_i^{n,-}w_i^{n,+}}{l_n}\mathbb{1}\left\{N_i^n\left(n^{2/3}s\right)\ge 1\right\}\right)\mathrm{d}s\end{aligned}$$

where we have changed variables with $u=n^{2/3}s$. Let ∂C^n denote the integrand, meaning

$$\partial C^n(s)=n^{1/3}\sum_{i=1}^{n}\frac{w_i^{n,-}w_i^{n,+}}{l_n}\mathbb{1}\left\{N_i^n\left(n^{2/3}s\right)\ge 1\right\},$$

for all $s\ge 0$. Then

$$\begin{aligned}\left(n^{-1/3}C^n\left(n^{2/3}t\right)-\tfrac{1}{2}\kappa\beta_{2,1}t^2\right)^2&=\left(\int_0^t\left(\partial C^n(s)-\kappa\beta_{2,1}s\right)\mathrm{d}s\right)^2\\&\le t\int_0^t\left(\partial C^n(s)-\kappa\beta_{2,1}s\right)^2\mathrm{d}s\end{aligned}$$

by Jensen's inequality. Hence,

$$\begin{aligned}\mathbb{E}\left[\sup_{t\le t_0}\left(n^{-1/3}C^n\left(n^{2/3}t\right)-\tfrac{1}{2}\kappa\beta_{2,1}t^2\right)^2\right]&\le t_0\int_0^{t_0}\mathbb{E}\left[\left(\partial C^n(s)-\kappa\beta_{2,1}s\right)^2\right]\mathrm{d}s\\&\le t_0^2\sup_{s\le t_0}\mathbb{E}\left[\left(\partial C^n(s)-\kappa\beta_{2,1}s\right)^2\right].\end{aligned}$$

Therefore, it suffices to prove that

$$\sup_{s\le t_0}\left|\mathbb{E}\left[\partial C^n(s)\right]-\kappa\beta_{2,1}s\right|\to 0 \quad \text{and} \quad \sup_{s\le t_0}\operatorname{Var}(\partial C^n(s))\to 0.$$

We have

$$\begin{aligned}\mathbb{E}[\partial C^n(s)]&=n^{1/3}\sum_{i=1}^{n}\frac{w_i^{n,-}w_i^{n,+}}{l_n}\left(1-e^{-n^{2/3}w_i^{n,-}s/l_n}\right)\\&=\frac{n}{l_n}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})}{l_n}s-n^{1/3}\sum_{i=1}^{n}\frac{w_i^{n,-}w_i^{n,+}}{l_n}R_1\left(n^{2/3}w_i^{n,-}s/l_n\right).\end{aligned}$$

Firstly, note that

$$\sup_{s\le t_0}\left|\frac{n}{l_n}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})}{l_n}s-\kappa\beta_{2,1}s\right|\le\left|\frac{n}{l_n}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})}{l_n}-\kappa\beta_{2,1}\right|t_0$$

which will converge to 0 by the assumptions in Eqs. (3.1a) and (3.1b). Secondly, using the bound in Eq. (3.4),

$$\sup_{s\le t_0}\left|n^{1/3}\sum_{i=1}^{n}\frac{w_i^{n,-}w_i^{n,+}}{l_n}R_1\left(n^{2/3}w_i^{n,-}s/l_n\right)\right|\le\frac{n^2}{l_n^2}\frac{w_{\max}^{n,-}}{n^{1/3}}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})}{l_n}t_0^2$$

which will converge to 0 by the assumptions in Eqs. (3.1a) to (3.1c). Hence,

$$\sup_{s\le t_0}\left|\mathbb{E}\left[\partial C^n(s)\right]-\kappa\beta_{2,1}s\right|\to 0$$

as $n\to\infty$. Finally,

$$\begin{aligned}\sup_{s\le t_0}\operatorname{Var}(\partial C^n(s))&=\sup_{s\le t_0}n^{2/3}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})^2}{l_n^2}\left(1-e^{-n^{2/3}w_i^{n,-}s/l_n}\right)e^{-n^{2/3}w_i^{n,-}s/l_n}\\&\le\sup_{s\le t_0}n^{4/3}\sum_{i=1}^{n}\frac{(w_i^{n,-})^3(w_i^{n,+})^2}{l_n^3}s\\&\le\frac{n^2}{l_n}\frac{(w_{\max}^{n,-})(w_{\max}^{n,+})}{n^{2/3}}\sum_{i=1}^{n}\frac{(w_i^{n,-})^2(w_i^{n,+})}{l_n}t_0\end{aligned}$$

which converges to 0 by the assumptions in Eqs. (3.1a) to (3.1c). □

Since the scaling limit for A^n has no jumps, we can immediately get joint convergence for

$$\left(n^{-1/3}X^n\left(n^{2/3}t\right),n^{-1/3}A^n\left(n^{2/3}t\right),n^{-1/3}Y^n\left(n^{2/3}t\right)\right)_{t\ge0}$$

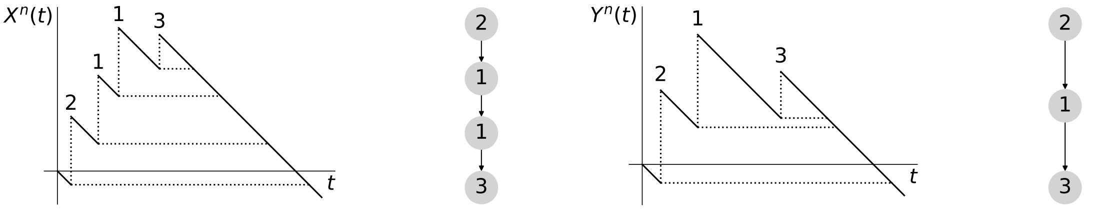
with respect to $\mathbb{D}([0,\infty),\mathbb{R}^3)$ using the relationship $Y^n = X^n - A^n$.

Corollary 3.7. *Under the assumptions in Eqs.* (3.1a) *to* (3.1d),

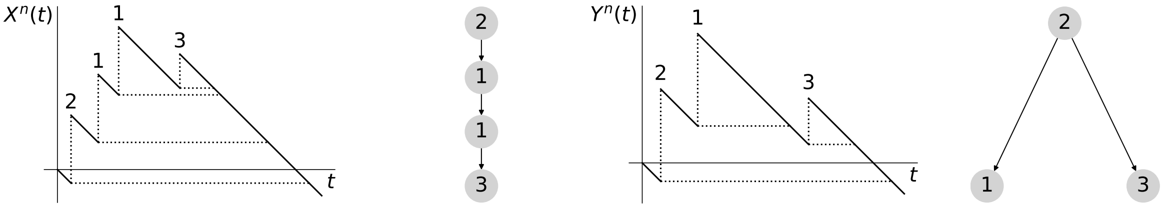
$$\left(n^{-1/3}X\left(n^{2/3}t\right), n^{-1/3}A\left(n^{2/3}t\right), n^{-1/3}Y\left(n^{2/3}t\right)\right)_{t\geq 0} \xrightarrow{(\mathrm{d})} (X(t), A(t), N(t))_{t\geq 0}$$

as $n \to \infty$ with respect to the Skorokhod topology on $\mathbb{D}([0,\infty),\mathbb{R}^3)$.

We would like to get a scaling limit for $n^{-1/3}H^n\left(n^{2/3}t\right)$. However, since Y is not a Lévy process, this does not immediately follow from the convergence of $n^{-1/3}Y^n\left(n^{2/3}t\right)$ to Y using [12, Proposition 2.2]. Instead, we will use the fact that the scaling limit of the height process corresponding to X is known. However, with this coupling there is no way to obtain the queueing forest for the queue with repeat customers from the queueing forest for the queue with duplicated customers removed. This is illustrated in Fig. 3.2, which shows two examples of X^n for queues with repeat customers. The queueing forest for X^n in both (a) and (b) are the same. In contrast, the queueing forest for Y^n in (a) is a line whereas the queueing forest for Y^n in (b) is a wedge. As the queueing forests are different, the corresponding height processes are also different. To fix this, as in Section 2.4.2, we borrow the idea of the red-blue coupling from [12].

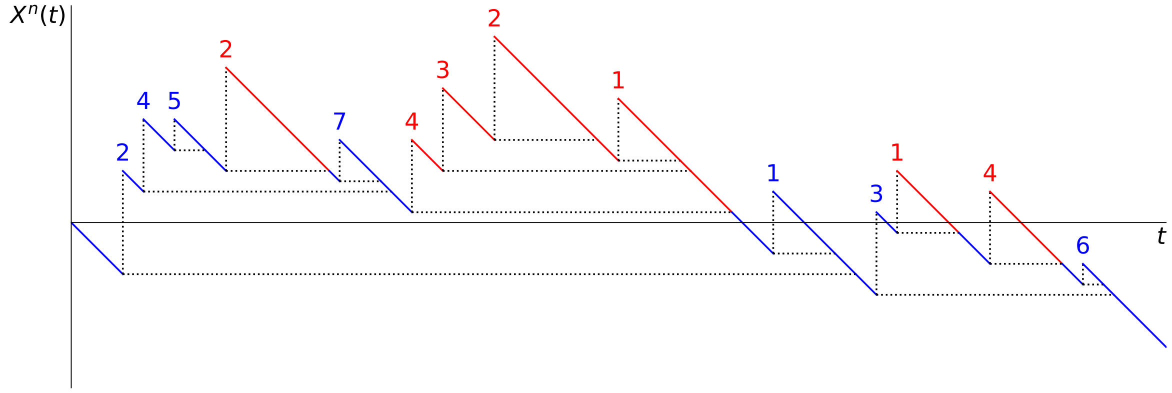


(a) The queueing forest for X^n is a line and the queuing forest for Y^n is a line.

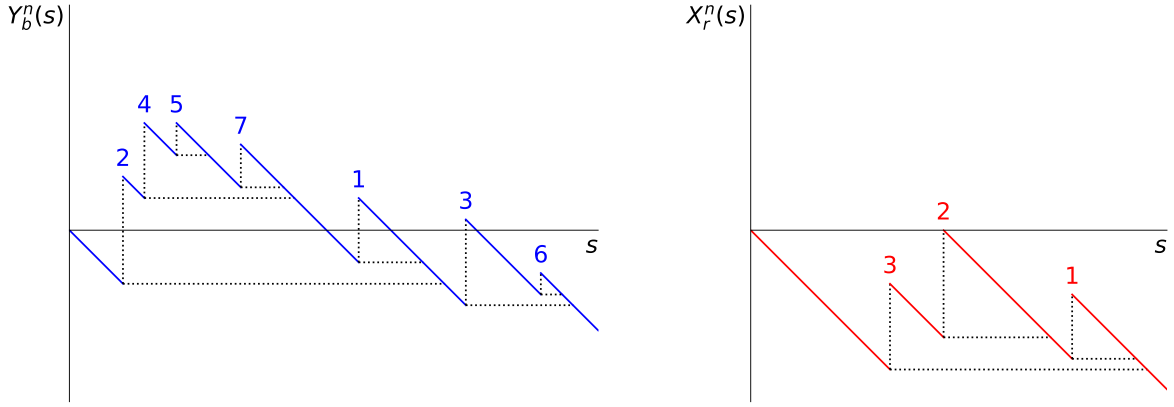


(b) The queueing forest for X^n is a line and the queueing forest for Y^n is a wedge.

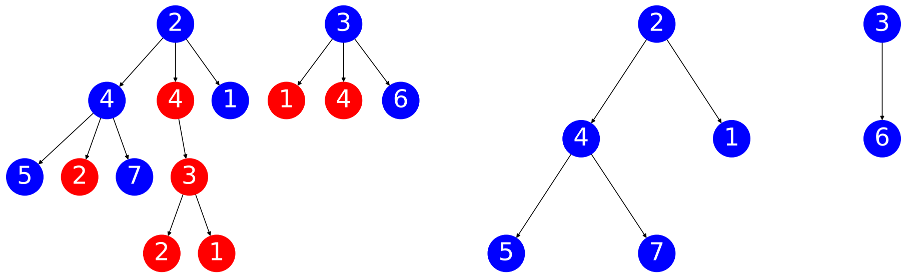
Fig. 3.2: Two examples of algebraic loads for which the queueing forests are the same, but the queueing forests after keeping only the first customer of each type are different.



(a) The combined red-blue algebraic load X^n.



(b) The separated blue algebraic load Y^n_b and the red algebraic load X^n_r. Note that jumps which occur at transitions from blue to red are not included in the red process.



(c) The red-blue queueing forest (d) The blue queueing forest

Fig. 3.3: An example of a red blue coloring of a forest associated with a queueing process, along with the algebraic load and the extracted red and blue processes.

Red-blue coupling

Consider the exploration tree. We colour the vertices of the graph in their order of arrival. A vertex of type i is coloured according to the following rules:

- If there is a blue vertex of type i which arrived earlier, colour the vertex red;
- If it is a child of a red vertex, colour the vertex red;
- Otherwise colour the vertex blue.

The blue vertices can be obtained following the vertices in order of exploration then discarding any vertex and its descendants when we see a repeated type. An example of this can be seen in Fig. 3.3. Then the forest of blue vertices and edges connecting them will have the same distribution as the forest for the finite queue.

In fact, this is not that different from the first arrival coupling. Instead of ignoring customers of repeated types, we just wait until they naturally leave the queue and skip the time it takes for them to do so. Because the queueing process labelled with arrival types is a Markov process and the arrival of a repeated type is a stopping time, the evolution of the queueing process after the repeated customer leaves is the same as if the customer never arrived in the first place. We will color the process X^n red at time t if a customer corresponding to a red vertex is being served at time t, and color X^n blue at time t otherwise (so when no customers are being served or a customer corresponding to a blue vertex is being served).

Consider splitting the queue into red and blue parts. First define

$$\Lambda_b^n(t) = \int_0^t \mathbb{1}\left\{X^n(t) \text{ is blue}\right\} \mathrm{d}t \quad \text{and} \quad \Lambda_r^n(t) = \int_0^t \mathbb{1}\left\{X^n(t) \text{ is red}\right\} \mathrm{d}t$$

which measures the amount of time elapsed for the red and blue processes respectively, such that $\Lambda_b^n(t) + \Lambda_r^n(t) = t$. We also define their respective right-continuous generalised inverses

$$\theta_b^n(s) = \inf\{s \geq 0 : \Lambda_b^n(s) \geq t\} \quad \text{and} \quad \theta_r^n(s) = \inf\{s \geq 0 : \Lambda_r^n(s) \geq t\}.$$

We can then extract the process

$$X_r^n(s) = \int_0^{\theta_r^n(s)} \mathbb{1}\left\{X^n(t) \text{ is red}\right\} \mathrm{d}X^n(t)$$

for $s \geq 0$, which will have the same law as the original process $(X^n(t))_{t\geq 0}$. The combined process X^n always transitions from blue to red at jumps. We do not include these jumps in the red process.

Further, we also extract the blue process

$$Y_b^n(s) = \int_0^{\theta_b^n(s)} \mathbb{1}\{X^n(t) \text{ is blue}\} \, \mathrm{d}X^n(t) = X^n(\theta_b^n(s)).$$

The nicer representation in the last equality is due to the fact X^n is always continuous at transitions from red to blue. These processes are also shown in Fig. 3.3.

The blue process $(Y_b^n(s))_{s\geq 0}$ can be obtained by carrying out the coupling in the previous section on an augmented process $(X_b^n(s))_{s\geq 0}$ which also has the same law as $(X^n(t))_{t\geq 0}$ and is independent of the red process $(X_r^n(s))_{s\geq 0}$.

It can be easier to view this in reverse: to first define the red and blue processes and then stitch them together to get a single process. We start the with the blue process

$$X_b^n(s) = -s + \sum_{i=1}^{n} w_i^{n,+} N_{b,i}^n(s)$$

where each $N_{b,i}^n$ is a independent Poisson process with rate $w_i^{n,-}/l_n$. Thus X_b^n and X^n have the same law. Further, define

$$A_b^n(s) = \sum_{i=1}^{n} w_i^{n,+}(N_{b,i}^n(s) - 1)_+.$$

Let the red process X_r^n be an independent copy of X_b^n. We can then define

$$\gamma_x^n = \inf\{s \geq 0 : X_r^n(s) \leq -x\}$$

which measures how long it takes for the queue corresponding to the red queue to be idle for time x.

To form X^n, we will run the blue process X_b^n until we see an excess jump. Then from that point on, we will run the red process X_r^n until X^n returns to its value before the excess jump. If time s has passed for the blue queue, then we need to run the red queue to compensate for all excess jumps in the blue queue by time s. The combined size of all such jumps is given by $A_b^n(s)$, and therefore we will have run the red queue for time $\gamma_{A_b^n(s)}^n$. Therefore, if $\theta_b^n(s)$ is the time the combined queue needs to run for the blue queue to have run for time s, then

$$\theta_b^n(s) = s + \gamma_{A_b^n(s)}^n.$$

Then we define $\Lambda_b^n(t)$ to be the right-continuous generalised inverse, i.e.

$$\Lambda_b^n(t) = \inf\left\{s \geq 0 : \theta_b^n(s) \geq t\right\}.$$

Finally, we can stitch together the red and blue process to get a single combined process

$$X^n(t) = X_b^n(\Lambda_b^n(t)) + X_r^n(t - \Lambda_b^n(t)).$$

Then (X^n, Y_b^n) is distributed according to the red-blue coupling described in this section. Further, (X_b^n, Y_b^n) is distributed according to the first arrival coupling. This is a similar idea to using filler vertices for the directed configuration model in Section 2.4.3, although the details of the coupling differ.

The convenience of this coupling is that if H_b^n and $\mathcal{H}^n$ are the height processes associated with the algebraic loads Y_b^n and X^n respectively, then we have the relation

$$H_b^n(s) = \mathcal{H}^n(\theta_b^n(s)).$$

We now set up the notation for their continuous analogues. Let (X_b, A_b, Y_b) and (X_r, A_r, Y_r) be independent copies of (X, A, Y), the limiting processes in Corollary 3.7. Let

$$\begin{aligned}
\gamma_x &= \inf\{s \geq 0 : X_r(s) \leq -x\}, \\
\theta_b(s) &= s + \gamma_{A_b(s)}, \\
\Lambda_b(t) &= \inf\{s \geq 0 : \theta_b(s) \geq t\}, \quad \text{and} \\
X(t) &= X_b(\Lambda_b(t)) + X_r(t - \Lambda_b(t)).
\end{aligned}$$

Since we have established the convergences in Corollary 3.7, we can exactly follow the proof of [12, Proposition 5.1] to establish the following lemma:

Lemma 3.8. *Under the assumptions in Eqs.* (3.1a) *to* (3.1d),

$$\left(n^{-1/3} Y_b^n\left(n^{2/3}t\right), n^{-1/3} H_b^n(t)\left(n^{2/3}t\right)\right) \xrightarrow{(\mathrm{d})} \left(X(\theta_b(t)), \mathcal{H}(\theta_b(t))\right)$$

The final lemma needed to prove Theorem 3.10 is the following:

Lemma 3.9. *We have almost surely that*

$$(\mathcal{H}(\theta_b(t)))_{t\geq 0} = \left(\frac{2}{\beta_{1,2}}(X(\theta_b(t)) - \inf_{s\in[0,t]} X(\theta_b(s)))\right)_{t\geq 0}$$

This is analogous to Lemma 2.47 for the directed configuration model. The proof is almost identical, just that we have a Brownian motion with drift instead of a Brownian motion, so we will not repeat it. We are finally ready to prove Theorem 3.10, which we restate below:

Theorem 3.10. *Then under the assumptions in Eqs.* (3.1a) *to* (3.1d),

$$\left(n^{-1/3}H^n\left(n^{2/3}t\right)\right)_{t\geq 0} \xrightarrow{\text{(d)}} H$$

where B is a standard Brownian motion,

$$Y(t) = -\alpha t - \frac{1}{2}\kappa\beta_{2,1}t^2 + \sqrt{\beta_{1,2}}B_t \quad \textit{and} \quad H(t) = \frac{2}{\beta_{1,2}}(Y(t) - \inf_{s\in[0,t]} Y(s)).$$

Proof. Since we know that $n^{-1/3}Y_b^n\left(n^{2/3}t\right)$ converges in law to Y_b, we know that $(X(\theta_b(t)))_{t\geq 0}$ and Y_b must have the same law. Thus, Lemma 3.8 and Lemma 3.9 combined prove Theorem 3.10. ☐

Appendix A

Ancestral back edges in the IRDG model

An *ancestral back edge* is a back edge which goes from a vertex i to an ancestor j or i. As for the directed configuration model, knowing how they arise is important to understand the structure of the SCCs since they immediately create a cycle.

Let $\mathcal{H}^{n,-}(t)$ be the sum of the $w_i^{n,-}$ for all the customers in the infinite queue at time t. Another way to think about this is to consider adding edge lengths to the forest associated with the infinite queue as follows: for each vertex representing customer of type i, assign length $w_i^{n,-}$ to all edges from the vertex to its children. Then $\mathcal{H}^{n,-}(t)$ is the distance from the root to the vertex corresponding to the customer being served at time t with these new edge lengths.

Note that the resulting forest is a BGW forest in which N^n the number of offspring of a vertex and the L^n length of the edge joining a vertex to its offspring are distributed in an i.i.d. fashion according to the law

$$(N^n, L^n) \sim \sum_{i=1}^{n} \frac{w_i^{n,-}}{l_n^-} \mathrm{Po}\left(w_i^{n,+} \frac{l_n^-}{l_n}\right) \otimes \delta_{w_i^{n,-}}$$

where $\mathrm{Po}(\lambda)$ is the Poisson distribution with mean λ. Let $(\hat{N}^n, \hat{L}^n)$ be (N^n, L^n) size-biased by N^n. Then note that

$$\mathbb{E}[N^n] = \sum_{i=1}^{n} \frac{w_i^{n,-} w_i^{n,+}}{l_n} = 1 + \Theta(n^{-1/3})$$

and that

$$\mathbb{E}[\hat{L}^n] = \frac{\mathbb{E}[N^n L^n]}{\mathbb{E}[N^n]} = \frac{1}{\mathbb{E}[N^n]} \sum_{i=1}^{n} \frac{(w_i^{n,-})^2 (w_i^{n,+})}{l_n} \to \beta_{2,1}$$

as $n \to \infty$. Then by applying [22, Proposition 4] the following holds:

Proposition A.1. *Suppose the assumptions in Eqs. (3.1a) to (3.1d) and Eq. (3.2) hold. Then, for all* $t_0 > 0$,

$$\sup_{t \in [0,t_0]} \left| n^{-1/3} \mathcal{H}^{n,-} \left(n^{2/3} t \right) - \beta_{2,1} n^{-1/3} \mathcal{H}^{n} \left(n^{2/3} t \right) \right| \xrightarrow{(p)} 0$$

as $n \to \infty$.

Proof. This is proven in the same manner as Proposition 2.51 using [22, Proposition 4]. The conditions of the theorem, translated into our setting, are the following:

1. $\mathbb{E}[N^n] = 1$.
2. $\sup_n \mathbb{E}[(N^n)^2] < \infty$.
3. $\sup_n \mathbb{P}\left(\hat{L}^n > x \right) = o(x^{-2})$ as $x \to \infty$.

The first condition is not satisfied here exactly, which introduces a factor of $\mathbb{E}[N^n]^{-\lfloor n^{2/3} T \rfloor}$ to normalise a measure change that is shown in [22, Proposition 3]. However since

$$\mathbb{E}[N^n]^{-\lfloor n^{2/3} T \rfloor} = \left(1 + o(n^{-2/3}) \right)^{-\lfloor n^{-2/3} T \rfloor} \to 1,$$

as $n \to \infty$, it does not affect the proof.

For the second condition, we can calculate that

$$\mathbb{E}[(N^n)^2] = \sum_{i=1}^{n} \frac{w_i^{n,-} w_i^{n,+}}{l_n} + \frac{l_n^-}{l_n} \sum_{i=1}^{n} \frac{(w_i^{n,-})(w_i^{n,+})^2}{l_n}.$$

This is uniformly bounded in n by the assumption in Eq. (3.1b).

Finally, the last condition is satisfied if $\sup_n \mathbb{E}[\hat{L}^n] < \infty$ by the Markov inequality. We have

$$\mathbb{E}[\hat{L}^n] = \frac{1}{\mathbb{E}[N^n]} \sum_{i=1}^{n} \frac{(w_i^{n,-})^3 w_i^{n,+}}{l_n}$$

which we have assumed to be uniformly bounded in n by Eq. (3.2). □

Let $H^{n,-}$ be the sum of the $w_i^{n,-}$ for the customers in the finite queue at time t. The next corollary then immediately follows from Proposition A.1 by using the red-blue coupling.

Corollary A.2. *Suppose the assumptions in Eqs.* (3.1a) *to* (3.1d) *and Eq.* (3.2) *hold. Then, as* $n \to \infty$,

$$\left(n^{-1/3} H^{n,-}\left(n^{2/3} t\right)\right)_{t \geq 0} \xrightarrow{\text{(d)}} H^{-}$$

with respect to the Skorokhod topology on $\mathbb{D}([0,\infty),\mathbb{R})$ *where* $H^{-} = \beta_{2,1} H$.

Appendix B

Multivariate triangular local limit theorem

The goal of this section is to prove Theorem 2.31. This can be deduced from Mukhin [48, Corollary 1]. However, Mukhin's result is more general than is needed to prove Theorem 2.31. 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

$$\Lambda = \bigcup_{m=1}^{\infty} \left\{ \textstyle\sum_{i=1}^{m} n_i \mathbf{x}_i^* : n_i \in \mathbb{Z} \text{ and } \mathbf{x}_i^* \in \operatorname{supp}(\mathbf{X}^*) \text{ for all } i = 1, \dots, m \right\}.$$

Now we restate Theorem 2.31.

Theorem 2.31. *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* $\mathbf{X}_n \xrightarrow{\text{(d)}} \mathbf{X}$ *as* $n \to \infty$*.*

2. *$(\|\mathbf{X}_n\|^2)_{n\geq 1}$ is a uniformly integrable sequence of random variables. Explicitly*

$$\lim_{L\to\infty}\sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\left\{\|\mathbf{X}_n\|^2 > L\right\}\right] = 0. \tag{2.14}$$

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 $N(0, \mathrm{Cov}(\mathbf{X}))$ distribution. This means that

$$\lim_{n\to\infty}\sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda}\left|n^{d/2}\mathbb{P}\Big(\textstyle\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\Big) - \det(\Lambda) f(\mathbf{x}_n(\mathbf{y}))\right| = 0.$$

Before we prove Theorem 2.31, 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 [54, P.67, T1].

Lemma B.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 B.2. *Suppose conditions (1) and (2) of Theorem 2.31 hold. Then, as $n \to \infty$,*

$$\mathbb{E}[\mathbf{X}_n] \to \mathbb{E}[\mathbf{X}] \quad \text{and} \quad \mathrm{Cov}(\mathbf{X}_n) \to \mathrm{Cov}(\mathbf{X}).$$

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. (2.14) 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. (2.14) 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 B.3. *Suppose we are in the setting of Theorem 2.31. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}_{n,i} - \mathbb{E}[\mathbf{X}_n]) \xrightarrow{\text{(d)}} N(0, \Sigma)$$

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{u} \cdot \hat{\mathbf{X}}_{n,i} \xrightarrow{\text{(d)}} N(0, \mathbf{u} \cdot \Sigma \mathbf{u})$$

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 [28, 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 $\epsilon > 0$, $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\Big[A_{n,i}^2 \mathbb{1}\left\{|A_{n,i}| > \epsilon\right\}\Big] = 0$.

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 B.2. To check condition (2), for all $\epsilon > 0$

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\Big[A_{n,i}^2 \mathbb{1}\left\{|A_{n,i}| > \epsilon\right\}\Big] &= \lim_{n\to\infty} \mathbb{E}\Big[(\mathbf{u} \cdot \hat{\mathbf{X}}_n)^2 \mathbb{1}\Big\{(\mathbf{u} \cdot \hat{\mathbf{X}}_n)^2 > \epsilon^2 n\Big\}\Big] \\
&\leq \|\mathbf{u}\|^2 \lim_{n\to\infty} \mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2 \mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2 > \frac{\epsilon^2}{\|\mathbf{u}\|^2} n\right\}\right] \\
&\leq \|\mathbf{u}\|^2 \lim_{n\to\infty} \sup_k \mathbb{E}\left[\|\hat{\mathbf{X}}_k\|^2 \mathbb{1}\left\{\|\hat{\mathbf{X}}_k\|^2 > \frac{\epsilon^2}{\|\mathbf{u}\|^2} n\right\}\right] \\
&= 0
\end{aligned}$$

by Lemma B.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 B.4. *Suppose we are in the setting of Theorem 2.31. 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* $n \geq N$

1. $|\phi_n(\mathbf{u})| \leq 1 - c\|\mathbf{u}\|^2$ *for all* $\mathbf{u} \in S(\delta)$, *and*
2. $|\phi_n(\mathbf{u})| \leq \rho$ *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 [28, P.116, Lemma 3.3.19]. By that lemma, there exists a constant $A > 0$ such that

$$\left|e^{ix} - \left(1 + ix - \tfrac{1}{2}x^2\right)\right| \leq A \min\{|x|, 1\} x^2$$

for all $x \in \mathbb{R}$. Then applying this with $x = \mathbf{u} \cdot (\mathbf{X}_n - \mathbb{E}[\mathbf{X}_n])$

$$|\phi_n(\mathbf{u})| \leq \left|1 - \tfrac{1}{2}\mathbf{u} \cdot \mathrm{Cov}(\mathbf{X}_n)\mathbf{u}\right| + R_n(\mathbf{u})$$

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 $\lambda_{\min n}$ and $\lambda_{\max n}$ be the minimum and maximum eigenvalues of $\mathrm{Cov}(\mathbf{X}_n)$ respectively. Then, by standard theory for quadratic forms,

$$\lambda_{\min n}\|\mathbf{u}\|^2 \leq \mathbf{u} \cdot \mathrm{Cov}(\mathbf{X}_n)\mathbf{u} \leq \lambda_{\max n}\|\mathbf{u}\|^2.$$

Moreover, let $\lambda_{\min}$ and $\lambda_{\max}$ 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 B.2. Therefore $\lambda_{\min n} \to \lambda_{\min}$ and $\lambda_{\max n} \to \lambda_{\max}$ as $n \to \infty$.

We have assumed that $\mathrm{Cov}(\mathbf{X})$ is non-degenerate thus $\lambda_{\min} > 0$. Hence, there exists N such that for all $n \geq N$,

$$\frac{1}{2}\lambda_{\min} \leq \lambda_{\min n} \leq \lambda_{\max n} \leq 2\lambda_{\max}.$$

There also exists $\delta_1 > 0$ sufficiently small that $\lambda_{\max}\|\mathbf{u}\|^2 < 1$ for all $\mathbf{u} \in S(\delta_1)$. Then for all $n \geq N$

and $\mathbf{u} \in S(\delta_1)$,

$$\left|1 - \tfrac{1}{2}\mathbf{u} \cdot \mathrm{Cov}(\mathbf{X}_n)\mathbf{u}\right| = 1 - \tfrac{1}{2}\mathbf{u} \cdot \mathrm{Cov}(\mathbf{X}_n)\mathbf{u} \leq 1 - \tfrac{1}{4}\lambda_{\min}\|\mathbf{u}\|^2. \tag{B.1}$$

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 B.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 $\mathbf{u} \in S(\delta_2)$

$$R_n(\mathbf{u}) \leq \frac{1}{8}\lambda_{\min}\|\mathbf{u}\|^2. \tag{B.2}$$

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 B.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 $\mathbf{X}_n \xrightarrow{\text{(d)}} \mathbf{X}$ 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 2.31

Proof of Theorem 2.31. 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 B.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 $\|\mathbf{s}\|^2 \leq \pi^2 dn$. Thus for all $n \geq N$ and $\mathbf{s} \in S(\pi\sqrt{n}) \setminus S(\delta\sqrt{n})$

$$|\phi_n(\mathbf{s}/\sqrt{n})|^n \leq e^{-Cn} \leq e^{-\frac{C}{\pi^2 d}\|\mathbf{s}\|^2}.$$

Hence for all $n \geq N$,

$$\left|\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi_n(\mathbf{s}/\sqrt{n})^n - \psi(\mathbf{s})\right| \leq e^{-c\|\mathbf{s}\|^2} + e^{-\frac{C}{\pi^2 d}\|\mathbf{s}\|^2} + |\psi(\mathbf{s})|$$

where, in particular, the right hand side is integrable. By Lemma B.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

$$\lim_{n\to\infty} \sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda} \left| n^{d/2}\mathbb{P}\left(\textstyle\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\right) - f(\mathbf{x}_n)\right| = 0,$$

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 2.31 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

$$\tilde{\Sigma} = \operatorname{Cov}(A^{-1}\mathbf{X}) = A^{-1}\operatorname{Cov}(\mathbf{X})(A^{-1})^T.$$

Therefore

$$\det(\tilde{\Sigma}) = \det(A)^{-2}\det(\operatorname{Cov}(\mathbf{X})) = \det(\Lambda)^{-2}\det(\operatorname{Cov}(\mathbf{X}))$$

and so

$$\mathbb{P}\left(\sum_{i=1}^{n}\mathbf{X}_{n,i} = \mathbf{y}\right) = \frac{\det(\Lambda)}{\sqrt{(2\pi n)^d \det(\operatorname{Cov}\mathbf{X})}}\exp\left(-\frac{1}{2}\mathbf{x}_n^T\operatorname{Cov}(\mathbf{X})^{-1}\mathbf{x}_n\right) + o(n^{-d/2}),$$

as required. □

Appendix C

Proof of technical lemmas

Proof of Lemma 2.45. 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 [31], 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 2.53. 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

$$\delta_1 = \inf\{f(s) : s \in [g_i + \epsilon, g_i + \sigma_i - \epsilon]\} - \inf\{f(s) : s \in (g_i - \epsilon, g_i + \epsilon)\}$$
$$\delta_2 = \inf\{f(s) : s \in [0, g_i - \epsilon]\} - \inf\{f(s) : s \in (g_i - \epsilon, g_i + \epsilon)\}$$
$$\delta_3 = \inf\{f(s) : s \in [0, g_i + \sigma_i - \epsilon]\} - \inf\{f(s) : s \in (g_i + \sigma_i - \epsilon, g_i + \sigma_i + \epsilon)\}.$$

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, $\sup\{|f_n(s) - f(s)| : s \leq g_i + \epsilon\} < \delta$, which implies conditions 2, 3, and 4 for such n.

Since i was arbitrary, and m is finite, we find that

$$(g_i^n, \sigma_i^n)_{1 \leq i \leq m} \to (g_i, \sigma_i)_{1 \leq i \leq m}$$

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 $g_j^n - g_i^n \geq \sigma_i^n$, 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

$$\#\left\{(g_i^n, \sigma_i^n) : 1 \leq i \leq m\right\} \to \#\left\{(g_i, \sigma_i) : 1 \leq i \leq m\right\}.$$

Then, the result follows. $\square$

Bibliography

[1] L. Addario-Berry, N. Broutin, and C. Goldschmidt. The continuum limit of critical random graphs. *Probability Theory and Related Fields*, 152(3):367–406, April 2012. ISSN 1432-2064. doi: 10.1007/s00440-010-0325-4.

[2] Louigi Addario-Berry, Anna Brandenberger, Jad Hamdan, and Céline Kerriou. Universal height and width bounds for random trees. *arXiv:2105.03195 [math]*, May 2021.

[3] D. Aldous. The Continuum random tree II: An overview. In M. T. Barlow and N. H.Editors 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.

[4] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Annals of Probability*, 25(2):812–854, April 1997. ISSN 0091-1798, 2168-894X. doi: 10.1214/aop/1024404421.

[5] Shankar Bhamidi and Sanchayan Sen. Geometry of the vacant set left by random walk on random graphs, Wright's constants, and critical random graphs with prescribed degrees. *Random Structures & Algorithms*, 56(3):676–721, 2020. ISSN 1098-2418. doi: 10.1002/rsa.20880.

[6] Shankar Bhamidi, Sanchayan Sen, and Xuan Wang. Continuum limit of critical inhomogeneous random graphs. *Probability Theory and Related Fields*, 169(1):565–641, October 2017. ISSN 1432-2064. doi: 10.1007/s00440-016-0737-x.

[7] Shankar Bhamidi, Souvik Dhara, Remco van der Hofstad, and Sanchayan Sen. Universality for critical heavy-tailed network models: Metric structure of maximal components. *Electronic Journal of Probability*, 25:1–57, January 2020. ISSN 1083-6489, 1083-6489. doi: 10.1214/19-EJP408.

[8] Shankar Bhamidi, Souvik Dhara, Remco van der Hofstad, and Sanchayan Sen. Global lower mass-bound for critical configuration models in the heavy-tailed regime. *Electronic Journal of Probability*, 27(none):1–29, January 2022. ISSN 1083-6489, 1083-6489. doi: 10.1214/22-EJP821.

[9] Mindaugas Bloznelis, Friedrich Götze, and Jerzy Jaworski. Birth of a strongly connected giant in an inhomogeneous random digraph. *Journal of Applied Probability*, 49(3):601–611, September 2012. ISSN 0021-9002, 1475-6072. doi: 10.1239/jap/1346955320.

[10] Béla Bollobás. A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. *European Journal of Combinatorics*, 1(4):311–316, December 1980. ISSN 0195-6698. doi: 10.1016/S0195-6698(80)80030-8.

[11] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007. ISSN 1098-2418. doi: 10.1002/rsa.20168.

[12] Nicolas Broutin, Thomas Duquesne, and Minmin Wang. Limits of multiplicative inhomogeneous random graphs and Lévy trees: Limit theorems. *Probability Theory and Related Fields*, July 2021. ISSN 1432-2064. doi: 10.1007/s00440-021-01075-z.

[13] Nicolas Broutin, Thomas Duquesne, and Minmin Wang. Limits of multiplicative inhomogeneous random graphs and Lévy trees: The continuum graphs. *The Annals of Applied Probability*, 32(4): 2448–2503, August 2022. ISSN 1050-5164, 2168-8737. doi: 10.1214/21-AAP1737.

[14] Xing Shi Cai and Guillem Perarnau. The giant component of the directed configuration model revisited. *Latin American Journal of Probability and Mathematical Statistics*, 18:1517, January 2021. doi: 10.30757/ALEA.v18-55.

[15] Xing Shi Cai and Guillem Perarnau. Minimum stationary values of sparse random directed graphs. *arXiv:2010.07246 [cs, math]*, February 2021.

[16] Xing Shi Cai and Guillem Perarnau. The diameter of the directed configuration model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 59(1):244–270, February 2023. ISSN 0246-0203. doi: 10.1214/22-AIHP1252.

[17] Junyu Cao and Mariana Olvera-Cravioto. Connectivity of a general class of inhomogeneous random digraphs. *Random Structures & Algorithms*, 56(3):722–774, October 2019. ISSN 1042-9832. doi: 10.1002/rsa.20892.

[18] Ningyuan Chen and Mariana Olvera-Cravioto. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013. ISSN 1946-5238. doi: 10.1214/12-SSY076.

[19] 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. *The Annals of Probability*, 51(1):1–69, January 2023. ISSN 0091-1798, 2168-894X. doi: 10.1214/22-AOP1587.

[20] Colin Cooper and Alan Frieze. The Size of the Largest Strongly Connected Component of a Random Digraph with a Given Degree Sequence. *Combinatorics, Probability and Computing*, 13 (3):319–337, May 2004. ISSN 1469-2163, 0963-5483. doi: 10.1017/S096354830400611X.

[21] 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. Springer-Verlag, New York, 2 edition, 2008. ISBN 978-0-387-21337-8. doi: 10.1007/978-0-387-49835-5.

[22] Loïc de Raphélis. Scaling limit of multitype Galton–Watson trees with infinitely many types. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):200–225, February 2017. ISSN 0246-0203. doi: 10.1214/15-AIHP713.

[23] Souvik Dhara, Remco van der Hofstad, Johan S. H. van Leeuwaarden, and Sanchayan Sen. Critical window for the configuration model: Finite third moment degrees. *Electronic Journal of Probability*, 22(none):1–33, January 2017. ISSN 1083-6489, 1083-6489. doi: 10.1214/17-EJP29.

[24] Souvik Dhara, Remco van der Hofstad, Johan S. H. van Leeuwaarden, and Sanchayan Sen. Heavy-tailed configuration models at criticality. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):1515–1558, August 2020. ISSN 0246-0203. doi: 10.1214/19-AIHP980.

[25] Serte Donderwinkel. Convergence of the height process of supercritical Galton-Watson forests with an application to the configuration model in the critical window. (arXiv:2105.12109), November 2021. doi: 10.48550/arXiv.2105.12109.

[26] Serte Donderwinkel and Zheneng Xie. Universality for the directed configuration model: Metric space convergence of the strongly connected components at criticality. *Electronic Journal of Probability*, 29:1 – 85, 2024. doi: 10.1214/24-EJP1131. URL https://doi.org/10.1214/24-EJP1131.

[27] Thomas Duquesne and Jean-François Le Gall. *Random Trees, Lévy Processes and Spatial Branching Processes*. Number 281 in Astérisque. Société mathématique de France, 2002.

[28] Rick Durrett. *Probability: Theory and Examples.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 5th edition, 2019. ISBN 978-1-108-47368-2. doi: 10.1017/9781108591034.

[29] P. Erdős and A. Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

[30] P. Erdős and A. Rényi. On the Evolution of Random Graphs. In *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, pages 17–61, 1960.

[31] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley, April 1986. ISBN 978-0-471-08186-9.

[32] E. N. Gilbert. Random Graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, December 1959. ISSN 0003-4851, 2168-8990. doi: 10.1214/aoms/1177706098.

[33] B. V. Gnedenko. On a local limit theorem of the theory of probability. *Uspekhi Mat. Nauk*, 3 (3(25)):187–194, 1948.

[34] Christina Goldschmidt and Robin Stephenson. The scaling limit of a critical random directed graph. *The Annals of Applied Probability*, 33(3):2024–2065, June 2023. ISSN 1050-5164, 2168-8737. doi: 10.1214/22-AAP1859.

[35] A. Graf. *On the Strongly Connected Components of Random Directed Graphs with Given Degree Sequences.* PhD thesis, University of Waterloo, 2016.

[36] Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes.* Grundlehren der mathematischen Wissenschaften №288. Springer, 2nd ed edition, 1987. ISBN 978-3-662-02516-1.

[37] Svante Janson and Malwina J. Luczak. A new approach to the giant component problem. *Random Structures & Algorithms*, 34(2):197–216, 2009. ISSN 1098-2418. doi: 10.1002/rsa.20231.

[38] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random Graphs.* John Wiley & Sons, September 2011. ISBN 978-1-118-03096-7.

[39] Adrien Joseph. The component sizes of a critical random graph with given degree sequence. *Annals of Applied Probability*, 24(6):2560–2594, December 2014. ISSN 1050-5164, 2168-8737. doi: 10.1214/13-AAP985.

[40] Richard M. Karp. The transitive closure of a random digraph. *Random Structures & Algorithms*, 1(1):73–93, 1990. ISSN 1098-2418. doi: 10.1002/rsa.3240010106.

[41] Ivan Kryven. Emergence of the giant weak component in directed random graphs with arbitrary degree distributions. *Physical Review E*, 94(1):012315, July 2016. doi: 10.1103/PhysRevE.94.012315.

[42] Ivan Kryven. Finite connected components in infinite directed and multiplex networks with arbitrary degree distributions. *Physical Review E*, 96(5):052304, November 2017. doi: 10.1103/PhysRevE.96.052304.

[43] Jean-François Le Gall. Random trees and applications. *Probability Surveys [electronic only]*, 2: 245–311, 2005. ISSN 1549-5787.

[44] Jean-François Le Gall. Random real trees. *Annales de la Faculté des sciences de Toulouse : Mathématiques*, 15(1):35–62, 2006. ISSN 2258-7519. doi: 10.5802/afst.1112.

[45] Tomasz Łuczak. The phase transition in the evolution of random digraphs. *Journal of Graph Theory*, 14(2):217–223, 1990. ISSN 1097-0118. doi: 10.1002/jgt.3190140210.

[46] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995. ISSN 1098-2418. doi: 10.1002/rsa.3240060204.

[47] Michael Molloy and Bruce Reed. The Size of the Giant Component of a Random Graph with a Given Degree Sequence. *Combinatorics, Probability and Computing*, 7(3):295–305, September 1998. ISSN 1469-2163, 0963-5483. doi: 10.1017/S0963548398003526.

[48] A. B. Mukhin. Local Limit Theorems for Lattice Random Variables. *Theory of Probability & Its Applications*, 36(4):698–713, January 1992. ISSN 0040-585X. doi: 10.1137/1136086.

[49] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997. ISBN 978-0-521-63396-3. doi: 10.1017/CBO9780511810633.

[50] Oliver Riordan. The Phase Transition in the Configuration Model. *Combinatorics, Probability and Computing*, 21(1-2):265–299, March 2012. ISSN 1469-2163, 0963-5483. doi: 10.1017/S0963548311000666.

[51] 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. *Journal of Applied Probability*, 53(1):146–161, March 2016. ISSN 0021-9002, 1475-6072.

[52] Alexander Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, June 1998. ISBN 978-0-471-98232-6.

[53] Bo Söderberg. General formalism for inhomogeneous random graphs. *Phys. Rev. E*, 66:066121, Dec 2002. doi: 10.1103/PhysRevE.66.066121. URL `https://link.aps.org/doi/10.1103/PhysRevE.66.066121`.

[54] Frank Spitzer. *Principles of Random Walk*. Graduate Texts in Mathematics. Springer-Verlag, New York, 2 edition, 1964. ISBN 978-1-4757-4229-9. doi: 10.1007/978-1-4757-4229-9.

[55] Remco van der Hofstad. *Random Graphs and Complex Networks*, volume 1. Cambridge University Press, 2017. ISBN 978-1-107-17287-6.