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 12, 2021

Contents

1 Introduction

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. We are interested in the limit under scaling of the strongly connected components as $n \to \infty$.

We require the degree distribution to satisfy the following properties. [Change to notation Zheneng]

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 Results

+explanation topology

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 [5] and the strategy of Goldschmidt and Stephenson in [11]. 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.

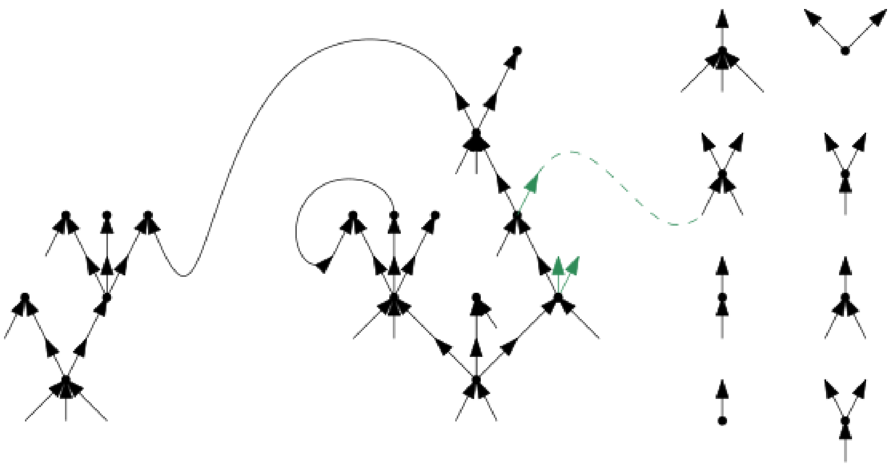


Figure 1: The green arrows represent unpaired out-half-edges of vertices that have been visited. One by one, in depth first order, these are paired to a uniform unpaired in-half-edge.

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 [6]. 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 1.
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 2. We refer to the out-forest corresponding to the exploration up to time k as $\hat{\mathcal{F}}_n(k)$.
A key fact is that the out-forest can be sampled without knowing what the endpoints 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$.

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

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

Then, our approach is as follows.

1. We find the limit under rescaling of $\hat{\mathcal{F}}_n(m_n)$ 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.

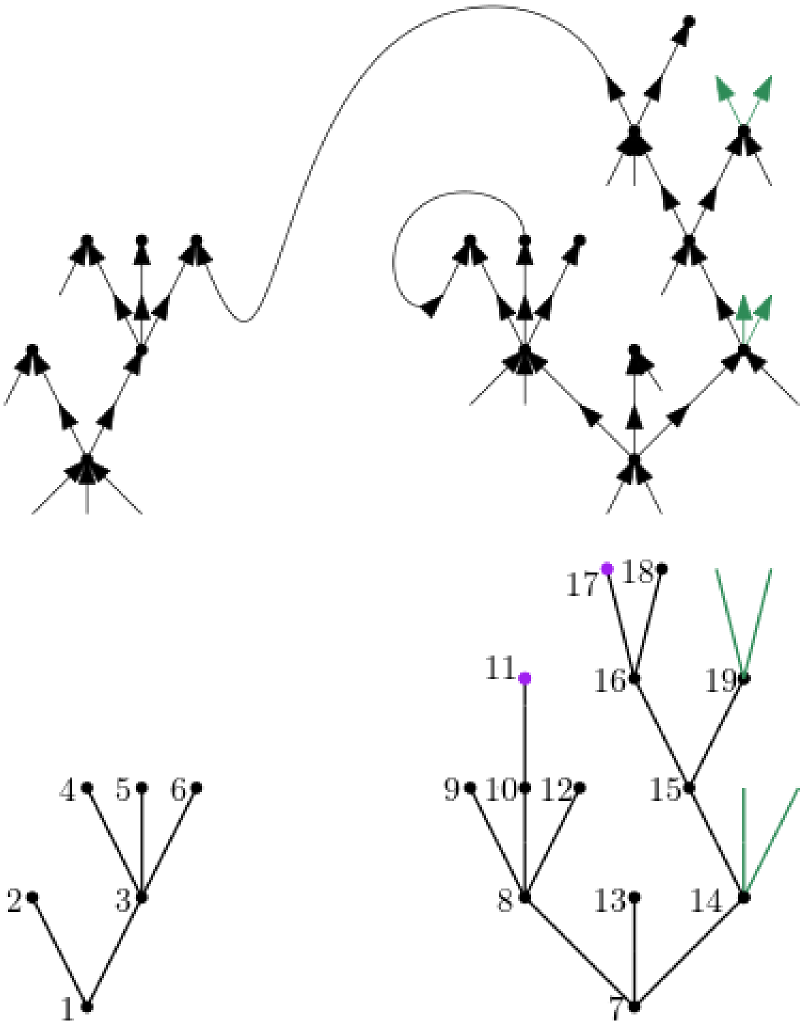


Figure 2: The out-forest is defined based on the exploration of the digraph. For each surplus edge, we add an extra leaf, which we colour purple. The labels of the vertices correspond to the time step in the exploration at which the vertex is added. The green edges lead to vertices of which the degree and colour have not yet been sampled.

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

3. We can recover the strongly connected components that appear in the exploration up to time m_n from $\hat{\mathcal{F}}_n(m_n)$ and the positions of the candidates up to time m_n. We do this by making vertex identifications and performing a cutting procedure. We show that these procedures carry over to the limit.

4. 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 The important edges in the discrete and continuum

If we forget about the directions of the edges, the graph is supercritical, and a condensation phenomenon takes place. 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. 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 xxx, 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 Subsubsection xxx and xxx we study the law of the out-forest and the candidates respectively, and we define a procedure to sample both. Then, in Subsubsection xxx, we define a cutting procedure that extracts the strongly connected components from the edges that remain. In Subsection xxx, we define the continuous counterpart of the sampling and cutting procedures. The resulting object will be the limit in distribution of the strongly connected components under rescaling.

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. [Change figures!] This is illustrated in Figure 3a. In Figure 3b 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* $(\hat{\mathcal{F}}_n(k), k \geq 1)$.

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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$.
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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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.

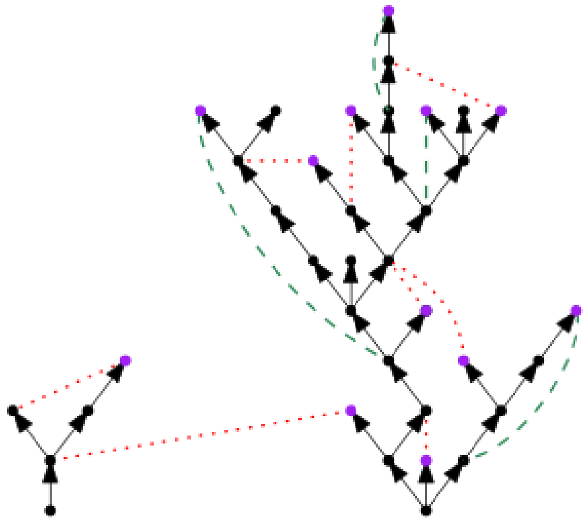
Corollary 2.3. *All edges that are part of a strongly connected component are either a candidate, or are contained in the subforest of $(\hat{\mathcal{F}}_n(k), k \geq 1)$ that is spanned by the tails of candidates and the component roots.*

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

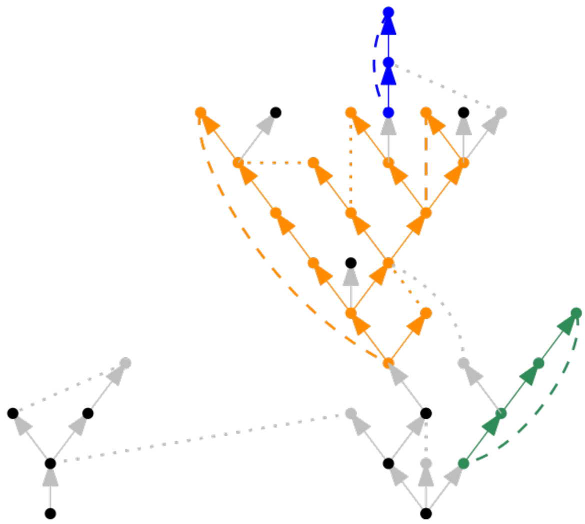
Corollary 2.3 implies that for every purple vertex, we only need to know whether it is a candidate, and if so, where its head is.

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



(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 ancentors. 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 3: We illustrate the different types of surplus edges and how they affect the structure of the strongly connected components.

of discovery are given by $(\hat{\mathbf{D}}_{n,1},\dots,\hat{\mathbf{D}}_{n,n})$. Up to time step k, suppose we have added the vertices corresponding to the first $m\le k$ elements of $(\hat{\mathbf{D}}_{n,1},\dots,\hat{\mathbf{D}}_{n,n})$ 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.

Lemma 2.4. *Suppose the sequence of degrees in order of discovery $(\hat{D}^n_1,\dots,\hat{D}^n_n)$ is given. Suppose, for $1\le l\le k$, that up to time l, $\hat{P}_n(l)$ surplus edges have been sampled. Then,*

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

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

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

define

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

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

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

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 $\hat{P}_n(k)$ leaves, and vertices with degrees $(\hat{D}^+_{n,1},\dots,\hat{D}^+_{n,k-\hat{P}_n(k)})$, so by definition of the Łukasiewicz path, its value is indeed equal to $\hat{S}^-_n(k)$ at time k. Moreover, up to time k, the total in-degree of 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 resoning, 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. □

2.1.3 Sampling the candidates

We will now study the law of the candidates conditional on $(\hat{\mathcal{F}}_n(k), k \geq 1)$. 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 with in-degree m, the edges connecting it to its children will have length $m-1$ (unless it is the root of the 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ and call the resulting forest with edge lengths $(\hat{\mathcal{F}}_n^\ell(k), k \geq 1)$. Denote the height process of $(\hat{\mathcal{F}}_n^\ell(k), k \geq 1)$ by $(\hat{H}_n^\ell(k), k \geq 1)$.

Recall that the first candidate in any component of $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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*

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

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 end point 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$.

Lemma 2.6. *Let $\mathcal{T}_g^n$ be a component of $(\hat{\mathcal{F}}_n(k), k \geq 1)$ with root $g+1$ and component length σ. Suppose the first ancestral surplus edge in $\mathcal{T}_g^n$ corresponds to purple vertex $c_1^n \in [g+2, g+\sigma]$. Let $c_1^n < k \leq g+\sigma$, and suppose the candidates found up to time k are given by $c_1^n, \dots, c_l^n$. Let T_k^n be the subtree of $\mathcal{T}_g^n$ spanned by $\{g+1, c_1^n, \dots, c_l^n, k\}$, and let $\ell_n(T_k^n)$ be its total length with edge lengths as defined by $(\hat{H}_n^\ell(m), m \in [g+1, g+\sigma])$. Then, the probability that k is a candidate is given by*

$$\frac{\ell_n\left(T_k\right) - l}{\hat{S}^-(k)} \mathbb{1}_{\{\hat{P}_n(k) = \hat{P}_n(k-1)+1\}}.$$

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. Observe that $\ell_n\left(T_k\right)$ 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 l of those have been paired: one for each element in $\{c_1^n, \dots, c_l^n\}$. This implies that $\ell_n\left(T_k\right) - l$ 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 $\mathcal{T}_g^n$ be a component of $(\hat{\mathcal{F}}_n(k), k \geq 1)$ with root $g+1$ and component length σ. Suppose its candidates are given by $\{c_1^n, \dots, c_N^n\}$. Then, for $1 \leq i \leq N$, suppose the heads of the surplus edges corresponding to $c_1^n, \dots, c_{i-1}^n$ are given by $d_1^n, \dots, d_{i-1}^n$ respectively. Then, the in-half-edge that c_i^n gets paired to is a uniform pick from the*

$$\ell\left(T_{c_i^n}\right) - (i-1)$$

unpaired in-half-edges of $T_{c_i^n}$ that remain. Call the corresponding vertex d_i^n.

Proof. Given that c_i is a candidate, its head will be in T_{c_i}. Then, the distribution follows. □

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

1. Sample the out-forest $(\hat{\mathcal{F}}_n(k), k \geq 1)$.
2. Fix $T > 0$. Define a counting process $(A_n(k), k \geq 1)$, 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

$$G_i^n = \min\left\{k \geq 1 : \hat{S}_n^+(k) = \min\{\hat{S}_n^+(l) : l \leq X_i^n\}\right\} \text{ for } i \geq 1$$
$$D_i^n = \min\left\{k \geq 1 : \min\left\{\hat{S}_n^+(l) : l \leq k\right\} < \min\left\{\hat{S}_n^+(l) : l \leq X_i^n\right\}\right\} \text{ for } i \geq 1,$$

such that for each $i \geq 1$, $\left(\hat{S}^+(k), k \in [G_i^n + 1, D_i^n]\right)$ encodes a tree. For each $[g, d] \in \{[G_i^n, D_i^n]\}$, let $\mathcal{T}_g^n$ be the tree in $(\hat{\mathcal{F}}_n(k), k \geq 1)$ with root $g+1$, and do the following.

(a) Set $c_1^n = \min\{m \geq 1 : A_n(m) = A_n(g) + 1\}$, and find the other candidates $\{c_2^n, \ldots, c_N^n\}$ according to the procedure described in the statement of Lemma 2.6.

(b) For $c_1^n, \ldots, c_N^n$, sample their heads $d_1^n, \ldots, d_N^n$ respectively according to the procedure described in the statement of Lemma 2.7.

(c) Let $T_{c_N^n}^n$ be the subtree of $\mathcal{T}_g^n$ spanned by $\{g+1, c_1^n, \ldots, c_N^n\}$, say $c_i^n \sim d_i^n$ for each $1 \leq i \leq N$, and set $\mathcal{M}_g^n := T_{c_N}^n / \sim$, which we note is a directed graph with surplus N.

Then, all strongly connected components of $(G_n(k), k \geq 1)$ are subgraphs of $\left\{\mathcal{M}_{G_i^n}^n, i \geq 1\right\}$. Observe that we may view $\mathcal{M}_{G_i^n}^n$ as a finite rooted directed multigraph $M_{G_i^n}^n$ whose edges are endowed with lengths: [Finish this] The following section discusses how to retrieve the strongly connected components from a directed graph $\mathcal{M}_g$. [Use Zheneng's notation for the directed graph resulting from the eDFS].

2.1.4 The cutting procedure

[To add by Zheneng]

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 [11].

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 [insert Le Gall reference]. 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 $f_{a,b} : [0, d(a, b)] \to \mathcal{T}$ such that $f_{a,b}(0) = a$ and $f_{a,b}(d(a, b)) = b$.

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

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

2.2.2 The limit object

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

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

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 $\left(\sigma_+\hat{B}_t, t \geq 0\right)$. See for instance [1].
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

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

such that for each i in $[A_T]$, $\left(\frac{2}{\sigma_+}\hat{R}_t, t \in [G_i, D_i]\right)$ encodes an $\mathbb{R}$-tree. For each element of $\{[G_i, D_i] : i \in A_T\}$ we will sample the candidates in the $\mathbb{R}$-tree. Fix i, set $[g, d] = [G_i, D_i]$, define $\sigma = d - g$. Let $c_1 = \inf\{s > 0 : A(s) = A(g) + 1\}$, such that $g \leq c_1 \leq g + \sigma$ by definition of $[g, d]$. Let $\mathcal{T}_g$ be the R-tree encoded by $\left(\frac{2}{\sigma_+}\hat{R}_t, t \in [g, d]\right)$ and let $p_g : [g, d] \to \mathcal{T}_g$ be the projection onto $\mathcal{T}_g$ given by the encoding. Set

$$||\mathcal{T}_g|| = \sup\left\{\frac{2}{\sigma_+}\hat{R}_t, t \in [g, d]\right\},$$

which we note is the height of $\mathcal{T}_g$.
Suppose we have found candidates $\{c_1, \ldots, c_l\}$. For $c_l \leq s \leq g + \sigma$, let T_s be the subtree of $\mathcal{T}_g$ spanned by $p_g(\{g, c_1, \ldots, c_l, s\})$, and let $\ell(T_s)$ be its total length. Then, let c_{l+1} be the first arrival time of a Poisson process on $[c_l, g + \sigma]$ of intensity

$$\frac{\sigma_{-+} + \nu_-}{\mu^2}\ell(T_s)ds.$$

If the process does not contain a point, let $\{c_1, \ldots, c_l\}$ be the candidates of $\mathcal{T}_g$, and set $N_g = l$. Otherwise, we repeat the inductive step for $\{c_1, \ldots, c_{l+1}\}$. If the induction does not terminate, we set $N_g = \infty$.
We claim that $\mathbb{P}(N_g = \infty) = 0$. Indeed, note that for $c_l \leq s \leq c_{l+1}$, $\ell(T_s) < (l+1)||\mathcal{T}_g||$. Therefore,

$$\mathbb{P}\left(N_g \geq l+1, c_{l+1} - c_l < t |\, N_g \geq l\right) \leq \mathbb{P}(E_{l+1} < t),$$

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

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

Then,

$$\mathbb{P}\left(N_g = \infty\right) = \mathbb{P}\left(N_g = \infty \text{ and } \sup\{c_i : i \in \mathbb{N}\} < g + \sigma\right) \leq \mathbb{P}\left(\sum_{i=2}^{\infty} E_k \leq g + \sigma - c_1\right).$$

However, $\sum_{i=2}^{\infty} E_k = \infty$ a.s., because the harmonic series diverges, so, indeed, $\mathbb{P}\left(N_g < \infty\right) = 1$.
Finally, for $1 \leq i \leq N_g$, let the head corresponding to c_i, which we call d_i, be a uniform pick from the length measure on T_{c_i}.
Then, define $c_i \sim d_i$ for $1 \leq i \leq N_g$, and set $M_g := T_{c_{N_g}}/\sim$, which we note is a directed $\mathbb{R}$-graph with surplus N_g.

2.2.3 The cutting procedure

3 Convergence of the out-forest

In this section, we will show that the Łukasiewicz path and height process corresponding to the forest $\hat{\mathcal{F}}_n(m_n)$ converge under rescaling, for $m_n = O(n^{2/3})$. The main result of this section is as follows. [I think it is best to come up with new notation for $\sum_{i=1}^m \hat{D}^+_{n,i}$. Maybe $\hat{Y}_n^+(m)$? I will use that for now, and we can easily replace it later.]

Theorem 3.1. *Like before, let $(\hat{\mathcal{F}}_n(k), k \geq 1)$ be the sequence of out-forests given by the exploration, where we set $\hat{\mathcal{F}}_n(k+1) = \hat{\mathcal{F}}_n(k)$ if all 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$. 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 $n \to \infty$.

Remark 3.2. *Note that the sign of the parabolic drift is the same as the sign of $\mu - \mathbb{E}[D^+(D^-)^2]$, which is non-positive. Indeed,*

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

is the variance of D^- under the law of $\mathbf{D}$ size-biased by D^+, which is non-negative. 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.

We prove Theorem 3.1 by studying two other forests that are related to $\hat{\mathcal{F}}_n(m_n)$ 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 the order in which the corresponding vertices are included in the exploration. 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 the Radon-Nikodym derivative ϕ_m^n is well-behaved 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 $(\mathcal{F}(k), k \geq 1)$. The convergence of the Łukasiewicz path under scaling of $(\mathcal{F}(k), k \geq 1)$ follows from Donsker's theorem.

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

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

3.1 The measure change and its convergence

[Chapter 6 of Zheneng's transfer]

3.2 Adding purple vertices to a Galton-Watson forest

In this subsection we define $(\mathcal{F}_n^p(k), k \geq 1)$ 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.
Let $Y^+(k) = \sum_{i=1}^k (Z_i^+ - 1)$ be the Łukasiewicz path corresponding to $(\mathcal{F}(k), k \geq 1)$, and set $Y^-(k) = \sum_{i=1}^k (Z_i^- - 1)$. Donsker's theorem and the strong law of large numbers imply that

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

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})^2$ as $n \to \infty$.
The following lemma motivates the definition of $(\mathcal{F}_n^p(k), k \geq 1)$.

Lemma 3.3. *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 Lukasiewicz 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 $(\hat{\mathbf{D}}_{n,1},\dots,\hat{\mathbf{D}}_{n,m})$ replaced by $(\mathbf{Z}_1,\dots,\mathbf{Z}_m)$, and $\sum_{i=1}^n \hat{D}_i^-$ replaced by its mean μn. $\square$

We will now define $(\mathcal{F}_n^p(k), k \geq 1)$ and its Łukasiewicz path $(S_n^+(k), k \geq 1)$ as a function of $(\mathcal{F}(k), k \geq 1)$, $(Y^-(k), k \geq 1)$ 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 $(\mathcal{F}_n^p(k), k \geq 1)$ be the forest with Łukasiewicz path $(S_n^+(k), k \geq 1)$ in which the k^{th} vertex is purple if and only if $P^n(k) - P^n(k-1) = 1$.

3.2.1 Convergence of the Łukasiewicz path

To show convergence of the Łukasiewicz path corresponding to $(\mathcal{F}_n^p(k), k \geq 1)$, 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 3.4. *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. $\square$

Lemma 3.5. *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 3.4, Lemma (1), 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, we would only sample $O(1)$ purple roots up to time $O(n^{2/3})$ with high probability, which does not affect $(P_n(k), k\leq m)$ on the scale that we are interested in.
Then, the convergence in (1), the tightness of $\left(n^{-1/3}I^+(\lfloor n^{2/3}t\rfloor)\right)_{n\geq 1}$ and Lemma 3.4 imply that

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

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 3.6. *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 (1), Lemma 3.5 and the expressions

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

and

$$S_n^{p,-}(k+1) = Y^-(k+1-P_n(k+1)) - P_n(k+1) - I^{p,+}(k) + 1.$$

□

3.2.2 Convergence of the height process

We will extend Corollary 3.6 to joint convergence under rescaling with the height process corresponding to $(\mathcal{F}_n^p(k), k \geq 1)$, which is the content of this subsubsection. We prove the following proposition.

Proposition 3.7. *Let $(H^+(k), k \geq 1)$ be the height process corresponding to $(\mathcal{F}^p(k), k \geq 1)$. Let $(B_t, t \geq 0)$ be a Brownian motion, and define*

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

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 $(\mathcal{F}^p_n(k), k \geq 1)$ is not a Galton-Watson forest, because the probability of sampling a purple leaf changes as time progresses. In fact, the sampling process does not even satisfy the Markov property, because the probability of sampling a purple vertex depends on the past of the process. The theory of convergence of height processes under rescaling for Galton-Watson forests is a well-developed (see e.g. [9]), but this is not the case for more general processes. We will adapt a technique that Broutin, Duquesne and Wang develoed in [3] to show convergence of the height process of inhomogeneous random graphs under rescaling. The key idea is that $(\mathcal{F}^p_n(k), k \geq 1)$ itself is not a Galton-Watson forest, but we can embed it in a Galton-Watson forest, say $(\mathcal{F}^{pr}(k), k \geq 1)$, which will be equal to $(\mathcal{F}^p_n(k), k \geq 1)$ with extra red vertices. We then show convergence under rescaling of the height process corresponding to $(\mathcal{F}^{pr}(k), k \geq 1)$, and use this to obtain height process convergence for $(\mathcal{F}^p_n(k), k \geq 1)$.
We start by defining $(\mathcal{F}^{pr}(k), k \geq 1)$. Informally, we obtain $(\mathcal{F}^{pr}(k), k \geq 1)$ by modifying $(\mathcal{F}^p_n(k), k \geq 1)$ 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 4.

The formal procedure is as follows. Suppose we are given $(Y^+(k), S^+(k), P_n(k), k \geq 1)$, which encode $(\mathcal{F}^p_n(k), k \geq 1)$.

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

$$(Y^{pr}(k), k \geq 1) = (Y^+(\Lambda_n(k)) + Y^{red}(k - \Lambda_n(k)), k \geq 1) \tag{3}$$

and we let $(\mathcal{F}^{pr}(k), k \geq 1)$ 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 $(\mathcal{F}^{pr}(k), k \geq 1)$.

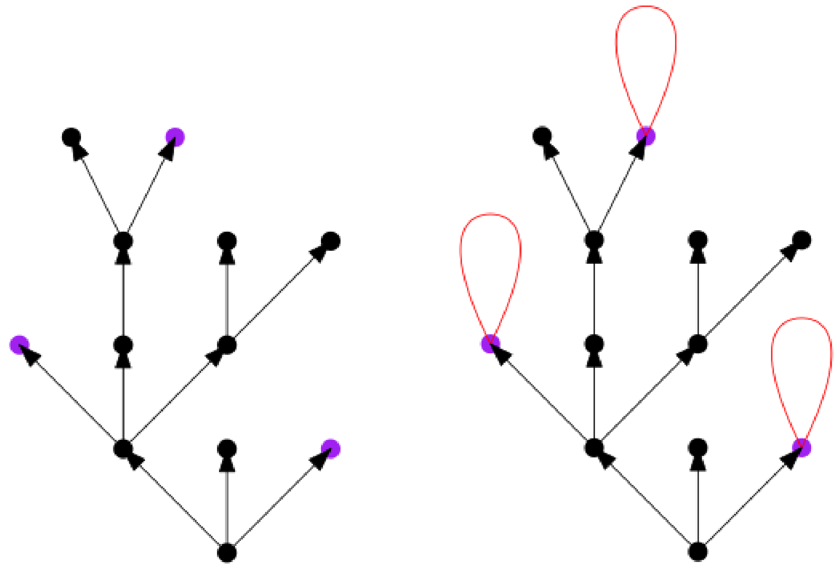


Figure 4: Given a component of $(\mathcal{F}_n^p(k), k \geq 1)$ (see left figure), we modify it by sampling independent red Galton-Watson trees with offspring distributed as Z^+ and identifying each purple vertex with a root of a red tree. The resulting tree (see right figure) is a Galton-Watson tree, and the resulting forest $(\mathcal{F}_n^{pr}(k), k \geq 1)$ is a Galton-Watson forest.

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

1. We claim that
$$\theta_n(k) = \min\{l : \mathcal{F}^p(k) \text{ is a subforest of } \mathcal{F}^{pr}(l)\}.$$
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 $\mathcal{F}^p(k)$. Then, $\theta_n(k)$ is the index in $(\mathcal{F}^{pr}(k), k \geq 1)$ 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 $(\mathcal{F}^{pr}(k), k \geq 1)$ 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 $(\mathcal{F}_n^p(k), k \geq 1)$. Moreover,
$$(S^+(k), k \geq 1) = (Y^{pr}(\theta_n(k)) - E(\theta_n(k)), k \geq 1), \tag{4}$$
where $E(k)$ counts the red children of the k^{th} vertex in $(\mathcal{F}^{pr}(k), k \geq 1)$.

Considering the construction above and Corollary 3.6, in order to prove Proposition 3.7, it is sufficient to show that there exists a process $(D_t, t \geq 0)$ such that

$$\begin{aligned} &\left(n^{-1/3}\left(Y^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right) - E\left(\lfloor n^{2/3}t\rfloor\right)\right), n^{-1/3}H^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t \geq 0\right) \\ &\xrightarrow{d} \left(\sigma_+ D_t, \frac{2}{\sigma_+}\left(D_t - \inf\left\{D_s, s \leq t\right\}\right), t \geq 0\right) \end{aligned} \tag{5}$$

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

The next lemma show that the pathwise construction of $(Y^{pr}(k), H^{pr}(k), k \geq 1)$ converges to the continuous counterpart of the pathwise construction.

Lemma 3.8. *Let $(B_t, t \geq 0)$ and $(B_t^{red}, t \geq 0)$ be two independent Brownian motions and let*

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

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

$$(B_t^{pr}, t \geq 0) := \left(B_{\Lambda(t)} + B_{t-\Lambda(t)}^{red}, t \geq 0\right). \tag{6}$$

Then, for

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

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

$$\left(n^{-1/3}Y^{pr}\left(\lfloor n^{2/3}t\rfloor\right), n^{-1/3}H^{pr}\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{d} \left(\sigma_+ B_t^{pr}, \frac{2}{\sigma_+}R_t^{pr}, t \geq 0\right) \tag{7}$$

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

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

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,

$$\left(n^{-1/3}Y^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), n^{-1/3}H^{pr}\left(\theta_n\left(\lfloor n^{2/3}t\rfloor\right)\right), t \geq 0\right) \xrightarrow{d} \left(\sigma_+ B_{\theta(t)}^{pr}, \frac{2}{\sigma_+}R_{\theta(t)}^{pr}, t \geq 0\right) \tag{8}$$

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

In the proof of Lemma 3.8 we use the following, straightforward, technical result.

Lemma 3.9. *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 [10] by Ethier and Kurtz, the result follows immediately. $\square$

of Lemma 3.8. 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 [13] gives us that for $(B_s^*, s \geq 0)$ a Brownian motion,

$$\left(n^{-1/3}Y^{pr}\left(\lfloor n^{2/3}s\rfloor\right), n^{-1/3}H^{pr}\left(\lfloor n^{2/3}s\rfloor\right), s \geq 0\right) \xrightarrow{d} \left(\sigma_+ B_s^*, \frac{2}{\sigma_+}\left(B_s^* - \inf\{B_u^* : u \leq s\}\right), s \geq 0\right) \tag{9}$$

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, [4] 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

$$\begin{aligned}&\left(n^{-1/3}Y^{red}\left(\lfloor n^{2/3}s\rfloor\right), n^{-2/3}\inf\left\{k : n^{-1/3}Y^{red}(k) \leq -x\right\}, s \geq 0, x \geq 0\right)\\ &\xrightarrow{d} \left(\sigma_+ B_s^{red}, \inf\left\{u : \sigma_+ B_u^{red} < -x\right\}, s \geq 0, x \geq 0\right)\end{aligned} \tag{10}$$

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 $n \to \infty$. Moreover, it is standard that

$$\left(\inf\left\{u : \sigma_+ B_u^{red} < -x\right\}, x \geq 0\right) \stackrel{d}{=} \left(\frac{1}{\sigma_+^2}L_x, x \geq 0\right), \tag{11}$$

with $(L_x, x \geq 0)$ a stable subordinator with exponent $1/2$. Since $(P_n(k), k \geq 1)$ is non-decreasing, applying Lemma 3.9, and combining the convergence in (10) with Lemma 3.5 gives that also

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

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in (10), and therefore,

$$\left(n^{-2/3}\theta_n\left(\lfloor n^{2/3}t\rfloor\right), t \geq 0\right) \xrightarrow{d} (\theta(t), t \geq 0) \tag{12}$$

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in (10). Recall that

$$\Lambda_n(k) = \max\{j : \theta_n(j) \leq k\} - P_n(\max\{j : \theta_n(j) \leq k\}).$$

By (11), the jumps of $(\theta(t), t \leq T)$ are dense, and since $(\theta_n(k), k \geq 1)$ is increasing for all n and $(\theta(t), t \geq 0)$ is increasing,

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

in $D(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$ jointly with the convergence in (10) and (12). Since $\max\{j : \theta_n(j) \leq \lfloor n^{2/3}t\rfloor\}$ is of order $n^{2/3}$, Lemma 3.4 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 (10) and (12).
To finish the proof, we examine the construction of $(Y^{pr}(k),k\geq 1)$ in (3) and the construction of $(B^{pr}_s,s\geq 0)$ in (6). Note that $\Lambda_n(k)$ and $k-\Lambda_n(k)$ are non-decreasing. Again, by Lemma 3.9, 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 (9) proves (7). The fact that $(\theta_n(k),k\geq 1)$ is non-decreasing and Lemma 3.9 then imply (8). □

Lemma 3.10. *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 (4), and by Lemma 3.8, 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 $(\mathcal{F}^{pr}(k),k\geq 1)$, 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 $n\to\infty$. □

The following lemma is the last ingredient in the proof of (5), and therefore also in the proof of Proposition 3.7.

Lemma 3.11. *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 [13], 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 $B^{pr}_{\theta(t-)} = B^{pr}_{\theta(t)}$. □

3.3 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 $(\mathcal{F}^p(k), k \geq 1)$, which is the content of Proposition 3.7, to prove Theorem 3.1. [Insert theorem of convergence Lukasiewicz path before adding purple vertices ($\hat{S}_n$) somewhere before this (maybe with proof of convergence measure change, maybe in this section?)(result is written up in StuffSerte.tex)]

Proof. Proof of *Theorem 3.1* Recall that $\hat{P}_n(k)$ denotes the number of purple vertices in $\hat{\mathcal{F}}_n(k)$. 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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 $(\mathcal{F}^p(k), k \geq 1)$, we would like to replace the term $\sum_{i=0}^n D_i^-$ in the denominator by μn. Therefore, define a new forest $(\hat{\mathcal{F}}'_n(k), k \geq 1)$ 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 $\hat{\mathcal{F}}'_n(k)$, and $\hat{I}'_n(k)$ is the number of components in $\hat{\mathcal{F}}'_n(k)$. 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, note that by earlier results and absolute continuity [Make this reference clearer once Zheneng's section has arrived]

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

is tight. Moreover, with a slight adaptation to the proof of Lemma 3.4, 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 $D_1^+, \ldots, D_n^+$ being i.i.d. random variables with mean μ and finite variance, [Argue that under measure change fluctuations are still $O(n^{1/2})$] $\left(n^{-1/2}\left(\sum_{i=0}^{n-1} D_i^- - \mu n\right)\right)_{n>0}$ 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$ and $(\hat{\mathcal{F}}'_n(k), k \geq 1)$ 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 $(\hat{\mathcal{F}}'_n(k), k \geq 1)$ 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:

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

[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 3.7, following the proof of Theorem 4.1 in [5] 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 XXX [law of measure changed BM (result is written up in StuffSerte.tex)] implies the convergence under rescaling of $(\hat{S}^+_n(k), \hat{H}_n(k), k \geq 0)$. By Proposition 3.7 , S^-_n converges in distribution to a deterministic process under scaling, which will not be effected by the measure change. This completes the proof. □

3.4 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 [12] to show that this implies that 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 $(\hat{\mathcal{F}}_n(k), k \geq 1)$. We call these anomalous edges.

Proposition 3.12. *Suppose $\beta < 3/4$. 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 [12] and of Proposition 5.3 of [5] 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 [12] and [5] to yield a correct proof.

Proof. For the proof of this lemma, we will use a method introducted by Joseph in [12] 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 [12]. 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. $\Delta_n := \int_{(0,n)\times\mathbb{N}^2} (k_1 - k_2)\hat{\Pi}_n(dt, k_1, k_2) = 0$,

the points of $\hat{\Pi}_n$ ordered by time are distributed as $(\hat{\mathbf{D}}_{n,1}, \ldots, \hat{\mathbf{D}}_{n,n})$ 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 at least n^β jumps at time $2n^\beta$. Indeed,

$$\mathbb{P}\left(\Pi_n\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right) < n^\beta \middle| \Delta_n = 0\right) \leq \frac{\mathbb{P}\left(\sum_{i=1}^{n^\beta} E_i > 2n^\beta\right)}{\mathbb{P}(\Delta_n = 0)}$$

for $(E_1, E_1, \dots)$ i.i.d. exponential random variables with rate 1, and we see that the numerator is $O(\exp(-n))$ by Cramér's Theorem, while the denominator is of $O(n^{-1/2})$ by the local limit theorem, so the ratio tends to 0.

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

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

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

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

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

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

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

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

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,

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

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

$$\mathbb{P}\left(B_n^1(\lfloor n^\beta \rfloor) + B_n^2(\lfloor n^\beta \rfloor) + B_n^3(\lfloor n^\beta \rfloor) > 0 \middle| \mathcal{G}^n\right) \xrightarrow{p} 0 \tag{13}$$

as $n \to \infty$. Then, the proposition follows from the bounded convergence theorem. By the observations above, it is sufficient to show that as $n \to \infty$,

$$\frac{1}{n} \int_{(0,2n^\beta) \times \mathbb{N}^2} k_1 k_2 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \tag{14}$$

$$\frac{1}{n} \int_{(0,2n^\beta) \times \mathbb{N}^2} k_1 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \tag{15}$$

$$\frac{1}{n} \int_{(0,2n^\beta) \times \mathbb{N}^2} k_1^2 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \tag{16}$$

$$\frac{1}{n} \int_{(0,2n^\beta) \times \mathbb{N}^2} k_2^2 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0, \text{ and} \tag{17}$$

$$\frac{1}{n} \int_{(0,2n^\beta) \times \mathbb{N}^2} k_2^3 \hat{\Pi}_n(dt, k_1, k_2) \xrightarrow{p} 0. \tag{18}$$

We will show 14. The proof of the other equations is analogous.
We start with the proof of 14. Note that, for $\hat{E}_t^n$ the expectation under $\hat{P}_t^n$, it is sufficient if we show that for some C,

$$\hat{E}_t^n \left[\hat{D}_t^- \hat{D}_t^+\right] < C \tag{19}$$

for all n and $t < 2n^\beta$, because by a similar argument as above,

$$\mathbb{P}\left(\hat{\Pi}_n\left((0,2n^\beta),\mathbb{N},\mathbb{N}\right) > 3n^\beta\right) \to 0$$

as $n \to \infty$. 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 $N_n = n-1$ or $N_n = n$, Δ_n is the sum of respectively $n-1$ or 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 $t < 2n^\beta$, (19) follows. We note that by definition of $\pi_n(dt,k_1,k_1)$,

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

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

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

as $s \to 0$. We refer the reader to the proof of Lemma A.1 for the details of a similar argument in the non-directed setting. This implies that

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

uniformly in all $t \leq 2n^\beta$, and (19) follows.
By applying the same techniques, (15), (16), (17) and (18) follow as well, which proves the statement.

$\square$

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 [12]. $\square$

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_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 $\hat{\mathcal{F}}_n\left(\lfloor n^{2/3}t\rfloor\right)$ 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 law of the set of components in $(\hat{\mathcal{F}}_n(k), k \geq 1)$ that contain a non-trivial strongly connected component is the same as the law of the set of components in $(\hat{\mathcal{F}}_n(k), k \geq 1)$ 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 law of the tail of the first ancestral surplus edge.

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 [8] 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 $(\mathcal{F}^{pr}(k), k \geq 1)$, as defined in Subsubsection 3.2.2.
We equip $(\mathcal{F}^{pr}(k), k \geq 1)$ 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 $(\mathcal{F}^{pr,\ell}(k), k \geq 1)$, and let $(H^{pr,\ell}(k), k \geq 1)$ be the corresponding height process.
We will translate the conditions of Theorem 1 in [8] to our setting and check them. The conditions are as follows.

1. $\mathbb{E}[Z^+] = 1$
2. $1 < \mathbb{E}[(Z^+)^2] < \infty$
3. $\mathbb{E}\left[Z^+ \mathbb{1}_{(Z^- -1-1)>x}\right] = o(x^{-2})$ as $x \to \infty$

Under these conditions, using the notation from Subsubsection 3.2.2,

$$\begin{aligned}&\left(n^{-1/3} Y^{pr}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{pr}\left(\lfloor tn^{2/3} \rfloor\right), n^{-1/3} H^{pr,\ell}\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right)\\ &\xrightarrow{d} \left(\sigma_+ B_s, \frac{2}{\sigma_+} R_s, \frac{2(\sigma_{+-} + \nu_-)}{\mu \sigma_+} R_s, t \geq 0\right)\end{aligned} \tag{20}$$

in $D(\mathbb{R}_+, \mathbb{R})^3$ as $n \to \infty$. Then, we observe that the rest of the argument in Subsubsection 3.2.2 and Subsection 3.3 can be extended to include the height process with edge lengths. This yields the result.
Therefore, to finish the proof, we need the conditions of Theorem 1 in [8] 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

$$\begin{aligned} A_n^{comp}(k) &= \sum_{i=1}^{k} \frac{\hat{H}_n^{\ell}(i)}{\hat{S}_n^{-}(i)} \mathbb{1}_{\{\hat{P}_n(i)-\hat{P}_n(i-1)=1\}} \\ &= \sum_{j=1}^{\hat{P}_n(k)} \frac{\hat{H}_n^{\ell}(\min\{l : \hat{P}_n(l) \geq k\})}{\hat{S}_n^{-}(\min\{l : \hat{P}_n(l) \geq k\})}, \end{aligned}$$

By Theorem 14.2.VIII of Daley and Vere-Jones [7], the claimed convergence under rescaling of $(A_n(k), k \geq 1)$ follows if we show that

$$\left(A_n^{comp}\left(\lfloor tn^{2/3} \rfloor\right), t \geq 0\right) \xrightarrow{d} \left(\frac{2(\sigma_{-+} + \nu_{-})}{\sigma_{+}\mu^2} \int_0^t \hat{R}_v dv, t \geq 0\right) \tag{21}$$

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 (21) 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 3.9, Slutsky's lemma and the continuous mapping theorem imply that

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

If we combine this with the convergence under rescaling of $(P_n(k), k \geq 1)$ and apply Lemma 3.9, some simple analysis then yields (21), which proves the statement. $\square$

4.2 Extracting the important components of 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 $(\hat{\mathcal{F}}_n(k), k \leq \lfloor Tn^{2/3} \rfloor)$ that contain ancestral surplus edges converges as well under rescaling. Lemma 4.3 is a statement about extracting excursions 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

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

Define $\sigma_n^i = d_i^n - g_i^n$, for $i \in [m]$, $n \geq 1$ and $\sigma^i = d_i - g_i$. 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 $\operatorname{ord}(S) \in (\mathbb{R}^3)^\infty$. Then,

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

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

Proof. First, note that g_i^n, d_i^n, g_i, and d_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 $g_i^n \to g_i$ and $d_i^n \to d_i$ as $n \to \infty$. Firstly, note that by the assumption that $f(x_i) - \inf\{f(s) : s \leq x_i\} > 0$ and the continuity of f, $g_i < x_i < d_i$. Fix $0 < \epsilon < \min\{x_i - g_i, d_i - x_i\}/2$. We claim that the following conditions are sufficient for $g_i^n \to g_i$ and $d_i^n \to d_i$ as $n \to \infty$

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

for all n large enough. Indeed, condition 1, 2 and 3 imply $|g_i^n - g_i| < \epsilon$, while condition 1, 2 and 4 imply $|d_i^n - d_i| < \epsilon$. 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

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

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

$$(g_i^n, d_i^n)_{1 \leq i \leq m} \to (g_i, d_i)_{1 \leq i \leq m}$$

in $\mathbb{R}^{2m}$ as $n \to \infty$.
We now claim that $g_i^n \to g_i$ and $g_j^n \to g_i$ implies that $g_i^n = g_j^n$ for n large enough. Indeed, by definition of g_i^n, g_j^n and σ_i^n, we see that $g_i^n < g_j^n$ implies that $g_j^n - g_i^n \geq \sigma_i^n$, and by the argument above, $\sigma_i^n \to \sigma_i > 0$, so $g_i^n - g_j^n \to 0$ can only hold if $g_i^n = g_j^n$ for n large enough. The equivalent statement can be proved for d_i^n, which implies that

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

Then, the result follows. $\square$

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*

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

Define $\Sigma_i^n = D_i^n - G_i^n$ and $\Sigma_i = D_i - G_i$. Then, for ord *defined as in the statement of Lemma 4.3, we get that*

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

in the ℓ_1-topology on $(\mathbb{R}^3)^\infty$ 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 $\hat{R}_{X_i} - \inf\{\hat{R}_s : s \leq X_i\} > 0$ 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

[Rewrite] 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 set of tail of the candidates on a particular component of $(\hat{\mathcal{F}}_n(k), k \geq 1)$.

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 $(G, \Sigma) \in \{(G_i, \Sigma_i) : i \leq A_T\}$, such that, for each n large enough, we can find a $(G_n, \Sigma_n) \in \left\{(G_i^n, \Sigma_i^n) : i \leq A_n\left(\lfloor T n^{2/3} \rfloor\right)\right\}$, such that $(G_n, \Sigma_n) \to (G, \Sigma)$. Set $C_1 = \inf\{t \in [G, G+\Sigma] : A(t) = A(G) + 1\}$, and similary, set $C_1^n = \min\{G_n < k \leq G_n + \Sigma_n : A_n(k) = A_n(G_n) + 1\}$, which are well-defined by definition of G, Σ, G_n and Σ_n. Then, by construction, $\{G_n + 1, \ldots, G_n + \Sigma_n\}$ encodes a component of $(\hat{\mathcal{F}}_n(k), k \geq 1)$. Call this component $\mathcal{T}_n^{G_n}$. We apply the procedure defined in 2.6 to find the tail of candidates in $\mathcal{T}_n^{G_n}$. Let $\mathbf{C}_n(G_n)$ denote the sequence of tails of candidates in $\mathcal{T}_n^{G_n}$. Similarly, $[G, G+\Sigma]$ encodes a component of $(\hat{\mathcal{F}}(t), t \geq 0)$. Call this component $\mathcal{T}^G$, and apply procedure in Subsubsection 2.2.2 to find the tails of candidates in $\mathcal{T}^G$, and denote its sequence of tails of candidates by $\mathbf{C}(G)$. Then, jointly with the convergence in Proposition 4.4,*

$$n^{-2/3}\mathbf{C}_n(G_n) \xrightarrow{d} \mathbf{C}(G)$$

in the ℓ_1 topology.

Proof. We will find a coupling such that $n^{-2/3}C_n(G_n) \xrightarrow{a.s.} C(G)$. By the convergence in Propositions 4.2 and 4.4, $n^{-2/3}C_1^n \xrightarrow{a.s.} C_1$. In general, let C_m^n denote the m^{th} candidate that is found in $\mathcal{T}_n^{G_n}$, and let C_m denote the m^{th} candidate that is found in $\mathcal{T}^G$. Suppose that, for some m, we have found a coupling such that

$$n^{-2/3}(C_1^n, \ldots, C_m^n) \xrightarrow{a.s.} (C_1, \ldots, C_m).$$

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

$$N_{comp,m+1}^n(k) = \sum_{i=C_m^n+1}^{k} \frac{\ell_n\left(T_i^n\right) - m}{\hat{S}^-(i)} \mathbb{1}\{P_n(i) = P_n(i-1) + 1\}$$

for $k \in [C_m^n + 1, G_n + \Sigma_n]$ and 0 otherwise, where T_i^n is the subtree of $\mathcal{T}_n^{G_n}$ spanned by $\{G_n + 1, C_1^n, \ldots, C_m^n, i\}$. Moreover, C_{m+1} is the first jump in a counting process $N_{m+1}(t)$ on $[G, \infty)$ with compensator

$$N_{comp,m+1}(t) = \frac{\sigma_{-+} + \nu_-}{\mu^2}|T_t|$$

for $t \in [C_m, G+\Sigma]$ and 0 otherwise, where T_t is the subtree of $\mathcal{T}^G$ spanned by $\{G, C_1, \ldots, C_m, t\}$, and $|T_t|$ is its length as encoded by $\left(\frac{2}{\sigma_+}\hat{R}_t, t \geq 0\right)$. Then, by the convergence under rescaling of $(\hat{H}_n^\ell(k), k \geq 1)$ and Proposition 4.4, we get that the metric structure of $\mathcal{T}_n^{G_n}$ with distances defined by $(\hat{H}_n^\ell(k), k \geq 1)$, and its projection onto $[n^{-2/3}(G_n+1), n^{-2/3}(G_n+\Sigma_n)]$, converge under rescaling to the metric structure of $\mathcal{T}^G$ with distances defined by

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

and its projection onto $[G, \Sigma]$. This implies that

$$\left(n^{-1/3}\ell_n\left(T^n_{\lfloor tn^{2/3}\rfloor}\right), C_m \leq t \leq G+\Sigma\right) \stackrel{a.s.}{\to} \left(\frac{\sigma_{-+}+\nu_-}{\mu^2}|T_t|, C_m \leq t \leq G+\Sigma\right)$$

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

$$\left(N^n_{comp,m+1}\left(\lfloor tn^{2/3}\rfloor\right), C_m \leq t \leq G+\Sigma\right) \stackrel{a.s.}{\to} \left(N_{comp,m+1}(t), C_m \leq t \leq G+\Sigma\right),$$

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

$$(N^n_{m+1}(\lfloor tn^{2/3}\rfloor), t \geq 0) \stackrel{d}{\to} (N_{m+1}(t), t \geq 0)$$

in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ as $n \to \infty$, and in particular, we can find a coupling such that $N_m(\infty) > 0$ if and only only if $N^n_m(\infty) > 0$ for all n large enough, and such that on this event,

$$n^{-2/3}C^n_{m+1} \stackrel{a.s.}{\to} C_{m+1}.$$

If $N_m(\infty) = 0$, set $\mathbf{C}(G) = (C_1, \ldots, C_m)$, $\mathbf{C}_n(G_n) = (C^n_1, \ldots, C^n_m)$, and the statement follows. If $N_m(\infty) > 0$, apply the induction step to $(C_1, \ldots, C_{m+1})$ and $(C^n_1, \ldots, C^n_{m+1})$. The fact that $|\mathbf{C}(G)| < \infty$ almost surely, as shown in Subsubsection 2.2.2, implies that the induction terminates. $\square$

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 $\mathbf{C}_n(G_n) = (C^n_1, \ldots, C^n_{M_n})$, $\mathbf{C}(G) = (C_1, \ldots, C_M)$, let D^n_i be the index of the vertex that the surplus edge corresponding to C^n_i connects to. Similarly, let D_i be the index of the vertex that the surplus edge corresponding to C_i connects to. Then,*

$$\begin{aligned}&\left(n^{-1/3}\mathcal{T}^{G_n}_n, n^{-2/3}(G_n+1), \left(n^{-2/3}C^n_1, n^{-2/3}D^n_1\right)\ldots, \left(n^{-2/3}C^n_{M_n}, n^{-2/3}D^n_{M_n}\right)\right)\\ &\stackrel{d}{\to} \left(\mathcal{T}^G, G, (C_1, D_1), \ldots, (C_M, D_M)\right)\end{aligned}$$

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

Proof. By definition, for $m \leq M_n$, D^n_m is the vertex corresponding to a uniform unpaired in-half-edge of the vertices of $\mathcal{T}^{G_n}_n(\{G_n+1, C^n_1, \ldots, C^n_m\})$. By

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

the law of D^n_m is asymptotically equal to the index of a uniform vertex on

$$\mathcal{T}_n^{G_n}\left(\{G_n+1, C_1^n, \ldots, C_m^n\}\right).$$

Note that, by Theorem 3.1 and Propositions 4.4, 4.5, we know that

$$\left(n^{-1/3}\mathcal{T}_n^{G_n}, n^{-2/3}G_n+1, n^{-2/3}C_1^n, \ldots, n^{-2/3}C_m^n\right) \stackrel{a.s.}{\to} \left(\mathcal{T}^G, G, C_1, \ldots, C_m\right)$$

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

$$\left|\mathcal{T}_n^{G_n}\left(\{G_n+1, C_1^n, \ldots, C_m^n\}\right)\right| = \left|\mathcal{T}_n^{G_n}\left(\{G_n+1, C_1^n, \ldots, C_m^n, D_m^n\}\right)\right|$$

passes to the limit, with $|\cdot|$ denoting the length in the tree as encoded by $(\hat{H}_n(k), k \geq 1)$, the limit in distribution of $n^{-2/3}D^n_m$ is a uniform point on

$$\mathcal{T}^G\left(G, C_1, \ldots, C_m\right),$$

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 $T^{n,mk}_{G_n}$ be the subtree of $\mathcal{T}_n^{G_n}$ spanned by $\{G_n+1, C_1^n, \ldots, C^n_{M_n}\}$, and similarly, let T_G^{mk} be the subtree of $\mathcal{T}^G$ spanned by $\{G, C_1, \ldots, C_M\}$. Then, also*

$$\left(n^{-1/3}T^{n,mk}_{G_n}, n^{-2/3}(G_n+1), \left(n^{-2/3}C_1^n, n^{-2/3}D_1^n\right)\ldots, \left(n^{-2/3}C^n_{M_n}, n^{-2/3}D^n_{M_n}\right)\right)$$
$$\to \left(T_G^{mk}, G, (C_1, D_1), \ldots, (C_M, D_M)\right)$$

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

$$n^{-1/3}\left|T^{n,mk}_{G_n}\right| \to \left|T_G^{mk}\right|$$

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 $T^{n,\mathrm{mk}}_{G_n}$, set $C_i^n \sim D_i^n$ for each $1 \leq i \leq M_n$, and set $\mathcal{M}^n_{G_n} := T^{n,\mathrm{mk}}_{G_n}/\sim$. Moreover, in T_G^{mk}, set $C_i \sim D_i$ for each $1 \leq i \leq M$, and set $\mathcal{M}_G := T_G^{\mathrm{mk}}/\sim$. View both as elements of $\vec{\mathcal{G}}$ in the natural way. To be precise, in $\mathcal{M}^n_{G_n}$, let the vertex set consist of G_n+1, D_i^n for $i \leq M_n$, and the branch points $C_i^n \wedge C_j^n$ for $i \neq j \leq M_n$. Similarly, in $\mathcal{M}_G$, let the vertex set consist of G, D_i for $i \leq M$, and the branch points $C_i \wedge C_j$ for $i \neq j \leq M$. 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, $n^{-1/3}\mathcal{M}^n_{G_n} \stackrel{a.s.}{\to} \mathcal{M}_G$ in $\vec{\mathcal{G}}$.*

Proof. The proof is analogous to the proof of Proposition 5.6 in [11]. □

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 $n^{-1/3}\mathcal{M}^n_{G_n}$, listed in decreasing order of length, converge to the strongly connected components in $\mathcal{M}_G$, listed in decreasing order of length, in $\vec{\mathcal{G}}$ almost surely as $n \to \infty$.*

Proof. This follows from Proposition 5.3 in [11]. This proposition requires that the lengths of the strongly connected components in $\mathcal{M}_G$ have different lengths almost surely, but this follows from the proof of Proposition 4.6 in [11], noting that our limit object is in the same universality class as theirs. $\square$

Corollary 4.10. *Let $T > 0$, and let $(C_i^T(n), i \geq 1)$ be the strongly connected components in [name for directed graph] with a candidate with head 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 [name for directed graph] with a candidate with head 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 the proof of Proposition 4.6 in [11]. $\square$

Finally, we claim that we can choose T large enough such that all SCC with large length are explored before time T. This is the content of the following lemma. The proof is in the same spirit as Lemma 9 in [2] 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 and whose time of first discovery is in $n^{2/3} I$. Then,*

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

Proof. Fix $\delta > 0$. Suppose there is 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

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

$\square$

Then, Theorem [main result] follows from Corollary 4.10 and Lemma 4.11.

References

[1] Addario-Berry, L., Broutin, N., and Goldschmidt, C. Critical random graphs: limiting constructions and distributional properties. *Electronic Journal of Probability 15* (2010), 741–775.

[2] Aldous, D. The Continuum random tree II: an overview. In *Stochastic Analysis*, M. T. Barlow and N. H. Bingham, Eds. Cambridge University Press, Cambridge, 1991, pp. 23–70.

[3] Broutin, N., Duquesne, T., and Wang, M. Limits of multiplicative inhomogeneous random graphs and Lévy trees: Limit theorems, 2020.

[4] Chaumont, L., and Doney, R. A. Invariance principles for local times at the maximum of random walks and Lévy processes. *Annals of Probability 38*, 4 (jul 2010), 1368–1389.

[5] Conchon-Kerjan, G., and Goldschmidt, C. The stable graph: the metric space scaling limit of a critical random graph with i.i.d. power-law degrees, 2020.

[6] Cooper, C., and Frieze, A. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combinatorics, Probability and Computing 13*, 3 (2004), 319–337.

[7] Daley, D., and Vere-Jones, D. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and Its Applications. Springer New York, 2007.

[8] de Raphélis, L. Scaling limit of multitype Galton–Watson trees with infinitely many types. *Ann. Inst. H. Poincaré Probab. Statist. 53*, 1 (02 2017), 200–225.

[9] Duquesne, T., and Le Gall, J.-F. Random Trees, Lévy Processes and Spatial Branching Processes, Astérisque (281).

[10] Ethier, S. N., and Kurtz, T. G. *Markov Processes: Characterization and Convergence*. Wiley, 1986.

[11] Goldschmidt, C., and Stephenson, R. The scaling limit of a critical random directed graph, 2019.

[12] Joseph, A. The component sizes of a critical random graph with given degree sequence. *Annals of Applied Probability 24*, 6 (dec 2014), 2560–2594.

[13] Le Gall, J.-F. Random trees and applications. *Probability Surveys 2* (2005), 245–311.