Metric space convergence under scaling of the strongly connected components of an uniform directed graph with an i.i.d. degree sequence

Serte Donderwinkel and Zheneng Xie

Version: May 19, 2021

Contents

1 Introduction

[Mention somewhere: Note that the SCC in the critical directed Erdős-Renyi model as studied in [?] has the same limit under rescaling as the model studied in this work with D^+ and D^- independent Poisson(1) random variables.]

1.1 The model

We study a uniform directed graph (or digraph) with a random degree sequence. We consider n vertices, to each of which we assign an in-degree and an out-degree. The degree tuples are independent and identically distributed. Let $\mathbf{D} = (D^-, D^+)$ be a random variable in $\mathbb{N} \times \mathbb{N}$ with this distribution, and for each $i \in [n]$, let $\mathbf{D}_i = (D_i^-, D_i^+)$ be the in- and out-degree of vertex i. In order for a graph with this degree sequence to exist, we require that $\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$, so we will condition on this event. Denote the uniform digraph with this degree sequence by $\vec{G}(n, (D^-, D^+))$. We are interested in the limit under scaling of the strongly connected components of $\vec{G}(n, (D^-, D^+))$ as $n \to \infty$.

We require the degree distribution to satisfy the following properties.

1. $\mathbb{E}[D^+] = \mathbb{E}[D^-] = \mathbb{E}[D^- D^+]$
2. $\mathbb{E}\left[(D^-)^3\right] < \infty$
3. $\mathbb{E}\left[(D^+)^2 D^-\right] < \infty$
4. $\mathbb{E}\left[D^+ (D^-)^3\right] < \infty$
5. $\mathbb{E}\left[(D^+)^3 D^-\right] < \infty$

We define the following parameters, that will determine the behaviour of the strongly connected components in the limit.

1. $\mu := \mathbb{E}[D^-] = \mathbb{E}[D^+] = \mathbb{E}[D^- D^+]$
2. $\nu_- := \frac{\mathbb{E}[(D^-)^2] - \mu}{\mu}$
3. $\sigma_- := \left(\frac{\mu \mathbb{E}[(D^-)^3] - \mathbb{E}[(D^-)^2]^2}{\mu^2} \right)^{1/2}$

4. $\sigma_+ := \left(\frac{\mathbb{E}[D^-(D^+)^2]-\mu}{\mu}\right)^{1/2}$

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

1.2 Strongly connected components

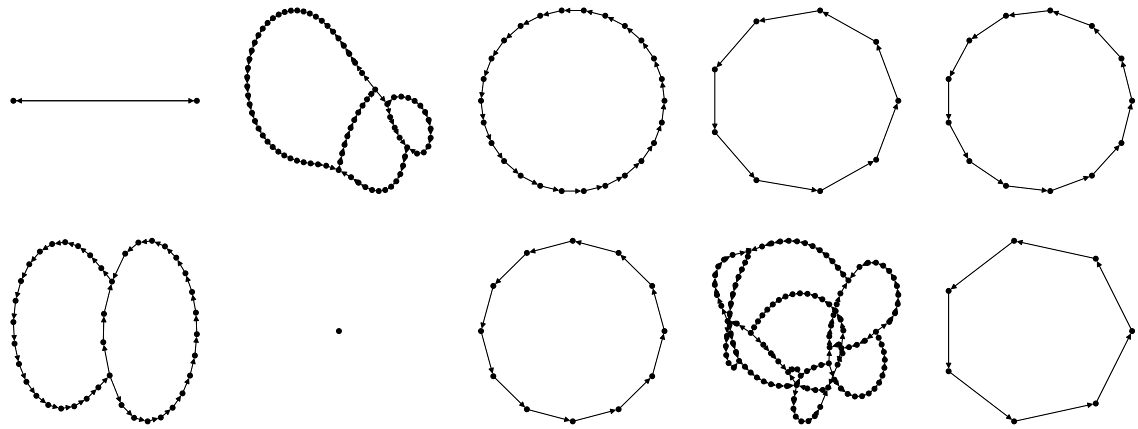


Figure 1: The largest SCC from samples of a directed configuration model with independent Poisson(1) in- and out-degrees

In fig. 1 [actually generate sccs from a poisson model] we see the largest SCC from samples of a directed configuration model. As can be seen, while the lengths of paths in the SCC are long, the actual structure of the SCC is often quite simple. In some cases the largest SCC is just a single cycle. This was confirmed in previous work by [insert reference for DER]. There it was shown that while the lengths of paths in the SCC will scale like $n^{1/3}$, the actual structure of the SCC will remain finite.

To formalise this idea we will introduce metric directed multigraphs (MDMs). These are simply weighted directed multigraphs, but in our context it is more appropriate to think of the weights as lengths hence the change in naming. Formally a directed multigraph is a tuple $M = (V, E, r)$ where

Our main result is as follows. Let $C_i(n)$ for $i \geq 1$ be the strongly connected components of $\vec{G}(n, (D^-, D^+))$, listed in decreasing order of size, breaking ties arbitrarily. We view these strongly connected components as MDMs, by assigning to each edge a length of 1, and then removing all vertices of degree 2 by merging neighbouring edges and adding up their lengths. Trivial strongly connected components, which consist of an isolated vertex, are considered loops of length 0. Complete the list with an infinite repeat of $\mathfrak{L}$, the loop of length 0. Then, the main theorem is as follows.

Theorem 1.1. *There exists a sequence $\mathcal{C} = (\mathcal{C}_i, i \in \mathbb{N})$ of random strongly connected MDMs such that, for each $i \geq 1$, $\mathcal{C}_i$ is either 3-regular or a loop, and such that*

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

as $n \to \infty$, with respect to the product $d_{\vec{\mathcal{G}}}$-topology.

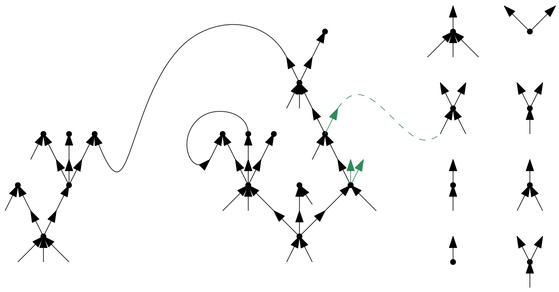


Figure 2: The green arrows represent unpaired out-half-edges of vertices that have been visited. One by one, in depth first order, these are paired to a uniform unpaired in-half-edge.

1.3 Previous work

1.4 Proof outline

The techniques we will use to investigate the graph model are a combination of the techniques introduced by Conchon-Kerjan and Goldschmidt in [**?**] and the strategy of Goldschmidt and Stephenson in [**?**]. The former work discusses the scaling limit of an undirected uniform graph with i.i.d. degrees at criticality, and the latter discusses the scaling limit of the strongly connected components of a directed Erdős-Renyi graph at criticality.
To investigate the structure of the strongly connected components of a uniform graph with degree sequence $(\mathbf{D}_1, \ldots, \mathbf{D}_n)$, conditioned on $\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$, we will use a version of the configuration model for digraphs that was introduced in [**?**]. The output of the configuration model, conditioned on being a simple graph, is a uniform digraph with the given degree sequence.
[Description of exploration written by Zheneng] This is illustrated in Figure 2.
The exploration algorithm naturally gives rise to a forest that we will refer to as the *out-forest*, which will play a key role in studying the limit under rescaling of the strongly connected components. An important motivation for studying the out-forest is the fact that the vertex set of any strongly connected component is contained in one of the components of the out-forest. We define the out-forest in such a way that every time step in the exploration corresponds to one vertex in the out-forest. At every time step in the exploration at which we find an unseen vertex, say with out-degree d^+, we add a vertex with d^+ children to the out-forest. At every time step at which we do not find a new vertex, but instead connect to a previously found vertex, we add a purple leaf to the out-forest. This is illustrated in Figure 3. We refer to the out-forest corresponding to the exploration up to time k as $\hat{F}_n(k)$.
A key fact is that the out-forest can be sampled without knowing what the heads of the surplus edges are, because this information does not affect the law of the out-forest. This allows us to build up the randomness of the exploration in the following layers.

1. We sample the out-forest $(\hat{F}_n(k), k \geq 1)$.

2. We visit the purple vertices in $(\hat{F}_n(k), k \geq 1)$, and for each vertex we sample whether it is the tail of a *candidate*, i.e. whether the corresponding surplus edge is possibly part of a strongly connected component.

3. We visit the tails of the candidates in depth-first order, and for each of them, sample where the head of the corresponding surplus edge is.

Then, our approach is as follows.

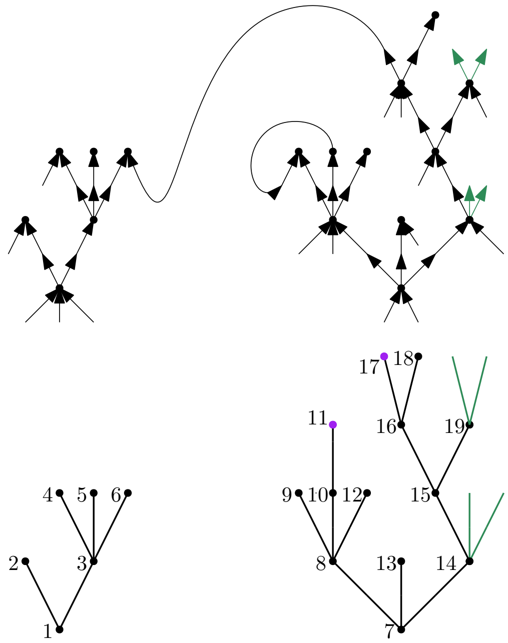


Figure 3: The out-forest is defined based on the exploration of the digraph. For each surplus edge, we add an extra leaf, which we colour purple. The labels of the vertices correspond to the time step in the exploration at which the vertex is added. The green edges lead to vertices of which the degree and colour have not yet been sampled.

1. We find the limit under rescaling of $\hat{F}_n(m_n)$ for $m_n = O(n^{2/3})$ conditional on $\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$ and simplicity of the digraph. We do this by studying the height process and Łukasiewicz path of the out-forest.

2. We show that the positions of the tails of the candidates converge.

3. We show that the positions of the heads of the candidates converge.

4. We identify the tails and heads of the candidates, and recover the strongly connected components from the resulting digraph with a cutting procedure. We show that this cutting procedure converges.

5. We show that for any $\delta > 0$, with high probability, all strongly connected components with length larger than $\delta n^{1/3}$ 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 strongly connected components by only considering the exploration up to time m_n. This finishes the proof of the convergence in the product topology.

2 Sampling the MDM in the discrete and the continuum

If we forget about the directions of the edges, the graph is supercritical, and the graph contains a unique giant component with surplus going to infinity as $n \to \infty$. 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 strongly connected components. Therefore, we will not try to sample the entire digraph, but focus on the information that we need to find the strongly connected components. We start by studying the discrete graph model, with the goal of identifying which edges can be part of a strongly connected components, and how to sample them. In Subsubsection 2.1.1, we establish necessary conditions for an edge to be part of a strongly connected component. The result implies that we only need to study the out-forest, and a small subset of the surplus edges, which we call *candidates*. In Subsubsections 2.1.2 and 2.1.3 we study the law of the out-forest and the candidates respectively, and we define a procedure to sample both. This yields a sequence of directed multigraphs with edge lengths that the strongly connected components are embedded in. In Subsection 2.2, we define the continuous counterpart of the sampling procedure. The resulting object will the limit under rescaling of the sequence of directed multigraphs with edge lengths that the strongly connected components are embedded in that was constructed in Subsubsections 2.1.2 and 2.1.3.

2.1 The discrete case

We will discuss the different type of edges that we can encounter in the exploration. By slight abuse of notation, we call the purple vertex that corresponds to a surplus edge its tail.

2.1.1 Necessary conditions for an edge to be part of an SCC

Amongst the surplus edges, *ancestral surplus edges*, which are surplus edges that point from a vertex to one of its ancestors, play a special rôle. All other surplus edges are called non-ancestral. This is illustrated in Figure 4a. In Figure 4b we show how surplus edges affect the structure of the strongly connected components. This is the content of Lemma 2.1.

Lemma 2.1. *The following facts hold for strongly connected components.*

1. *The vertices of a strongly component are contained in one of the components of $(\hat{F}_n(k), k \geq 1)$.*

2. *Ancestral surplus edges are always part of a strongly connected component.*

3. *A non-ancestral surplus edge is only part of a strongly connected component if its head is an ancestor of the tail of a surplus edge that is part of a strongly connected component.*

4. *An edge in $(\hat{F}_n(k), k \geq 1)$ is only part of a strongly connected component if its head is an ancestor of the tail of a a surplus edge that is part of a strongly connected component.*

5. *For any non-trivial strongly connected component, the first surplus edge of the SCC that is explored is an ancestral surplus edge, and a component of $(\hat{F}_n(k), k \geq 1)$ contains a strongly connected component 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 strongly connected component. Without loss of generality, v is explored first in depth-first order in the out-direction. By v and w being part of the same strongly connected component, 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 rooted at v. This implies that they are part of the same component of $(\hat{F}_n(k), k \geq 1)$.
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 strongly connected component and that the ancestral surplus edge from v to w is in this strongly connected component as well.
To prove 3, suppose we sample a non-ancestral surplus edge from v to w that is part of a

strongly connected component. Then, by 2, there is a path from w to v present at the time of sampling (v, w). Let (x, y) be the first surplus edge on this path. This implies that (x, y) is in the same strongly connected component as v and w. Moreover, the path from w to x consists of edges in the out-forest, x is a descendant of w.
Next, for 4, suppose (v, w) is an edge of $(\hat{F}_n(k), k \geq 1)$ that is part of a strongly connected component. This means that there is a path from w to v. Let (x, y) be the first edge on this path such that y is not a descendant of w. Then, (x, y) is a surplus edge that is part of the same strongly connected component as v and w, and (v, w) is on the path from the root to x.
Finally, 2 and 3 imply 5. □

Lemma 2.1 motivates the following definition.

Definition 2.2. *A surplus edge is a* candidate *if either*

- *It is an ancestral surplus edge, or*
- *One of the descendants of its head is the tail of a candidate.*

The following corollary is at the core of our strategy to study the strongly connected components.

Corollary 2.3. *All edges that are part of a strongly connected component are either a candidate, or are contained in the subforest of $(\hat{F}_n(k), k \geq 1)$ that is spanned by the tails of candidates and the component roots.*

Proof. This follows from Definition 2.2 and parts 1, 2, 3 and 5 of Lemma 2.1. □

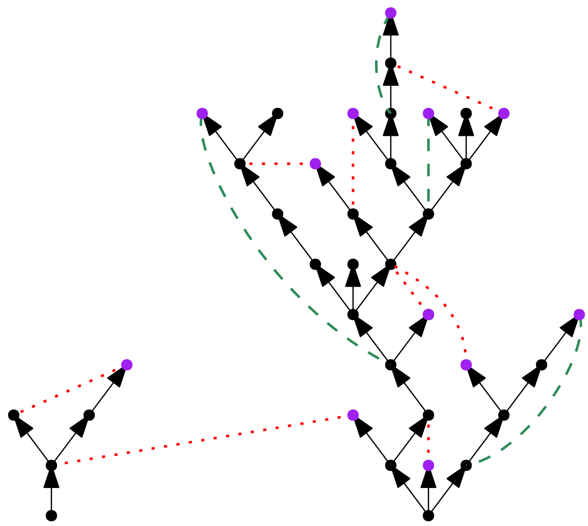
Corollary 2.3 implies that for every purple vertex, we only need to know whether it is a candidate, and if so, where its head is.

2.1.2 Sampling the out-forest

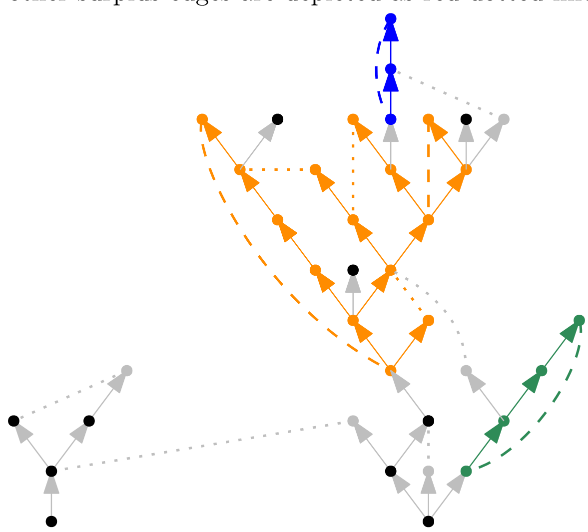
This subsubsection discusses how to obtain the out-forest conditional on the order in which the vertices are discovered. We will study the law of the degrees in order of discovery in Subsection 3.1. Informally, the out-forest is obtained in the following way. Suppose the degrees in order of discovery are given by $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,n})$. Up to time step k, suppose we have added the vertices corresponding to the first $m \leq k$ elements of $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,n})$ to the forest. We call these vertices *discovered*. Then, at time $k + 1$,

1. If we have finished a component of the out-forest, let the next component have a root with out-degree $\hat{D}^+_{n,m+1}$.
2. Otherwise,
   (a) With probability proportional to the total in-degree of the undiscovered vertices, i.e. $\sum_{i=m+1}^{n} \hat{D}^-_{n,i}$, let the next vertex in depth-first order be a black vertex with out-degree $\hat{D}^+_{n,m+1}$.
   (b) With probability proportional to the unpaired in-half-edges of the l discovered vertices, let the next vertex in depth-first order be a purple leaf, and reduce the number of unpaired in-edges of the l discovered vertices by 1.

We make this rigourous in the following lemma.



(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 (green dashed) point from a vertex v to one of its ancestors. They are always part of a strongly connected component. The other surplus edges are depicted as red dotted lines.



(b) The non-trivial strongly connected components embedded in the components of the out-forest are depicted in orange, blue and green. The trivial strongly connected components are black. The grey edges are not part of a strongly connected component, and the grey vertices correspond to purple leaves that are not part of a strongly connected component.

Figure 4: We illustrate the different types of surplus edges and how they affect the structure of the strongly connected components.

Lemma 2.4. *Suppose the sequence of degrees in order of discovery $(\hat{D}_1^n, \ldots, \hat{D}_n^n)$ is given. Suppose, for $1 \leq l \leq k$, that up to time l, $\hat{P}_n(l)$ surplus edges have been sampled. Then,*

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

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

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

define

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

such 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

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

Therefore, we do not need to know the position of the heads of the surplus edges 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 $k - \hat{P}_n(k)$ vertices have been discovered. Thus, up to time k, the out-forest contains $\hat{P}_n(k)$ purple leaves, and black vertices with degrees $(\hat{D}_{n,1}^+, \ldots, \hat{D}_{n,k-\hat{P}_n(k)}^+)$, so by definition of the Łukasiewicz path, its value is indeed equal to $\hat{S}_n^-(k)$ at time k. Moreover, up to time k, the total in-degree of discovered vertices is equal to $\sum_{i=1}^{k-\hat{P}_n(k)} \hat{D}_{n,i}^-$. At every time step, we pair 1 in-half-edge of a discovered vertex, unless we start a new component. $-\hat{I}_n^+(k)$ corresponds to the number of out-components that are finished 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 follows. We note that this probability does not depend on the position of the heads of the surplus edges, which implies that we can sample the out-forest without this information. $\square$

2.1.3 Sampling the candidates

We will now study the law of the candidates conditional on $(\hat{F}_n(k), k \geq 1)$. Like before, for each k, let $\hat{P}_n(k)$ denote the number of purple vertices amongst the first k vertices in the out-forest, and let $\hat{S}^-(k)$ denote the number of unpaired in-half-edges of discovered vertices at time k. We will first identify the tails of the candidates amongst the purple vertices, and then we will sample the position of their heads.
If the vertex visited at time k is purple, 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 purple vertex visited 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 tree: for a vertex v with in-degree m, the edges connecting it to its children will have length $m-1$ (unless v is the root of an out-component, then the edges connecting to its children will have length m). The height of vertex w in this forest with edge lengths corresponds to the number of in-half-edges that can be used to form an ancestral surplus edge with tail w. We add lengths to all edges in $(\hat{F}_n(k), k \geq 1)$ and call the resulting forest with edge lengths $(\hat{F}^{\ell}(k), k \geq 1)$. Denote the height process of $(\hat{F}^{\ell}(k), k \geq 1)$ by $(\hat{H}^{\ell}_n(k), k \geq 1)$.
Recall that the first candidate in any component of $(\hat{F}_n(k), k \geq 1)$ is an ancestral surplus edge. The following lemma illustrates the importance of $\hat{H}^{\ell}$ in finding the first ancestral surplus edges in the out-components.

Lemma 2.5. *Consider the exploration of $(\hat{F}_n(k), k \geq 1)$ at time k. If no ancestral surplus edge has been sampled in the current component, then the probability that k is the tail of an ancestral surplus edge is given by*

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

This event is independent of the position of the heads of the surplus edges that were found before time k.

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 visited since the discovery of x up to time k are descendants of x, because $(\hat{F}_n(k), k \geq 1)$ 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 purple vertices, k is the tail of a surplus edge if and only if k is purple, 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 visited vertex is equal to $1/\hat{S}^-_n(k)$. The stated probability follows. The independence on the position of the heads of earlier surplus edges is immediate. □

We now illustrate how to find the other candidates in a component of $(\hat{F}_n(k), k \geq 1)$.

Lemma 2.6. *Let $T^n_{l_n}$ be a component of $(\hat{F}_n(k), k \geq 1)$ with root $l_n + 1$ and component length σ_n. Suppose the first ancestral surplus edge in $T^n_{l_n}$ corresponds to purple vertex $V^n_1 \in [l_n + 2, l_n + \sigma_n]$. Let $V^n_1 < k \leq l_n + \sigma_n$, and suppose the candidates found up to time k are given by $V^n_1, \dots, V^n_m$. Let $T^{n,mk}_k$ be the subtree of $T^n_{l_n}$ spanned by $\{l_n + 1, V^n_1, \dots, V^n_m, k\}$, and let $\ell(T^{n,mk}_k)$ be its total length with edge lengths as defined by $(\hat{H}^{\ell}(m), m \in [l_n + 1, l_n + \sigma_n])$. Then, the probability that k is a candidate is given by*

$$\frac{\ell\left(T^{n,mk}_k\right) - l}{\hat{S}^-(k)} \mathbb{1}_{\{\hat{P}_n(k) = \hat{P}_n(k-1) + 1\}}.$$

Proof. Note that if k is purple, it gets paired to a uniform pick from the $\hat{S}^-(k)$ unpaired in-half-edges of discovered vertices. By Definition 2.2, in that case, k is a candidate if and only if its head 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 $\{V_1^n, \dots, V_m^n\}$. 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 purple vertex corresponds to 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 candidates can be found as follows.

Lemma 2.7. *Let $T_{l_n}^n$ be a component of $(\hat{F}_n(k), k \geq 1)$ with root $l_n + 1$ and component length σ_n. Suppose its candidates are given by $\{V_1^n, \dots, V_{N_n}^n\}$. Then, for $1 \leq i \leq N_n$, suppose the heads of the surplus edges corresponding to $V_1^n, \dots, V_{i-1}^n$ are given by $W_1^n, \dots, W_{i-1}^n$ respectively. Then, the in-half-edge that V_i^n gets paired to is a uniform pick from the*

$$\ell\left(T_{V_i^n}^{n,mk}\right) - (i-1)$$

unpaired in-half-edges of $T_{V_i^n}^{n,mk}$ that remain. Call the corresponding vertex W_i^n.

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. ☐

Lemmas 2.4, 2.5, 2.6, and 2.7 justify the following sampling procedure.

1. Sample the out-forest $(\hat{F}_n(k), k \geq 1)$.
2. Fix $T > 0$ and define a counting process $(A_n(k), k \leq \lfloor Tn^{2/3} \rfloor)$, with the probability of an increment at time k given by
$$a_k = \frac{\hat{H}_n^\ell(k)}{\hat{S}_n^-(k)} \mathbb{1}_{\{\hat{P}_n(k) - \hat{P}_n(k-1) = 1\}}.$$
3. For $i \geq 1$, set $X_i^n = \min\{k : A_n(k) = i\}$. Define
$$L_i^n = \min\left\{k \geq 1 : \hat{S}_n^+(k) = \min\{\hat{S}_n^+(l) : l \leq X_i^n\}\right\} \text{ for } i \geq 1$$
$$\Sigma_i^n = \min\left\{k \geq 1 : \min\left\{\hat{S}_n^+(l) : l \leq L_i^n + k\right\} < \min\left\{\hat{S}_n^+(l) : l \leq X_i^n\right\}\right\} \text{ for } i \geq 1,$$
such that for each $i \geq 1$, $\left(\hat{S}^+(k), k \in [L_i^n + 1, L_i^n + \Sigma_i^n]\right)$ encodes the tree containing X_i^n. For each $(l_n, \sigma_n) \in \{(L_i^n, \Sigma_i^n)\}$, let $T_{l_n}^n$ be the tree in $(\hat{F}_n(k), k \geq 1)$ with root $l_n + 1$, and do the following.
   (a) Set $V_1^n = \min\{m \geq 1 : A_n(m) = A_n(g) + 1\}$, and find the other candidates $\{V_2^n, \dots, V_{N_n}^n\}$ according to the procedure described in the statement of Lemma 2.6.
   (b) For $V_1^n, \dots, V_{N_n}^n$, sample their heads $W_1^n, \dots, W_{N_n}^n$ respectively according to the procedure described in the statement of Lemma 2.7.

(c) Let $T^{n,\mathrm{mk}}(l_n)$ be the subtree of $T^n_{l_n}$ spanned by $\{l_n+1, V_1^n, \dots, V_{N_n}^n\}$, say $V_i^n \sim W_i^n$ for each $1 \leq i \leq N_n$, and set $M^n_{l_n} := T^{n,\mathrm{mk}}_{l_n}/\sim$, which we note is a directed graph with surplus N_n.

Then, all strongly connected components of $\vec{G}(n,(D^-,D^+))$ of which the first candidate is sampled before time $\lfloor Tn^{2/3} \rfloor$ are subgraphs of $\left\{M^n_{L_i^n}, i \geq 1\right\}$. Call these strongly connected components, ordered by decreasing size $(C_i^T(n), i \geq 1)$, completed with an infinite repeat of $\mathfrak{L}$. Observe that we may view $M^n_{L_i^n}$ as a finite rooted directed multigraph $M^n_{L_i^n}$ whose edges are endowed with lengths. To be precise, in $M^n_{L_i^n}$, let the vertex set consist of L_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 $(C_i^T(n), i \geq 1)$ by ordering the SCCs in $\left\{M^n_{L_i^n}, i \geq 1\right\}$ by decreasing size, and completing the list with an infinite repeat of $\mathfrak{L}$. See Figure [picture Robin] for a sketch of this procedure.

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 slight modification of the procedure defined in Subsubsection 3.2.2 of [?].

2.2.1 $\mathbb{R}$-trees and their encoding

The continuum analogue of discrete trees are given by $\mathbb{R}$-trees. A survey paper on $\mathbb{R}$-trees can be found in [?]. An $\mathbb{R}$-tree is a compact metric space $(\mathcal{T}, d)$ such that for every $a, b \in \mathcal{T}$ the following two properties hold:

1. There exists a unique isometry $i_{a,b} : [0, d(a,b)] \to \mathcal{T}$ such that $i_{a,b}(0) = a$ and $i_{a,b}(d(a,b)) = b$.

2. If $q : [0,1] \to \mathcal{T}$ is any continuous map such that $q(0) = a$ and $q(1) = b$ then the image of q is the same as the image of $i_{a,b}$.

Let $[\![a,b]\!]$ denote the image of $i_{a,b}$. This is the unique path between a and b.

$\mathbb{R}$-trees are often encoded by continuous excursions which can be seen as a continuous analogue of the height function of a tree. Let $f : [0,\sigma] \to [0,\infty)$ be a continuous excursion, meaning f is continuous, $f(0) = f(\sigma) = 0$ and $f(x) > 0$ for all $x \in (0,\sigma)$. Using f we can define a pseudo-metric

$$d_f(x,y) = f(x) + f(y) - 2 \min_{s \in [x \wedge y, x \vee y]} f(s).$$

Using this we can define the quotient space

$$\mathcal{T}_f = [0,\sigma]/\{d_f = 0\}.$$

The space $\mathcal{T}_f$ equipped with the metric d_f is the $\mathbb{R}$-tree encoded by the excursion f. Let $p_f : [0,\sigma] \to \mathcal{T}_f$ be the natural projection function. Then $\mathcal{T}_f$ inherits a distinguished root vertex $\rho = p(0) = p(\sigma)$.

2.2.2 The limit object

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

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

Remark 2.8. *We note that the coefficient of the parabolic drift is negative. Indeed, the sign of the parabolic drift is the same as the sign of $\mu - \mathbb{E}[D^+(D^-)^2]$, and we note that*

$$\frac{\mathbb{E}[(D^-)^2 D^+]}{\mathbb{E}[D^+]} - \left(\frac{\mathbb{E}[D^+D^-]}{\mathbb{E}[D^+]}\right)^2 = \frac{\mathbb{E}[(Z^+)^2]}{\mu} - 1$$

is the variance of D^- under the law of $\mathbf{D}$ size-biased by D^+, which is positive. Hence $\mathbb{E}[D^+(D^-)^2]/\mu \geq 1$, which shows that $(\hat{B}_t)_{t \geq 0}$ is a Brownian motion with a downwards parabolic drift.

Define

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

Then, it is standard 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 $\left(\sigma_+\hat{B}_t, t \geq 0\right)$. This fact also follows from the argument in Section 3. Call this forest $(\hat{\mathcal{F}}(t), t \geq 0)$.
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, fix $T > 0$, such that $A_T < \infty$ almost surely. For i in $[A_T]$, set $X_i = \min\{t : A_T = i\}$. Define

$$L_i = \inf\left\{t \geq 0 := \hat{B}_t = \inf\{\hat{B}_s : s \leq X_i\}\right\} \text{ for } i \in [A_T] \text{ and}$$
$$\Sigma_i = \inf\left\{t \geq 0 : \inf\{\hat{B}_s : s \leq L_i + t\} < \inf\{\hat{B}_s : s \leq X_i\}\right\} \text{ for } i \in [A_T],$$

such that for each i in $[A_T]$, $\left(\frac{2}{\sigma_+}\hat{R}_t, t \in [L_i, L_i + \Sigma_i]\right)$ encodes the $\mathbb{R}$-tree in $(\hat{\mathcal{F}}(t), t \geq 0)$ that contains X_i. For each element of $\{(L_i, \Sigma_i) : i \in A_T\}$ we will sample the candidates in the $\mathbb{R}$-tree. Fix i, and set $(l, \sigma) = [L_i, \Sigma_i]$. Let $V_1 = \inf\{s > 0 : A(s) = A(l) + 1\}$, such that $l \leq V_1 \leq l + \sigma$ by definition of (l, σ). Let $\mathcal{T}_l$ be the R-tree encoded by $\left(\frac{2}{\sigma_+}\hat{R}_t, t \in [l, l+\sigma]\right)$ and let $p_l : [l, l+\sigma] \to \mathcal{T}_l$ be the projection onto $\mathcal{T}_l$ given by the encoding. Set

$$||\mathcal{T}_l|| = \sup\left\{\frac{2}{\sigma_+}\hat{R}_t, t \in [l, l+\sigma]\right\},$$

which we note is the height of $\mathcal{T}_l$.
Suppose we have found candidates $\{V_1, \ldots, V_m\}$. For $V_m \leq s \leq l + \sigma$, let T_s^{mk} be the subtree of $\mathcal{T}_l$ spanned by $p_l(\{l, V_1, \ldots, V_m, s\})$, 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 $[V_m, l + \sigma]$ 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 $\mathcal{T}_l$, 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 claim that $\mathbb{P}(N = \infty) = 0$. Indeed, note that $V_m \leq s \leq V_{m+1}$ implies that $|T_s^{\mathrm{mk}}| < (m+1)||\mathcal{T}_l||$. Therefore,

$$\mathbb{P}\left(N \geq l+1, V_{m+1} - V_m < t \middle| N \geq l\right) \leq \mathbb{P}(E_{m+1} < t),$$

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

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

Then,

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

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 $1 \leq i \leq N$, let the head corresponding to V_i, which we call W_i, be a uniform pick from the length measure on $T_{V_i}^{\mathrm{mk}}$.

Let T_l^{mk} be the subtree of $\mathcal{T}^l$ spanned by $\{l, V_1, \ldots, V_N\}$. Then, in T_l^{mk}, set $V_i \sim W_i$ for each $1 \leq i \leq N$, and set $\mathcal{M}_l := T_l^{\mathrm{mk}}/\sim$. View $\mathcal{M}_l$ as an element of $\vec{\mathcal{G}}$ in the natural way, and call it M_l. To be precise, let the vertex set of M_l consist of l, W_i for $i \leq N$, and the branch points $V_i \wedge V_j$ for $i \neq j \leq N$. The directions are inherited from $\mathcal{T}^l$, by considering all edges directed away from the root. Remove all edges that do not lie in a strongly connected component of M_l and delete any isolated vertices that are thus created. Then, for any vertices of degree 2, merge the neighbouring edges and sum their lengths. This creates a collection $\mathcal{C}_l$ of strongly connected MDMs. Doing this for each $(l, \sigma) \in \{[L_i, \Sigma_i]\}$ yields the collection of strongly connected MDMs $\mathcal{C}$ that has the law of the limit in Theorem 1.1.

2.2.3 Properties of the limit object

We note that the limit object is encoded by 3 parameters: the real 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 point as described in 2.2.2. The limit object that is studied in [?] 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 [?], 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 properties also hold for our limit object. Let $\mathcal{M} := \bigcup_{L_i} \mathcal{M}_{L_i}$.

Corollary 2.9. *1. The number of complex connected components of $\mathcal{M}$ has finite expectation.*

2. The number of loops of $\mathcal{M}$ is a.s. finite.

Proof. The proof is analogous to the proof of Theorem 4.5 in [?]. □

Corollary 2.10. *The strongly connected components of $\mathcal{M}$ all have different lengths almost surely.*

Proof. The proof is analogous to the proof of Proposition 4.6 in [?]. $\square$

Write $\mathcal{C}$ for the strongly connected components 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 Subsubsection 2.2.2. Write $\mathcal{C}_{\text{cplx}}$ for the list of complex components of $\mathcal{C}$ in decreasing order of length. For sequences $(K_1, \ldots, K_j)$ and $(J_1, \ldots, J_k)$ of directed multigraphs, write $(J_1, \ldots, J_k) \equiv (K_1, \ldots, K_j)$ if $j = k$ and J_i is isomorphic to K_i for each $i \leq j$. Extend this notation naturally to the case where one or both of the sequences has edge lengths by ignoring the edge lengths.

Corollary 2.11. *Let $K_1, \ldots, K_j$ be a finite sequence consisting of 3-regular strongly connected directed multigraphs or loops. We have*

$$\mathbb{P}\left[\mathcal{C}_l \equiv (K_1, \ldots, K_j)\right] > 0.$$

Assuming that $K_1, \ldots, K_j$ are all complex, we also have that

$$\mathbb{P}\left[\mathcal{C}_{cplx} \equiv (K_1, \ldots, K_j)\right] > 0.$$

Let $(e_i, 1 \leq i \leq K)$ be an arbitrary ordering of the edges of $K_1, \ldots, K_j$. Then, conditionally on $\mathcal{C}_l \equiv (K_1, \ldots, K_j)$, (resp. $\mathcal{C}_{cplx} \equiv (K_1, \ldots, K_j)$), $\mathcal{C}_l$ (resp. $\mathcal{C}_{cplx}$) gives lengths $(\ell(e_i), 1 \leq i \leq K)$ to these edges, and their joint distribution has full support in

$$\left\{ \mathbf{x} = (x_1, \ldots, x_K) \in \mathbb{R}_+^K : \forall 1 \leq i \leq k-1, \sum_{j: e_j \in E(K_i)} x_j \geq \sum_{j: e_j \in E(K_{i+1})} x_j \right\}.$$

Proof. The proof is analogous to the proof of Theorem 6.1 in [?]. $\square$

3 Convergence of the out-forest

In this section, we will show that the Łukasiewicz path and height process corresponding to the forest $\hat{F}_n(m_n)$ converge under rescaling, for $m_n = O(n^{2/3})$. The main result of this section is as follows.

Theorem 3.1. *Like before, let $(\hat{F}_n(k), k \geq 1)$ be the sequence of out-forests given by the exploration, where we set $\hat{F}_n(k+1) = \hat{F}_n(k)$ if all in-half-edges have been paired at time k. Let $(\hat{S}_n^+(k), \hat{H}_n(k), k \geq 1)$ be the Łukasiewicz path and height process corresponding to $(\hat{F}_n(k), k \geq 1)$. Let $\hat{S}_n^-(k)$ denote the number of unpaired in-half-edges of vertices that are seen at time k. Let $\hat{P}_n(k)$ be the number of purple vertices seen in the first k time steps. Moreover, let $(B_t)_{t \geq 0}$ be a Brownian motion, and define*

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

Set

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

Then,

$$\left(n^{-1/3}\hat{S}_n^+\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}\hat{H}_n\left(\lfloor n^{2/3}t\rfloor\right), t\geq 0\right) \xrightarrow{d} \left(\sigma_+\hat{B}_t, \frac{2}{\sigma_+}\hat{R}_t, t\geq 0\right)$$

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

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

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

We prove Theorem 3.1 by studying two other forests that are related to $\hat{F}_n(m_n)$ via a change of measure.
The proof is structured as follows.

1. Let $(\hat{\mathbf{D}}_{n,1},\ldots,\hat{\mathbf{D}}_{n,n})$ denote the degree tuples in order of discovery. Fix m and let $(\mathbf{Z}_1,\ldots,\mathbf{Z}_m)$ be i.i.d. elements of $\mathbb{N}\times\mathbb{N}$, $\mathbf{Z}_i := (Z_i^-, Z_i^+)$ such that

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

   Then, we show that the law of $(\hat{\mathbf{D}}_{n,1},\ldots,\hat{\mathbf{D}}_{n,m})$ conditional on $\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$ is absolutely continuous to $(\mathbf{Z}_1,\ldots,\mathbf{Z}_m)$, and we show convergence of the Radon-Nikodym derivative ϕ_m^n for $m = O(n^{2/3})$. This is the content of Subsection 3.1.

2. Then, 1 is the motivation to sample i.i.d. copies of $\mathbf{Z}$, $(\mathbf{Z}_i, i\geq 1)$, and to study a Galton-Watson forest with offspring distributed as Z_1^+. Call this forest $(F(k), k\geq 1)$. The convergence of the Łukasiewicz path of $(F(k), k\geq 1)$ under rescaling follows from Donsker's theorem.

3. In Subsection 3.3, we modify $(F(k), k\geq 1)$ to include purple leaves. We add extra randomness, similarly to the procedure described in Lemma 2.4, such that at some time steps, a purple leaf is added. We call the resulting forest $(F_n^p(k), k\geq 1)$. We respect the order of the degrees in $(F(k), k\geq 1)$, in the sense that for any k, the k^{th} black vertex in $(F_n^p(k), k\geq 1)$ has the same number of children as the k^{th} vertex in $(F(k), k\geq 1)$. $(F_n^p(k), k\geq 1)$ depends on n, because the probability of finding a purple vertex depends on n. We then show that the Łukasiewicz path and height process corresponding to $(F_n^p(k), k\geq 1)$ converge under rescaling, jointly with the convergence of the Łukasiewicz path and height process corresponding to $(F(k), k\geq 1)$ under rescaling up to time $O(n^{2/3})$.

4. We use the measure change to translate the convergence of the encoding processes of $(F_n^p(k), k\geq 1)$ under rescaling to convergence of the encoding processes of $(\hat{F}_n(k), k\geq 1)$ under rescaling up to time $O(n^{2/3})$. This yields Theorem 3.1.

3.1 The measure change and its convergence

Lemma 3.2. *For all* $m\leq n$ *there exists a function* $\phi_m^n : (\mathbb{N}\times\mathbb{N})^m \to \mathbb{R}$ *such that for all bounded measurable functions* $u : (\mathbb{N}\times\mathbb{N}) \to \mathbb{R}$*,*

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

Let us define

$$\Phi(n,m) = \phi_m^n(\mathbf{Z}_1, \dots, \mathbf{Z}_m).$$

Then for all $T > 0$, *[phrase the convergence in a good way].*

3.2 Convergence of the degrees in order of discovery

Recall that

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

encodes the out-degrees in order of discovery, while

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

encodes the in-degrees in order of discovery. We will study these processes via the measure change that we defined in Subsection 3.1. Recall that

$$Y^+(k) = \sum_{i=1}^{k} (Z_i^+ - 1),$$

and

$$Y^-(k) = \sum_{i=1}^{k} (Z_i^- - 1).$$

We will first examine the joint distribution under rescaling of these processes, which is the content of the following lemma.

Lemma 3.3. *We have that*

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

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

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

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^2$, *as* $n \to \infty$, *with* $(B_t^1)_{t \geq 0}$ *and* $(B_t^2)_{t \geq 0}$ *two independent standard Brownian motions.*

Proof. The first statement follows from the fact that $(Y^-(k), k \geq 1)$ is a random walk with steps of mean ν^-. Then, by Donsker's Theorem in two dimensions (see e.g. Theorem 7.1.4 in [**?**]), we get that

$$\left(n^{-1/3}\left(Y^-\left(\lfloor n^{2/3}t\rfloor\right) - n^{2/3}\nu_- t\right), n^{-1/3}Y^+\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right)$$

converges to a Gaussian process with covariance matrix

$$\begin{pmatrix} \sigma_-^2 & \sigma_{-+} \\ \sigma_{-+} & \sigma_+^2 \end{pmatrix} t,$$

which proves the first statement. □

We are now able to examine the effect of the change of measure on $(Y^-(k), Y^+(k), k \geq 1)$. This is the content of the following theorem.

Theorem 3.4. *For $T > 0$,*

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

in the Skorokhod topology as $n \to \infty$, where $(\hat{B}_t, t \geq 0)$ is distributed as follows. For F a suitable test function, and for $(B_t)_{t\geq 0}$ a Brownian motion,

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

Proof. Firstly, we use the results of Proposition **??** [refer to result Zheneng] and Lemma 3.3, and repeat the proof of Theorem 4.1 in [**?**] to find that for $Y^{[n,-]}(t) := n^{-1/3}Y^-\left(\lfloor n^{2/3}t\rfloor\right) - n^{1/3}\nu_- t$ and $Y^{[n,+]}(t) := n^{-1/3}Y^+\left(\lfloor n^{2/3}t\rfloor\right)$, for F a bounded continuous test-function,

$$\begin{aligned}
&\mathbb{E}\left[F(Y^{[n,-]}(t), Y^{[n,+]}(t), 0 \leq t \leq T)\right] \\
&\to \mathbb{E}\left[\Phi(T) F\left(\frac{\sigma_{-+}}{\sigma_+}B_t^1 + \left(\sigma_-^2 - \frac{\sigma_{-+}^2}{\sigma_+^2}\right)^{1/2} B_t^2, \sigma_+ B_t^1, 0 \leq t \leq T\right)\right]
\end{aligned}$$

as $n \to \infty$. Hence, for F a suitable test function,

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

Combining this with the first statement of Lemma 3.3 yields the result. ☐

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

Lemma 3.5. *For $(\hat{B}_t, 0 \leq t \leq T)$ as in the statement of Theorem 3.4, we have that*

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

for $(B_t)_{t\geq 0}$ a Brownian motion.

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

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

for $(B_t)_{t\geq 0}$ a Brownian motion. 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. □

3.3 Adding purple vertices to a Galton-Watson forest

In this subsection we define $(F_n^p(k), k \geq 1)$ and show that its Łukasiewicz path and height process converge under rescaling. Moreover, we will show that this convergence holds jointly with the convergence under rescaling of the number of purple vertices seen up to time k and the number of seen, but unused in-half-edges up to time k. The following lemma motivates the definition of $(F_n^p(k), k \geq 1)$.

Lemma 3.6. *Consider an eDFS of a configuration model on n vertices, with the total number of in-half-edges equal to μn. Suppose the number of unpaired in-half-edges of discovered vertices at step k in the exploration is equal to $S_n^-(k)$, suppose $(S_n^+(l), 1 \leq l \leq k)$ encodes the Łukasiewicz path of the out-forest up to time k, and set*

$$I_n^+(k) = \inf\left\{S_n^+(l), 1 \leq l \leq k\right\}.$$

Then, the probability that, in the $(k+1)^{th}$ time step, we sample a surplus edge is given by

$$p_{k+1} := \frac{S_n^-(k)}{\mu n - k - I^+(k) + 1}\mathbb{1}_{\{I^+(k) = I^+(k-1)\}}.$$

Proof. This is a slight adaptation of Lemma 2.4, with $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,m})$ replaced by $(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$, and $\sum_{i=1}^n \hat{D}_i^-$ replaced by its mean μn. □

We will now define $(F_n^p(k), k \geq 1)$ and its Łukasiewicz path $(S_n^+(k), k \geq 1)$ as a function of $(Y^-(k), Y^+(k), k \geq 1)$ and extra randomness.

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}, 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 $(F_n^p(k), k \geq 1)$ be the forest with Łukasiewicz path $(S_n^+(k), k \geq 1)$ in which the k^{th} vertex is purple if and only if $P^n(k) - P^n(k-1) = 1$.

3.3.1 Convergence of the Łukasiewicz path

To show convergence of the Łukasiewicz path corresponding to $(F_n^p(k), k \geq 1)$, we will first examine the limit of $(P_n(k), k \geq 1)$ under rescaling. We will first prove tightness, after which we will show convergence.

Lemma 3.7. *For every $T > 0$,*
$$\left(n^{-1/3} P_n\left(\lfloor n^{2/3} T \rfloor\right)\right)_{n \geq 1}$$
is tight.

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,
$$p_k \leq \frac{Y^-(k) + k}{\mu n - k},$$
and note that 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 proves the statement. □

Lemma 3.8. *We have that*

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

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

Proof. Fix $T>0$. Recall that

$$p_{k+1} = \frac{S_n^-(k)}{\mu n - k - I^+(k)+1}\mathbb{1}_{\{I^+(k)=I^+(k-1)\}}.$$

Define $M^+(k)=\min\{Y^+(l): l\leq k\}$ such that $0\geq I^+(k)\geq M^+(k)-P_n(k)$. Then, by Lemma 3.7, Lemma 3.3, and the continuous mapping theorem, $\left(n^{-1/3}I^+(\lfloor n^{2/3}t\rfloor)\right)_{n\geq 1}$ is tight for all $t>0$. We will now argue that the indicator, which ensures that the roots are never purple, does not have an effect on $(P_n(k), k\leq m)$ on the scale that we are interested in. Let $t>0$, and set $m=\lfloor n^{2/3}t\rfloor$. 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 for all $t\geq 0$. This means that if we allow the roots to be purple, with high probability, we would only sample $O(1)$ purple roots up to time $O(n^{2/3})$. This does not affect $(P_n(k), k\leq m)$ on the scale that we are interested in.
Then, the convergence in Lemma 3.3, the tightness of $\left(n^{-1/3}I^+(\lfloor n^{2/3}t\rfloor)\right)_{n\geq 1}$ and Lemma 3.7 imply that

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

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 the colours of the vertices. Then,

$$M_n(k) := \sum_{i=1}^{k}(\mathbb{1}_{\{P_n(i)-P_n(i-1)=1\}} - p_i)$$

is a martingale in $\mathcal{G}$. We claim that $(n^{-1/3}M_n(\lfloor n^{2/3}t \rfloor), t \geq 0)$ converges to $(0, t \geq 0)$ 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,

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

which proves the statement.

□

The convergence of P_n under rescaling implies the convergence of S^+ and S^- under rescaling, which is the content of the following corollary.

Corollary 3.9. *Let $(B_t, t \geq 0)$ be a Brownian motion. We have that*

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

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

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

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

Proof. This follows from Lemmas 3.3 and 3.8 and the expressions

$$S_n^{p,+}(k+1) = Y^+\left[k+1-P_n(k+1)\right] - P_n(k+1),$$

and

$$S_n^{p,-}(k+1) = Y^-\left[k+1-P_n(k+1)\right] - P_n(k+1) - I^+(k) + 1.$$

□

3.3.2 Convergence of the height process

We will extend Corollary 3.9 to joint convergence under rescaling with the height process corresponding to $(F_n^p(k), k \geq 1)$, which is the content of this subsubsection. We prove the following proposition.

Proposition 3.10. *Let $(H^+(k), k \geq 1)$ be the height process corresponding to $(F^p(k), k \geq 1)$. Let $(B_t, t \geq 0)$ be a Brownian motion, and define*

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

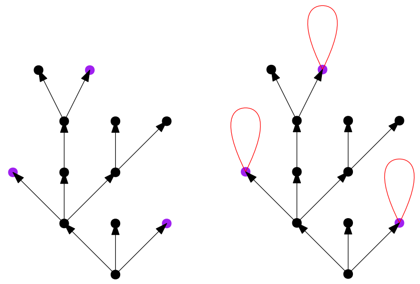


Figure 5: Given a component of $(F_n^p(k), k \geq 1)$ (see left figure), we modify it by sampling independent red Galton-Watson trees with offspring distributed as Z^+ and identifying each purple vertex with a root of a red tree. The resulting tree (see right figure) is a Galton-Watson tree, and the resulting forest $(F_n^{pr}(k), k \geq 1)$ is a Galton-Watson forest.

Set

$$(R_t^+, t \geq 0) = \left(B_t^+ - \inf\left\{B_s^+ : s \leq t\right\}, t \geq 0\right).$$

Then,

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

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^3$*, and*

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

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

The difficulty in proving this proposition is the fact that $(F_n^p(k), k \geq 1)$ is not a Galton-Watson forest, because the probability of sampling a purple vertex changes as time progresses. The theory of convergence of height processes under rescaling is well-developed for Galton-Watson processes (see e.g. [?]), but this is not the case for more general processes. We will adapt a technique that Broutin, Duquesne and Wang develoed in [?] to show convergence of the height process of inhomogeneous random graphs under rescaling. The key idea is that $(F_n^p(k), k \geq 1)$ itself is not a Galton-Watson forest, but we can embed it in a Galton-Watson forest, say $(F^{pr}(k), k \geq 1)$, which will be equal to $(F_n^p(k), k \geq 1)$ with extra red vertices. We then show convergence under rescaling of the height process corresponding to $(F^{pr}(k), k \geq 1)$, and use this to obtain height process convergence for $(F_n^p(k), k \geq 1)$.
We start by defining $(F^{pr}(k), k \geq 1)$. Informally, we obtain $(F^{pr}(k), k \geq 1)$ by modifying $(F_n^p(k), k \geq 1)$ in such a way that the sub-tree rooted at a purple vertex has the same law as a sub-tree rooted at a black vertex. We do this by sampling extra Galton Watson trees with offspring distributed as Z^+, of which we colour all vertices red, and identifying their roots with the purple vertices. The resulting forest is a black-purple-red Galton-Watson forest in which the black-purple forest is embedded. This is illustrated in Figure 5.

The formal procedure is as follows. Suppose we are given $(Y^+(k), S^+(k), P_n(k), k \geq 1)$, which encode $(F_n^p(k), k \geq 1)$.

1. Let $(Y^{red}(k), k \geq 1)$ be an independent copy of $(Y^+(k), k \geq 1)$. $(Y^{red}(k), k \geq 1)$ will encode the red pendant subtrees.
2. Define $\theta_n(k) = k + \min\{j : Y^{red}(j) = -P_n(k-1)\} - P_n(k-1)$.

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

$$(Y^{pr}(k), k \geq 1) = (Y^+(\Lambda_n(k)) + Y^{red}(k - \Lambda_n(k)), k \geq 1) \tag{2}$$

and we let $(F^{pr}(k), k \geq 1)$ be the Galton-Watson process encoded by $(Y^{pr}(k), k \geq 1)$, in which $P_n(\max\{j : \theta_n(j) \leq k\})$ of the first k vertices are blue, $\Lambda_n(k)$ of the first k vertices are black, and the rest is red. We let $(H^{pr}(k), k \geq 1)$ be the height process corresponding to $(F^{pr}(k), k \geq 1)$.

We claim that the forest consisting of the black and blue vertices in $F^{pr}(\theta_n(k))$ is, by construction, equal to $F^p(k)$. Moreover, $(F^{pr}(k), k \geq 1)$ is a Galton-Watson forest. We make the following observations.

1. We claim that

$$\theta_n(k) = \min\{l : F^p(k) \text{ is a subforest of } F^{pr}(l)\}.$$

Indeed, note that $\min\{j : Y^{red}(j) = -P_n(k-1)\}$ is equal to the number of vertices in the first $P_n(k-1)$ trees in the forest encoded by Y^{red}, so

$$\min\{j : Y^{red}(j) = -P_n(k-1)\} - P_n(k-1)$$

is equal to the number of red vertices we add to $F^p(k)$. Then, $\theta_n(k)$ is the index in $(F^{pr}(k), k \geq 1)$ of the k^{th} black or purple vertex.

2. Note that $\Lambda_n(k)$ is the number of blue vertices amongst the first k vertices. This follows from the fact that $\max\{j : \theta_n(j) \leq k\}$ is the number of blue or purple vertices amongst the first k vertices.

3. By the argument above, $(\Lambda_n(k), k \geq 1)$ only takes steps of size 0 or 1. Both $(Y^+(k), k \geq 1)$ and $(Y^{red}(k), k \geq 1)$ are random walks with steps distributed as $Z^+ - 1$, so, by construction, $(Y^{pr}(k), k \geq 1)$ is a random walk with steps distributed as $Z^+ - 1$, so $(F^{pr}(k), k \geq 1)$ is a Galton-Watson forest with offspring distributed as Z^+.

4. By construction, $(H^{pr}(\theta_n(k)), k \geq 1)$ is the height process corresponding to $(F_n^p(k), k \geq 1)$. Moreover,

$$(S^+(k), k \geq 1) = (Y^{pr}(\theta_n(k)) - E(\theta_n(k)), k \geq 1), \tag{3}$$

where $E(k)$ counts the red children of the k^{th} vertex in $(F^{pr}(k), k \geq 1)$.

Considering the construction above and Corollary 3.9, in order to prove Proposition 3.10, it is sufficient to prove the following proposition.

Proposition 3.11. *There exists a process $(D_t, t \geq 0)$ such that*

$$\left(n^{-1/3}\left[Y^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right) - E\left(\lfloor n^{2/3}t\rfloor\right)\right], n^{-1/3}H^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t \geq 0\right)$$
$$\xrightarrow{d} \left(\sigma_+ D_t, \frac{2}{\sigma_+}\left(D_t - \inf\{D_s, s \leq t\}\right), t \geq 0\right)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^2$ as $n \to \infty$ and $\left(\frac{2}{\sigma_+}\left(D_t - \inf\{D_s, s \leq t\}\right), t \geq 0\right)$ is the height process corresponding to $(\sigma_+ D_t, t \geq 0)$.

We postpone the proof to Appendix A.

3.4 Applying the measure change to prove Theorem 3.1

We will combine the convergence of the measure change under rescaling, which is the content of Theorem XXX, and the convergence of the encoding processes of $(F^p(k), k \geq 1)$, which is the content of Proposition 3.10, to prove Theorem 3.1.

Proof. Proof of *Theorem 3.1* Recall that $\hat{P}_n(k)$ denotes the number of purple vertices in $\hat{F}_n(k)$. Set $\hat{I}_n(k) = \min\{\hat{S}_n^+(l) : l \leq k\}$. Then, as shown in Lemma 2.4, the probability that the $(k+1)^{th}$ vertex in $(\hat{F}_n(k), k \geq 1)$ is purple is given by

$$q_{k+1} := \frac{\hat{S}^-(k)}{\sum_{i=0}^n D_i^- - k - \hat{I}_n(k)} \mathbb{1}_{\{\hat{I}_n(k-1) = \hat{I}_n(k)\}}.$$

In order to use the results on $(F^p(k), k \geq 1)$, we would like to replace the term $\sum_{i=0}^n D_i^-$ in the denominator by μn. Therefore, define a new forest $(\hat{F}_n'(k), k \geq 1)$ in which the probability that the $(k+1)^{th}$ vertex is a purple leaf is

$$q_{k+1}' := \frac{\hat{S}^-(k)}{\mu n - k - \hat{I}_n'(k)} \mathbb{1}_{\{\hat{I}_n'(k-1) = \hat{I}_n'(k)\}},$$

where $\hat{P}_n'(k)$ is the number of purple vertices in $\hat{F}_n'(k)$, and $\hat{I}_n'(k)$ is the number of components in $\hat{F}_n'(k)$. We claim that there exists a coupling such that

$$\sum_{i=1}^{\lfloor n^{2/3} T \rfloor} |q_i - q_i'| \xrightarrow{p} 0$$

as $n \to \infty$. Indeed, by the convergence in Theorem 3.4,

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

is tight. Moreover, with a slight adaptation to the proof of Lemma 3.7, we can show that $\left(n^{-1/3} \hat{P}_n' \left(\lfloor n^{2/3} T \rfloor \right) \right)_{n>0}$ is tight. This, combined with the convergence under rescaling of $(\hat{Y}_n^+(k), k \geq 1)$, implies that also $\left(n^{-1/3} \hat{I}_n' \left(\lfloor n^{2/3} T \rfloor \right) \right)_{n>0}$ is tight. By $D_1^-, \dots, D_n^-$ being i.i.d. random variables with mean μ and finite variance, $\left(n^{-1/2} \left(\sum_{i=0}^{n-1} D_i^- - \mu n \right) \right)_{n>0}$ is tight. By using the trivial identity $a/b - c/d = (b(a-c) - c(d-b))/bd$, this implies that $\left(n^{2/3} \max_{k \leq \lfloor n^{2/3} T \rfloor} |q_k - q_k'| \right)_{n>0}$ is tight, which implies that there exists a coupling such that $\left(\max_{k \leq \lfloor n^{2/3} T \rfloor} |\hat{P}_n(k) - \hat{P}_n'(k)| \right)_{n>1}$ and $\left(\max_{k \leq \lfloor n^{2/3} T \rfloor} |\hat{I}_n(k) - \hat{I}_n'(k)| \right)_{n>1}$ are tight, which implies that, again by $a/b - c/d = (b(a-c) - c(d-b))/bd$, $\left(n^{5/6} \max_{k \leq \lfloor n^{2/3} T \rfloor} |q_k - q_k'| \right)_{n>0}$ is tight, which implies that

$$\sum_{i=0}^{\lfloor n^{2/3} T \rfloor} |q_i - q_i'| \xrightarrow{p} 0$$

as $n \to \infty$. Therefore, under the right coupling,

$$\mathbb{P} \left(\max_{k \leq \lfloor n^{2/3} T \rfloor} |\hat{P}_n(k) - \hat{P}'(k)| > 0 \right) \to 0.$$

In other words, we can couple $(\hat{F}_n(k), k \geq 1)$ and $(\hat{F}'_n(k), k \geq 1)$ in such a way that we do not see the difference on the scale that we are interested in. Therefore, we can show convergence under rescaling of the encoding processes of $(\hat{F}'_n(k), k \geq 1)$ instead. To avoid further complicating notation, we will from now on refer to its encoding processes as

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

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

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

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

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

[Change letter measure change to match notation Zheneng] Let $f : D([0,T], \mathbb{R})^3 \to \mathbb{R}$ be a bounded, continuous test-function. Then,

$$\begin{aligned}
\mathbb{E}\left[f\left(\hat{Y}^+_{(n)}, \hat{S}^+_{(n)}, \hat{H}_{(n)}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[f\left(\hat{Y}^+_{(n)}, \hat{S}^+_{(n)}, \hat{H}_{(n)}\right)\middle| \hat{Y}^+_{(n)}\right]\right] \\
&= \mathbb{E}\left[\Phi(n, \lfloor n^{2/3}t \rfloor)\mathbb{E}\left[f\left(Y^+_{(n)}, S^+_{(n)}, H^+_{(n)}\right)\middle| Y^+_{(n)}\right]\right] \\
&= \mathbb{E}\left[\Phi(n, \lfloor n^{2/3}t \rfloor) f\left(Y^+_{(n)}, S^+_{(n)}, H^+_{(n)}\right)\right],
\end{aligned}$$

where we use that $\mathbb{E}\left[f\left(\hat{Y}^+_{(n)}, \hat{S}^+_{(n)}, \hat{H}_{(n)}\right)\middle| \hat{Y}^+_{(n)}\right]$ is a bounded, adapted function of $\hat{Y}^+_{(n)}$, and that $\Phi(n, \lfloor n^{2/3}t \rfloor)$ is the measure change from $Y^+_{(n)}$ to $\hat{Y}^+_{(n)}$. Then, using Proposition XXX, Proposition XXX [Refer to results Zheneng of existence of the measure change and convergence of the measure change] and Proposition 3.10, following the proof of Theorem 4.1 in [?] gives us that

$$\begin{aligned}
&\mathbb{E}\left[f\left(\hat{Y}^+_{(n)}, \hat{S}^+_{(n)}, \hat{H}_{(n)}\right)\right] \\
&\to \mathbb{E}\left[\Phi(t) f\left(\sigma_+ B_t, \sigma_+ B_t^+, \frac{2}{\sigma_+} R_t^+, 0 \leq t \leq T\right)\right].
\end{aligned}$$

Since

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

Lemma 3.5 implies the convergence under rescaling of $(\hat{S}_n^+(k), \hat{H}_n(k), k \geq 0)$. By Proposition 3.10 , S_n^- converges in distribution to a deterministic process under scaling, which will not be effected by the measure change. This completes the proof. □

3.5 Convergence of the out-forest holds conditional on the multigraph being simple

We will now show that the parts of the multigraph we observe up until the timescale in which we are interested are with high probability simple. We will then use an argument by Joseph [?] to show that this implies that Theorem 3.1 holds conditional on the resulting multigraph being simple. We let $B_n(k)$ 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 k^{th} vertex of $(\hat{F}_n(k), k \geq 1)$. We call these anomalous edges.

Proposition 3.12. *Suppose $\beta < 1$. Then we have*

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

as $n \to \infty$.

Remark 3.13. *We adapt the proof of Lemma 7.1 of [?] and of Proposition 5.3 of [?] to the directed setting. An extra complication is caused by the conditioning on*

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

We remark that in both papers, the proof of the aforementioned result is not fully correct, because the authors use a wrong expression for the probability of sampling an anomalous edge. However, the argument below can be adapted to the setting of [?] and [?] to yield a correct proof.

Proof. Note that we can only show convergence of the Radon-Nikodym derivative up to time $O(n^{2/3})$, so it is not straightforward to use the measure change to proof results on a time scale $O(n^\beta)$. Therefore, for the proof of this lemma, we will use a different method, that was introduced by Joseph in [?] referred to as *Poissonization*. We note that $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,n})$ (before conditioning on $\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$) are distributed as the jumps ordered by jump time in a Poisson process Π^0 with intensity measure π^0 on $\mathbb{R}_+ \times \mathbb{N}^2$ such that

$$\pi^0(dt, k_1, k_2) = n\mathbb{P}(D^- = k_1, D^+ = k_2)k_1 \exp(-k_1 t)dt$$

conditioned on $\Pi^0(\mathbb{R}, \mathbb{N}, \mathbb{N}) = n$. The intensity of this process is not constant in t, so we perform a time change. Define

$$\mathcal{L}_{\mathbf{D}}(x, y) = \mathbb{E}\left[\exp(-xD^- - yD^+)\right],$$

and set

$$\psi(t) = (1 - \mathcal{L}(\cdot, 0))^{-1},$$

such that for

$$\pi_n(dt, k_1, k_2) := \mathbb{P}(D^- = k_1, D^+ = k_2)k_1 \exp\left(-k_1\psi(t/n)\right)\psi'(t/n)dt$$

on $(0, n) \times \mathbb{N}^2$, we see that for $t \in (0, n)$, there exists a probability measure P_t on $\mathbb{N}^2$ such that

$$\pi_n(dt, k_1, k_2) = P_t(D^+ = k_1, D^+ = k_2)dt.$$

This is a trivial adaptation of Lemma 4.1 of [?]. Let Π_n be a decorated point process with intensity π_n. Now, let $\hat{\Pi}_n$ be a random measure, which is a decorated point process with intensity π_n, conditioned on

1. $N_n := \hat{\Pi}_n((0,n), \mathbb{N}, \mathbb{N}) = n$, and
2. $\Delta_n := \int_{(0,n)\times\mathbb{N}^2} (k_1 - k_2)\hat{\Pi}_n(dt, k_1, k_2) = 0$.

Then, the points of $\hat{\Pi}_n$ ordered by time are distributed as $(\hat{\mathbf{D}}_{n,1}, \dots, \hat{\mathbf{D}}_{n,n})$ conditioned on $\sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$. Let $\hat{\pi}_t^n$ be the marginal density of $\hat{\Pi}^n$ at time t, such that there is a probability measure $\hat{P}_t^n$ on $\mathbb{N}^2$ and a measure λ_t^n on $\mathbb{R}_+$ such that

$$\hat{\pi}_t^n(dt, k_1, k_2) = \lambda_t^n(dt) \times \hat{P}_t^n(D^- = k_1, D^+ = k^2).$$

Note that due to the conditioning, $\lambda_t^n(dt)$ can be unequal to dt. However, we claim that, also after conditioning, with high probability, we will have seen between n^β and $3n^\beta$ jumps at time $2n^\beta$. Indeed,

$$\begin{aligned}\mathbb{P}\left(\Pi_n\left((0,2n^\beta), \mathbb{N}, \mathbb{N}\right) \notin (n^\beta, 3n^\beta)\middle| \Delta_n = 0, N_n = n\right) &\leq \frac{\mathbb{P}\left(\sum_{i=1}^{n^\beta} E_i > 2n^\beta \text{ or } \sum_{i=1}^{3n^\beta} E_i < 2n^\beta\right)}{\mathbb{P}(\Delta_n = 0, N_n = n)} \\ &= O(n^{1/2}\exp(-n))\end{aligned} \tag{4}$$

for $(E_1, E_1, \dots)$ i.i.d. exponential random variables with rate 1, where the order follows from Cramér's Theorem and the local limit theorem.

We will use this set-up to show that with high probability, we do not sample anomalous edges in the first n^β time steps of the eDFS. We distinguish between the following types of anomalous edges.

Self-loops occur when the out-half-edge of a vertex is paired to an in-half-edge of the same vertex. Let $B_n^1(k)$ be the number of self-loops that are found up to time k. For v explored up to time $\lfloor n^\beta \rfloor$, a vertex with in-degree d_v^- and out-degree d_v^+, there are $d_v^- d_v^+$ possible combinations of an in-half-edge and an out-half-edge that form a self-loop connected to v. Any of these combinations of half-edges is paired with probability bounded above by

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

Parallel edges occur when an out-half-edge of a vertex is paired to an in-half-edge of one of its previously explored children. Let $B_n^2(k)$ be the number of parallel edges that are found up to time k. For any vertex v with in-degree d_v^-, and a parent $p(v)$ with out-degree $d_{p(v)}^+$, there are at most $d_v^- d_{p(v)}^+$ possible combinations of an in-half-edge and an out-half-edge that form a parallel edge from $p(v)$ to v. Again, any of these combinations of half-edges is paired with probability bounded above by

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

The last type of anomalous edges is a surplus edge with multiplicity greater than 1. Let $B_n^3(k)$ be the number of surplus edges with multiplicity greater than 1 that are found up to time k. For a vertex w with out-degree d_w^+ and a vertex v with in-degree d_v^-, a multiple surplus edge from w to v can only occur if v is discovered before w. In that case, there are at most $(d_w^+)^2(d_v^-)^2$ possible pairs of combinations of half-edges, and each of these pairs appears with probability bounded above by

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

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

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

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

$$\mathbb{P}\left(B_n^1(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) \leq \frac{\sum_{i=1}^{\lfloor n^\beta \rfloor} \hat{D}_i^- \hat{D}_i^+}{\sum_{i=\lfloor n^\beta \rfloor + 1}^n \hat{D}_i^-} \wedge 1,$$

$$\mathbb{P}\left(B_n^2(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) \leq \frac{\sum_{i=1}^{\lfloor n^\beta \rfloor} \hat{D}_i^- \mathbb{E}\left[\hat{D}_{p(i)}^+ \middle| \mathcal{G}^n\right]}{\sum_{i=\lfloor n^\beta \rfloor + 1}^n \hat{D}_i^-} \wedge 1,$$

$$\mathbb{P}\left(B_n^3(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) \leq \frac{\sum_{i=1}^{\lfloor n^\beta \rfloor} \sum_{j<i} (\hat{D}_i^+)^2 (\hat{D}_j^-)^2}{\left(\sum_{i=\lfloor n^\beta \rfloor + 1}^n \hat{D}_i^-\right)^2} \wedge 1,$$

where we note that $p(i)$ is not adapted to $\mathcal{G}^n$, because ancestral relations in the tree also depend on surplus edges. However, we observe that by the Cauchy-Schwarz inequality,

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

where the last inequality follows from the conditional Jensen inequality and the fact that a vertex with out-degree d^+ that is discovered before time n^β is the parent of at most d^+ vertices that are discovered before time n^β.

We will show that

$$\mathbb{P}\left(B_n^1(\lfloor n^\beta \rfloor) + B_n^2(\lfloor n^\beta \rfloor) + B_n^3(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) \xrightarrow{p} 0 \tag{5}$$

as $n \to \infty$. Then, the proposition follows from the bounded convergence theorem. By the observations above, and the fact that

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

it is sufficient to show that as $n \to \infty$,

$$\frac{1}{n} \int_{(0, 2n^\beta) \times \mathbb{N}^2} k_1 k_2 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \tag{6}$$

$$\frac{1}{n} \int_{(0, 2n^\beta) \times \mathbb{N}^2} k_1 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \tag{7}$$

$$\frac{1}{n} \int_{(0, 2n^\beta) \times \mathbb{N}^2} k_1^2 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \tag{8}$$

$$\frac{1}{n} \int_{(0, 2n^\beta) \times \mathbb{N}^2} k_2^2 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \text{ and} \tag{9}$$

$$\frac{1}{n} \int_{(0, 2n^\beta) \times \mathbb{N}^2} k_2^3 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0. \tag{10}$$

We will show 6. The proof of the other equations is analogous.
We start with the proof of 6. Note that by (4), for $\hat{E}_t^n$ the expectation under $\hat{P}_t^n$, it is sufficient if we show that for some C,

$$\hat{E}_t^n\left[\hat{D}_t^- \hat{D}_t^+\right] < C \tag{11}$$

for all n and $t < 2n^\beta$. We note that

$$\hat{E}_t^n\left[\hat{D}_t^- \hat{D}_t^+\right] = E_t^n\left[\hat{D}^- \hat{D}^+ | \Delta_n = 0, N_n = n\right] = E_t^n\left[\hat{D}^- \hat{D}^+ \frac{\mathbb{P}\left[\Delta_n = 0, N_n = n | \hat{D}_t^-, \hat{D}_t^+\right]}{\mathbb{P}\left[\Delta_n = 0, N_n = n\right]}\right].$$

By the fact that Π_n is a decorated point process, we see that for k_1, k_2 in $\mathbb{N}$,

$$\mathbb{P}\left[\Delta_n = 0, N_n = n | \, \hat{D}_t^- = k_1, \hat{D}_t^+ = k_2\right] = \mathbb{P}\left[\Delta_n = k_2 - k_1, N_n = n-1\right],$$

such that, since $N_n \sim \text{Poisson}(n)$, and since on $N_n = n-1$ (resp. $N_n = n$), Δ_n is the sum of $n-1$ (resp. n) i.i.d. mean 0 random variables with finite variance, there exists a C' such that

$$\frac{\mathbb{P}\left[\Delta_n = 0, N_n = n | \hat{D}_t^- = k_1, \hat{D}_t^+ = k_2\right]}{\mathbb{P}\left[\Delta_n = 0, N_n = n\right]} < C'$$

for all k_1, k_2, t and n. Therefore, if we show that for some C''

$$E_t^n\left[\hat{D}^- \hat{D}^+\right] < C''$$

for all n and $t < 2n^\beta$, (11) follows. We note that by definition of $\pi_n(dt, k_1, k_1)$,

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

By definition of $\mathcal{L}_{\mathbf{D}}(x,y)$ and $\psi(s)$, we find that

$$\begin{aligned}
\frac{d^3}{dx^2 dy}\mathcal{L}_{\mathbf{D}}(x,y)_{(s,0)} &= -\mathbb{E}[(D^-)^2 D^+] + o(1), \\
\frac{d}{dx}\mathcal{L}_{\mathbf{D}}(x,y)_{(s,0)} &= -\mathbb{E}[D^-] + o(1), \text{ and} \\
\psi(s) &= \frac{s}{\mu} + o(s)
\end{aligned}$$

as $s \to 0$. We refer the reader to the proof of Lemma A.1 in [?] for the details of a similar argument in the undirected setting. This implies that

$$E_t^n\left[\hat{D}^- \hat{D}^+\right] = \frac{\mathbb{E}[(D^-)^2 D^+]}{\mathbb{E}[D^-]} + o(1)$$

uniformly in all $t \leq 2n^\beta$, and (11) follows.
By applying the same techniques, (7), (8), (9) and (10) follow as well, which proves the statement.

□

Corollary 3.14. *Theorem 3.1 holds conditional on the resulting directed multigraph being simple.*

Proof. Let $\rho(n) = \inf\{k \geq 1 : B_n(k) > 0\}$, and note that the event that the multigraph formed by the configuration model on n vertices is simple is equal to $\{\rho(n) = \infty\}$. Proposition 3.12 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 conditionally on $\{\rho(n) = \infty\}$. This follows from the proof of Theorem 3.2 in [**?**]. $\square$

All results that follow are obtained by studying the exploration up to time $O(n^{2/3})$, so will also be true conditioned on the resulting directed multigraph being simple.

4 Convergence of the strongly connected components under rescaling

In this section, we will use the convergence of the out-forest that we obtained in Section 3 to show that the strongly connected components ordered by decreasing length converge under rescaling in the $d_{\vec{\mathcal{G}}}$-product topology. The structure of the argument is as follows.

4.1 Convergence of the out-components that contain an ancestral surplus edge

In this subsection, we will prove that the components of $\hat{F}_n\left(\lfloor n^{2/3}t \rfloor\right)$ that contain an ancestral surplus edge converge under rescaling. Recall the definition of $(A_n(k), k \geq 1)$ from Subsection 2.1.3, and recall that the components in $(\hat{F}_n(k), k \geq 1)$ that contain a non-trivial strongly connected component correspond to the components in $(\hat{F}_n(k), k \geq 1)$ 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 first ancestral surplus edge in the component.
We first study the convergence of $(\hat{H}^{\ell}_n(k), k \geq 1)$ under rescaling. This is an extension of Theorem 3.1.

Lemma 4.1. *Let $(B_t, t \geq 0)$ be a Brownian motion, and define*

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

and

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

Then,

$$\begin{aligned}&\left(n^{-1/3}\hat{S}^+_n\left(\lfloor n^{2/3}t \rfloor\right), n^{-1/3}\hat{H}_n\left(\lfloor n^{2/3}t \rfloor\right), n^{-1/3}\hat{H}^{\ell}_n\left(\lfloor n^{2/3}t \rfloor\right), t \geq 0\right)\\ &\xrightarrow{d} \left(\sigma_+\hat{B}_t, \frac{2}{\sigma_+}\hat{R}_t, \frac{2(\sigma_{-+} + \nu_-)}{\sigma_+\mu}\hat{R}_t, t \geq 0\right)\end{aligned}$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^3$, jointly with

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

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

Proof. We use Theorem 1 in [?] by de Raphélis, that shows convergence of the height process of a Galton-Watson forest with edge-lengths under a few conditions on the degree and edge length distribution. We will apply this result to the black-purple-red Galton-Watson forest $(F^{pr}(k), k \geq 1)$, as defined in Subsubsection 3.3.2.
We equip $(F^{pr}(k), k \geq 1)$ with edge lengths in the following manner. For a purple or red vertex with out degree d^+, sample its in-degree with law Z^- conditioned on $Z^+ = d^+$. The in-degree of the black 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 to its children will have length d^-). Call the resulting forest with edge lengths $(F^{pr,\ell}(k), k \geq 1)$, and let $(H^{pr,\ell}(k), k \geq 1)$ be the corresponding height process.
We will translate the conditions of Theorem 1 in [?] 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 Subsubsection 3.3.2,

$$\begin{aligned}&\left(n^{-1/3} Y^{pr}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{pr}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{pr,\ell}\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right)\\ &\xrightarrow{d} \left(\sigma_+ B_s, \frac{2}{\sigma_+} R_s, \frac{2(\sigma_{+-} + \nu_-)}{\mu \sigma_+} R_s, t \geq 0\right)\end{aligned} \tag{12}$$

in $D(\mathbb{R}_+, \mathbb{R})^3$ as $n \to \infty$. Then, we observe that the rest of the argument in Subsubsection 3.3.2 and Subsection 3.4 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 [?] 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 condition follow directly from the assumptions, and the third condition is implied by $\mathbb{E}[D^+(D^-)^3] < \infty$. □

Proposition 4.2. *We have that, jointly with the convergence in Lemma 4.1,*

$$\left(A_n\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right) \xrightarrow{d} (A_t, t \geq 0),$$

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 $D(\mathbb{R}_+, \mathbb{R})$.

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 Theorem 14.2.VIII of Daley and Vere-Jones [?], the claimed convergence under rescaling of $(A_n(k), k \geq 1)$ follows if we show that

$$\left(A_n^{comp}\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right) \xrightarrow{d} \left(\frac{2(\sigma_{-+} + \nu_-)}{\sigma_+ \mu^2} \int_0^t \hat{R}_v dv, t \geq 0\right) \tag{13}$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in Lemma 4.1. Therefore, we will now prove that (13) holds. By

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

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$, we get that

$$\begin{aligned} \left(n^{-2/3} \min\{l \geq 1 : n^{-1/3} \hat{P}_n(l) \geq t\}, t \geq 0\right) &\xrightarrow{p} \left(\min\left\{s > 0 : \frac{\nu_-}{2\mu} s^2 \geq t\right\}, t \geq 0\right) \\ &=: \left(p^{-1}(t), t \geq 0\right) \end{aligned}$$

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, Lemma 4.1, Lemma A.2, Slutsky's lemma and the continuous mapping theorem imply that

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

If we combine this with the convergence under rescaling of $(P_n(k), k \geq 1)$ and apply Lemma A.2, some simple analysis then yields (13), which proves the statement. □

4.2 Finding the important components in the out-forest

In this subsection, we will show that, conditional on the convergence under rescaling in Proposition 4.2, the sequence of components in $(\hat{F}_n(k), k \leq \lfloor Tn^{2/3} \rfloor)$ that contain ancestral surplus edges converges as well under rescaling. Lemma 4.3 is a statement about extracting excursion intervals from deterministic functions with marks, which we will apply to the sample paths of $(\hat{S}_n^+(k), k \geq 1)$ and the increase times of $(A_n(k), k \geq 1)$. The lemma tells us that if the sample paths and increase times converge under rescaling, the beginnings and endpoints of the excursions above the running infimum that contain the increase times converge as well.

Lemma 4.3. *Let $(f_n(t), t \geq 0)$ for $n \geq 1$, and $(f(t), t \geq 0)$ be functions in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$, such that*

$$(f_n(t), t \geq 0) \to (f(t), t \geq 0)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$. Assume that $(f(t), t \geq 0)$ is continuous, that $f(t) \to -\infty$ as $t \to \infty$, 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 $\mathbb{R}^m$ such that for all $i \in [m]$, $x_i^n \to x_i$ in $\mathbb{R}$ as $n \to \infty$, and such that $f(x_i) - \inf\{f(s) : s \leq x_i\} > 0$ for all $i \in [m]$. Define

$$\begin{aligned} l_i^n &= \inf\{t \geq 0 : f_n(t) = \inf\{f_n(s) : s \leq x_i^n\}\} \text{ for } i \in [m],\ n \geq 1 \\ \sigma_i^n &= \inf\{t \geq 0 : \inf\{f_n(s) : s \leq l_i^n + t\} < \inf\{f_n(s) : s \leq x_i^n\}\} \text{ for } i \in [m],\ n \geq 1 \\ l_i &= \inf\{t \geq 0 : f(t) = \inf\{f(s) : s \leq x_i\}\},\ \text{and} \\ \sigma_i &= \inf\{t \geq 0 : \inf\{f(s) : s \leq l_i + t\} < \inf\{f(s) : s \leq x_i\}\}. \end{aligned}$$

For $S = \{(a_i, b_i), i \in [m]\}$, let $\operatorname{ord}(S)$ be a sequence consisting of the elements of S put in increasing order of a_i, with ties broken arbitrarily, and concatenated with $(0,0)_{i \geq 1}$ such that $\operatorname{ord}(S) \in (\mathbb{R}^2)^\infty$. Then,

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

in $(\mathbb{R}^2)^\infty$ in the ℓ_1-topology as $n \to \infty$.

Proof. First, note that l_i^n, σ_i^n, l_i, and σ_i are well-defined for all $i \in [m]$, $n \geq 1$ by $f(t) \to -\infty$ as $t \to \infty$ and convergence of f_n to f.
Fix i. We will first show that $l_i^n \to l_i$ 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, $l_i < x_i < l_i + \sigma_i$. Fix $0 < \epsilon < \min\{x_i - l_i, l_i + \sigma_i - x_i\}/2$. We claim that the following conditions are sufficient for $l_i^n \to l_i$ and $\sigma_i^n \to \sigma_i$ as $n \to \infty$. For all n large enough,

1. $l_i + \epsilon < x_i^n < l_i + \sigma_i - \epsilon$
2. $\inf\{f_n(s) : s \in (l_i - \epsilon, l_i + \epsilon)\} < \inf\{f_n(s) : s \in [l_i + \epsilon, l_i + \sigma_i - \epsilon]\}$,
3. $\inf\{f_n(s) : s \in (l_i - \epsilon, l_i + \epsilon)\} < \inf\{f_n(s) : s \in [0, l_i - \epsilon]\}$,
4. $\inf\{f_n(s) : s \in (l_i + \sigma_i - \epsilon, l_i + \sigma_i + \epsilon)\} < \inf\{f_n(s) : s \in [0, l_i + \sigma_i - \epsilon]\}$

Indeed, condition 1, 2 and 3 imply $|l_i^n - l_i| < \epsilon$, while condition 1, 2 and 4 imply $|(l_i^n + \sigma_i^n) - (l_i + \sigma_i)| < \epsilon$. Note that condition 1 holds for n large enough by definition of ϵ and convergence of x_i^n to x_i. To show the other conditions, define

$$\begin{aligned} \delta_1 &= \inf\{f(s) : s \in [l_i + \epsilon, l_i + \sigma_i - \epsilon]\} - \inf\{f(s) : s \in (l_i - \epsilon, l_i + \epsilon)\} \\ \delta_2 &= \inf\{f(s) : s \in [0, l_i - \epsilon]\} - \inf\{f(s) : s \in (l_i - \epsilon, l_i + \epsilon)\} \\ \delta_3 &= \inf\{f(s) : s \in [0, l_i + \sigma_i - \epsilon]\} - \inf\{f(s) : s \in (l_i + \sigma_i - \epsilon, l_i + \sigma_i + \epsilon)\} \end{aligned}$$

By uniqueness of local minima and the definition of l_i and σ_i, $\delta := \min\{\delta_1, \delta_2, \delta_3\}/3 > 0$. Then, note that for n large enough, $\sup\{|f_n(s) - f(s)| : s \leq l_i + \epsilon\} < \delta$, which implies conditions 2, 3, and 4 for such n.
Since i was arbitrary, and m is finite, we find that

$$(l_i^n, \sigma_i^n)_{1 \leq i \leq m} \to (l_i, \sigma_i)_{1 \leq i \leq m}$$

in $\mathbb{R}^{2m}$ as $n \to \infty$.
We now claim that $l_i^n \to l_i$ and $l_j^n \to l_i$ implies that $l_i^n = l_j^n$ for n large enough. Indeed, by definition of l_i^n, l_j^n and σ_i^n, we see that $l_i^n < l_j^n$ implies that $l_j^n - l_i^n \geq \sigma_i^n$, and by the argument above, $\sigma_i^n \to \sigma_i > 0$, so $l_i^n - l_j^n \to 0$ can only hold if $l_i^n = l_j^n$ for n large enough. This implies that

$$\#\left\{(l_i^n, \sigma_i^n) : 1 \leq i \leq m\right\} \to \#\left\{(l_i, \sigma_i) : 1 \leq i \leq m\right\}.$$

Then, the result follows. $\square$

We now apply this result to our process to extract the excursion intervals that contain the marks representing ancestral backedges.

Proposition 4.4. *Fix $T > 0$. Use notation as before. For $i \in \left[A_n\left(\lfloor Tn^{2/3} \rfloor\right)\right]$, set $X_i^n = \min\{k : A_n(k) = i\}$. Similarly, for i in $[A_T]$, set $X_i = \min\{t : A_T = i\}$. Define, for $n \geq 1$,*

$$L_i^n = \min\left\{k \geq 1 : \hat{S}_n^{p,+}(k) = \min\{\hat{S}_n^{p,+}(l) : l \leq X_i^n\}\right\} \text{ for } i \in \left[A_n\left(\lfloor Tn^{2/3} \rfloor\right)\right]$$
$$\Sigma_i^n = \min\left\{k \geq 1 : \min\left\{\hat{S}_n^{p,+}(l) : l \leq L_i^n + k\right\} < \min\left\{\hat{S}_n^{p,+}(l) : l \leq X_i^n\right\}\right\} \text{ for } i \in \left[A_n\left(\lfloor Tn^{2/3} \rfloor\right)\right]$$
$$L_i = \inf\left\{t \geq 0 : \sigma_+ \hat{B}_t = \inf\{\sigma_+ \hat{B}_s : s \leq X_i\}\right\} \text{ for } i \in [A(T)] \text{ and}$$
$$\Sigma_i = \inf\left\{t \geq 0 : \inf\{\sigma_+ \hat{B}_s : s \leq L_i + t\} < \inf\{\sigma_+ \hat{B}_s : s \leq X_i\}\right\} \text{ for } i \in [A(T)].$$

Then, for ord *defined as in the statement of Lemma 4.3, we get that*

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

in the ℓ_1-topology on $(\mathbb{R}^2)^\infty$ as $n \to \infty$, jointly with the convergence in Proposition 4.2.

Proof. We work on a probability space where the convergence in Proposition 4.2 holds almost surely, and claim that we can apply Lemma 4.3 to the sample paths of $\left(n^{-1/3} \hat{S}_n^p\left(\lfloor n^{2/3} t \rfloor\right), t \geq 0\right)$ with marks

$$\left(n^{-2/3} X_n^i\right)_{1 \leq i \leq A_n\left(\lfloor Tn^{2/3} \rfloor\right)}.$$

We check the conditions. Firstly, note that by $A_n\left(\lfloor Tn^{2/3} \rfloor\right) \to A(T)$ almost surely as $n \to \infty$, we can pick n large enough such that $A_n\left(\lfloor Tn^{2/3} \rfloor\right) = A(T)$, where we ignore events of 0 probability. Furthermore, we observe that $(\hat{B}_t, t \geq 0)$ is a Brownian motion with negative parabolic drift, so the sample paths of $(\sigma_+ \hat{B}_t, t \geq 0)$ are continuous and drift to $-\infty$ almost surely. By the local absolute continuity of $(\hat{B}_t, t \geq 0)$ to a Brownian motion, its local minima are almost surely unique. By

$$\left(A_n\left(\lfloor tn^{2/3} \rfloor\right), t \leq T\right) \xrightarrow{a.s.} (A(t), t \geq 0)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$, we observe that for all $i \in [A_T]$, $n^{-2/3} X_i^n \to X_i$ almost surely in $\mathbb{R}$ as $n \to \infty$. The fact that $\hat{R}_{X_i} > 0$ for all i almost surely follows from the intensity of $(A_t, t \geq 0)$ at time t being proportional to $\hat{R}_t$. This allows us to apply Lemma 4.3, and the convergence follows. $\square$

4.3 Convergence of the set of candidates

By Proposition 4.4, 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 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 Subsubsection 2.6.

The following proposition shows convergence under rescaling of the sequence of tails of the candidates on a particular component of $(\hat{F}_n(k), k \geq 1)$.

Proposition 4.5. *Fix $T > 0$. We work on a probability space where the convergence in Propositions 4.2 and 4.4 holds almost surely. Let $(l, \sigma) \in \{(L_i, \Sigma_i) : i \leq A_T\}$, such that, for each n large enough, we can find a $(l_n, \sigma_n) \in \left\{(L_i^n, \Sigma_i^n) : i \leq A_n\left(\lfloor Tn^{2/3} \rfloor\right)\right\}$, such that $(l_n, \sigma_n) \to (l, \sigma)$. Set $V_1 = \inf\{t \in [l, l+\sigma] : A(t) = A(l) + 1\}$, and similary, set $V_1^n = \min\{l_n < k \leq l_n + \sigma_n : A_n(k) = A_n(l_n) + 1\}$, which are well-defined by definition of l, σ, l_n and σ_n. Then, by construction, $\{l_n + 1, \ldots, l_n + \sigma_n\}$ encodes a component of $(\hat{F}_n(k), k \geq 1)$. Call this component $T_{l_n}^n$. We apply the procedure defined in 2.6 to find the tail of candidates in $T_{l_n}^n$. Let $\mathbf{C}_n(l_n)$ denote the sequence of tails of candidates in $T_{l_n}^n$. Similarly, $[l, l+\sigma]$ encodes a component of $(\hat{\mathcal{F}}(t), t \geq 0)$. Call this component $\mathcal{T}_l$, and apply procedure in Subsubsection 2.2.2 to find the tails of candidates in $\mathcal{T}_l$, and denote its sequence of tails of candidates by $\mathbf{V}(l)$. Then, jointly with the convergence in Proposition 4.4,*

$$n^{-2/3}\mathbf{V}_n(l_n) \xrightarrow{d} \mathbf{V}(l)$$

in the ℓ_1 topology.

Proof. We will find a coupling such that $n^{-2/3}\mathbf{V}_n(l_n) \xrightarrow{a.s.} \mathbf{V}(l)$. By the convergence in Propositions 4.2 and 4.4, $n^{-2/3}V_1^n \xrightarrow{a.s.} V_1$. In general, let V_m^n denote the m^{th} candidate that is found in $T_{l_n}^n$, and let V_m denote the m^{th} candidate that is found in $\mathcal{T}_l$. Suppose that, for some m, we have found a coupling such that

$$n^{-2/3}(V_1^n, \ldots, V_m^n) \xrightarrow{a.s.} (V_1, \ldots, V_m). \tag{14}$$

Then, V_{m+1}^n is distributed as the position of the first jump of a counting process $K_{m+1}^n(k)$ on $[0, \infty)$ with compensator

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

for $k \in [V_m^n + 1, l_n + \sigma_n]$ and 0 otherwise, where $T_i^{n,\mathrm{mk}}$ is the subtree of $T_{l_n}^n$ spanned by $\{l_n + 1, V_1^n, \ldots, V_m^n, i\}$. Moreover, 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 $t \in [V_m, L + \Sigma]$ and 0 otherwise, where T_s is the subtree of $\mathcal{T}_l$ spanned by $\{l, V_1, \ldots, V_m, s\}$, and $|T_s|$ is its length as encoded by $\left(\frac{2}{\sigma_+}\hat{R}_t, t \geq 0\right)$. By the convergence under rescaling of

$(\hat{H}^\ell(k), k \geq 1)$ in Lemma 4.1, and by Proposition 4.4, we get that the metric structure of $T^n_{l_n}$ with distances defined by $(\hat{H}^\ell(k), k \geq 1)$, and its projection onto $[n^{-2/3}(l_n+1), n^{-2/3}(l_n+\sigma_n)]$, converge under rescaling to the metric structure of $\mathcal{T}_l$ with distances defined by

$$\left(\frac{2(\sigma_{-+}+\nu_-)}{\sigma_+\mu}\hat{R}_t, t \geq 0\right)$$

and its projection onto $[l, \sigma]$. This, combined with (14) implies that

$$\left(n^{-1/3}\ell\left(T^{n,\mathrm{mk}}_{\lfloor tn^{2/3}\rfloor}\right), V_m \leq t \leq l+\sigma\right) \xrightarrow{a.s.} \left(\frac{\sigma_{-+}+\nu_-}{\mu^2}|T^{\mathrm{mk}}_t|, V_m \leq t \leq l+\sigma\right)$$

in $\mathbb{D}([V_m, l+\sigma], \mathbb{R}_+)$ as $n \to \infty$. Then, a similar argument as used in the proof of Proposition 4.2 implies that

$$\left(K^n_{comp,m+1}\left(\lfloor tn^{2/3}\rfloor\right), V_m \leq t \leq l+\sigma\right) \xrightarrow{a.s.} \left(K_{comp,m+1}(t), V_m \leq t \leq l+\sigma\right),$$

$\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ as $n \to \infty$, which implies that

$$(K^n_{m+1}(\lfloor tn^{2/3}\rfloor), t \geq 0) \xrightarrow{d} (K_{m+1}(t), t \geq 0)$$

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 only if $K^n_m(\infty) > 0$ for all n large enough, and such that on this event,

$$n^{-2/3}V^n_{m+1} \xrightarrow{a.s.} V_{m+1}.$$

If $K_m(\infty) = 0$, set $\mathbf{V}(l) = (V_1, \ldots, V_m)$, $\mathbf{V}_n(l_n) = (V^n_1, \ldots, V^n_m)$, and the statement follows. If $K_m(\infty) > 0$, apply the induction step to $(V_1, \ldots, V_{m+1})$ and $(V^n_1, \ldots, V^n_{m+1})$. The fact that $|\mathbf{V}(l)| < \infty$ almost surely, as shown in Subsubsection 2.2.2, implies that the induction terminates. □

The following proposition shows that the law of the heads of the candidates converges as well under rescaling, and that the convergence holds in the pointed Gromov-Hausdorff topology.

Proposition 4.6. *Suppose the convergence in Propositions 4.2, 4.4 and 4.5 holds almost surely. Then, for $\mathbf{V}_n(l_n) = (V^n_1, \ldots, V^n_{N_n})$, $\mathbf{V}(l) = (V_1, \ldots, V_N)$, let W^n_i be the index of the vertex that the surplus edge corresponding to V^n_i connects to. Similarly, let W_i be the index of the vertex that the surplus edge corresponding to V_i connects to. Then,*

$$\begin{aligned}&\left(n^{-1/3}T^n_{l_n}, n^{-2/3}(l_n+1), \left(n^{-2/3}V^n_1, n^{-2/3}W^n_1\right)\ldots, \left(n^{-2/3}V^n_{N_n}, n^{-2/3}W^n_{N_n}\right)\right)\\&\xrightarrow{d} (\mathcal{T}_l, l, (V_1, W_1), \ldots, (V_N, W_N))\end{aligned}$$

in the $2M+1$-pointed Gromov-Hausdorff topology.

Proof. For S a subset of the vertices of $T^n_{l_n}$, let $T^n_{l_n}(S)$ denote the subtree of $T^n_{l_n}$ spanned by S. By definition, for $m \leq N_n$, W^n_m is the vertex corresponding to a uniform unpaired in-half-edge of the vertices in $T^n_{l_n}(\{l_n+1, V^n_1, \ldots, V^n_m\})$. By

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

the law of W_m^n converges to the law of a uniform vertex in $T_{l_n}^n(\{l_n+1, V_1^n, \dots, V_m^n\})$. Note that, by Theorem 3.1 and Propositions 4.4, 4.5, we know that

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

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

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

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 $\mathcal{T}_l$ spanned by $(l, V_1, \dots, V_m)$, which is equal to the law of W_m, which proves the statement.

□

The proof of Propositions 4.5 and 4.6 implies the following corollary.

Corollary 4.7. *We can work on a probability space where the convergence in Propositions 4.5 and 4.6 holds almost surely. Let $T^{n,mk}(l_n)$ be the subtree of $T_{l_n}^n$ spanned by $\{l_n+1, V_1^n, \dots, V_{N_n}^n\}$, and similarly, let $T^{mk}(l)$ be the subtree of $\mathcal{T}_l$ spanned by $\{l, V_1, \dots, V_N\}$. Then, also*

$$\begin{aligned}&\left(n^{-1/3}T^{n,mk}(l_n), n^{-2/3}(l_n+1), \left(n^{-2/3}V_1^n, n^{-2/3}W_1^n\right) \dots, \left(n^{-2/3}V_{N_n}^n, n^{-2/3}W_{N_n}^n\right)\right)\\ &\to \left(T^{mk}(l), l, (V_1, W_1), \dots, (V_N, W_N)\right)\end{aligned}$$

almost surely in the $2M+1$-pointed Gromov-Hausdorff topology as $n \to \infty$. Also the total length in the trees converges, i.e.

$$n^{-1/3}\left|T^{n,mk}(l_n)\right| \to \left|T^{mk}(l)\right|$$

almost surely as $n \to \infty$.

We now identify the vertices that are part of a candidate as described in Subsubsection 2.1.3. In $T^{n,\mathrm{mk}}(l_n)$, set $V_i^n \sim W_i^n$ for each $1 \leq i \leq N_n$, and set $M_{l_n}^n := T^{n,\mathrm{mk}}(l_n)/\sim$. Moreover, in $T^{\mathrm{mk}}(l)$, set $V_i \sim W_i$ for each $1 \leq i \leq N$, and set $\mathcal{M}_l := T^{\mathrm{mk}}(l)/\sim$. View both as elements of $\vec{\mathcal{G}}$ in the natural way. To be precise, in $M_{l_n}^n$, let the vertex set consist of l_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 $\mathcal{M}_l$, let the vertex set consist of l, W_i for $i \leq N$, and the branch points $V_i \wedge V_j$ for $i \neq j \leq N$. Then, the following proposition follows.

Proposition 4.8. *On the probability space where the convergence in Propositions 4.5 and 4.6 holds almost surely, $n^{-1/3}M_{l_n}^n \overset{a.s.}{\to} \mathcal{M}_l$ in $\vec{\mathcal{G}}$.*

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

Corollary 4.9. *On the probability space where the convergence in Propositions 4.5 and 4.6 holds almost surely, the strongly connected components in $n^{-1/3}M_{l_n}^n$, listed in decreasing order of length, converge to the strongly connected components in $\mathcal{M}_l$, listed in decreasing order of length, in $\vec{\mathcal{G}}$ almost surely as $n \to \infty$.*

Proof. This follows from Proposition 5.3 in [?]. This proposition requires that the lengths of the strongly connected components in $\mathcal{M}_l$ have different lengths almost surely, which is the content of Corollary 2.10. □

Corollary 4.10. *Let $T > 0$, and let $(C_i^T(n), i \geq 1)$ be the strongly connected components in $(G_n(k), k \geq 1)$ with a candidate with label at most $\lfloor Tn^{2/3} \rfloor$ ordered by length. Similarly, let $(\mathcal{C}_i^T, i \geq 1)$ be the strongly connected components in $\vec{G}(n, (D^-, D^+))$ with a candidate with label at most T ordered by length. Then,*

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

in the $\vec{\mathcal{G}}$ product topology as $n \to \infty$.

Proof. This follows from Proposition 4.4, Corollary 4.9, and the fact that all SCCs in the limit object have a different length by Corollary 2.10. □

Finally, we claim that we can choose T large enough such that all SCCs with large size are explored before time T. This is the content of the following lemma. The proof is in the same spirit as Lemma 9 in [**?**] by Aldous.

Lemma 4.11. *For $\delta > 0$ and I an interval, let $SCC(n, I, \delta)$ denote the number of SCCs whose vertices have at total of at least $\delta n^{1/3}$ in-edges (including those which are not part of the SCC) and whose time of first discovery is in $n^{2/3} I$. Then,*

$$\lim_{s \to \infty} \limsup_n \mathbb{P}\left(SCC(n, (s, \infty), \delta) \geq 1\right) = 0 \text{ for all } \delta > 0.$$

Proof. Fix $\delta > 0$. Suppose there is a strongly connected component C with $vn^{1/3}$ total in-edges. Conditional on this fact, the in-edges that are paired up to the first in-edge of C is paired, are uniform picks (without replacement) from the total set of in-edges. Denote the time of discovery of the first in-edge of C times $n^{-2/3}$ by Ξ_n. Then, $\Xi_n \xrightarrow{d} \mathrm{Exp}(v)$. Fix $\epsilon > 0$. We see 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, such 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 4.6. Then, note that it is clear from the description of the limit process that, for s large enough, with probability at most $\epsilon/2$, an SCC with total length at least $\frac{\mu}{\sigma_{-+} + \nu_-}\delta$ is discovered after time s. By 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 size of a SCC is bounded from below by the total number of in-edges of vertices in the SCC. Then, Theorem 1.1 follows from Corollary 4.10 and Lemma 4.11.

A Proof of Proposition 3.11

Recall the notation from Subsubsection 3.3.2. We will show that there exists a process $(D_t, t \geq 0)$ such that

$$\begin{aligned}&\left(n^{-1/3}\left[Y^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right)-E\left(\lfloor n^{2/3}t\rfloor\right)\right], n^{-1/3}H^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t\geq 0\right)\\&\xrightarrow{d}\left(\sigma_+D_t, \frac{2}{\sigma_+}\left(D_t-\inf\{D_s, s\leq t\}\right), t\geq 0\right)\end{aligned}$$

in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})^2$ as $n\to\infty$ and $\left(\frac{2}{\sigma_+}\left(D_t-\inf\{D_s, s\leq t\}\right), t\geq 0\right)$ is the height process corresponding to $(\sigma_+D_t, t\geq 0)$.
The next lemma show that the pathwise construction of $(Y^{pr}(k), H^{pr}(k), k\geq 1)$ converges to the continuous counterpart of the pathwise construction.

Lemma A.1. *Let $(B_t, t\geq 0)$ and $(B^{red}_t, t\geq 0)$ be two independent Brownian motions and let*

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

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

$$(B^{pr}_t, t\geq 0) := \left(B_{\Lambda(t)} + B^{red}_{t-\Lambda(t)}, t\geq 0\right). \tag{15}$$

Then, for

$$(R^{pr}_t, t\geq 0) := (B^{pr}_t - \inf\{B^{pr}_s, s\leq t\}, t\geq 0),$$

$((2/\sigma_+)R^{pr}_t, t\geq 0)$ is the height process corresponding to $(\sigma_+B^{pr}_t, t\geq 0)$. Moreover,

$$\left(n^{-1/3}Y^{pr}\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}H^{pr}\left(\lfloor n^{2/3}t\rfloor\right), t\geq 0\right)\xrightarrow{d}\left(\sigma_+B^{pr}_t, \frac{2}{\sigma_+}R^{pr}_t, t\geq 0\right) \tag{16}$$

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

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

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

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

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

$$\left(n^{-1/3}Y^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), n^{-1/3}H^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t\geq 0\right)\xrightarrow{d}\left(\sigma_+B^{pr}_{\theta(t)}, \frac{2}{\sigma_+}R^{pr}_{\theta(t)}, t\geq 0\right) \tag{17}$$

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

In the proof of Lemma A.1 we use the following, straightforward, technical result.

Lemma A.2. *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 $n \to \infty$.

Proof. Using the characterization of convergence in the Skorokhod topology given in Proposition 3.6.5 in [?] by Ethier and Kurtz, the result follows immediately. □

Proof of Lemma A.1. Firstly, note that by $(Y^{pr}(k), k \geq 1)$ encoding a critical Galton-Watson forest with offspring variance σ_+^2, the proof of Theorem 1.8 in [?] gives us that for $(B_s^*, s \geq 0)$ a Brownian motion,

$$\left(n^{-1/3} Y^{pr}\left(\lfloor n^{2/3} s \rfloor\right), n^{-1/3} H^{pr}\left(\lfloor n^{2/3} s \rfloor\right), s \geq 0\right) \xrightarrow{d} \left(\sigma_+ B_s^*, \frac{2}{\sigma_+}\left(B_s^* - \inf\{B_u^* : u \leq s\}\right), s \geq 0\right) \tag{18}$$

in $D(\mathbb{R}_+, \mathbb{R})^2$ as $n \to \infty$, and $\left(\frac{2}{\sigma_+}(B_s^* - \inf\{B_u^*, u \leq s\}), s \geq 0\right)$ is the height process corresponding to $(\sigma_+ B_s^*, s \geq 0)$. Moreover, [?] and the fact that $(Y^+(k), k \geq 1) \stackrel{d}{=} (Y^{red}(k), k \geq 1) \stackrel{d}{=} (Y^{pr}(k), k \geq 1)$ imply that

$$\begin{aligned} &\left(n^{-1/3} Y^{red}\left(\lfloor n^{2/3} s \rfloor\right), n^{-2/3} \inf\left\{k : n^{-1/3} Y^{red}(k) \leq -x\right\}, s \geq 0, x \geq 0\right) \\ &\xrightarrow{d} \left(\sigma_+ B_s^{red}, \inf\left\{u : \sigma_+ B_u^{red} < -x\right\}, s \geq 0, x \geq 0\right) \end{aligned} \tag{19}$$

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

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

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$. Since $(P_n(k), k \geq 1)$ is non-decreasing, applying Lemma A.2, and combining the convergence in (19) with Lemma 3.8 gives that also

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

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in (19), and therefore,

$$\left(n^{-2/3} \theta_n\left(\lfloor n^{2/3} t \rfloor\right), t \geq 0\right) \xrightarrow{d} (\theta(t), t \geq 0) \tag{20}$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in (19). 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

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

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in (19) and (20). Since $\max\{j : \theta_n(j) \leq \lfloor n^{2/3} t \rfloor\}$ is of order $n^{2/3}$, Lemma 3.7 implies that also

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

in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ jointly with the convergence in (19) and (20).
To finish the proof, we examine the construction of $(Y^{pr}(k),k\geq 1)$ in (2) and the construction of $(B^{pr}_s,s\geq 0)$ in (15). Note that $\Lambda_n(k)$ and $k-\Lambda_n(k)$ are non-decreasing. Again, by Lemma A.2, this implies that

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

in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ jointly with all earlier mentioned converging random variables. Combining this with the convergence in (18) proves (16). The fact that $(\theta_n(k),k\geq 1)$ is non-decreasing and Lemma A.2 then imply (17). $\square$

Lemma A.3. *We have that*

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

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

Proof. By (3), and by Lemma A.1, it is sufficient to show that for any $T>0$,

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

We remind the reader that $E(k)$ counts the number of red children of k^{th} vertex in $(F^{pr}(k),k\geq 1)$, so

$$n^{-1/3}\max_{k\leq\lfloor n^{2/3}T\rfloor}E(k)\leq n^{-1/3}\max_{k\leq\theta_n(\lfloor n^{2/3}T\rfloor)}(Y^{red}(k)-Y^{red}(k-1)+1),$$

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

$$\left(n^{-1/3}Y^{red}\left(\lfloor n^{2/3}t\rfloor\right),t\geq 0\right)$$

converges in distribution to a continuous process in $D(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$. $\square$

The following lemma is the last ingredient in the proof of Proposition 3.11.

Lemma A.4. *We have that with probability* 1,

$$\left(\frac{2}{\sigma_+}\left(B^{pr}_{\theta(t)}-\inf\{B^{pr}_s:s\leq\theta(t)\}\right),t\leq T\right)=\left(\frac{2}{\sigma_+}\left(B^{pr}_{\theta(t)}-\inf\{B^{pr}_{\theta(s)}:s\leq t\}\right),t\leq T\right),$$

which is continuous, and it is the height process corresponding to $\left(\sigma_+B^{pr}_{\theta(t)},t\leq T\right)$.

Proof. From [?], we know that $\left(\frac{2}{\sigma_+}R^{pr}_t,t\geq 0\right)$ is the height process corresponding to $(\sigma_+B^{pr}_t,t\geq 0)$. By definition of the height process, we prove the statement if we show that with probability 1, $(B^{pr}_{\theta(t)},t\geq 0)$ is continuous, and for all $t\geq 0$, for all s such that $\theta(t-)<s<\theta(t)$, $B^{pr}_s>B^{pr}_{\theta(t)}$.
Recall that $(B_t,t\geq 0)$ and $(B^{red}_t,t\geq 0)$ are two independent Brownian motions,

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

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

$$(B^{pr}_t, t \geq 0) := \left(B_{\Lambda(t)} + B^{red}_{t-\Lambda(t)}, t \geq 0\right).$$

Firstly, note that the jumps of θ correspond to excursions above the infimum of B^{red}. With probability 1, for all these excursions, the minimum on the excursion is only attained at the endpoints. This can be seen by uniqueness of local minima of Brownian motion. We will assume that this holds.

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 $\theta(t)$ this follows by $(\theta(u) : u \geq 0)$ being strictly increasing. This implies that

$$s - \Lambda(s) < \theta(t) - \Lambda(\theta(t)) = \inf\left\{u \geq 0 : \sigma_+ B^{red}_u < -\frac{\nu_-}{2\mu}t^2\right\}.$$

By our assumption on the minima on the excursions above the infimum of B^{red}, this implies that

$$B^{red}_{s-\Lambda(s)} > -\frac{\nu_-}{2\mu}t^2 = B^{red}_{\theta(t)-\Lambda(\theta(t))},$$

where the last equality follows from continuity of B^{red}. Combining this with $\Lambda(s) = \Lambda(\theta(t))$ implies that $B^{pr}_s > B^{pr}_{\theta(t)}$.

Finally,

$$B^{pr}_{\theta(t-)} = B_{\Lambda(\theta(t-))} + B^{red}_{\theta(t-)-\Lambda(\theta(t-))} = B_t + B^{red}_{\theta(t-)-t}$$

and by continuity of $(B^{red}_s, s \geq 0)$,

$$\begin{aligned} B^{red}_{\theta(t-)-t} &= B^{red}\left(\lim_{s\uparrow t} \inf\{u : B^{red}_u < -\frac{\nu_-}{2\mu}s^2\}\right) \\ &= \lim_{s\uparrow t} B^{red}\left(\inf\left\{u : B^{red}_u < -\frac{\nu_-}{2\mu}s^2\right\}\right) \\ &= -\frac{\nu_-}{2\mu^2}t^2 \\ &= B^{red}_{\theta(t)-t}, \end{aligned}$$

so $B^{pr}_{\theta(t-)} = B^{pr}_{\theta(t)}$. □

B Multivariate triangular local limit theorem

For an i.i.d. sequence $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of random variables, central limit theorems address probabilities of the form

$$\Pr\left(n^{-\alpha} \textstyle\sum_{i=1}^n (\mathbf{X}_i - \mathbb{E}[\mathbf{X}]) \in A\right).$$

where the $\alpha > 0$ is chosen such that the above probabilities converge to a sensible limit and A is some Borel subset. More abstractly, the α is chosen so that $n^{-\alpha}\sum_{i=1}(\mathbf{X}_i - \mathbb{E}[\mathbf{X}])$ converges to a stable law. On the other hand local limit theorems address probabilities of the form

$$\Pr\left(\textstyle\sum_{i=1}^n (\mathbf{X}_i - \mathbb{E}[\mathbf{X}]) \in A\right).$$

without the $n^{-\alpha}$ scaling. Such probabilities are going to decay to 0, but we can still make statements quantifying how they decay to 0, utilizing the density of the limiting stable distribution for the central limit theorem. Local limit theorems are what we are going to need to know the asymptotic behaviour of

$$\Pr(\Delta_n = 0) = \Pr\left(\sum_{i=1}^{n}(D_i^- - D_i^+) = 0\right).$$

In fact to prove lemma 3.2 we will need to introduce a sequence of exponentially tilted measures $\Pr_n$ and thus require a local limit theorem for i.i.d. triangular arrays. We will also need to simultaneously control the behaviour of a separate sum and thus we need a bivariate local limit theorem.

Historically, **?**] covers the stable case in one dimension. For multidimensional random variables, **?**] covers the lattice case and **?**] covers the non-lattice case. **?**] proves a local limit theorem for the bivariate case which allows each component to use a different scaling. A local limit theorem specifically for integer valued random vectors is covered by **?**].

?] prove a triangular local limit theorem for one dimensional random variables converging to a stable limit. In the one-dimensional case where the random variables possess densities, **?**] address more general non i.i.d. triangular arrays where the distribution can vary across rows.

For such more general i.i.d. triangular arrays, **?**] prove a local limit theorem in the multivariate case under the assumption that the moment generating functions of the sums are holomorphic. This builds upon work done by **? ?**] for the non-triangular case.

Unfortunately the assumptions required to apply the results in the literature are too restrictive for use in our case. In particular we don't assume that our random variables have finite fourth moments, and therefore a Cramér type assumption is too restrictive. Instead it will turn out that the common distribution for each row of our i.i.d. triangular array will be converging to the non-tilted case. We will utilize this, along with a uniform integrability type criterion, to prove a general multivariate local limit theorem for lattice valued random variables.

B.1 Lattices

Before we state the theorem, we introduce some definitions and theory about lattices. This section will borrow heavily from the textbook by **?**]. Suppose we are working in the space $\mathbb{R}^d$. Then a set Λ is a *lattice* if there exists a basis $\{\mathbf{a}_1, \dots, \mathbf{a}_d\}$ of $\mathbb{R}^d$ such that

$$\Lambda = \left\{ \sum_{i=1}^{d} n_i \mathbf{a}_i : (n_1, \dots, n_d) \in \mathbb{Z}^d \right\}.$$

The basis can be summarised by the $n \times n$ matrix A with columns $\mathbf{a}_1, \dots, \mathbf{a}_d$, meaning $A_{ij} = a_i^{(j)}$. Then $\{\mathbf{a}_1, \dots, \mathbf{a}_d\}$ is a basis of $\mathbb{R}^d$ if and only if A is invertible.

The vectors $\mathbf{a}_1, \dots, \mathbf{a}_d$ define a d-dimensional parallelepiped

$$\left\{ \sum_{i=1}^{d} \lambda_i \mathbf{a}_i : (\lambda_1, \dots, \lambda_d) \in [0,1]^d \right\}$$

which is sometimes called the *fundamental region* of the lattice. The choice of basis generating a lattice Λ is not unique. This can be seen in fig. 6 where three different choices of the

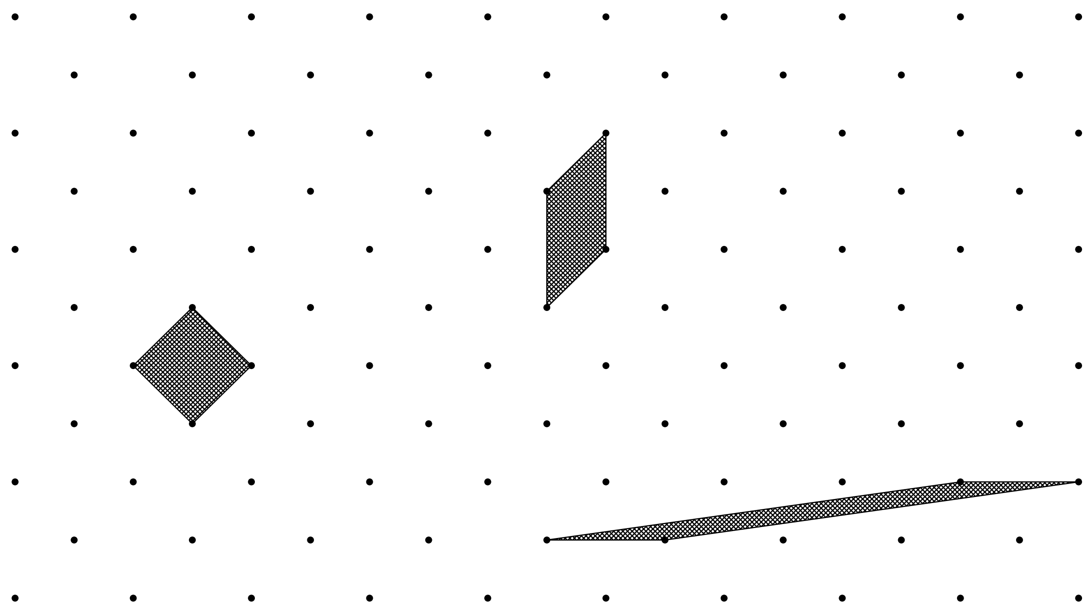


Figure 6: A lattice with three possible choices of fundamental region shaded

fundamental region are drawn. It turns out two matrices correspond to the same lattice if and only if they are obtainable from each other through multiplication by a unimodular matrix. A square matrix U is *unimodular* if it has integer entries and has determinant ± 1. [**?** , Theorem 4.3] shows that the inverse of a unimodular matrix will also be unimodular, and in particular also has integer entries. The following is then a rephrasing of [**?** , Corollary 4.3a]:

Lemma B.1. *Let A and B be invertible matrices. Then the columns of A and the columns of B generate the same lattice if and only if there exists a unimodular matrix U such that $A = UB$.*

Suppose a lattice Λ is generated by the columns of A. Define

$$\det(\Lambda) = |\det(A)|.$$

Then the preceding lemma shows this is well defined even though the choice of A is not unique. The quantity $\det(\Lambda)$ is the volume of any choice of the fundamental region.

A canonical choice of matrix generating a lattice is obtainable when the points of the lattice are rational. An invertible matrix B is in *Hermite normal form* if it is lower triangular with non-negative entries. Then the following is a rephrasing of [**?** , Corollary 4.3b]:

Lemma B.2. *Let A be an invertible matrix with rational entries. Then there exists a unique unimodular matrix U such that UA is in Hermite normal form.*

This gives us the following corollary:

Lemma B.3. *Suppose $\Lambda \subseteq \mathbb{Z}^d$ is a lattice. Then there exists a lower triangular matrix*

$$A = \begin{pmatrix} p_1 & & & 0 \\ p_{2,1} & \ddots & & \\ \vdots & \ddots & \ddots & \\ p_{d,1} & \cdots & p_{d,d-1} & p_d \end{pmatrix}$$

with non-negative integer entries such that Λ is generated by the columns of A.

We say a $\mathbb{R}^d$-valued random variable $\mathbf{X}$ is *non-degenerate* if $\mathbf{X}$ is not supported in a $(d-1)$-dimensional affine subspace of $\mathbb{R}^d$. Further $\mathbf{X}$ is *lattice valued* if there a translation of a lattice containing the support of $\mathbf{X}$.

If $\mathbf{X}$ is non-degenerate and lattice valued, we wish to define the smallest lattice on which $\mathbf{X}$ lives. Let $\mathcal{S}$ be the support of $\mathbf{X}$ and fix an arbitrary $\mathbf{c} \in \mathcal{S}$. Then the *main lattice* of $\mathbf{X}$ is the smallest lattice containing $\mathcal{S} - \mathbf{c}$. Since X is non-degenerate, such a lattice is unique and given by

$$\Lambda = \left\{ \sum_{i=1}^{m} k_i(\mathbf{x}_i - \mathbf{c}) : m \geq 1 \text{ and } k_i \in \mathbb{Z}, \mathbf{x}_i \in \mathcal{S} \text{ for all } i = 1, \dots, m \right\}.$$

The strong aperiodicity condition we imposed on $D^- - D^+$ in **??** can then be stated as saying $D^- - D^+$ has main lattice $\mathbb{Z}$.

The main lattice of $\mathbf{X}$ is related to the periodicity of its characteristic function. The following lemma is an adaptation of [**?** , P.67, T1].

Lemma B.4. *Suppose the main lattice of $\mathbf{X}$ is $\mathbb{Z}^d$ and let ϕ be the characteristic function of $\mathbf{X}$, such that $\phi(\mathbf{u}) = e^{i\mathbf{u}\cdot\mathbf{X}}$. Then $|\phi(\mathbf{u})| = 1$ if and only if every coordinate of $\mathbf{u}$ is a multiple of 2π.*

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 therefore $|\phi(\mathbf{u})| = 1$.

For the converse, by Jensen's inequality

$$|\mathbb{E}[e^{i\mathbf{u}\cdot\mathbf{X}}]| \leq \mathbb{E}[|e^{i\mathbf{u}\cdot\mathbf{X}}|] = 1.$$

More importantly to achieve equality, it must be the case that $e^{i\mathbf{u}\cdot\mathbf{X}}$ is almost surely constant. Therefore there exists $t \in \mathbb{R}$ such that

$$\mathbf{u} \cdot \mathbf{x} \in t + 2\pi\mathbb{Z}$$

for all $\mathbf{x} \in \mathcal{S}$. Then fixing an arbitrary $\mathbf{c} \in \mathcal{S}$,

$$\mathbf{u} \cdot (\mathbf{x} - \mathbf{c}) \in 2\pi\mathbb{Z}$$

for all $\mathbf{x} \in \mathcal{S}$. Since the fundamental lattice of $\mathbf{X}$ is $\mathbb{Z}^d$, there exists $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathcal{S}$ and $k_1, \dots, k_m \in \mathbb{Z}$ such that

$$\sum_{i=1}^{m} k_i(\mathbf{x}_i - \mathbf{c}) = (1, 0, \dots, 0).$$

Therefore

$$u^{(1)} = \sum_{i=1}^{m} k_i \mathbf{u} \cdot (\mathbf{x}_i - \mathbf{c}) \in 2\pi\mathbb{Z}.$$

Repeating this argument for the other coordinates of $\mathbf{u}$ shows all coordinates of $\mathbf{u}$ are multiples of 2π. ∎

B.2 Triangular local limit theorem

Throughout this chapter $\|\mathbf{u}\|$, for $\mathbf{u} \in \mathbb{R}^d$, will denote the Euclidean norm

$$\|\mathbf{u}\| = \left(\sum_{i=1}^{d} \left\{ u^{(i)} \right\}^2 \right)^{1/2}.$$

Theorem B.5. *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 non-degenerate $\mathbb{R}^d$-valued random variable $\mathbf{X}$ such that*

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

2. *For all n, $\mathbf{X}_n$ and $\mathbf{X}$ have a common main lattice Λ.*
3. *$(\|\mathbf{X}_n\|^2)_{n \geq 1}$ is a uniformly integrable sequence of random variables. Explicitly*

$$\lim_{L \to \infty} \sup_{n} \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\left\{ \|\mathbf{X}_n\|^2 > L \right\} \right] = 0. \tag{21}$$

Then $\mathbf{X}$ has finite second moment. Also if $\mathbf{c}_n$ is defined such that $\mathbf{c}_n + \Lambda$ contains the support of $\sum_{i=1}^{n} \mathbf{X}_{n,i}$ then uniformly for $\mathbf{y} \in \mathbf{c}_n + \Lambda$

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

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

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

We now make a few remarks. Firstly $\operatorname{Cov}(\mathbf{X})$ is non-degenerate, meaning it is invertible, since we assume that $\mathbf{X}$ is non-degenerate. In particular this ensures $N(0, \operatorname{Cov}(\mathbf{X}))$ has a density f, explicitly given by

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\operatorname{Cov}(\mathbf{X}))}} \exp\left(-\frac{1}{2} \mathbf{x} \cdot \operatorname{Cov}(\mathbf{X})^{-1} \mathbf{x} \right).$$

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. (21) is still well defined.

Finally, while the local limit theorem holds for all $\mathbf{y}$ on the appropriate lattice, the error term will dominate if $\mathbf{y}$ deviates too far from the mean of $\sum_{i=1}^{n} \mathbf{X}_i$. Specifically consider a sequence $\mathbf{y}_n$ such that

$$\|\mathbf{y}_n - n\mathbb{E}[\mathbf{X}_n]\| = \omega(\sqrt{n}).$$

Let $\lambda_{\max}$ be the maximum eigenvalue of $\operatorname{Cov}(\mathbf{X})$. Then

$$\left(\frac{\mathbf{y}_n - n\mathbb{E}[\mathbf{X}_n]}{\sqrt{n}} \right) \cdot \operatorname{Cov}(\mathbf{X})^{-1} \left(\frac{\mathbf{y}_n - n\mathbb{E}[\mathbf{X}_n]}{\sqrt{n}} \right) \geq \frac{1}{\lambda_{\max}} \frac{\|\mathbf{y}_n - n\mathbb{E}[\mathbf{X}_n]\|^2}{n} \to \infty$$

as $n \to \infty$. Thus the local limit theorem only tells us that

$$\Pr\left(\sum_{i=1}^n \mathbf{X}_i = \mathbf{y}_n\right) = o\left(n^{-d/2}\right)$$

meaning we lose the precise characterisation of the leading order term in how this probability decays.

Before we prove our main theorem, we first prove a series of lemmas. Firstly we show the condition in eq. (21) still holds when we centre the random variables.

Lemma B.6. *Suppose we are in the setting of theorem B.5. Define $\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbb{E}[\mathbf{X}_n]$. Then*

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

Proof. The condition in eq. (21) shows that $M = \sup_n \mathbb{E}[\|X_n\|^2] < \infty$. By Jensen's inequality, $\sup_n \|\mathbb{E}[\mathbf{X}_n]\|^2 \leq M$. Then

$$\begin{aligned}\|\hat{\mathbf{X}}_n\|^2 = \|\mathbf{X}_n - \mathbb{E}[\mathbf{X}_n]\|^2 &\leq 2\|\mathbf{X}_n\|^2 + 2\|\mathbb{E}[\mathbf{X}_n]\|^2 \\ &\leq 2\|\mathbf{X}_n\|^2 + 2M.\end{aligned}$$

Therefore

$$\begin{aligned}\sup_n \mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2 \mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2 > L\right\}\right] \leq{}& 2\sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\left\{\|\mathbf{X}_n\|^2 \geq \tfrac{1}{2}L - M\right\}\right] \\ &+ 2M \sup_n \Pr\left(\|\mathbf{X}_n\|^2 \geq \tfrac{1}{2}L - M\right).\end{aligned}$$

By eq. (21)

$$\lim_{L\to\infty} \sup_n \mathbb{E}\left[\|\mathbf{X}_n\|^2 \mathbb{1}\left\{\|\mathbf{X}_n\|^2 \geq \tfrac{1}{2}L - M\right\}\right] = 0.$$

By Markov's inequality

$$\sup_n \Pr\left(\|\mathbf{X}_n\|^2 \geq \tfrac{1}{2}L - M\right) \leq \frac{2M}{L - 2M} \to 0$$

as $L \to \infty$. The claimed result follows. □

Next we show that the condition in eq. (21) ensures that the means and covariances of the $\mathbf{X}_n$ converge to those of $\mathbf{X}$.

Lemma B.7. *Suppose we are in the setting of theorem B.5. Then $\mathbf{X}$ has finite second moments. Further*

$$\lim_{n\to\infty} \mathbb{E}[\mathbf{X}_n] = \mathbb{E}[\mathbf{X}] \quad \textit{and} \quad \lim_{n\to\infty} \mathrm{Cov}(\mathbf{X}_n) = \mathrm{Cov}(\mathbf{X}).$$

Proof. By Skorokhod's representation theorem WLOG the $\mathbf{X}_n$ and $\mathbf{X}$ are in the same probability space and $\mathbf{X}_n \to \mathbf{X}$ almost surely as $n \to \infty$. Now that the $\mathbf{X}_n$ are in the same probability space, the assumption in eq. (21) shows that $\{\|\mathbf{X}_n\|^2\}_{n\in\mathbb{N}}$ is a uniformly integrable sequence. In particular by Vitali's convergence theorem this shows $\mathbb{E}[\|\mathbf{X}\|^2] = \lim_n \mathbb{E}[\|\mathbf{X}\|^2]$ and thus is finite. Then

$$\|\mathbf{X}_n - \mathbf{X}\|^2 \leq 2\|\mathbf{X}_n\|^2 + 2\|\mathbf{X}\|^2$$

thus $\{\|\mathbf{X}_n - \mathbf{X}\|^2\}_{n\in\mathbb{N}}$ is also an uniformly integrable sequence. Hence

$$\lim_{n\to\infty} \mathbb{E}[\|\mathbf{X}_n - \mathbf{X}\|^2] = 0$$

by Vitali's convergence theorem again, showing that $\mathbf{X}_n \to \mathbf{X}$ in L^2. Therefore the corresponding means and covariances converge, as required. $\square$

Our proof of the local limit theorem will use characteristic functions. The following lemma shows that we have a normal central limit theorem. This will be applied in the form of uniform convergence of characteristic functions on compact sets.

Lemma B.8. *Suppose we are in the setting of theorem B.5. Then*

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

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 [**?** , 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.7. To check condition (2), for all $\epsilon > 0$

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

by lemma B.6. $\square$

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.9. *Suppose we are in the setting of theorem B.5. Moreover assume that the common main lattice Λ is $\mathbb{Z}^d$. Let $\phi_n(\mathbf{u}) = \mathbb{E}[\exp(\mathbf{u}\cdot(\mathbf{X}_n - \mathbb{E}[\mathbf{X}_n])]$ 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 [? , 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.7. 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 $\mathbf{u} \in S(\delta_1)$,

$$\left|1 - \tfrac{1}{2}\mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u}\right| = 1 - \tfrac{1}{2}\mathbf{u}\cdot\mathrm{Cov}(\mathbf{X}_n)\mathbf{u} \leq 1 - \tfrac{1}{4}\lambda_{\min}\|\mathbf{u}\|^2. \tag{22}$$

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 $\|\hat{\mathbf{X}}_n\|^2 > L^2$

$$\begin{aligned}\sup_n E_n &\leq L^2\min\{\|L\mathbf{u}\|, 1\} + \sup_n \mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2\mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2 > L^2\right\}\right] \\ &\to \sup_n \mathbb{E}\left[\|\hat{\mathbf{X}}_n\|^2\mathbb{1}\left\{\|\hat{\mathbf{X}}_n\|^2 > L^2\right\}\right]\end{aligned}$$

as $\mathbf{u} \to 0$. This holds for all $L > 0$, hence taking the limit $L \to \infty$ and using lemma B.6 we obtain that $\lim_{L\to\infty} E_n = 0$. Thus there exists δ_2 such that for all $\mathbf{u} \in S(\delta_2)$

$$R_n(\mathbf{u}) \leq \frac{1}{8}\lambda_{\min}\|\mathbf{u}\|^2. \tag{23}$$

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.4. In particular $|\phi(\mathbf{u})| < 1$ for all $\mathbf{u} \in S(\pi) \setminus S(\delta)$. ϕ is continuous and $S(\pi) \setminus S(\delta)$ is compact. Therefore there exists $\epsilon > 0$ such that $\sup_{\mathbf{u}\in S(\pi)\setminus S(\delta)}|\phi(\mathbf{u})| \leq 1 - \epsilon$.

Since $\mathbf{X}_n \xrightarrow{(\mathrm{d})} \mathbf{X}$ as $n \to \infty$, $\phi_n \to \phi$ uniformly on compact sets. Therefore there exists N such that for all $n \geq N$

$$\sup_{\mathbf{u}\in S(\pi)\setminus S(\delta)} |\phi_n(\mathbf{u})| \leq \rho = 1 - \tfrac{1}{2}\epsilon. \qquad \square$$

We are finally ready to prove theorem B.5

Proof of theorem B.5. 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 $\mathbf{y} \in \mathbf{c}_n + \mathbb{Z}^d$, $\sum_{i=1}^n \mathbf{X}_{n,i} - \mathbf{y} \in \mathbb{Z}^d$ therefore

$$\begin{aligned}
\Pr\left(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\right) &= \mathbb{E}\left[\frac{1}{(2\pi)^d}\int_{S(\pi)} e^{i\mathbf{u}\cdot(\sum_{i=1}^n \mathbf{X}_{n,i} - \mathbf{y})}\,\mathrm{d}\mathbf{u}\right] \\
&= \frac{1}{(2\pi)^d}\int_{S(\pi)} \phi_n(\mathbf{u})^n e^{-i\mathbf{u}\cdot(\mathbf{y}-n\mathbb{E}[\mathbf{X}_n])}\,\mathrm{d}\mathbf{u},
\end{aligned}$$

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 $\mathbf{s} = \sqrt{n}\mathbf{u}$,

$$n^{d/2}\Pr\left(\sum_{i=1}^n \mathbf{X}_{n,i} = \mathbf{y}\right) = \frac{1}{(2\pi)^d}\int_{S(\pi\sqrt{n})} \phi_n(\mathbf{s}/\sqrt{n})^n e^{-i\mathbf{s}\cdot\mathbf{x}_n}\,\mathrm{d}\mathbf{s}.$$

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

$$\begin{aligned}
&\sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda}\left|n^{d/2}\Pr\left(\sum_{i=1}^n \mathbf{X}_{n,i}=\mathbf{y}\right)-f(\mathbf{x}_n)\right| \\
&\qquad = \sup_{\mathbf{y}\in\mathbf{c}_n+\Lambda}\left|\int_{\mathbb{R}^d}\left(\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi_n(\mathbf{s}/\sqrt{n})^n-\psi(\mathbf{s})\right)e^{-i\mathbf{s}\cdot\mathbf{x}_n}\,\mathrm{d}\mathbf{s}\right| \\
&\qquad \le \int_{\mathbb{R}^d}\left|\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi_n(\mathbf{s}/\sqrt{n})^n-\psi(\mathbf{s})\right|\mathrm{d}\mathbf{s}.
\end{aligned}$$

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.9. For all $n \ge N$ and for all $\mathbf{s} \in S(\delta\sqrt{n})$,

$$|\phi_n(\mathbf{s}/\sqrt{n})|^n \le (1-c\|s\|^2/n)^n \le e^{-c\|s\|^2}.$$

Let $C = -\log(\rho)$. Note if $\mathbf{s} \in S(\pi\sqrt{n})$ then $\|\mathbf{s}\|^2 \le \pi^2 dn$. Thus for all $n \ge N$ and $\mathbf{s} \in S(\pi\sqrt{n}) \setminus S(\delta\sqrt{n})$

$$|\phi_n(\mathbf{s}/\sqrt{n})|^n \le e^{-Cn} \le e^{-\frac{C}{\pi^2 d}\|\mathbf{s}\|^2}.$$

Hence for all $n \ge N$,

$$\left|\mathbb{1}_{S(\pi\sqrt{n})}(\mathbf{s})\phi_n(\mathbf{s}/\sqrt{n})^n-\psi(\mathbf{s})\right| \le e^{-c\|\mathbf{s}\|^2}+e^{-\frac{C}{\pi^2 d}\|\mathbf{s}\|^2}+|\psi(\mathbf{s})|$$

where, in particular, the right hand side is integrable. By lemma B.8,

$$\phi_n(\mathbf{s}/\sqrt{n})^n \to \psi(\mathbf{s})$$

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

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

as required.

Finally we generalise to any main lattice Λ. Suppose that Λ is generated by the columns of the invertible matrix A. Then A, viewed as a linear transform, is a isomorphism mapping $\mathbb{Z}^d$ to Λ. Then $\tilde{\mathbf{X}}_n = A^{-1}\mathbf{X}_n$ and $\tilde{\mathbf{X}} = A^{-1}\mathbf{X}$ have common main lattice $\mathbb{Z}^d$. We can check that $\tilde{\mathbf{X}}_n$ satisfies the assumptions of theorem B.5. Let $\tilde{\Sigma} = \mathrm{Cov}(\tilde{\mathbf{X}})$. Then uniformly for $\mathbf{x}$ in the translation of Λ containing the support of $\sum_{i=1}^n \mathbf{X}_{n,i}$,

$$\begin{aligned}
\Pr\left(\sum_{i=1}^n \mathbf{X}_{n,i}=\mathbf{y}\right) &= \Pr\left(\sum_{i=1}^n \tilde{\mathbf{X}}_{n,i}=A^{-1}\mathbf{y}\right) \\
&= \frac{1}{\sqrt{(2\pi n)^d\det\tilde{\Sigma}}}\exp\left(-\frac{1}{2}(A^{-1}\mathbf{x}_n)^T\tilde{\Sigma}^{-1}(A^{-1}\mathbf{x}_n)\right)+o(n^{-d/2}) \\
&= \frac{1}{\sqrt{(2\pi n)^d\det\tilde{\Sigma}}}\exp\left(-\frac{1}{2}\mathbf{x}_n^T(A\tilde{\Sigma}A^T)^{-1}\mathbf{x}_n\right)+o(n^{-d/2}).
\end{aligned}$$

Moreover

$$\tilde{\Sigma} = \mathrm{Cov}(\tilde{\mathbf{X}}) = \mathrm{Cov}(A^{-1}\mathbf{X}) = A^{-1}\,\mathrm{Cov}(\mathbf{X})(A^{-1})^T.$$

Therefore

$$\det(\tilde{\Sigma}) = \det(A)^{-2}\det(\mathrm{Cov}(\mathbf{X})) = \det(\Lambda)^{-2}\det(\mathrm{Cov}(\mathbf{X}))$$

and so

$$\Pr\left(\sum_{i=1}^{n}\mathbf{X}_{n,i} = \mathbf{y}\right) = \frac{\det(\Lambda)}{\sqrt{(2\pi n)^d \det(\mathrm{Cov}\,\mathbf{X})}}\exp\left(-\frac{1}{2}\mathbf{x}_n^T\,\mathrm{Cov}(\mathbf{X})^{-1}\mathbf{x}_n\right) + o(n^{-d/2}),$$

as required. □

B.3 Non-triangular local limit theorem

The local limit theorem for non-triangular arrays of lattice valued random variables, proven by **?**] and **?**], is a direct corollary of the triangular case. For sake of completeness, we state it below.

Theorem B.10. *Let $\mathbf{X}$ be a non-degenerate $\mathbb{R}^d$ valued random variable. Assume the following holds:*

1. *$\mathbf{X}$ is lattice valued with main lattice Λ.*
2. *$\mathbf{X}$ has finite second moments with covariance matrix Σ.*

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. copies of $\mathbf{X}$ and $\mathbf{c}$ be an arbitrary vector in the support of $\mathbf{X}$. Then uniformly for $\mathbf{y} \in n\mathbf{c} + \Lambda$,

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

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

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

C Analysis of the measure change

C.1 Exact form of the measure change

Lemma C.1. *For all $r \le n$ and test functions $u : (\mathbb{N}\times\mathbb{N})^r \times \mathbb{N} \to \mathbb{R}$,*

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

where

$$\psi_r(\mathbf{k}_1, \ldots, \mathbf{k}_r) = \frac{1}{p^r}\prod_{i=1}^{r}\frac{(r-i+1)\mu}{\sum_{j=i}^r k_i^-}.$$

Proof. For any $\mathbf{k}_1, \ldots, \mathbf{k}_m \in \mathbb{N}^+ \times \mathbb{N}$ for all i and $s \in \mathbb{N}$.

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

where the second summation is taken over all bijections $\sigma : [r] \to I$. We examine a single summand.

$$
\begin{aligned}
&\mathbb{P}\left(\mathbf{D}_{\Sigma_n(1)} = \mathbf{k}_1, \dots, \mathbf{D}_{\Sigma_n(r)} = \mathbf{k}_r, \textstyle\sum_{i\in\mathcal{I}_n^c} D_i^+ = s, \mathcal{I}_n = I, \Sigma_n = \sigma\right) \\
=&\mathbb{P}\left(\mathbf{D}_{\Sigma_n(j)} = \mathbf{k}_j \text{ for } j = 1, \dots, r, \textstyle\sum_{i\in\mathcal{I}_n^c} D_i^+ = s, D_i^- = 0 \text{ for } i \in I^c, \Sigma_n = \sigma\right) \\
=&\prod_{i=1}^r \frac{k_i^-}{\sum_{j=i}^r k_j^-} \times \prod_{i=1}^r \nu_{\mathbf{k}_i} \times \mathbb{P}\left(\textstyle\sum_{i\in\mathcal{I}_n^c} D_i^+ = s, D_i^- = 0 \text{ for } i \in I^c\right).
\end{aligned}
$$

We have

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

Also

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

Therefore

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

Finally dividing by $\mathbb{P}(R_n = r) = \binom{n}{r}p^r(1-p)^{n-r}$ gives the desired measure change. □

Lemma C.2. *For all $m \le n$ and test functions $u : (\mathbb{N}\times\mathbb{N})^m \to \mathbb{R}$,*

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

where

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

Proof. By lemma C.1, for all $r \ge m$

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

where

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

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

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

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

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

therefore

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

where

$$A_n(\mathbf{k}_1,\ldots,\mathbf{k}_m) = \left\{\Delta_{n-m} = \sum_{i=1}^{m}(k_i^+ - k_i^-)\right\}.$$

Hence

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

where

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

We wish to view the sum as an expectation over R_{n-m}. In order to do this we rewrite the expression such that we are taken a sum over the probabilities of a $B(n-m,p)$ 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

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

Finally dividing by $\mathbb{P}(R_n \geq m, \Delta_n = 0)$ yields the desired form of ϕ^n_m. □

C.2 Asymptotic bound of the measure change from below

In lemma 3.2 we see the correct scaling to in order to get a sensible scaling limit is $m = \Theta(n^{2/3})$. This will be used to prove the out-components scale as $n^{2/3}$. The following result, which is an analogue of [? , Lemma 6.7], describes the asymptotic behaviour of the measure change when $m = \Theta(n^{2/3})$.

Lemma C.3. *Define*

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

Suppose we are given $\epsilon \in (0, 1/6)$ *and that* $\mathbf{k}_1, \ldots, \mathbf{k}_m$ *are such that*

$$\max_{i=1,\ldots,m}|s^{\pm}(i)| \leq m^{\frac{1}{2}+\epsilon}. \tag{24}$$

Then in the regime $m = \Theta(n^{2/3})$

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

where the $o(1)$ *term is independent of* $\mathbf{k}_1, \ldots, \mathbf{k}_m$ *satisfying our assumptions.*

The $1/6$ upper bound on ϵ is used to show certain terms that arise in the proof will decay to 0, but the exact value of $1/6$ is unimportant. Rather ϵ should be thought of as a positive constant we can make arbitrarily small to make the proof work. We also explain the condition in eq. (24). In lemma 3.2 we evaluate ϕ^n_m as

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

Thus the condition in eq. (24) corresponds to the event

$$\max_{i=1,\ldots,m}\left|\sum_{j=1}^{i}(Z_j^{\pm} - \nu_{\pm})\right| \leq m^{1/2+\epsilon}.$$

This is saying that the centered random walks corresponding to Z_i^+ and Z_i^- do not deviate by more than $m^{1/2+\epsilon}$ in the first m steps. This will occur with high probability and thus eq. (24) is not a restrictive condition to take.

The fact that we only prove a lower bound may seem strange at first. The idea is that since we are dealing with a measure change, as long as the lower bound will have expectation 1 in the limit, this shows that we have not lost a significant amount of mass. The following lemma taken from [? , Lemma 4.8] makes this formal.

Lemma C.4. *Let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be two sequences of non-negative random variables such that $X_n \geq Y_n$ and $\mathbb{E}[X_n] = 1$ for all n. Suppose there exists another non-negative random variable X such that $Y_n \xrightarrow{(d)} X$ and $\mathbb{E}[X] = 1$. Then $X_n \xrightarrow{(d)} X$ and $(X_n)_{n\geq 1}$ is uniformly integrable.*

C.2.1 Exponential tilting

Note that

$$\mathbb{E}[Z^- - Z^+] = \frac{1}{\mu}\mathbb{E}[D^-D^+ - (D^-)^2]$$

and thus is, in general, non-zero even if $\mathbb{E}[D^- - D^+] = 0$. The deviation of the $Z_i^\pm$ around their mean is controlled by the assumption in eq. (24). If $\mathbf{k}_1, \ldots, \mathbf{k}_n$ satisfy eq. (24) and $m = \Theta(n^{2/3})$ then

$$\sum_{i=1}^{m}(k_i^- - k_i^+) = s^-(m) - s^+(m) + (\nu_+ - \nu_-)m = \Theta(n^{2/3}).$$

In contrast Δ_{n-m} is centered, so

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

is looking at the event that Δ_{n-m} takes a value at distance $n^{2/3}$ away from its mean. As remarked in appendix B, the local limit theorem provides no information at distances $\omega(n^{1/2})$ from the mean since the error term will dominate.

To shift the mean of Δ_{n-m} we will introduce a sequence of exponentially tilted measures. 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 C.5. *Define an measure $\mathbb{P}_\theta$, for $\theta \geq 0$, by its Radon–Nikodym derivative*

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \exp\left(-\theta D^- - \alpha(\theta)\right) \quad \textit{where} \quad \alpha(\theta) = \log \mathbb{E}\left[e^{-\theta D^-}\right].$$

Then as $\theta \downarrow 0$ we have

$$\begin{aligned}
\alpha(\theta) &= -\mu\theta + \tfrac{1}{2}\operatorname{Var}(D^-)\theta^2 - \tfrac{1}{6}\mathbb{E}\left[(D^- - \mu)^3\right]\theta^3 + o(\theta^3), \\
\mathbb{E}_\theta[D^-] &= \mu - \operatorname{Var}(D^-)\theta + \mathcal{O}(\theta^3), \\
\textit{and} \quad \mathbb{E}_\theta[D^+] &= \mu - \operatorname{Cov}(D^-, D^+)\theta + \mathcal{O}(\theta^3).
\end{aligned}$$

Proof. Since $\mathbb{E}\left[|D^-|^3\right] < \infty$ and D^- is non-negative, by the dominated convergence theorem

$$\mathbb{E}\left[(D^-)^3 \exp(-\theta D^-)\right] = \mathbb{E}\left[(D^-)^3\right] + o(1) \tag{25}$$

as $\theta \downarrow 0$. Integrating eq. (25) with respect to θ and applying Fubini's theorem gives

$$\begin{aligned}
\int_0^\theta \mathbb{E}\left[(D^-)^3 e^{-\theta' D^-}\right] \mathrm{d}\theta' &= \int_0^\theta \left(\mathbb{E}\left[(D^-)^3\right] + o(1)\right) \mathrm{d}\theta \\
\iff \mathbb{E}\left[\int_0^\theta (D^-)^3 e^{-\theta' D^-}\, \mathrm{d}\theta'\right] &= \mathbb{E}\left[(D^-)^3\right]\theta + o(\theta) \\
\iff \mathbb{E}\left[(D^-)^2\right] - \mathbb{E}\left[(D^-)^2 e^{-\theta D^-}\right] &= \mathbb{E}\left[(D^-)^3\right]\theta + o(\theta) \\
\iff \mathbb{E}\left[(D^-)^2 e^{-\theta D^-}\right] &= \mathbb{E}\left[(D^-)^2\right] - \mathbb{E}\left[(D^-)^3\right]\theta + o(\theta).
\end{aligned}$$

Repeating this method yields

$$\mathbb{E}\left[D^- e^{-\theta D^-}\right] = \mu - \mathbb{E}\left[(D^-)^2\right]\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^3\right]\theta^2 + o(\theta^2), \tag{26}$$

$$\text{and}\quad \mathbb{E}\left[e^{-\theta D^-}\right] = 1 - \mu\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^2\right]\theta^2 - \tfrac{1}{6}\mathbb{E}\left[(D^-)^3\right]\theta^3 + o(\theta^3). \tag{27}$$

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

$$\mathbb{E}\left[D^+ e^{-\theta D^-}\right] = \mu\theta - \mathbb{E}\left[D^- D^+\right]\theta + \tfrac{1}{2}\mathbb{E}\left[(D^-)^2 D^+\right]\theta^2 + o(\theta^2). \tag{28}$$

eq. (27) gives the expansion of the normalising constant of the measure change. Combining this with eq. (26) and eq. (28) yields the expansions for $\mathbb{E}_\theta[D^-]$ and $\mathbb{E}_\theta[D^+]$ respectively. Taking the logarithm of eq. (27) gives the expansion of the cumulant generating function $\alpha(\theta)$. □

To achieve the centering of Δ_{n-m} we desire, let us define a sequence of tilted measures $\mathbb{P}_n$ defined by their Radon–Nikodym derivative

$$\frac{d\mathbb{P}_n}{d\mathbb{P}} = \exp\left(-\theta_n \Xi^-_{n-m} - (n-m)\alpha(\theta_n)\right) \tag{29}$$

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 and each $\mathbf{D}_i$ has the law of $\mathbf{D}$ under $\mathbb{P}_{\theta_n}$. Applying lemma C.5 we can compute that

$$\mathbb{E}_n[\Delta_{n-m}] = m(\nu_+ - \nu_-) + \mathcal{O}(n^{1/3}).$$

Hence

$$\begin{aligned}\textstyle\sum_{i=1}^m (k_i^+ - k_i^-) - \mathbb{E}_n[\Delta_{n-m}] &= s^-(m) - s^+(m) + \left[m(\nu_+ - \nu_-) - \mathbb{E}_n[\Delta_{n-m}]\right] \\ &= \mathcal{O}(n^{1/3+\epsilon})\end{aligned}$$

which is within the $\mathcal{O}(n^{1/2})$ range from the mean where the local limit theorem yields useful information. Thus $\theta_n = \frac{m}{\mu n}$ is the correct tilting to take.

C.2.2 Local limit results

We wish to determine the local limit behaviour of Δ_n under the untilted measure and Δ_{n-m} under the tilted measures. In the latter case, we wish to show the behaviour remains the same even when conditioning on Ξ^-_{n-m} having fluctuations of order $\mathcal{O}(n^{1/2+\epsilon})$ about its tilted mean. The way we show this is to first show a bivariate local limit theorem.

$(D^- - D^+, D^-)$ is $\mathbb{Z}^2$ valued and thus certainly lattice valued. Further in the degenerate case where $D^- = D^+$ almost surely, the total in-degree and total out-degree will be equal almost surely and thus the conditioned and non-conditioned case are the same. So the proof of lemma 3.2 will be the same as for [? , Proposition 4.3]. Hence WLOG assume $(D^- - D^+, D^-)$ is non-degenerate. Let Λ be the main lattice of $(D^- - D^+, D^-)$. Then by lemma B.3 there exist $p, q, r \in \mathbb{Z}_{\geq 0}$ such that Λ is generated by the columns of

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

We have assumed that $D^- - D^+$ has main lattice $\mathbb{Z}$ and therefore $p = 1$. Conditional on $D^- - D^+$ taking some value, D^- will have main lattice $q\mathbb{Z}$. Let σ^2 be the variance of $D^- - D^+$ and Σ be the covariance matrix of $(D^- - D^+, D^-)$.

The $\mathbb{P}(\Delta_n = 0)$ divisor in the definition of ϕ_m^n can be handled by the standard local limit theorem.

Lemma C.6. *As $n \to \infty$*

$$\mathbb{P}(\Delta_n = 0) = \frac{1}{\sqrt{2\pi\sigma^2 n}} (1 + o(1)) .$$

Proof. Let $X = D^- - D^+$. This is assumed to have finite variance σ and main lattice $\mathbb{Z}$. Moreover X is centered and 0 is in the main lattice $\mathbb{Z}$. Thus the result follows by theorem B.10. $\square$

Next we show the local limit theorem holds for $(\Delta_{n-m}, \Xi_{n-m}^-)$ under the tilting.

Lemma C.7. *Let $\mathbf{c}_n \in \mathbb{Z}^2$ be such that $\mathbf{c}_n + \Lambda$ contains the support of $(\Delta_{n-m}, \Xi_{n-m}^-)$. Then uniformly for $(x, y) \in \mathbf{c}_n + \Lambda$*

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

as $n \to \infty$.

Proof. Let $\mathbf{X}_{n1}, \ldots, \mathbf{X}_{nn}$ have the same distribution as

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

under the tilted measure $\mathbb{P}_n$. Then since $\theta_n = o(1)$, it is simple to show that $\mathbf{X}_{n1}$ tends weakly in law to $\mathbf{X} = (D^- - D^+, D^-)$ under the non-tilted measure. The measure change does not change the support of the random variables and thus all $\mathbf{X}_{n1}$ and $\mathbf{X}$ have the same main lattice Λ. Finally we check the uniform integrability condition. Since $\theta_n = o(1)$, $M = -\inf_n \alpha(\theta_n) < \infty$. Then

$$\begin{aligned}\sup_n \mathbb{E}[\|\mathbf{X}_{n1}\|^2 \mathbb{1}\{\|\mathbf{X}_{n1}\|^2 \geq L\}] &= \sup_n \mathbb{E}[e^{-\theta_n D^- - \alpha(\theta_n)} \|\mathbf{X}\|^2 \mathbb{1}\{\|\mathbf{X}\|^2 \geq L\}] \\ &\leq e^M \mathbb{E}[\|\mathbf{X}\|^2 \mathbb{1}\{\|\mathbf{X}\|^2 \geq L\}] \\ &\to 0\end{aligned}$$

as $L \to \infty$ since $\mathbf{X}$ is assumed to have finite second moment. Thus the desired result follows from theorem B.5. There is a slight change in that we have $n - m$ terms in the summations for Δ_{n-m} and Ξ_{n-m}^- rather than n terms, but since $m = \Theta(n^{2/3})$ this will not matter asymptotically. $\square$

Now we show the $\mathbb{P}(\Delta_{n-m} = \sum_{i=1}^m (k_i^+ - k_i^-))$ will have the same asymptotic behaviour as $\mathbb{P}(\Delta_n = 0)$ even when we condition on Ξ_{n-m}^- not 'varying too much' about its tilted mean. We only prove a lower bound, but this is sufficient for proving lemma C.3.

Lemma C.8. *Under the assumptions of lemma C.3,*

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

Proof. For convenience let

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

Define

$$a_n = \sum_{i=1}^m (k_i^+ - k_i^-) - \mathbb{E}_n[\Delta_{n-m}].$$

Also let

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

L_n has a simpler representation. Fix any $y_0 \in L_n$. Then

$$\Lambda \text{ is generated by the columns of } \begin{pmatrix} 1 & 0 \\ r & q \end{pmatrix} \implies L_n = y_0 + q\mathbb{Z}.$$

Fix arbitrary $M > 0$. Then

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

for all n sufficiently large. By lemma C.7, using that the error is uniform, we have that

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

To factorise this we make a change of variables. There exists c 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 C.5,

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

Therefore

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

by the assumption in eq. (24). 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 $a_n = \mathcal{O}(n^{1/3+\epsilon})$, for n sufficiently large

$$\begin{aligned} \sum_{\substack{y\in L_n \\ |y|\leq Mn^{1/2}}} \frac{q}{\sqrt{2\pi\tau^2 n}} \exp\left(-\frac{1}{2\tau^2}\frac{(y-ca_n)^2}{n}\right) &\geq \sum_{\substack{z\in L_n - ca_n \\ |z|\leq \frac{1}{2}Mn^{1/2}}} \frac{q}{\sqrt{n}}\, g\left(\frac{z}{\sqrt{n}}\right) && (30) \\ &= \sum_{\substack{z\in \tilde{L}_n \\ |z|\leq \frac{1}{2}M}} \frac{q}{\sqrt{n}}\, g(z) && (31) \end{aligned}$$

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. (31) 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

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

as required. □

C.2.3 Proving an asymptotic lower bound

We are now ready to prove lemma C.3.

Proof of lemma C.3. Firstly

$$\prod_{i=1}^{m} \frac{(n-i+1)\mu}{\sum_{k=i}^{m} k_i^- + \Xi_{n-m}^-} = \exp(X_n - Y_n)$$

where

$$X_n = \sum_{i=1}^{m} \log\left(1 - \frac{i-1}{n}\right) \quad \text{and} \quad Y_n = \sum_{i=1}^{m} \log\left(\frac{\sum_{k=i}^{m} k_i^- + \Xi_{n-m}^-}{\mu n}\right).$$

Note

$$\sum_{j=i}^{m} k_j^- = s^-(m) - s^-(i-1) + (m-i+1)\nu_-$$

and define

$$\Omega_n^- = \Xi_{n-m}^- - (n-m)\mu + (\nu_- + \mu)m.$$

Then Ω_n^- is almost the centering of Ξ_{n-m}^- under $\mathbb{P}_n$. We have

$$\begin{aligned} Y_n &= \sum_{i=1}^{m} \log\left(\frac{s^-(m) - s^-(i-1) + (m-i+1)\nu_- + \Omega_n^- + \mu n - \nu_- m}{\mu n}\right) \\ &= \sum_{i=1}^{m} \log\left(1 + A_n^i + B_n + C_n^i\right) \end{aligned}$$

where

$$A_n^i = -\frac{1}{\mu n}\left[s^-(i-1) - s^-(m)\right], \quad B_n = \frac{1}{\mu n}\Omega_n^-, \quad C_n^i = -\frac{\nu_-}{\mu n}(i-1).$$

When expanding $\log(1 + A_n^i + B_n + C_n^i)$, the summation contributes order $m = O(n^{2/3})$. Thus we keep terms of order $\Omega(n^{-2/3})$ in the expansion. Write

$$\mathcal{E}_n = \left\{|\Omega_n^-| \le 2n^{\frac{1}{2}+\epsilon}\right\}$$

On the event $\mathcal{E}_n$, we can check that the A_n, B_n, C_n^i and $(C_n^i)^2$ terms are the only ones in the expansion which have order $\Omega(n^{-2/3})$. Moreover

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

Therefore

$$\begin{aligned} Y_n &= \sum_{i=1}^{m} (A_n^i + B_n + C_n^i - \tfrac{1}{2}(C_n^i)^2) + o(1) \\ &= -\frac{1}{\mu}\frac{1}{n}\sum_{i=0}^{m}\left(s^-(i) - s^-(m)\right) + \frac{1}{\mu}\frac{m}{n}\Omega_n^- - \frac{\nu_-}{2\mu}\frac{m^2}{n} - \frac{\nu_-^2}{6\mu^2}\frac{m^3}{n^2} + o(1), \end{aligned}$$

where we use that $\sum_{i=1}^{m}(s^-(i-1)-s^-(m))=\sum_{i=0}^{m}(s^-(i)-s^-(m))$.

Similarly we can expand X_n as

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

Thus

$$\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{k=i}^{m}k_i^-+\Xi_{n-m}^-}\geq\exp\left(\frac{1}{\mu}\frac{1}{n}\sum_{i=1}^{m}(s^-(i)-s^-(m))\right.$$
$$\left.-\frac{1}{\mu}\frac{m}{n}\Omega_n^-+\frac{\nu_--\mu}{2\mu}\frac{m^2}{n}+\frac{\nu_-^2-\mu^2}{6\mu^2}\frac{m^3}{n^2}\right)\mathbb{1}_{\mathcal{E}_n}$$

In addition using lemma C.5, the measure change can be expanded as

$$\frac{\mathrm{d}\mathbb{P}_n}{\mathrm{d}\mathbb{P}}=\exp\left(-\frac{1}{\mu}\frac{m}{n}\Omega_n^-+\frac{\nu_--\mu}{2\mu}\frac{m^2}{n}+\frac{\nu_-^2-\mu^2}{6\mu^2}\frac{m^3}{n^2}+\frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}+o(1)\right).$$

Hence

$$\begin{aligned}
&\phi_m^n(\mathbf{k}_1,\dots,\mathbf{k}_m)\\
=&\frac{1}{\mathbb{P}(\Delta_n=0)}\mathbb{E}\left[\prod_{i=1}^{m}\frac{(n-i+1)\mu}{\sum_{k=i}^{m}k_i^-+\Xi_{n-m}^-}\mathbb{1}_{A_n}\right]\\
\geq&\frac{1}{\mathbb{P}(\Delta_n=0)}\mathbb{E}_n\left[\exp\left(\frac{1}{\mu n}\sum_{i=1}^{m}\left(s^-(i)-s^-(m)\right)-\frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}+o(1)\right)\mathbb{1}_{\mathcal{E}_n\cap A_n}\right]\\
\geq&\exp\left(\frac{1}{\mu n}\sum_{i=1}^{m}\left(s^-(i)-s^-(m)\right)-\frac{\sigma_-}{6\mu^2}\frac{m^3}{n^2}\right)(1+o(1))\frac{\mathbb{P}_n(\mathcal{E}_n\cap A_n)}{\mathbb{P}(\Delta_n=0)}.
\end{aligned}$$

By lemma C.5 we have that

$$\Omega_n^-=\Xi_{n-m}^--\mathbb{E}_n[\Xi_{n-m}^-]+\mathcal{O}(n^{1/3}).$$

In particular for all sufficiently large n,

$$\mathbb{P}_n(\mathcal{E}_n\cap A_n)\geq\mathbb{P}_n\left(|\Xi_{n-m}^--\mathbb{E}[\Xi_{n-m}^-]|\leq n^{1/2+\epsilon},\Delta_{n-m}=\sum_{i=1}^{m}(k_i^+-k_i^-)\right).$$

Thus by lemma C.6 and lemma C.8,

$$\frac{\mathbb{P}_n(\mathcal{E}_n\cap A_n)}{\mathbb{P}(\Delta_n=0)}\geq 1+o(1)$$

as $n\to\infty$, which gives the desired final result. □

C.3 Convergence of the measure change