

THE SCALING LIMIT OF A CRITICAL RANDOM DIRECTED GRAPH

BY CHRISTINA GOLDSCHMIDT^{1,a} , AND ROBIN STEPHENSON^{2,b} 

¹Department of Statistics and Lady Margaret Hall, University of Oxford, [a goldschm@stats.ox.ac.uk](mailto:goldschm@stats.ox.ac.uk)

²School of Mathematics and Statistics, University of Sheffield, [b robin.stephenson@normalesup.org](mailto:robin.stephenson@normalesup.org)

We consider the random directed graph $\vec{G}(n, p)$ with vertex set $\{1, 2, \dots, n\}$ in which each of the $n(n-1)$ possible directed edges is present independently with probability p . We are interested in the strongly connected components of this directed graph. A phase transition for the emergence of a giant strongly connected component is known to occur at $p = 1/n$, with critical window $p = 1/n + \lambda n^{-4/3}$ for $\lambda \in \mathbb{R}$. We show that, within this critical window, the strongly connected components of $\vec{G}(n, p)$, ranked in decreasing order of size and rescaled by $n^{-1/3}$, converge in distribution to a sequence $(\mathcal{C}_1, \mathcal{C}_2, \dots)$ of finite strongly connected directed multigraphs with edge lengths which are either 3-regular or loops. The convergence occurs in the sense of an ℓ^1 sequence metric for which two directed multigraphs are close if there are compatible isomorphisms between their vertex and edge sets which roughly preserve the edge lengths. Our proofs rely on a depth-first exploration of the graph which enables us to relate the strongly connected components to a particular spanning forest of the undirected Erdős–Rényi random graph $G(n, p)$, whose scaling limit is well understood. We show that the limiting sequence $(\mathcal{C}_1, \mathcal{C}_2, \dots)$ contains only finitely many components which are not loops. If we ignore the edge lengths, any fixed finite sequence of 3-regular strongly connected directed multigraphs occurs with positive probability.

1. Introduction and main result. Many real-world networks are inherently directed in nature. Consider, for example, the World Wide Web: hyperlinks point from one webpage to another but the link in the other direction is not necessarily present. However, this structurally important feature is often ignored in modelling, and the corresponding mathematical literature is much less well-developed. In this paper, we consider the simplest possible model of a random directed graph and endeavour to understand the way in which its directed connectivity properties change as we adjust its parameters.

Let $\vec{G}(n, p)$ be a random directed graph with vertex set $[n] := \{1, \dots, n\}$ and random edge set where each of the $n(n-1)$ possible edges (i, j) , $i \neq j$, is present independently with probability p . We are interested in the *strongly connected components* of $\vec{G}(n, p)$, that is the maximal subgraphs for which there exists a directed path from a vertex to any other.

The usual Erdős–Rényi random graph, $G(n, p)$, in which each of the $n(n-1)/2$ possible *undirected* edges is present independently with probability p , will play an important role in our results. It is well known that $G(n, p)$ undergoes a phase transition [7]: if $np \rightarrow c > 1$ as $n \rightarrow \infty$ then $G(n, p)$ has a unique giant component with high probability, while if $np \rightarrow c < 1$ as $n \rightarrow \infty$ then the components of $G(n, p)$ are of size $O_{\mathbb{P}}(\log n)$. In the so-called *critical window*, where $p = \frac{1}{n} + \lambda n^{-4/3}$, Aldous [3] proved that the sequence of sizes of the largest components possesses a distributional limit when renormalised by $n^{2/3}$.

Previous work by Karp [14] and Łuczak [17] has shown that $\vec{G}(n, p)$ undergoes a similar phase transition to that of $G(n, p)$: if $np \rightarrow c > 1$ as $n \rightarrow \infty$, then $\vec{G}(n, p)$ has a unique

MSC2020 subject classifications: Primary 60C05, 05C80; secondary 05C20, 60F05.

Keywords and phrases: random directed graph, phase transition, scaling limit.

giant strongly connected component with high probability, while if $np \rightarrow c < 1$ as $n \rightarrow \infty$, then the sizes of all the strongly connected components are $o_{\mathbb{P}}(n)$. (In $\vec{G}(n, p)$, in contrast to the situation for $G(n, p)$, Łuczak [17] shows that the sizes of the subcritical components are tight.) These results were strengthened by Łuczak and Seierstad [18], who showed that $\vec{G}(n, p)$ has, in fact, the same critical window as $G(n, p)$.

THEOREM 1.1 (Łuczak and Seierstad [18]). *Let $\gamma_n = (np - 1)n^{1/3}$ and assume $\gamma_n = o(n^{1/3})$ as $n \rightarrow \infty$.*

1. *If $\gamma_n \rightarrow \infty$ then the largest strongly connected component of $\vec{G}(n, p)$ has size $(4 + o_{\mathbb{P}}(1))\gamma_n^2 n^{1/3}$ and the second largest has size $O_{\mathbb{P}}(\gamma_n^{-1} n^{1/3})$.*
2. *If $\gamma_n \rightarrow -\infty$ then the largest strongly connected component of $\vec{G}(n, p)$ has size $O_{\mathbb{P}}(|\gamma_n^{-1}| n^{1/3})$.*

However, in contrast to $G(n, p)$, Łuczak and Seierstad also show that within the critical window, the complex strongly connected components (that is, those which do not just consist of a single directed cycle) occupy only $O_{\mathbb{P}}(n^{1/3})$ vertices in total. This shows that the critical components are very much “thinner” objects than in the setting of $G(n, p)$, where the complex components occupy $O_{\mathbb{P}}(n^{2/3})$ vertices.

In a recent work [6], Coulson shows that, on rescaling by $n^{-1/3}$, the size of the largest strongly connected component of $\vec{G}(n, p)$ in the critical window is tight, with explicit upper and lower tail bounds.

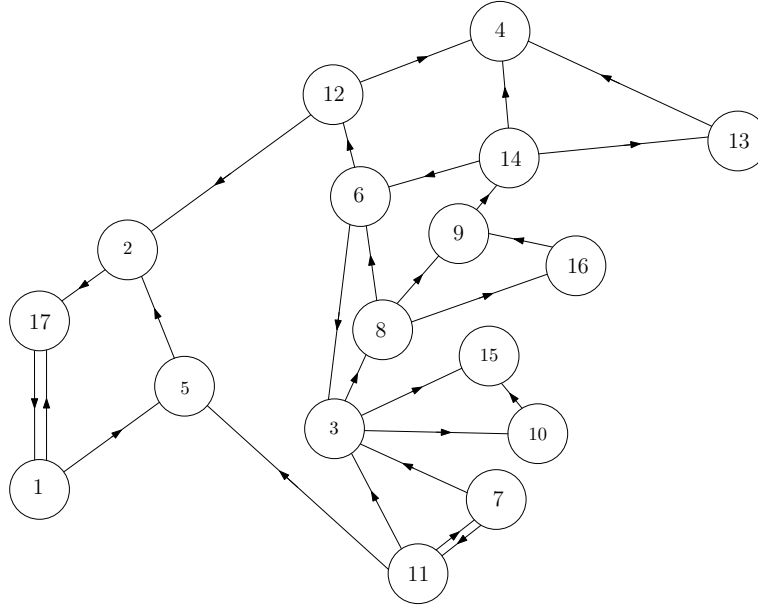


FIG 1. A directed graph on [17]. Its strongly connected components have vertex sets $\{3, 6, 8, 9, 14, 16\}$, $\{1, 2, 5, 17\}$, $\{7, 11\}$, $\{4\}$, $\{10\}$, $\{12\}$, $\{13\}$ and $\{15\}$.

In this paper, we investigate the behaviour within the critical window in more detail, and in particular we prove a scaling limit for the strongly connected components. We do this by relating a particular subgraph of $\vec{G}(n, p)$ to a spanning forest of $G(n, p)$, and the convergence of that spanning forest (thought of as a collection of discrete metric spaces, one per component) to a collection of random \mathbb{R} -trees. Similar tools have already been used to study the components of $G(n, p)$ in the same critical window, leading to the main theorem of [2].

THEOREM 1.2 (Addario-Berry, Broutin and Goldschmidt [2]). *Let $p = p(n) = \frac{1}{n} + \lambda n^{-4/3}$ for fixed $\lambda \in \mathbb{R}$. Let $(A_1(n), A_2(n), \dots)$ be the connected components of $G(n, p)$, each considered as a metric space by endowing the vertex-set with the graph distance. Then*

$$\left(\frac{A_i(n)}{n^{1/3}}, i \in \mathbb{N} \right) \xrightarrow{(d)} (\mathcal{A}_i, i \in \mathbb{N}),$$

where $\mathcal{A} = (\mathcal{A}_i, i \in \mathbb{N})$ is a random sequence of compact metric spaces, and the convergence is in distribution for the ℓ^4 metric for sequences of compact metric spaces based on the Gromov–Hausdorff distance.

Let us immediately give a description of the scaling limit \mathcal{A} , since it plays an important role in the sequel. Define $W^\lambda(t) = W(t) + \lambda t - t^2/2$ for $t \geq 0$, where W is a standard Brownian motion, and let $(\sigma_i, i \in \mathbb{N})$ be the collection of excursion lengths above the running infimum of W^λ , ranked in decreasing order. For $\sigma > 0$, let $\tilde{e}^{(\sigma)}$ be a Brownian excursion with length σ biased by the exponential of its area, so that if $e^{(\sigma)}$ is a Brownian excursion of length σ then, for any non-negative measurable test function g , we have

$$\mathbb{E} \left[g \left(\tilde{e}^{(\sigma)} \right) \right] = \frac{\mathbb{E} \left[\exp \left(\int_0^\sigma e^{(\sigma)}(u) du \right) g \left(e^{(\sigma)} \right) \right]}{\mathbb{E} \left[\exp \left(\int_0^\sigma e^{(\sigma)}(u) du \right) \right]}.$$

Let \mathcal{T}_σ be the \mathbb{R} -tree encoded by $2\tilde{e}^{(\sigma)}$ (see Section 3.2.1 below for a description of how this is done). We make some additional point-identifications in this tree. Let (t_1, \dots, t_K) be the points of a Poisson random measure on $[0, \sigma]$ with intensity $\tilde{e}^{(\sigma)}(t)dt$. The point $t_j \in [0, \sigma]$ corresponds to a point x_j in \mathcal{T}_σ at distance $2\tilde{e}^{(\sigma)}(t_j)$ from the root. For all $1 \leq j \leq K$, we identify x_j with a uniformly chosen point on its path to the root. Write \mathcal{G}_σ for the resulting metric space. Finally, conditionally on $(\sigma_i, i \in \mathbb{N})$, the metric spaces $\mathcal{A}_1, \mathcal{A}_2, \dots$ are independent and, for each $i \in \mathbb{N}$, \mathcal{A}_i has the law of \mathcal{G}_{σ_i} .

While metric spaces provide the natural setting in which to consider scaling limits of undirected graphs, this is no longer the case in the directed setting: we need some extra structure to encode the orientations. Let us make some useful definitions.

By a *directed multigraph*, we mean a triple (V, E, r) where

- V and E are finite sets.
- $r = (r_1, r_2)$ is a function from E to $V \times V$, with $r_1(e)$ and $r_2(e)$ for $e \in E$ being respectively the *tail* and *head* of the directed edge e .

We will refer to the case where $V = \{v\}$, $E = \{e\}$ and $r_1(e) = r_2(e) = v$ as a *loop*. $X = (V, E, r, \ell)$ is a *metric directed multigraph* (henceforth MDM) if (V, E, r) is a directed multigraph and ℓ is a function from E to $[0, \infty)$ which assigns each edge a length. A special role will be played by the degenerate case of a loop whose single edge is assigned length 0, which we denote by \mathfrak{L} . The *length* $\text{len}(X)$ of X is given by $\sum_{e \in E} \ell(e)$.

We now define a distance between MDMs $X = (V, E, r, \ell)$ and $X' = (V', E', r', \ell')$ in such a way that they are close if there is a graph isomorphism from X to X' which changes the lengths very little. Specifically, let $\text{Isom}(X, X')$ be the set of graph isomorphisms from X to X' , that is pairs of bijections f from V to V' and g from E to E' such that, for all $e \in E$, $r'(g(e)) = (f(r_1(e)), f(r_2(e)))$. Then set

$$d_{\bar{\mathcal{G}}}(X, X') = \inf_{(f, g) \in \text{Isom}(X, X')} \sup_{e \in E} |\ell(e) - \ell'(g(e))|.$$

Note that if X and X' do not have the same graph structure, then $\text{Isom}(X, X')$ is empty and $d_{\vec{\mathcal{G}}}(X, X')$ is set to infinity. Let $\vec{\mathcal{G}}$ be the set of (isometry classes of) MDMs. Then $(\vec{\mathcal{G}}, d_{\vec{\mathcal{G}}})$ is a Polish space (as a countable union of powers of $[0, +\infty)$).

Let $C_i(n)$ for $i \geq 1$ be the strongly connected components of $\vec{G}(n, p)$, listed in decreasing order of size, breaking ties by increasing order of the lowest labelled vertex. We view these strongly connected components as MDMs, by assigning to each edge a length of 1, and then removing all vertices with degree 2 and merging their corresponding edges into paths of length greater than 1. In the case of a strongly connected component which consists of a single directed cycle with $k \geq 2$ vertices, we think of it as a loop of length k . Similarly, we think of isolated vertices as loops of length 0. Finally, since there are at most n components, we complete the list with an infinite repeat of \mathcal{L} , the loop of length 0.

We can now state our main theorem.

THEOREM 1.3. *Suppose $p = p(n) = \frac{1}{n} + \lambda n^{-4/3} + o(n^{-4/3})$. There exists a sequence $\mathcal{C} = (C_i, i \in \mathbb{N})$ of random strongly connected MDMs such that, for each $i \geq 1$, C_i is either 3-regular or a loop, and such that*

$$(1) \quad \left(\frac{C_i(n)}{n^{1/3}}, i \in \mathbb{N} \right) \xrightarrow{(d)} (C_i, i \in \mathbb{N})$$

with respect to the distance d defined by

$$d(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{\infty} d_{\vec{\mathcal{G}}}(A_i, B_i),$$

for $\mathbf{A}, \mathbf{B} \in \vec{\mathcal{G}}^{\mathbb{N}}$.

In particular, the limit object \mathcal{C} has finite total length. We will show later that \mathcal{C} has only finitely many complex components (i.e. components which are not loops). So Theorem 1.3 implies the convergence in distribution of the number of complex components of $\vec{G}(n, p)$, their rescaled numbers of vertices, and their excesses (where the excess of a component is given by its number of edges minus its number of vertices). This, in particular, significantly strengthens Theorems 13 and 14 of [18]. Finally, we also show that, if we ignore the edge lengths, then any fixed finite sequence of 3-regular strongly connected directed multigraphs occurs with positive probability.

We defer a proper description of \mathcal{C} , which is rather involved, to Section 4 below. As is the case for $(\mathcal{A}_i, i \in \mathbb{N})$, the $(C_i, i \in \mathbb{N})$ are derived from the \mathbb{R} -trees encoded by the excursions of W^λ . However, the strongly connected components $(C_i, i \in \mathbb{N})$ are much simpler objects than $(\mathcal{A}_i, i \in \mathbb{N})$ which, for example, have a rich fractal structure coming from their relationship to the Brownian continuum random tree. A closer analogy is obtained by instead looking at the scaling limit of the subgraph of $G(n, p)$ consisting only of edges and vertices which lie in cycles. Each component of the graph has a *core*, which is defined to be the maximal subgraph of minimum degree 2, and consists of the vertices and edges which lie in cycles, as well as those in paths joining cycles. (The core can be obtained by successively deleting leaves and their incident edges from the graph until no leaves remain.) The core is empty if there are no cycles. If the core is non-empty, removing those of its edges which do not lie in a cycle yields one or more components, which represent the cycle structure of the original component.

It is possible to define an analogous notion of a core for each component $\mathcal{A}_i, i \in \mathbb{N}$ of the scaling limit of the critical random undirected graph, created by the point-identifications we make in the \mathbb{R} -trees encoded by the excursions of W^λ . Indeed, for each $i \geq 1$, $\text{core}(\mathcal{A}_i)$ is

a connected undirected multigraph with edge lengths which is empty if there are no point-identifications, is a loop if there is a single point-identification, and is otherwise 3-regular almost surely. For each $i \geq 1$, if $\text{core}(\mathcal{A}_i)$ is non-empty and we remove from it any points not contained in cycles, we obtain a collection of one or more multigraphs with edge-lengths which are again either loops or 3-regular. Let us refer to these as the *cycle-components*. The MDMs $(\mathcal{C}_i, i \in \mathbb{N})$ are similarly obtained by making (a different collection of) point-identifications in the \mathbb{R} -trees encoded by the excursions of W^λ . In this context, a single \mathbb{R} -tree may give rise to one or more strongly connected components, or indeed none. The fact that we obtain an ℓ^1 convergence in Theorem 1.3, comes from the property that for very small σ , an \mathbb{R} -tree with the same distribution as \mathcal{T}_σ is very unlikely to produce any strongly connected components at all.

It would be interesting to know if the distributions of the undirected version of $(\mathcal{C}_i, i \in \mathbb{N})$ and of the decreasing ordering of all the cycle-components coming from $(\text{core}(\mathcal{A}_i), i \in \mathbb{N})$ are mutually absolutely continuous. We leave this as an open problem.

The rest of this paper is structured as follows. In Section 2, we introduce some standard terminology and then describe the depth-first exploration which we use in order to understand the directed graph $\vec{G}(n, p)$. A key role is played by a particular class of edges known as back edges, and we discuss back edges in both the discrete and continuum settings in Section 3. In Section 4, we prove some useful properties of the scaling limit \mathcal{C} . Section 5 contains the proof of Theorem 1.3. In Section 6, we prove the further properties of the scaling limit which were mentioned immediately after the main theorem.

2. Some graph theory.

2.1. *Basic terminology.* We recall here some elementary graph theoretic terminology which we will use throughout the paper.

Directed graphs and strongly connected components. Let \vec{G} be a directed graph. For a directed edge (x, y) of \vec{G} , we say that x is the *tail* of the edge and y is its *head*. For two vertices x and y , we also say that x is a *parent* of y (and y is a *child* of x) if there is an edge from x to y , and that x is an *ancestor* of y (and y is a *descendant* of x) if there is a directed path from x to y .

A directed graph \vec{G} is *strongly connected* if for every pair $\{u, v\}$ of distinct vertices of \vec{G} there exists a directed path from u to v and a directed path from v to u . For a general directed graph \vec{G} , its *strongly connected components* are the maximal strongly connected subgraphs. The strongly connected components partition the vertex set but note that, unlike for undirected graphs, edges of \vec{G} may lead from one strongly connected component to another.

Trees and plane trees. A discrete tree is a connected undirected graph T with no cycles. For two vertices x and y in T , we write $\llbracket x, y \rrbracket$ for the unique path between x and y . Our trees will often be rooted at a specified vertex ρ . This allows us to think of T as a *directed* graph, by orienting all of its edges away from ρ . We write $|T|$ for the size of the vertex set of T and $\|T\|$ for the *height* of T , that is the largest distance between ρ and another vertex.

A *planar ordering*, also known as *topological sort*, of a rooted tree T is any total order $>$ on its vertex set such that every directed edge (u, v) of T is increasing, in the sense that $v > u$ (decreasing edges are defined similarly). A rooted plane tree is then a rooted tree endowed with a planar ordering.

Directed multigraphs. Recall the definition of a directed multigraph from the introduction. Directed multigraphs have the same notion of ancestor and descendant as directed graphs,

and have strongly connected components in the same way. Note that the loop is strongly connected. The *excess* of a strongly connected directed multigraph (V, E, r) is defined to be $|V| - |E|$. If the excess is strictly positive then we say that the multigraph is *complex*.

2.2. The exploration process. The strongly connected components of any directed graph can be found in time which is linear in the sum of the sizes of the vertex and edge sets. Several linear-time algorithms, including Tarjan's algorithm [23] and the so-called path-based algorithms (see [8] for an example), rely on a *depth-first search*, that is a procedure which consists in exploring the graph in such a way that, after we visit a vertex, we visit all of its as-yet unseen descendants before backtracking. Broadly speaking, as we traverse the graph, some information is kept in the form of a stack, which allows us to determine the strongly connected components.

For our study of $\vec{G}(n, p)$, we use a variant of these ideas to give a simple algorithm which does not directly yield the strongly connected components, but instead gives a specific *plane* spanning forest which will be a key part of the structure of the strongly connected components. In order to find our spanning forest, we use the now-standard ordered depth-first search exposed, for example, in [2], but with the modification that we only allow ourselves to follow edges in the direction of their orientation. The standard ordering on the vertex set $[n]$ is used to induce a planar ordering on the out-neighbours of a vertex. Let us give a precise definition of the construction and, along the way, remind the reader of the depth-first exploration for undirected graphs. Let \vec{G} (resp. G) be any directed graph (resp. undirected graph) on $[n]$. Inductively on $i \in \{0, \dots, n\}$, we define an ordered list \mathcal{O}_i of open vertices (the *stack*) which have been seen but not yet explored, and a set \mathcal{E}_i of explored vertices:

- $i = 0$: let $\mathcal{O}_0 = (1)$ and $\mathcal{E}_0 = \emptyset$.
- Induction step: given \mathcal{O}_i and \mathcal{E}_i , let v_i be the first vertex of \mathcal{O}_i and let $\mathcal{E}_{i+1} = \mathcal{E}_i \cup \{v_i\}$. Let \mathcal{N}_i be the set of out-neighbours (resp. neighbours) of v_i which are not in $\mathcal{O}_i \cup \mathcal{E}_i$. Construct \mathcal{O}_{i+1} by removing v_i from \mathcal{O}_i , and adding in the elements of \mathcal{N}_i in increasing order, so that the smallest element of \mathcal{N}_i is now at the start of \mathcal{O}_{i+1} . If, however, this leads to $\mathcal{O}_{i+1} = \emptyset$, then add to it the smallest element of $\{1, \dots, n\} \setminus \mathcal{E}_{i+1}$.

This procedure builds a directed spanning forest $\mathcal{F}_{\vec{G}}$ of \vec{G} , by saying that two vertices x and y are linked by an edge from x to y if there exists i for which $x = v_i$ and $y \in \mathcal{N}_i$. This is illustrated in Figure 2, for the graph given by Figure 1. We call $\mathcal{F}_{\vec{G}}$ the *forward depth-first forest* of \vec{G} .

We also obtain a total order of $[n]$, given by (v_0, \dots, v_{n-1}) , which is a planar ordering of $\mathcal{F}_{\vec{G}}$, in the sense that it is a topological sort of each of its trees and it also functions as a total order on the set formed by the trees. We write $v_i \prec v_j$ iff $i < j$. The edges of \vec{G} may now be partitioned into two categories: the *forward edges*, which are increasing for this order, and the *back edges*, which are decreasing. The forward edges can themselves also be separated into two sets: those which are edges of $\mathcal{F}_{\vec{G}}$, and those which are not, which we call *surplus edges*. (In the case of the undirected graph G , we still get a forest \mathcal{F}_G , but all edges of G are either part of the forest or are surplus edges.)

The combination of forward edges and back edges is what creates the strongly connected components of \vec{G} . Notice in particular that, since there are no forward edges going between different trees of $\mathcal{F}_{\vec{G}}$, each strongly connected component lies within a single such tree. Moreover, since strongly connected components are made of cycles, any strongly connected component with at least two vertices must contain at least one forward and one back edge. We call an *ancestral* back edge one which goes from a vertex to one of its ancestors. See Figure 2 for an illustration. The ancestral back edges will play a particularly important role in the sequel.

LEMMA 2.1. *Each strongly connected component contains either a surplus edge or an ancestral back edge.*

PROOF. Suppose there are no surplus edges. Any strongly connected component has a least element x_0 for the depth-first ordering. Then the vertex x_0 can only have in-edges from vertices which are later in the ordering, and must have at least one such. Take the smallest of these, namely the edge (x_1, x_0) with the smallest x_1 for the ordering. There must be a path from x_0 to x_1 . We claim that this path can use only forward edges of the tree. Suppose not. Then x_1 does not belong to the subtree rooted at x_0 . But then any path from x_0 to x_1 must use at least one surplus edge, which contradicts the assumption that there are none. It follows that x_0 is an ancestor of x_1 , and so the edge (x_1, x_0) is ancestral. \square

We deduce from this a useful bound: the number of strongly connected components of \vec{G} is smaller than the sum of its numbers of surplus edges and ancestral back edges.

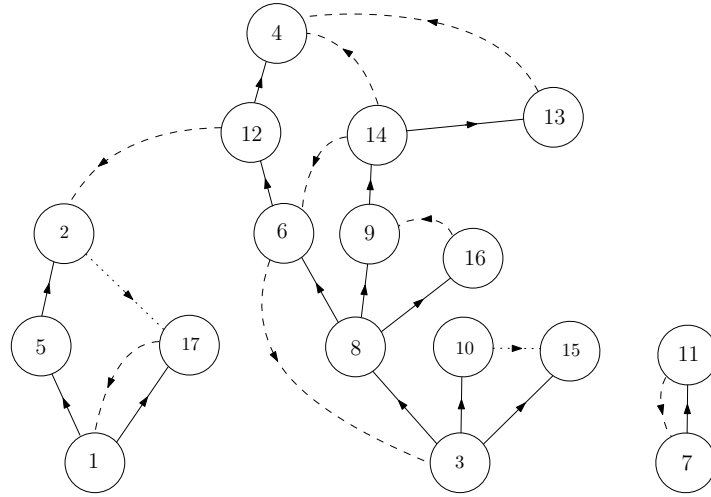


FIG 2. *The planar embedding of exploration forest of the graph in Figure 1. Surplus edges and back edges are then dotted, and respectively straight and curved. Notice that there are back edges of both kinds: ancestral, such as $(6, 3)$, and non-ancestral, such as $(14, 6)$. The strongly connected components have vertex sets $\{3, 6, 8, 9, 14, 16\}$, $\{1, 2, 5, 17\}$, $\{7, 11\}$, $\{4\}$, $\{10\}$, $\{12\}$, $\{13\}$ and $\{15\}$.*

Note that the surplus edges of G are taken from the set of edges *permitted* by $\mathcal{F}_{\vec{G}}$, which are the pairs (u, v) such that there exists i such that u and v are both in \mathcal{O}_i . In this case, v is a sibling of an ancestor of u which occurs later in the planar ordering.¹ In fact, given $\mathcal{F}_{\vec{G}}$, we can add or remove any permitted edge to \vec{G} , and this will not change $\mathcal{F}_{\vec{G}}$. The same holds true for back edges. Thus, conditionally on $\mathcal{F}_{\vec{G}(n,p)}$, the permitted surplus edges and back edges of $\vec{G}(n,p)$ appear independently with probability p . This leads to the following proposition, which allows us to relate $\vec{G}(n,p)$ to $G(n,p)$ by their explorations.

¹Note also that $\mathcal{F}_{\vec{G}}$ determines the exploration process fully, so defining the permitted edges using the \mathcal{O}_i is unambiguous.

PROPOSITION 2.2. *For any directed graph \vec{G} on $[n]$ we call \vec{G}_{fwd} the undirected graph obtained by removing the back edges of \vec{G} and keeping the forward edges, but ignoring their orientation. We then have the following:*

1. $\mathcal{F}_{\vec{G}(n,p)} \stackrel{(d)}{=} \mathcal{F}_{G(n,p)}$
2. $(\vec{G}(n,p))_{\text{fwd}} \stackrel{(d)}{=} G(n,p)$
3. *One can couple $G(n,p)$ and $\vec{G}(n,p)$ in the following way: first sample $G(n,p)$, which creates in particular a depth-first ordering on $\{1, \dots, n\}$. Then let $(\vec{G}(n,p))_{\text{fwd}} = G(n,p)$, and add to it each of the possible back edges (v_i, v_j) for $j < i$ independently with probability p .*

PROOF. The proof of the first statement is straightforward by induction: notice that, in the explorations of both $\vec{G}(n,p)$ and $G(n,p)$, for all i , given \mathcal{O}_i and \mathcal{E}_i , the neighbourhood \mathcal{N}_i contains each element of $\{1, \dots, n\} \setminus (\mathcal{O}_i \cup \mathcal{E}_i)$ independently with probability p . Thus, each step of the forward exploration of $\vec{G}(n,p)$ has the same distribution as the corresponding step of the depth-first exploration of $G(n,p)$, and in particular the forests they build have the same distribution.

The second statement is obtained by observing that, both for $\vec{G}(n,p)$ and $G(n,p)$, given the exploration forest, each permitted surplus edge is present independently with probability p . Similarly, the third statement follows from the fact that, given $\vec{G}_{\text{fwd}}(n,p)$, each back edge is present independently with probability p . \square

This proposition motivates the study of a process which adds back edges to trees. The next section will formalise this, especially for the continuum trees which arise in the scaling limit of $G(n, 1/n + \lambda n^{-4/3})$.

3. Back edges on discrete and continuum trees. We show that, when considering a plane tree with additional back edges, we can safely ignore a portion of the back edges and keep the same strongly connected components. We then adapt this idea to give a procedure for building a random finite set of backward identifications on a continuum tree, which is how \mathcal{C} will be built.

3.1. *The discrete case.* Let $T = (V(T), E(T))$ be a finite rooted plane tree, with root ρ and $|V(T)| = n$. We think of T as a directed graph, by orienting all the edges away from ρ . Let us write $(v_0, v_1, \dots, v_{n-1})$ for the vertices of T listed in the depth-first order explained in Section 2.2. Recall that for two vertices x and y of T , $\llbracket x, y \rrbracket$ is the path between x and y .

Consider a set B of additional edges between elements of $V(T)$ which go backwards for the depth-first order, so that any element of B is of the form (v_i, v_j) with $v_j \prec v_i$ (i.e. $j < i$). Such an edge is called *ancestral* if v_j is an ancestor of v_i in T . It is useful to have an ordering on the elements of B , and we do this by declaring that $(v_i, v_j) \prec (v_k, v_l)$ if $i < k$ or if $i = k$ and $j < l$. (This is just the lexicographic ordering on the pairs of indices (i, j) and (k, l) .) Note that the elements of B are thus listed in the order in which we would encounter them when performing a depth-first search of the directed graph $(V(T), E(T) \cup B)$.

We now extract a subset $B^* = \{(x_i, y_i), i \leq N\}$ of B inductively as follows (as illustrated in Figure 3).

- Let (x_1, y_1) be the first ancestral back edge in B .
- Assume that we are now given (x_j, y_j) for $j \leq i$. For $x \in V(T)$ such that $x \succ x_i$, define $T(i, x) = \bigcup_{j=1}^i \llbracket \rho, x_j \rrbracket \cup \llbracket \rho, x \rrbracket$. Let (x_{i+1}, y_{i+1}) be the smallest element (x, y) of $B \setminus \{(x_j, y_j), j \in \{1, \dots, i\}\}$ such that $y \in T(i, x)$. If there is no such element, then we end the procedure, and set $N = i$.

Observe that all of the ancestral edges in B are elements of B^* , since if (x, y) is ancestral then $y \in [[\rho, x]]$. Moreover, notice that, in the inductive part of the definition of B^* , we could equivalently have defined (x_{i+1}, y_{i+1}) to be smallest element (x, y) of $B \setminus \{(x_j, y_j), j \in \{1, \dots, i\}\}$ which is either ancestral or such that there is a directed path from y towards an ancestral back edge. This implies that B^* is exactly the set of back edges which are either themselves ancestral, or which lead to ancestral back edges.

PROPOSITION 3.1. *Let X be the directed graph obtained by taking T (with edges directed away from ρ) and adding all the edges of B . Let X^* be the subgraph of X where we remove any element of B that is not in B^* . Then X and X^* have the same strongly connected components.*

PROOF. Let (x, y) be an element of B which is in a strongly connected component of X . We want to prove that it is in B^* . Since there are no surplus edges, by Lemma 2.1, the strongly connected component must contain an ancestral back edge. Thus, it is possible to reach that ancestral edge starting from y and following edges of X . Hence, (x, y) is in B^* . It follows that X and X^* have the same strongly connected components. \square

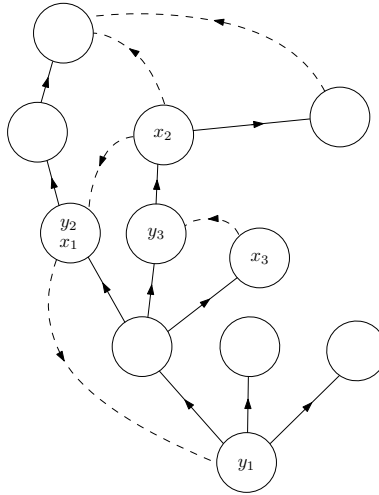


FIG 3. The construction of Section 3.1 applied to the second tree of Figure 2. Notice that the two topmost back edges are not contained in a strongly connected component, and they are not of the form (x_i, y_i) .

This seemingly innocuous lemma is, in fact, a key tool for us. Indeed, if T is taken to be a large tree of $\mathcal{F}_{\vec{G}(n,p)}$ (meaning it has size of order $n^{2/3}$), and B is the set of back edges of $\vec{G}(n, p)$ which join elements of T , then B has size of order $n^{1/3}$. However, as we will see later, the number of back edges in X^* remains of order 1, and in fact the (x_i, y_i) will, in the scaling limit, converge to continuous analogues. This means that the reduction from X to X^* , while not changing the strongly connected components, allows us to ignore the majority of the back edges at no cost. (Let us emphasise, however, that in general there are back-edges in X^* which are *not* contained in any strongly connected component.)

3.2. The continuum case.

3.2.1. *\mathbb{R} -trees and notation.* We recall here some basic terminology about \mathbb{R} -trees; more information concerning their use in probability may be found in the survey paper [16]. An \mathbb{R} -tree is any metric space (\mathcal{T}, d) such that

- For all $x, y \in \mathcal{T}$, there exists a unique distance-preserving map $\phi_{x,y}$ from $[0, d(x, y)]$ into \mathcal{T} such $\phi_{x,y}(0) = x$ and $\phi_{x,y}(d(x, y)) = y$. We write $[[x, y]]$ for the image of $\phi_{x,y}$.
- For all continuous and one-to-one functions $c: [0, 1] \rightarrow \mathcal{T}$, we have $c([0, 1]) = [[c(0), c(1)]]$.

Our \mathbb{R} -trees will typically be *rooted*, which means we distinguish a point of \mathcal{T} called the root, usually denoted ρ . For $x, y \in \mathcal{T}$, we say that x is an *ancestor* of y , or that y is a *descendant* of x , if $x \in [[\rho, y]]$, and we call the point $x \wedge y$ such that $[[\rho, x]] \cap [[\rho, y]] = [[\rho, x \wedge y]]$ the *most recent common ancestor* of x and y . The *degree* of a point $x \in \mathcal{T}$ is the number of connected components of $\mathcal{T} \setminus \{x\}$. If x has degree 1 we call it a *leaf*. The *height* $\|\mathcal{T}\|$ of \mathcal{T} is the largest distance from ρ to another point.

The \mathbb{R} -trees we encounter will all be encoded by functions. For $\sigma \in (0, \infty)$, a function $f: [0, \sigma] \rightarrow \mathbb{R}_+$ is called an *excursion function* if it is continuous and $f(x) = 0$ if and only if $x = 0$ or σ . Let $\hat{f}: [0, \sigma]^2 \rightarrow \mathbb{R}_+$ be defined by $\hat{f}(x, y) = \min_{s \in [x \wedge y, x \vee y]} f(s)$. Then f encodes a

pseudo-distance d_f on $[0, \sigma]$, defined by $d_f(x, y) = f(x) + f(y) - 2\hat{f}(x, y)$, and an \mathbb{R} -tree \mathcal{T}_f , defined by

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

The natural projection from $[0, \sigma]$ to \mathcal{T}_f will be called p_f , and we let the root of \mathcal{T}_f be $p_f(0) = p_f(\sigma)$. \mathcal{T}_f also inherits a natural total order from $[0, \sigma]$ which we call the *planar order*.

In the sequel, we will assume that the functions f we consider have unique local minima. Under this assumption, the resulting \mathbb{R} -tree \mathcal{T}_f is binary (meaning its points all have degree at most 3). We also assume that the push-forward of the Lebesgue measure on $[0, \sigma]$ onto \mathcal{T}_f via p_f is concentrated on the leaves so that, in particular, if U is uniform on $[0, \sigma]$ then $p_f(U)$ is a leaf with probability 1. These properties hold almost surely for the *random excursion functions* we will use in Subsection 4.2.

3.2.2. *Constructing the identifications.* We now describe a random process which will give us a finite number of point identifications which go backward for this planar ordering: pairs of points of the form (x, y) with $x > y$ for the planar ordering and an “arrow with zero length” pointing from x to y . Specifically, we will define times $(s_i, i \in \{1, \dots, N\})$ in $[0, \sigma]$, their projections $x_i = p_f(s_i)$ and points y_i such that y_i is in the subtree $\mathcal{T}_i = \bigcup_{j=1}^i [[\rho, x_j]]$.

We start with the base case $i = 1$. Let s_1 be the first point of a Poisson point process on $[0, \sigma]$ with intensity $f(x)dx$, and $x_1 = p_f(s_1)$. Then let y_1 be a uniform random point on the segment $[[\rho, x_1]]$. If the Poisson point process has no points on $[0, \sigma]$, we let $N = 0$.

Now let us assume that (s_j, x_j, y_j) have been built for $j \in \{1, \dots, i\}$. For $t \in [s_i, \sigma]$, consider the subtree $\mathcal{T}_i(t) = \mathcal{T}_i \cup [[\rho, p_f(t)]]$ and its length $\ell_i(t)$. Then, straightforwardly,

$$(2) \quad \ell_i(t) = \sum_{j=1}^i \left(f(s_j) - \hat{f}(s_{j-1}, s_j) \right) + f(t) - \hat{f}(t, s_i).$$

Now let s_{i+1} be the first arrival time of an independent Poisson point process on $[s_i, \sigma]$ of intensity $\ell_i(t)dt$, let $x_{i+1} = p_f(s_{i+1})$, and let y_{i+1} be uniform on the finite length space $\mathcal{T}_i(s_i)$, independently of everything else. If the Poisson point process has no points, we let $N = i$.

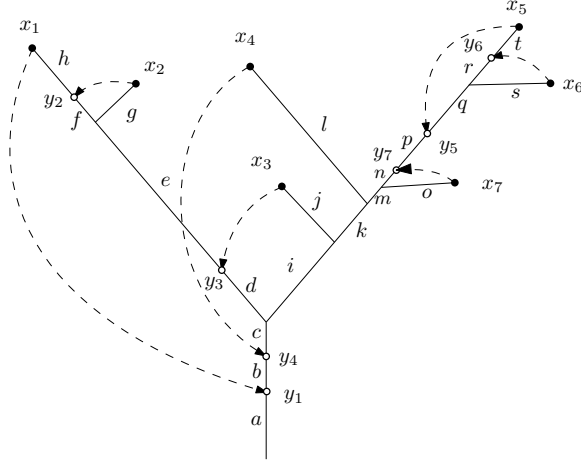


FIG 4. Some identifications on a tree. Lengths of the segments are represented by a, \dots, t for the next figures.

Observe that we necessarily have $s_1 \leq s_2 \leq \dots \leq \sigma$ (so that, in particular, the first back edge in the planar ordering is always ancestral). It is, however, in principle possible for the sequence $(s_i)_{i \geq 1}$ to accumulate, with the consequence that there are infinitely many back edges. In that case, we set $N = \infty$. However, this in fact occurs with probability 0. Indeed, by (2), we have that $\ell_i(t) \leq (i+1)\|f\|$, with $\|\cdot\|$ denoting the usual supremum norm. Thus, for all $i \in \mathbb{N}$ and $t \geq 0$,

$$\mathbb{P}[N \geq i \text{ and } s_i - s_{i-1} \leq t \mid N \geq i-1] \leq \mathbb{P}[E_i \leq t]$$

where $(E_i, i \in \mathbb{N})$ are independent exponential variables with respective parameters $(i\|f\|, i \in \mathbb{N})$, and $s_0 = 0$ by convention. Hence we also have

$$\mathbb{P}[N = \infty] = \mathbb{P}\left[N = \infty \text{ and } \sup_{i \in \mathbb{N}} s_i \leq \sigma\right] \leq \mathbb{P}\left[\sum_{i=1}^{\infty} E_i \leq \sigma\right].$$

However, $\sum_{i=1}^{\infty} E_i = \infty$ a.s. by the divergence of the harmonic series and standard properties of sums of exponential variables (see e.g. Theorem 2.3.2 of [20]), and so the above probability is 0.

We end this section with two elementary technical points.

LEMMA 3.2. *The joint distribution of N and the $(s_i, i \leq N)$ can be written explicitly as:*

$$\begin{aligned} \mathbb{P}[N = n, s_i \in dt_i \forall i \leq n] = \\ \prod_{k=1}^n \left(\sum_{i=1}^k \left(f(t_i) - \hat{f}(t_{i-1}, t_i) \right) \right) \\ \times \exp \left(- \int_0^\sigma \left(f(t) - \hat{f}(t_{I(t)}, t) + \sum_{i=1}^{I(t)} \left(f(t_i) - \hat{f}(t_{i-1}, t_i) \right) \right) dt \right) dt_1 \dots dt_n, \end{aligned}$$

where $t_0 = 0$ and, for $t \in [0, \sigma]$, $I(t) = \max\{i : t_i < t\}$.

REMARK 3.3. It will at times be convenient to consider the pairs (x_i, y_i) which are *ancestral*, i.e. such that y_i is an ancestor of x_i . Thus we let $((x_i^a, y_i^a), i \leq N^a)$ be those

pairs, and $(s_i^a, i \leq N_a)$ the corresponding times in $[0, \sigma]$. Note that these are the points of a Poisson point process with intensity $f(x)dx$ on $[0, \sigma]$. In particular, N^a is a Poisson variable with parameter $\int_0^\sigma f(t)dt$, and conditionally on $N^a = 1$, the time s_1^a has density proportional to f on $[0, \sigma]$.

On the other hand, if we let $((x_i^b, y_i^b), i \leq N^b)$ be the pairs which are *not* ancestral, then the corresponding times $(s_i^b, i \leq N^b)$ can, conditionally on $(s_i^a, i \leq N^a)$, also be seen as the points of a Poisson point process. We will not need this full description, but the following expression will be useful:

$$\mathbb{P}[N^b = 0 \mid N^a = 1, s_1^a] = \exp \left(- \int_{s_1^a}^\sigma (f(t) - \hat{f}(s_1^a, t)) dt \right).$$

3.2.3. The resulting strongly connected components. Let $\mathcal{T}_f^{\text{mk}} = \mathcal{T}_N$ be the subtree spanned by the root and the marked leaves, and quotient it by the equivalence relation \sim which identifies x_i and y_i for $i \in \{1, \dots, N\}$, to obtain a rooted metric space

$$\mathcal{M}_f = \mathcal{T}_f^{\text{mk}} / \sim.$$

Since $\mathcal{T}_f^{\text{mk}}$ has only finitely many leaves, we may also view \mathcal{M}_f as a finite rooted directed multigraph M_f whose edges are endowed with lengths: the vertices of M_f are the images of the (y_i) and of the branchpoints of \mathcal{T}_f , and the directions are inherited from $\mathcal{T}_f^{\text{mk}}$ (which we always think of as having edges directed away from the root). We observe that, with the exception of the root (which can be a leaf), the vertices of M_f all have degree at least 3. Now remove all edges which do not lie in a strongly connected component of M_f and delete any isolated vertices thus created. This yields a collection of strongly connected components of minimum degree 2. If there remain vertices of degree precisely two, we repeatedly apply the following merging operation. Pick an arbitrary vertex of degree 2 and merge its two incident edges as long as they are different edges, summing their lengths. This yields a collection \mathcal{C}_f of strongly connected MDMs, as illustrated in Figure 5.

4. The scaling limit.

4.1. Excursions of Brownian motion with parabolic drift. Let $(W(t), t \geq 0)$ be a standard Brownian motion. For $\lambda \in \mathbb{R}$ and $t \geq 0$, let $W^\lambda(t) = W(t) + \lambda t - t^2/2$ and let $\underline{W}^\lambda(t) = \inf_{0 \leq s \leq t} W^\lambda(s)$. Let $B^\lambda(t) = W^\lambda(t) - \underline{W}^\lambda(t)$, and let Γ^λ be the set of excursions of B^λ . For an excursion $\gamma \in \Gamma^\lambda$, let $|\gamma|$ denote its length.

PROPOSITION 4.1. 1. For $\alpha \in \{2, 3\}$, we have $\mathbb{E} \left[\sum_{\gamma \in \Gamma^\lambda} |\gamma|^\alpha \right] < \infty$.
2. $\sum_{\gamma \in \Gamma^\lambda} |\gamma|^{3/2} = \infty$ a.s.

The $\alpha = 2$ case of the first statement is Lemma 25 of Aldous [3], which we extend here to $\alpha = 3$. (Our method of proof also works for all $\alpha > 3/2$ but we omit the details for the sake of brevity.) We first need a standard result on moments of hitting times of Brownian motion with constant drift.

LEMMA 4.2. For $\mu > 0$ and $b > 0$, let $T(b, \mu) = \inf\{t \geq 0 : W(t) - \mu t = -b\}$. Then we have

$$\mathbb{E}[T(b, \mu)] = \frac{b}{\mu} \quad \text{and} \quad \mathbb{E} \left[(T(b, \mu))^2 \right] = \frac{b(1 + b\mu)}{\mu^3}.$$

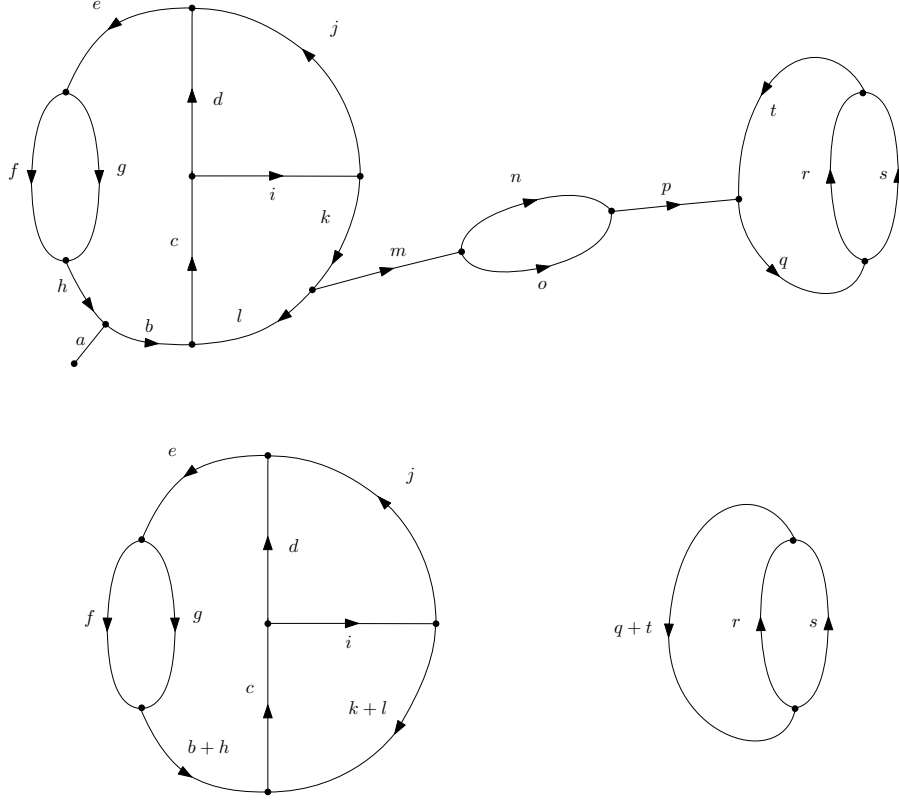


FIG 5. A representation of \mathcal{M}_f obtained from the identifications given in Figure 4, and the resulting strongly connected components.

PROOF. The Laplace transform of $T(b, \mu)$ is given by

$$\mathbb{E} \left[e^{-\theta T(b, \mu)} \right] = \exp \left(b\mu - b\sqrt{\mu^2 + 2\theta} \right), \quad \theta > 0,$$

(see, for example, Exercise 5.10 in Chapter 3 of [13]) and the first two moments of $T(b, \mu)$ follow from differentiating twice. \square

PROOF OF PROPOSITION 4.1 (STATEMENT 1.) We adapt the proof of Lemma 25 of Aldous [3]. The time-interval $[0, \infty)$ may be split into the union of a countable collection of open intervals during each of which the reflected Brownian motion with drift B^λ is away from zero (the *excursion intervals*), and the complementary set of times (when the process is at 0), which has zero Lebesgue measure. (For a careful proof of these excursion-theoretic facts, see Section 3.2 of Goldschmidt and Conchon-Kerjan [5].) Let γ be an excursion of B^λ , and let l and r be its endpoints. We have

$$|\gamma|^3 = 3 \int_l^r (r - t)^2 dt.$$

For $t \geq 0$, write $H_t = \min\{s > 0 : B^\lambda(t + s) = 0\}$. Then we have

$$\sum_{\gamma \in \Gamma^\lambda} |\gamma|^3 = 3 \int_0^\infty H_t^2 dt$$

and so we only need to prove that $\int_0^\infty \mathbb{E}[H_t^2] dt < \infty$. To do this we split the integral into $\int_0^\tau \mathbb{E}[H_t^2] dt$ and $\int_\tau^\infty \mathbb{E}[H_t^2] dt$ where $\tau = \max(0, 2\lambda + 1)$.

For $t > \max(0, \lambda)$ and $s \geq 0$, we have $\lambda s - (t + s)^2/2 + t^2/2 \leq (\lambda - t)s$. Thus, conditionally on $B^\lambda(t)$, the process $(B^\lambda(t + s), s \geq 0)$ may, until it hits 0, be coupled with a Brownian motion with initial value $B^\lambda(t)$ and drift $\lambda - t$, in such a way that the latter is larger than or equal to the former. Hence H_t is stochastically dominated by $T(B^\lambda(t), t - \lambda)$, which leads to

$$\mathbb{E}[H_t | B^\lambda(t)] \leq \frac{B^\lambda(t)}{t - \lambda} \quad \text{and} \quad \mathbb{E}[(H_t)^2 | B^\lambda(t)] \leq \frac{B^\lambda(t)(1 + (t - \lambda)B^\lambda(t))}{(t - \lambda)^3}.$$

In particular, we have

$$\int_\tau^\infty \mathbb{E}[H_t^2] dt \leq \int_\tau^\infty \frac{\mathbb{E}[B^\lambda(t)] + (t - \lambda)\mathbb{E}[(B^\lambda(t))^2]}{(t - \lambda)^3} dt.$$

However, it is also established in the proof of Lemma 25 of [3] that, for $t > 2\lambda$, the random variable $B^\lambda(t)$ is stochastically dominated by an exponential random variable with parameter $t - 2\lambda$, implying that $\mathbb{E}[B^\lambda(t)] \leq 1/(t - 2\lambda)$ and $\mathbb{E}[(B^\lambda(t))^2] \leq 2/(t - 2\lambda)^2$. In consequence, $\int_\tau^\infty \mathbb{E}[H_t^2] dt < \infty$.

To bound $\int_0^\tau \mathbb{E}[H_t^2] dt$, notice that we have $H_t \leq \tau - t + H_\tau \leq \tau + H_\tau$ for $t \leq \tau$. Hence,

$$\begin{aligned} \mathbb{E}[H_t^2] &\leq \tau^2 + 2\tau\mathbb{E}[H_\tau] + \mathbb{E}[H_\tau^2] < \infty \\ &\leq \tau^2 + 2\tau \frac{\mathbb{E}[B^\lambda(\tau)]}{\tau - \lambda} + \frac{\mathbb{E}[B^\lambda(\tau)]}{(\tau - \lambda)^3} + \frac{\mathbb{E}[(B^\lambda(\tau))^2]}{(\tau - \lambda)^2}, \end{aligned}$$

and this uniform upper bound is finite since $B^\lambda(\tau)$ is stochastically dominated by an exponential variable which has moments of all orders. It follows that we do indeed have $\int_0^\tau \mathbb{E}[H_t^2] dt < \infty$. \square

The second statement in Proposition 4.1 also requires a preliminary lemma, which allows us to work directly with B^λ instead of powers of its excursion lengths.

LEMMA 4.3. *If $\int_0^\infty B^\lambda(t) dt = \infty$ a.s. then $\sum_{\gamma \in \Gamma^\lambda} |\gamma|^{3/2} = \infty$ a.s.*

PROOF. Let $(\sigma_1, \sigma_2, \dots)$ be the lengths of the excursions of B^λ listed in decreasing order, and let $\tilde{e}_1, \tilde{e}_2, \dots$ be the excursions themselves, so that

$$\int_0^\infty B^\lambda(t) dt = \sum_{i \geq 1} \int_0^{\sigma_i} \tilde{e}_i(x) dx.$$

Now, we have by [2, Section 5] that $\tilde{e}_i \stackrel{d}{=} \tilde{e}^{(\sigma_i)}$ for $i \geq 1$, and the excursions $(\tilde{e}_i, i \geq 1)$ are conditionally independent given their lengths. Moreover, calling \mathbf{e} a normalised Brownian excursion, we have using Brownian scaling

$$\mathbb{E} \left[\int_0^s \tilde{e}^{(s)}(x) dx \right] = s^{3/2} \frac{\mathbb{E} \left[\int_0^1 \mathbf{e}(x) dx \exp \left(s^{3/2} \int_0^1 \mathbf{e}(x) dx \right) \right]}{\mathbb{E} \left[\exp \left(s^{3/2} \int_0^1 \mathbf{e}(x) dx \right) \right]} \underset{s \rightarrow 0}{\sim} s^{3/2} \mathbb{E} \left[\int_0^1 \mathbf{e}(x) dx \right].$$

It follows that, almost surely

$$\mathbb{E} \left[\sum_{i \geq 1} \int_0^{\sigma_i} \tilde{e}_i(x) dx \mid (\sigma_j)_{j \geq 1} \right] = \sum_{i \geq 1} \mathbb{E} \left[\int_0^{\sigma_i} \tilde{e}^{(\sigma_i)}(x) dx \mid \sigma_i \right] < \infty$$

if and only if

$$\sum_{i \geq 1} \sigma_i^{3/2} \mathbb{E} \left[\int_0^1 \mathbf{e}(x) du \right] < \infty,$$

which itself occurs if and only if $\sum_{i \geq 1} \sigma_i^{3/2} < \infty$. But by assumption we have $\int_0^\infty B^\lambda(t) dt = \infty$ a.s., and so $\sum_{\gamma \in \Gamma^\lambda} |\gamma|^{3/2} = \sum_{i \geq 1} \sigma_i^{3/2} = \infty$ a.s. \square

The proof of the second statement in Proposition 4.1 is due to Éric Brunet.

PROOF OF PROPOSITION 4.1 (STATEMENT 2). Recall that

$$W^\lambda(t) = W(t) + \lambda t - t^2/2,$$

$\underline{W}^\lambda(t) = \inf_{0 \leq s \leq t} W^\lambda(s)$ and $B^\lambda(t) = W^\lambda(t) - \underline{W}^\lambda(t)$. By Lemma 4.3, it is sufficient to show that $\int_0^\infty B^\lambda(t) dt = \infty$ a.s. We will construct a lower bound for B^λ , built on the same probability space, whose integral we can more easily show to be infinite.

Let $(Z(t), t \geq 0)$ be defined by $Z(0) = 0$ and

$$dZ(t) = dW(t) - \left(3t + \frac{Z(t)}{t} \right) dt.$$

Now define $\theta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ by $\theta(s) = (3s)^{1/3}$ and let

$$Y(s) = \theta(s)Z(\theta(s)).$$

Then some stochastic calculus gives that

$$\begin{aligned} dY(s) &= \theta'(s)Z(\theta(s)) + \theta(s) \left(dW(\theta(s)) - \left(3\theta(s) + \frac{Z(\theta(s))}{\theta(s)} \right) \theta'(s) ds + d\langle \theta(\cdot), Z(\theta(\cdot)) \rangle_s \right) \\ &= \theta(s) dW(\theta(s)) - 3ds. \end{aligned}$$

Note that the first equality used Ito's Lemma and the second used $(\theta'(s))^2 \theta(s) = 1$ as well as $d\langle \theta(\cdot), Z(\theta(\cdot)) \rangle_s = 0$. Again since $(\theta'(s))^2 \theta(s) = 1$, by the Dubins–Schwarz theorem ([22, Theorem 1.6]) we get that

$$\widetilde{W}(s) := \int_0^s \theta(r) dW(\theta(r))$$

is a standard Brownian motion and so Y is a Brownian motion with drift -3 :

$$Y(s) = \widetilde{W}(s) - 3s, \quad s \geq 0.$$

As a consequence, there exists a random time $S_1 \geq 0$ such that, for all $s \geq S_1$,

$$-\frac{9}{2}s < Y(s) < 0.$$

Letting $T_1 = \theta(S_1)$, since $\theta^{-1}(t) = t^3/3$, we get

$$-\frac{3}{2}t^2 < Z(t) < 0$$

for $t \geq T_1$. In particular, for $t \geq T_1 \vee 2|\lambda|$, we have the following bounds on the drift term in the SDE defining Z :

$$-\left(3t + \frac{Z(t)}{t} \right) \leq -\frac{3}{2}t \leq \lambda - t.$$

Using the fact that Z and W^λ are constructed from the same Brownian motion W , we get that after time $T_2 := T_1 \vee 2|\lambda|$, Z has smaller increments than W^λ .

Now choose a time $T_3 > T_2$ large enough such that the minima of Z and W^λ on $[0, T_3]$ are both attained after T_2 . Then, for $t \geq T_3$, the minimum of W^λ on $[0, t]$ is attained at some $u \in [T_2, t]$, and so

$$\begin{aligned} B^\lambda(t) &= W^\lambda(t) - \inf_{u \in [T_2, t]} W^\lambda(u) = \sup_{u \in [T_2, t]} (W^\lambda(t) - W^\lambda(u)) \geq \sup_{u \in [T_2, t]} (Z(t) - Z(u)) \\ &\geq Z(t) - \underline{Z}(t), \end{aligned}$$

where \underline{Z} is the running infimum of Z . Moreover, using the fact that $Y(s) < 0$ for $s \geq S_1$, we get

$$\begin{aligned} Z(t) - \underline{Z}(t) &= \frac{1}{t} Y(t^3/3) - \inf_{u \in [T_2, t]} \frac{1}{u} Y(u^3/3) \\ &\geq \frac{1}{t} Y(t^3/3) - \frac{1}{t} \inf_{u \in [T_2, t]} Y(u^3/3) = \frac{1}{t} (Y(t^3/3) - \underline{Y}(t^3/3)), \end{aligned}$$

where \underline{Y} is the running infimum of Y . This yields $B^\lambda(t) \geq \frac{1}{t} (Y(t^3/3) - \underline{Y}(t^3/3))$ for $t \geq T_3$. In particular,

$$\int_0^\infty B^\lambda(t) dt \geq \int_{T_3}^\infty (Y(t^3/3) - \underline{Y}(t^3/3)) \frac{dt}{t} = \int_{\frac{1}{3}T_3^3}^\infty (Y(s) - \underline{Y}(s)) \frac{ds}{3s}.$$

We now show that the final integral is infinite. Observe that the reflected drifting Brownian motion $Y - \underline{Y}$ is a positive recurrent Markov process, its stationary distribution being exponential with parameter 6 (see, for example, p.94 of Harrison [11]). In particular, the sets of times at which it is above 2 and below 1 are both unbounded, and we can define two sequences $(\tau_n)_{n \geq 0}$ and $(\eta_n)_{n \geq 1}$ of stopping times as follows. Let $\tau_0 = \frac{1}{3}T_3^3$ and then, for $n \geq 1$,

$$\begin{aligned} \eta_n &= \inf\{s > \tau_{n-1} : Y(s) - \underline{Y}(s) > 2\}, \\ \tau_n &= \inf\{s > \eta_n : Y(s) - \underline{Y}(s) < 1\}. \end{aligned}$$

On a downcrossing interval $[\eta_n, \tau_n]$, we trivially have $Y(s) - \underline{Y}(s) \geq 1$, so

$$\int_{\frac{1}{3}T_3^3}^\infty (Y(s) - \underline{Y}(s)) \frac{ds}{s} \geq \sum_{n=1}^\infty \frac{(\tau_n - \eta_n)}{\tau_n}.$$

By the strong Markov property, we have that $\{\tau_n - \eta_n : n \geq 2\}$ are i.i.d., and their expectation is finite by the aforementioned positive recurrence of $Y - \underline{Y}$. The law of large numbers then implies that, as $n \rightarrow \infty$,

$$\frac{\tau_n}{n} \rightarrow \mathbb{E}[\tau_2 - \tau_1] < \infty \quad \text{a.s.}$$

We also have that $\{\tau_n - \eta_n : n \geq 2\}$ are i.i.d. and so $\sum_{n \geq 1} (\tau_n - \eta_n)/n$ diverges a.s. It follows that

$$\sum_{n \geq 1} \frac{(\tau_n - \eta_n)}{\tau_n} = \infty \quad \text{a.s.} \quad \square$$

4.2. *Bounds for a single tree.* Let $\sigma > 0$. We now let $f = 2\tilde{\mathbf{e}}^{(\sigma)}$ be twice a tilted Brownian excursion with length σ , whose distribution is determined by

$$\mathbb{E}[g(\tilde{\mathbf{e}}^{(\sigma)})] = \frac{\mathbb{E}\left[g(\sqrt{\sigma}\mathbf{e}(\cdot/\sigma)) \exp\left(\sigma^{3/2} \int_0^1 \mathbf{e}(x) dx\right)\right]}{\mathbb{E}\left[\exp\left(\sigma^{3/2} \int_0^1 \mathbf{e}(x) dx\right)\right]},$$

for any non-negative measurable function g , where \mathbf{e} is a standard Brownian excursion. The choice is guided by the fact mentioned in the introduction, that these functions (including the factor 2) encode the limiting continuum random trees in which we are interested. We perform the construction detailed in Section 3.2, defining the \mathbb{R} -tree \mathcal{T}_σ (we now replace the subscript f by σ since henceforth all of our coding functions will be of this type), performing N_σ identifications, N_σ^a of them being ancestral and N_σ^b being non-ancestral, and thus build the MDM \mathcal{M}_σ . The following proposition will enable us to control the number of strongly connected components of \mathcal{M}_σ .

PROPOSITION 4.4. *Let $c = \mathbb{E}[\int_0^1 \mathbf{e}(t) dt] = F'(0)$ where $F(z) = \mathbb{E}[e^{z \int_0^1 \mathbf{e}(t) dt}]$ is the moment generating function of the Airy distribution, which is an entire function [12]. We have the following asymptotics: as $\sigma \rightarrow 0$,*

1. $\mathbb{P}[N_\sigma^a = 0] = 1 - 2c\sigma^{3/2} + O(\sigma^3)$
2. $\mathbb{P}[N_\sigma^a = 1, N_\sigma^b = 0] = 2c\sigma^{3/2} + O(\sigma^3)$
3. $\mathbb{P}[N_\sigma^a \geq 2 \text{ or } N_\sigma^b \geq 1] = O(\sigma^3)$.

Moreover,

4. $\sup_{\sigma > 0} \sigma^{-3} \mathbb{E}[N_\sigma^a \mathbf{1}_{\{N_\sigma^a \geq 2\}}] < \infty$.

PROOF. Instead of working with $\tilde{\mathbf{e}}^{(\sigma)}$, we express the probabilities in terms of a standard Brownian excursion \mathbf{e} and its area $\mathcal{A} = \int_0^1 \mathbf{e}(t) dt$. For 1, recall that, conditionally on $\tilde{\mathbf{e}}^{(\sigma)}$, we have that N_σ^a has a Poisson distribution with parameter $\int_0^\sigma 2\tilde{\mathbf{e}}^{(\sigma)}(x) dx$. Therefore,

$$\mathbb{P}[N_\sigma^a = 0] = \frac{\mathbb{E}\left[e^{-\sigma^{3/2} \int_0^1 2\mathbf{e}(t) dt} e^{\sigma^{3/2} \int_0^1 \mathbf{e}(t) dt}\right]}{\mathbb{E}\left[e^{\sigma^{3/2} \int_0^1 \mathbf{e}(t) dt}\right]} = \frac{F(-\sigma^{3/2})}{F(\sigma^{3/2})} = 1 - 2c\sigma^{3/2} + O(\sigma^3).$$

We begin the proofs of 2. and 3. by computing

$$\mathbb{P}[N_\sigma^a = 1] = \frac{\mathbb{E}\left[2\sigma^{3/2} \mathcal{A} e^{-\sigma^{3/2} \mathcal{A}}\right]}{\mathbb{E}\left[e^{\sigma^{3/2} \mathcal{A}}\right]} = \frac{2\sigma^{3/2} F'(-\sigma^{3/2})}{F(\sigma^{3/2})} = \frac{2\sigma^{3/2}(c + O(\sigma^{3/2}))}{1 + O(\sigma^{3/2})} = 2c\sigma^{3/2} + O(\sigma^3).$$

Next, recalling Remark 3.3, we write

$$\begin{aligned} \mathbb{P}[N_\sigma^a = 1, N_\sigma^b = 0] &= \mathbb{E}\left[e^{-2 \int_0^\sigma \tilde{\mathbf{e}}^{(\sigma)}(t) dt} \int_0^\sigma 2\tilde{\mathbf{e}}^{(\sigma)}(x) \exp\left(-\int_x^1 2(\tilde{\mathbf{e}}^{(\sigma)}(y) - \hat{\mathbf{e}}^{(\sigma)}(x, y)) dy\right) dx\right] \\ &= \frac{1}{\mathbb{E}[e^{\sigma^{3/2} \mathcal{A}}]} \mathbb{E}\left[e^{-\sigma^{3/2} \mathcal{A}} \sigma^{3/2} \int_0^1 2\mathbf{e}(x) \exp\left(-\sigma^{3/2} \int_x^1 2(\mathbf{e}(y) - \hat{\mathbf{e}}(x, y)) dy\right) dx\right]. \end{aligned}$$

Using $|1 - e^{-u}| \leq u$ for $u \geq 0$, we obtain

$$\begin{aligned} \mathbb{P}[N_\sigma^a = 1] - \mathbb{P}[N_\sigma^a = 1, N_\sigma^b = 0] &\leq 4\sigma^3 \mathbb{E}\left[e^{-\sigma^{3/2} \mathcal{A}} \int_0^1 \mathbf{e}(x) dx \int_x^1 (\mathbf{e}(x) - \hat{\mathbf{e}}(x, y)) dy\right] \\ &\leq 4\sigma^3 \mathbb{E}\left[\int_0^1 (\mathbf{e}(x))^2 dx\right]. \end{aligned}$$

Now note that $\int_0^1 (e(x))^2 dx$ has finite expectation because it is smaller than $(\sup e)^2$, which is indeed integrable ($\sup e$ has sub-Gaussian tails, see [15]). So the above quantity is $O(\sigma^3)$. This finishes the proof of 2. and 3.

Finally, since N_σ^a is integer-valued, we have

$$\begin{aligned} \mathbb{E}[N_\sigma^a \mathbf{1}_{\{N_\sigma^a \geq 2\}}] &= \mathbb{E}[N_\sigma^a] - \mathbb{P}[N_\sigma^a = 1] \\ &= \frac{\mathbb{E}[2\sigma^{3/2} \mathcal{A} e^{\sigma^{3/2} \mathcal{A}}]}{\mathbb{E}[e^{\sigma^{3/2} \mathcal{A}}]} - \mathbb{P}[N_\sigma^a = 1] \\ &= \frac{2\sigma^{3/2} (F'(\sigma^{3/2}) - F'(-\sigma^{3/2}))}{F(\sigma^{3/2})}. \end{aligned}$$

This proves that $\mathbb{E}[N_\sigma^a \mathbf{1}_{\{N_\sigma^a \geq 2\}}] = O(\sigma^3)$ as $\sigma \rightarrow 0$, but we also want the bound as σ tends to infinity. To this end, we write

$$\mathbb{E}[N_\sigma^a \mathbf{1}_{\{N_\sigma^a \geq 2\}}] \leq 2\sigma^{3/2} \frac{F'(\sigma^{3/2})}{F(\sigma^{3/2})}$$

and simply aim to prove that $F'(x) = O(xF(x))$ as $x \rightarrow \infty$. Quoting [12, Section 7], we have

$$F(x) = \sum_{n=0}^{\infty} a_n x^n,$$

where

$$a_n \underset{n \rightarrow \infty}{\sim} \frac{3\sqrt{2}}{(n-1)!} \left(\frac{n}{12e} \right)^{n/2}.$$

The desired domination will follow from the fact that $\frac{(n+1)a_{n+1}}{a_{n-1}}$ (the ratio of the coefficients of x^n in $F'(x)$ and $xF(x)$) is uniformly bounded for $n \geq 1$, which is true, since the sequence in fact converges:

$$\frac{(n+1)a_{n+1}}{a_{n-1}} \sim \frac{n+1}{n(n-1)} \frac{1}{12e} \frac{(n+1)^{(n+1)/2}}{(n-1)^{(n-1)/2}} \sim \frac{1}{12e} \frac{(n+1)^{n/2}}{(n-1)^{n/2}} \rightarrow \frac{1}{12}.$$

This completes the proof. \square

4.3. Some properties of the scaling limit. Let $(\sigma_1, \sigma_2, \dots)$ be the lengths of the excursions of B^λ , listed in decreasing order. For each $i \geq 1$, let \mathcal{D}_i be an independent copy of \mathcal{M}_{σ_i} and let $\mathcal{D} = \bigcup_{i=1}^{\infty} \mathcal{D}_i$. We think of \mathcal{D} as a countable MDM.

THEOREM 4.5. 1. *The number of complex connected components of \mathcal{D} has finite expectation.*

2. *The number of loops of \mathcal{D} is a.s. infinite.*

PROOF. We start with part 1. For each $i \geq 1$, let K_i be the number of complex components in \mathcal{D}_i . Each complex component contains at least one ancestral identification and so $K_i \leq N_{\sigma_i}^a$. Furthermore, if there is exactly one ancestral identification, there must also be at least one which is non-ancestral in order to obtain a complex component, so that $\mathbb{P}[K_i = 1 \mid \sigma_i] \leq \mathbb{P}[N_{\sigma_i}^a = 1, N_{\sigma_i}^b \geq 1 \mid \sigma_i] + \mathbb{P}[N_{\sigma_i}^a \geq 2 \mid \sigma_i]$. Hence, by 3. and 4. from Proposition 4.4,

$$\begin{aligned} \mathbb{E}[K_i \mid \sigma_i] &= \mathbb{P}[K_i = 1 \mid \sigma_i] + \mathbb{E}[K_i \mathbf{1}_{K_i \geq 2} \mid \sigma_i] \\ &\leq \mathbb{P}[N_{\sigma_i}^a = 1, N_{\sigma_i}^b \geq 1 \mid \sigma_i] + \mathbb{P}[N_{\sigma_i}^a \geq 2 \mid \sigma_i] + \mathbb{E}[N_{\sigma_i}^a \mathbf{1}_{\{N_{\sigma_i}^a \geq 2\}} \mid \sigma_i] \\ &\leq C\sigma_i^3 \end{aligned}$$

for some $C > 0$. Thus,

$$\mathbb{E} \left[\sum_{i=1}^{\infty} K_i \right] \leq C \mathbb{E} \left[\sum_{i=1}^{\infty} \sigma_i^3 \right] < \infty.$$

For part 2, notice first that, since we now know that \mathcal{D} has finitely many complex components a.s., it is sufficient to show that there are infinitely many ancestral identifications, i.e. that $\sum_{i=1}^{\infty} N_{\sigma_i}^a = \infty$ a.s. But since $\mathbb{P}[N_{\sigma_i}^a \geq 1 | (\sigma_j, j \in \mathbb{N})]$ is asymptotically equivalent to $2c\sigma_i^{3/2}$ by Proposition 4.4, and Proposition 4.1 gives $\sum_i \sigma_i^{3/2} = \infty$ a.s., the claim follows from an application of the Borel-Cantelli lemma. \square

The following property of \mathcal{D} is not surprising, but nonetheless requires proof.

PROPOSITION 4.6. *The strongly connected components of \mathcal{D} all have different lengths a.s.*

This follows straightforwardly from the following lemma, in which we work on a single tree.

LEMMA 4.7. *Let $\sigma > 0$.*

1. *For all $x > 0$, $\mathbb{P}[\mathcal{M}_\sigma \text{ has a strongly connected component of length } x] = 0$.*
2. *$\mathbb{P}[\mathcal{M}_\sigma \text{ has two strongly connected components with equal lengths}] = 0$.*

PROOF. Let $f = 2\tilde{e}^{(\sigma)}$ be the excursion function encoding the tree \mathcal{T}_σ from which \mathcal{M}_σ is obtained, let the selected leaves be $(x_i, i \in \{1, \dots, N\})$, and let $\mathcal{C}_1^{(\sigma)}, \dots, \mathcal{C}_K^{(\sigma)}$ be the strongly connected components of \mathcal{M}_σ , listed in the order of appearance of their first elements in the planar ordering of \mathcal{T}_σ . For each $k \in \mathbb{N}$, on the event where $k \leq K$, let $E_k = \{i \in \{1, \dots, N\} : x_i \in \mathcal{C}_k^{(\sigma)}\}$ be the set of indices of the leaves implicated in the construction of the k th strongly connected component, let u_k be the most recent common ancestor of those leaves. Let $\rho_k = \sup\{x \in [\rho, u_k] : x \text{ is a branchpoint and } x \notin \mathcal{C}_k^{(\sigma)}\}$, $\mathcal{T}_k = \bigcup_{i \in E_k} [\rho_k, x_i]$, and finally let $n_k = \#\{i \in E_k, y_i \in [\rho_k, u_k]\}$, be the number of heads along the line-segment separating ρ_k from u_k . Notice then that the length of $\mathcal{C}_k^{(\sigma)}$ is exactly that of \mathcal{T}_k , minus the initial part between ρ_k and the first y_i to be encountered. However, since the y_i are chosen uniformly from the length measure, this means that $[\rho_k, u_k]$ is split according to a Dirichlet distribution with $n_k + 1$ components. More specifically, we have

$$\text{len}(\mathcal{T}_k) - \text{len}(\mathcal{C}_k^{(\sigma)}) = (f(u_k) - f(\rho_k))\Delta_1^k$$

where, conditionally on n_k , Δ_1^k is the first component of a vector $\Delta^k = (\Delta_1^k, \dots, \Delta_{n_k+1}^k)$ which has Dirichlet(1, ..., 1) distribution. (See Figure 6 for an illustration.) Since Dirichlet distributions have a density, we obtain

$$\mathbb{P}[\text{len}(\mathcal{C}_k^{(\sigma)}) = x \mid k \leq K, \text{len}(\mathcal{T}_k), f(u_k), f(\rho_k), n_k] = 0,$$

and integrating and taking the union over all k gives us 1.

To prove 2, consider two distinct integers k and l . If $k \leq K$ and $l \leq K$, let

$$A_k = \{\rho_k \notin \mathcal{C}_l^{(\sigma)}\}$$

and

$$A_l = \{\rho_l \notin \mathcal{C}_k^{(\sigma)}\}.$$

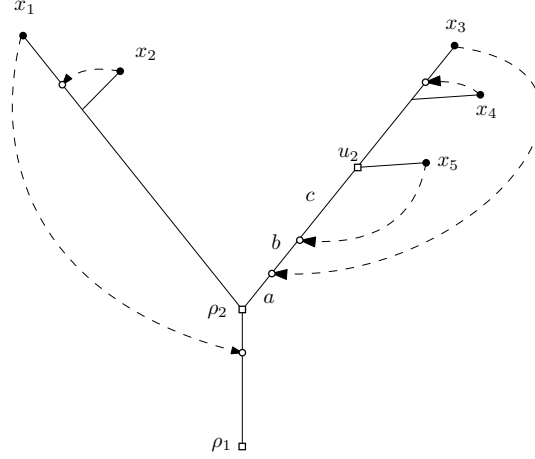


FIG 6. For point (i), focusing on the second component, \mathcal{T}_2 contains the leaves x_3, x_4, x_5 , and the length of its initial segment $f(u_2) - f(\rho_2)$ is split by a $\text{Dirichlet}(1, 1, 1)$ into (a, b, c) . For (ii), conditioning on ρ_2 being in \mathcal{T}_1 , then this split is still $\text{Dirichlet}(1, 1, 1)$.

Observe that $\mathbb{P}[A_k \cup A_l] = 1$, since $\mathcal{C}_k^{(\sigma)}$ and $\mathcal{C}_l^{(\sigma)}$ do not intersect. Now, on the event A_l , \mathcal{T}_k and \mathcal{T}_l intersect either at point ρ_k or not at all, and we can still write

$$\text{len}(\mathcal{T}_k) - \text{len}(\mathcal{C}_k^{(\sigma)}) = (f(u_k) - f(\rho_k))\Delta_1^l$$

where, conditionally on \mathcal{T}_l , n_k and the event A_l , Δ_1^l is the first component of a $\text{Dirichlet}(1, \dots, 1)$ vector. (Again see Figure 6.) This means that the length of $\mathcal{C}_k^{(\sigma)}$ has a (conditional) density, and integrating, we get

$$\mathbb{P} \left[\text{len}(\mathcal{C}_k^{(\sigma)}) = \text{len}(\mathcal{C}_l^{(\sigma)}) \mid A_l, k, l \leq K \right] = 0.$$

Symmetrising then yields that

$$\mathbb{P} \left[\text{len}(\mathcal{C}_k^{(\sigma)}) = \text{len}(\mathcal{C}_l^{(\sigma)}) \mid k, l \leq K \right] = 0,$$

and taking a countable union yields 2. \square

PROOF OF PROPOSITION 4.6. We label the strongly connected components of \mathcal{D} in such a way that, for $i \in \mathbb{N}$, those which belong to \mathcal{D}_i are called $\mathcal{C}_{i,1}, \dots, \mathcal{C}_{i,K_i}$. Consider $\mathcal{C}_{i,k}$ and $\mathcal{C}_{j,l}$ for i, j, k, l in \mathbb{N} . We can assume $i \neq j$ as the case where $i = j$ has already been treated. Conditionally on the excursion lengths $(\sigma_i, i \in \mathbb{N})$, $\mathcal{C}_{i,k}$ and $\mathcal{C}_{j,l}$ are independent and we have $\mathbb{P}[\text{len}(\mathcal{C}_{i,k}) = x] = 0$ for all $x > 0$. Thus we have $\mathbb{P}[\text{len}(\mathcal{C}_{i,k}) = \text{len}(\mathcal{C}_{j,l}) \mid \text{len}(\mathcal{C}_{j,l})] = 0$, and integrating to remove the conditioning yields $\mathbb{P}[\text{len}(\mathcal{C}_{i,k}) = \text{len}(\mathcal{C}_{j,l})] = 0$. This completes the proof. \square

5. Convergence of the strongly connected components. For $n \in \mathbb{N}$, let $p = p(n)$ be such that $p = 1/n + \lambda n^{-4/3} + o(n^{-4/3})$ as $n \rightarrow \infty$. Recall that $(C_i(n), i \in \mathbb{N})$ are the strongly connected components of $\vec{G}(n, p)$, listed in decreasing order of size (with ties broken by using the increasing order of smallest vertex-label), where we treat isolated vertices as copies of the loop of zero length, and additionally append infinitely many copies of the loop of zero length. Let $(\mathcal{C}_i, i \in \mathbb{N})$ be the strongly connected components of \mathcal{D} , listed in decreasing order of length.

We restate the main theorem.

THEOREM 1.3. *We have*

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

with respect to the distance d defined by

$$d(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{\infty} d_{\vec{\mathcal{G}}}(A_i, B_i),$$

for $\mathbf{A}, \mathbf{B} \in \vec{\mathcal{G}}^{\mathbb{N}}$.

The aim of this section is to prove this theorem. We begin by discussing some topological issues related to $d_{\vec{\mathcal{G}}}$. We then prove a series of preliminary results, before finally turning to the proof of Theorem 1.3.

5.1. *The relationship between $d_{\vec{\mathcal{G}}}$ and the Gromov–Hausdorff distance.* Recall from the introduction the definition of metric directed multigraphs (MDMs), and that the distance between two such objects $X = (V, E, r, \ell)$ and $X' = (V', E', r', \ell')$ is defined by

$$d_{\vec{\mathcal{G}}}(X, X') = \inf_{(f,g) \in \text{Isom}(X, X')} \sup_{e \in E} |\ell(e) - \ell'(g(e))|.$$

Elements of $\vec{\mathcal{G}}$ can also be viewed as metric spaces, by thinking of each edge with positive length as a line segment, identifying vertices joined by edges with length 0, and forgetting the orientation of the edges. This means that we can also compare them using the Gromov–Hausdorff distance d_{GH} (see Appendix A.1 for a short introduction). The resulting topology is however weaker, as the following lemma shows.

LEMMA 5.1. *For $X \in \vec{\mathcal{G}}$ and $X' \in \vec{\mathcal{G}}$, we have*

$$d_{\text{GH}}(X, X') \leq \frac{1}{2}|E| d_{\vec{\mathcal{G}}}(X, X')$$

PROOF. If X and X' do not have the same graph structure, then $d_{\vec{\mathcal{G}}}(X, X') = \infty$ and the statement holds trivially. If they do have the same graph structure then, up to applying an optimal isomorphism (f, g) , we can assume that they have the same vertex and edge sets, i.e. $X = (V, E, r, \ell)$ and $X' = (V, E, r, \ell')$, where the length assignments ℓ and ℓ' are such that $\sup_{e \in E} |\ell(e) - \ell'(e)| = d_{\vec{\mathcal{G}}}(X, X')$. We let ϕ be the natural bijection from X to X' when viewed as metric spaces, which acts identically on V and follows the edges “linearly”. Viewing ϕ as a correspondence (see Appendix A.2), its distortion can be bounded above by

$$\sum_{e \in E} |\ell(e) - \ell'(e)| \leq |E| \sup_{e \in E} |\ell(e) - \ell'(e)| = |E| d_{\vec{\mathcal{G}}}(X, X'). \quad \square$$

In the case of trees, it is possible to recover a convergence for $d_{\vec{\mathcal{G}}}$ from a *pointed* Gromov–Hausdorff convergence (see Appendix A.1 for a definition). (Variants of these ideas have been used in several places in the literature, and we do not claim that the following proposition is particularly novel. We have not, however, found a convenient reference.)

PROPOSITION 5.2. *Fix $k \in \mathbb{N}$. For $n \in \mathbb{N}$, let $(\mathcal{T}_n, n \in \mathbb{N})$ (resp. \mathcal{T}) be \mathbb{R} -trees with roots ρ_n (resp. ρ) and k selected distinct leaves $(x_{i,n}, 1 \leq i \leq k)$ (resp. $(x_i, i \leq k)$). Then let*

$$\overline{\mathcal{T}}_n = \bigcup_{i=1}^k [[\rho_n, x_{i,n}]]$$

be the subtree spanned by the k selected leaves and the root (and define $\bar{\mathcal{T}}$ similarly). View it as an element of $\vec{\mathcal{G}}$ by taking as vertices the root, the leaves, and all the branch points, orienting each edge away from ρ_n (resp. ρ) and giving each edge the length of its corresponding metric path.

Suppose that $(\mathcal{T}_n, \rho_n, x_{1,n}, \dots, x_{k,n})$ converges to $(\mathcal{T}, \rho, x_1, \dots, x_k)$ for the $(k+1)$ -pointed Gromov–Hausdorff topology, and that $\bar{\mathcal{T}}$ is binary. Then $\bar{\mathcal{T}}_n$ converges to $\bar{\mathcal{T}}$ for $d_{\vec{\mathcal{G}}}$. Specifically, $\bar{\mathcal{T}}_n$ and $\bar{\mathcal{T}}$ are seen as elements of $\vec{\mathcal{G}}$ by taking as vertices their roots, leaves, and branchpoints and directing edges away from the root. The map which sends ρ_n to ρ and $x_{i,n}$ to x_i for each $1 \leq i \leq k$ then extends uniquely to a graph isomorphism, under which the length of each edge in $\bar{\mathcal{T}}_n$ converges to that of the corresponding edge in $\bar{\mathcal{T}}$.

PROOF. For notational convenience, we let $x_{0,n} = \rho_n$ and $x_0 = \rho$ in the following proof. The pointed Gromov–Hausdorff convergence which is assumed to hold in the statement of the proposition implies that, for any i and j in $\{0, \dots, k\}$, we have that $d(x_{i,n}, x_{j,n})$ converges to $d(x_i, x_j)$ as $n \rightarrow \infty$. Indeed, by taking a suitable embedding for each n , we can see that

$$d(x_i, x_j) - 2\eta \leq d(x_{i,n}, x_{j,n}) \leq d(x_i, x_j) + 2\eta$$

for any $\eta > d_{\text{GH}}((\mathcal{T}_n, \rho_n, x_{1,n}, \dots, x_{k,n}), (\mathcal{T}, \rho, x_1, \dots, x_k))$. Next, using the formula $d(\rho, x \wedge y) = \frac{1}{2}(d(\rho, x) + d(\rho, y) - d(x, y))$ (which is valid in any \mathbb{R} -tree), we see that these convergences extend to distances of the form $d(\rho_n, x_{i,n} \wedge x_{j,n})$ and $d(x_{i,n}, x_{i,n} \wedge x_{j,n})$.

We will now prove the proposition by induction, adding the leaves one by one. Specifically, for $l \leq k$, let \mathcal{T}_n^l and \mathcal{T}^l be the subtrees of \mathcal{T}_n and \mathcal{T} spanned by the root and the first l selected leaves. We prove that the map mentioned at the end of the proposition, restricted to \mathcal{T}_n^l , gives convergence in $d_{\vec{\mathcal{G}}}$, by induction on l .

The base case $l = 1$ is simple, since all of the trees we consider are just single line segments, and $d(\rho_n, x_{1,n})$ converges to $d(\rho, x_1)$ as mentioned above. So let us now focus on the induction step, and assume the proposition at rank l for some $l \leq k - 1$. Observe then that we only need to prove two things: first, that the new branchpoint of \mathcal{T}_n^{l+1} is, for n large enough, added to the same edge as the new branchpoint of \mathcal{T}^{l+1} and, second, that the height of this branchpoint converges. Indeed, the first point will give the desired graph isomorphism, while the second, combined with the convergence of $d(\rho_n, x_{l+1,n})$, will give convergence of the lengths of all the edges.

Let $\llbracket y, z \rrbracket$ be the edge of \mathcal{T}^l which contains the new branchpoint of \mathcal{T}^{l+1} , with y being an ancestor of z . Since \mathcal{T} is binary, this new branchpoint is equal to neither y nor z ; let us call it b . This implies that $|d(x_{l+1,n}, y) - d(x_{l+1,n}, z)| = |d(y, b) - d(z, b)| < d(y, b) + d(b, z) = d(y, z)$. Now, letting y_n and z_n be the points in \mathcal{T}_n^l corresponding to y and z respectively through the graph isomorphism, and using that distances in \mathcal{T}_n^l converge to those in \mathcal{T}^l , we have that $|d(x_{l+1,n}, y_n) - d(x_{l+1,n}, z_n)| < d(y_n, z_n)$ for n sufficiently large. This implies that the new branchpoint lies between y_n and z_n . (Indeed, if the new branchpoint did not lie in $\llbracket y_n, z_n \rrbracket$, then we would have that the path from the new branchpoint to one of y_n and z_n would need to traverse the whole edge $\llbracket y_n, z_n \rrbracket$ and then we would have $|d(x_{l+1,n}, y_n) - d(x_{l+1,n}, z_n)| = d(y_n, z_n)$.) Finally, by observing that this branchpoint can consistently be written as $x_{i,n} \wedge x_{l+1,n}$, with i chosen such that x_i is a descendant of y , our initial remark on convergence of distances concludes the induction and the proof. \square

Note that Proposition 5.2 fails if $\bar{\mathcal{T}}$ is not binary. Indeed, it would then be possible for $\bar{\mathcal{T}}_n$ to never have the same graph structure, thus preventing the convergence for $d_{\vec{\mathcal{G}}}$.

PROPOSITION 5.3. *If the connected components of an MDM X all have different total lengths, and $(X_n, n \in \mathbb{N})$ is a sequence which converges to X for $d_{\vec{G}}$, then the strongly connected components of X_n , listed in decreasing order of length and seen as elements of \vec{G} , converge to those of X .*

PROOF. Writing $X = (V, E, r, \ell)$, let (C_1, \dots, C_k) be the strongly connected components of X , ordered by decreasing length. For $n \in \mathbb{N}$ large enough, one may assume we have $X_n = (V, E, r, \ell_n)$, where $\ell_n(C_i) \rightarrow \ell(C_i)$ as $n \rightarrow \infty$ for all i . In particular, for n large enough, $\ell_n(C_i)$ is strictly decreasing in i , and so (C_1, \dots, C_k) is also the ordered sequence of the strongly connected components of X_n , which completes the proof. \square

5.2. The components originating from a single tree. The first part of the proof will consist in proving the convergence of the components originating from a single tree. For $m \in \mathbb{N}$, we take a plane tree T_m which has the distribution of a tree component of $\mathcal{F}_{\vec{G}(n,p)}$ conditioned to have size m . We are interested in $m \sim \sigma n^{2/3}$ so that, in particular, we have $mp^{2/3} \rightarrow \sigma$ as $m \rightarrow \infty$. From [2], up to an unimportant relabelling of the vertices, T_m has the same distribution as a uniform random labelled tree on $[m]$, biased by $(1-p)^{-a(T_m)}$, where $a(T_m)$ is the number of permitted edges in T_m . We give this tree a planar embedding by rooting at the vertex labelled 1 (we also refer to this root as ρ_m) and then simply using the increasing order on the labels of the children of any vertex. Let $H^m : \{0, \dots, m-1\} \rightarrow \mathbb{Z}_+$ be the height function of T_m , such that $H^m(i)$ is the height of the i -th vertex in the planar order, starting with $H^m(0) = 0$. We recall that $\|T_m\| = \max_{0 \leq i \leq m-1} H^m(i)$ is the height of the tree T_m . Theorem 15 of [2] states that

$$(3) \quad ((m/\sigma)^{-1/2} H^m(\lfloor (m/\sigma)t \rfloor), 0 \leq t \leq \sigma) \xrightarrow{(d)} (2\tilde{e}^{(\sigma)}(t), 0 \leq t \leq \sigma)$$

uniformly as $m \rightarrow \infty$. By Lemma A.2 this has the straightforward consequence that

$$(4) \quad \left(\left(\frac{\sigma}{m} \right)^{1/2} T_m, \rho_m \right) \xrightarrow{(d)} (\mathcal{T}_\sigma, \rho)$$

as $m \rightarrow \infty$, for the 1-pointed Gromov–Hausdorff distance (see (19) for a definition).

As in Proposition 2.2, we include each of the $\binom{m}{2}$ possible back edges and $a(T_m)$ possible surplus edges independently with probability p , and let X_m be the resulting directed graph. The aim of this section is to show that the rescaled strongly connected components of X_m converge in distribution to those of \mathcal{M}_σ . In order to do this, we will use the structure of back edges outlined in Section 3.1. Specifically, let $((x_{i,m}, y_{i,m}), m \in \mathbb{N}, i \leq N_m)$ be the back edges obtained with this procedure, and let X_m^* be the subgraph of X_m obtained by removing (a) all surplus edges and (b) all back edges which are not of the above form. We will first show that the strongly connected components of X_m^* converge in distribution to those of \mathcal{M}_σ , and then that X_m and X_m^* have the same strongly connected components with high probability. In particular, we show that the surplus edges with high probability do not play any role in creating the strongly connected components.

5.2.1. Convergence of the marked points. Our next step is to improve the convergence of the rooted tree $\left(\left(\frac{\sigma}{m} \right)^{1/2} T_m, \rho_m \right)$ to include the marked points $(x_{i,m})$. Since the number of marked points is random, we use a pointed Gromov–Hausdorff distance d_{GH}^* which allows for a variable number of marks (see (20) for a formal definition).

Recall the relevant notation for the limit object from Section 3.2.2. In particular, we write $f = 2\tilde{e}^{(\sigma)}$, we let $s_i, 1 \leq i \leq N$ be the points of the Poisson process and we let $x_i = p_f(s_i)$ be their projections onto the tree \mathcal{T}_σ encoded by f .

PROPOSITION 5.4. *We have*

$$\left(\left(\frac{\sigma}{m} \right)^{1/2} T_m, \rho_m, (x_{i,m}, i \leq N_m) \right) \xrightarrow{(d)} \left(\mathcal{T}_\sigma, \rho, (x_i, i \leq N) \right)$$

as $m \rightarrow \infty$, for the topology generated by d_{GH}^* .

PROOF. By Skorokhod's representation theorem, there exists a probability space on which the convergence (3) occurs almost surely, which entails an almost sure convergence in (4) also. We need to add the other marked points to the convergence. Let $k_{i,m}$ be the index of $x_{i,m}$ in the planar ordering of T_m . We will show by induction on i that $\mathbb{P}[N_m \geq i] \rightarrow \mathbb{P}[N \geq i]$ and that, conditionally on $N_m \geq i$, the rescaled index $(\frac{m}{\sigma})^{-1} k_{i,m}$ converges in distribution to s_i . We can then use Lemma A.2 to transfer this convergence of the rescaled indices $(\frac{m}{\sigma})^{-1} k_{i,m}$ to that of the $x_{i,m}$. Together, these elements will suffice to give the claimed convergence in the sense of d_{GH}^* .

We start with $i = 1$. Since the number of ancestral back edges originating at the k -th point of T_m has distribution $\text{Bin}(H^m(k), p)$, and $pH^m(\lfloor xm \rfloor) \sim (\frac{m}{\sigma})^{-1} f(x)$, it is straightforward to see that $\mathbb{P}[N_m \geq 1] \rightarrow \mathbb{P}[N \geq 1]$ and that, conditionally on $N_m \geq 1$, $(\frac{m}{\sigma})^{-1} k_{1,m}$ converges in distribution to the first point of a Poisson point process with intensity $f(x)dx$, also conditioned to have at least one point, which is precisely s_1 . (See the proof of Lemma 19 of [2] for a more detailed version of an essentially identical argument.)

The induction step uses the same idea. Assume that the claimed convergence holds up to rank i . Recall that $\mathcal{T}_i = \cup_{j=1}^i [\rho, x_i]$. The index $k_{i+1,m}$ is found by giving to each $k \geq k_{i,m}$ a $\text{Bin}(l_i(k), p)$ number of marks, where $l_i(k)$ is the number of possible heads of a back edge originating at the k -th point in the planar ordering of T_m . If we let $T_m(i)$ be the subtree of T_m spanned by the root and $x_{1,m}, \dots, x_{i,m}$, then $l_i(k) = |T_m(i)| + H^m(k) - \min\{H^m(l) : l \in \{k_{i,m}, \dots, k\}\}$. By the induction hypothesis and Proposition 5.2, we have that $T_m(i)$ converges to \mathcal{T}_i for $d_{\mathcal{T}}$ and so, in particular, $|T_m(i)| \sim (\frac{m}{\sigma})^{1/2} \text{len}(\mathcal{T}_i)$. So we have that the instantaneous rate at which points occur satisfies

$$pl_i(\lfloor xm \rfloor) \sim \left(\frac{m}{\sigma} \right)^{-1} \left(\text{len}(\mathcal{T}_i) + f(x) - \hat{f}(s_i, x) \right).$$

It follows that $\mathbb{P}[N_m \geq i+1 | N_m \geq i] \rightarrow \mathbb{P}[N \geq i+1 | N \geq i]$ and that, conditionally on $N_m \geq i+1$, we have $(\frac{m}{\sigma})^{-1} k_{i+1,m} \xrightarrow{(d)} s_{i+1}$. Hence, by induction we have that

$$N_m \xrightarrow{(d)} N$$

and

$$(s_{i,m}, 1 \leq i \leq N_m) \xrightarrow{(d)} (s_i, 1 \leq i \leq N)$$

as $m \rightarrow \infty$. The proposition then follows by applying Lemma A.2. \square

PROPOSITION 5.5. *Let $T_m^{\text{mk}} := \cup_{i=1}^{N_m} [\rho_m, x_{i,m}]$ and $\mathcal{T}_\sigma^{\text{mk}} := \cup_{i=1}^N \mathcal{T}_i$ be the marked subtrees of T_m and \mathcal{T}_σ respectively. Then*

$$\left(\left(\frac{\sigma}{m} \right)^{1/2} T_m^{\text{mk}}, \rho_m, ((x_{i,m}, y_{i,m}), i \leq N_m) \right) \xrightarrow{(d)} \left(\mathcal{T}_\sigma^{\text{mk}}, \rho, ((x_i, y_i), i \leq N) \right),$$

as $m \rightarrow \infty$ for the topology generated by d_{GH}^* and, moreover,

$$\left(\frac{\sigma}{m} \right)^{1/2} |T_m^{\text{mk}}| \xrightarrow{(d)} \text{len}(\mathcal{T}_\sigma^{\text{mk}}).$$

PROOF. The pointed Gromov–Hausdorff convergence of Proposition 5.4 can be restricted to the marked subtrees by using the same embeddings. Using Skorokhod’s representation theorem, we may thus work on a probability space where

$$\left(\left(\frac{\sigma}{m} \right)^{1/2} T_m^{\text{mk}}, \rho_m, (x_{i,m}, i \leq N_m) \right) \xrightarrow{m \rightarrow \infty} \left(\mathcal{T}_\sigma^{\text{mk}}, \rho, (x_i, i \leq N) \right)$$

almost surely for d_{GH}^* . (In particular, almost surely $N_m = N$ for all m sufficiently large.) We will add the y terms to this convergence by using their distributions and a correspondence. The basic idea is that $y_{i,m}$ has the uniform distribution on the finite set $T_m(i) \setminus \{x_{i,m}\}$, which converges to the normalised length measure on \mathcal{T}_i .

By Proposition 5.2, T_m^{mk} has the same graph structure as $\mathcal{T}_\sigma^{\text{mk}}$ for m sufficiently large. Let us make this more precise by denoting by T^{mk} the graph which has as vertices the root, branchpoints and leaves of $\mathcal{T}_\sigma^{\text{mk}}$, and E^{mk} its set of edges. Then $\mathcal{T}_\sigma^{\text{mk}}$ can be seen as a MDM by giving to each edge $e \in \mathsf{E}^{\text{mk}}$ its length $\ell(e)$, while T_m^{mk} is the graph obtained by splitting each edge $e \in \mathsf{E}^{\text{mk}}$ into a path of length $k_m(e) \in \mathbb{N}$. Note that we have (again by Proposition 5.2) that

$$\left(\frac{\sigma}{m} \right)^{1/2} k_m(e) \rightarrow \ell(e).$$

This allows us to build a function ϕ_m which naturally injects T_m^{mk} into $\mathcal{T}_\sigma^{\text{mk}}$. Specifically, the vertices of T_m^{mk} which also belong to T^{mk} are mapped so as to preserve the structure, and the vertices of degree 2 in T_m^{mk} which subdivide the edge e are mapped “linearly”, dividing the corresponding edge of $\mathcal{T}_\sigma^{\text{mk}}$ into $k_m(e)$ segments of equal length. The inverse of this injection then may be naturally extended to a projection ψ_m from $\mathcal{T}_\sigma^{\text{mk}}$ onto T_m^{mk} by letting $\psi_m(x)$ be the most recent ancestor of x which belongs to $\phi_m(T_m^{\text{mk}})$. This allows us to define a correspondence \mathcal{R} between T_m^{mk} and $\mathcal{T}_\sigma^{\text{mk}}$ by letting $\psi_m(x) \mathcal{R} x$ for all $x \in \mathcal{T}_\sigma^{\text{mk}}$.

The distortion of \mathcal{R} is then

$$\text{dis } \mathcal{R} = \sup_{x, y \in \mathcal{T}_\sigma^{\text{mk}}} \left| d(x, y) - \left(\frac{\sigma}{m} \right)^{1/2} d(\psi_m(x), \psi_m(y)) \right|.$$

Now, for any $x, y \in \mathcal{T}_\sigma^{\text{mk}}$,

$$\begin{aligned} & \left| d(x, y) - \left(\frac{\sigma}{m} \right)^{1/2} d(\psi_m(x), \psi_m(y)) \right| \\ & \leq d(x, \phi_m(\psi_m(x))) + d(y, \phi_m(\psi_m(y))) \\ & \quad + \left| d(\phi_m(\psi_m(x)), \phi_m(\psi_m(y))) - \left(\frac{\sigma}{m} \right)^{1/2} d(\psi_m(x), \psi_m(y)) \right| \\ & \leq 2 \sup_{z \in \mathcal{T}_\sigma^{\text{mk}}} d(z, \phi_m(\psi_m(z))) + \sup_{u, v \in \mathcal{T}_\sigma^{\text{mk}}} \left| \left(\frac{\sigma}{m} \right)^{1/2} d(u, v) - d(\phi_m(u), \phi_m(v)) \right| \\ & \leq 2 \sup_{e \in \mathsf{E}^{\text{mk}}} \frac{\ell(e)}{k_m(e)} + \sum_{e \in \mathsf{E}^{\text{mk}}} \left| \left(\frac{\sigma}{m} \right)^{1/2} k_m(e) - \ell(e) \right|. \end{aligned}$$

This upper bound is easily seen to tend to 0 as $m \rightarrow \infty$, and so $\text{dis } \mathcal{R} \rightarrow 0$ also.

For $i \leq N$, let ν_i be the normalised length measure on \mathcal{T}_i , and let $\nu_{i,m}$ be the uniform measure on $T_m(i) \setminus \{x_i\}$. We aim to apply Lemma A.1. Notice that $\nu_i(\phi_m^{-1})$ is the probability measure on $T_m(i)$ which gives a weight proportional to $\deg(v)$ to any vertex v . We want to construct a coupling between this and $\nu_{i,m}$. Let $B_i(m)$ be the set containing the branchpoints of $T_m(i)$ as well as $x_{i,m}$. Then the measures $\nu_{i,m}$ and ν_i are both equal to the uniform measure

when conditioned on $T_m(i) \setminus B_i(m)$. Since the cardinality of $B_i(m)$ does not change as m increases, we obtain that both $\nu_{i,m}(B_i(m))$ and $\nu_i(\phi_m^{-1}(B_i(m)))$ tend to 0. Then for any $\varepsilon > 0$, if m is large enough, there exists a coupling (U_m, V_m) of $\nu_{i,m}$ and $\nu_i(\phi_m^{-1})$ such that $U_m = V_m$ with probability at least $1 - \varepsilon$. By construction, we can write $V_m = \psi_m(W_m)$, and (U_m, W_m) is a coupling of $\nu_{i,m}$ and ν_i such that $U_m \mathcal{R} W_m$ with probability at least $1 - \varepsilon$.

It follows that

$$\left(\left(\frac{\sigma}{m} \right)^{1/2} T_m^{\text{mk}}, \rho_m, (x_{i,m}, i \leq N_m), (\nu_{i,m}, i \leq N_m) \right) \xrightarrow{(d)} \left(\mathcal{T}_\sigma^{\text{mk}}, \rho, (x_i, i \leq N), (\nu_i, i \leq N) \right)$$

for the topology generated by d_{GHP}^* . The proposition then follows by applying Lemma A.1. \square

5.2.2. Convergence of the marked graph. Let $X_m^* = T_m^{\text{mk}}$ along with all back edges $(x_{i,m}, y_{i,m})$ for $i \leq N_m$, and recall that $\mathcal{M}_\sigma = \mathcal{T}_\sigma^{\text{mk}} / \sim$, where \sim is the equivalence relation which identifies x_i with y_i for $i \leq N$. We view these objects as elements of $\vec{\mathcal{G}}$, in a way which will fit the metric on $\vec{\mathcal{G}}$. Specifically, we take the vertex set of X_m^* to consist of ρ , the heads $y_{i,m}$ of the back edges for $i \leq N_m$, and the branch points $x_{i,m} \wedge x_{j,m}$ for $i \neq j \leq N_m$. We take the vertices of \mathcal{M}_σ to be ρ, y_i for $i \leq N$ (note that post-identification we have $x_i = y_i$), and the branch points $x_i \wedge x_j$ for $i \neq j \leq N$. Because the Brownian continuum random tree is almost surely binary and the law of \mathcal{T}_σ is absolutely continuous with respect to that of the Brownian continuum random tree, $\mathcal{T}_\sigma^{\text{mk}}$ is also binary almost surely. It follows that \mathcal{M}_σ has $2N$ vertices and, as we will see, the same must also be true for X_m^* for sufficiently large m .

PROPOSITION 5.6. $(\frac{\sigma}{m})^{1/2} X_m^* \xrightarrow{(d)} \mathcal{M}_\sigma$ in $\vec{\mathcal{G}}$.

PROOF. Using Skorokhod's representation theorem, we may assume that the convergence of Proposition 5.5 holds almost surely. Recall that, by Proposition 5.2, $((\frac{\sigma}{m})^{1/2} T_m^{\text{mk}}, \rho, (x_{i,m}, i \leq N_m))$ converges in $\vec{\mathcal{G}}$ (taking the root, $(x_{i,m})$ and branch points as vertices). In particular the elements of this sequence have the same underlying graph structure for all m large enough.

For m large enough, no $x_{i,m}$ is an ancestor of a $x_{j,m}$ or $y_{j,m}$, so the graph structure of X_m^* can be obtained from that of T_m^{mk} by removing $x_{i,m}$ and instead connecting the edge ending in $x_{i,m}$ back into $y_{i,m}$, for each i . Since $y_{i,m}$ converges to y_i in the Gromov–Hausdorff sense it will, in particular, always be on the same edge of T_m^{mk} for m sufficiently large. Thus the combinatorial structure is constant for m large, and the same as that of \mathcal{M}_σ .

Once we know the combinatorial structure, the lengths of all the edges then also converge since they can be expressed in terms of the distances between the root, the $(x_{i,m})$ and the $(y_{i,m})$.

Using also Propositions 5.3 and 4.6, we then obtain that the connected components of $(\frac{\sigma}{m})^{1/2} X_m^*$, listed in decreasing order of size, converge in the sense of $d_{\vec{\mathcal{G}}}$ to those of \mathcal{M}_σ , listed in decreasing order of length. \square

5.2.3. Surplus edges do not contribute. As mentioned earlier, we now want to prove that the surplus edges contribute to the strongly connected components of X_m with vanishingly small probability. Specifically, we aim to prove the following proposition.

PROPOSITION 5.7. $\mathbb{P}[X_m \text{ and } X_m^* \text{ have different strongly connected components}] \rightarrow 0$ as $m \rightarrow \infty$.

Let $R(m)$ be the number of surplus edges in X_m . For $1 \leq i \leq R(m)$, let $\alpha_{i,m}$ and $\beta_{i,m}$ be the tail and head respectively of the i -th surplus edge in increasing planar order of their tails. Let $W_i(m)$ be the number of vertices descending from $\beta_{i,m}$ in T_m . Proposition 5.7 will follow if we can establish that the family $\left(\sum_{i=1}^{R(m)} W_i(m), m \in \mathbb{N}\right)$ is tight, namely if

$$(5) \quad \lim_{K \rightarrow \infty} \limsup_{m \rightarrow \infty} \mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) > K \right] = 0.$$

Indeed, for a strongly connected component of X_m to feature a surplus edge, we need at least one back edge to originate from a descendant of some $\beta_{i,m}$ (since any surplus edge in a strongly connected component is part of a cycle and must thus lead to a back edge after following tree edges or further surplus edges). By Proposition 2.2, conditionally on $\sum_{i=1}^{R(m)} W_i(m) \leq K$, the probability of this event is smaller than the probability that a $\text{Bin}(mK, p)$ variable is non-zero. Assuming (5) and fixing $\varepsilon > 0$, we may find a K sufficiently large that

$$\limsup_{m \rightarrow \infty} \mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) > K \right] \leq \varepsilon/3,$$

and m large enough such that $\mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) \geq K \right] \leq \varepsilon/2$ and $1 - (1-p)^{mK} \leq \varepsilon/2$ (recall that $p \sim \sigma^{3/2} m^{-3/2}$). Then

$$\begin{aligned} & \mathbb{P}[X_m \text{ and } X_m^* \text{ have different strongly connected components}] \\ & \leq \mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) \geq K \right] + 1 - (1-p)^{mK} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

As we have already mentioned, it is shown in [2] that T_m is a biased version of the uniform labelled tree \mathbb{T}_m on $[m]$ (with a canonical planar embedding): for non-negative measurable test functions f ,

$$(6) \quad \mathbb{E}[f(T_m)] = \frac{\mathbb{E}[(1-p)^{-a(\mathbb{T}_m)} f(\mathbb{T}_m)]}{\mathbb{E}[(1-p)^{-a(\mathbb{T}_m)}]}.$$

We recall that $a(T)$ denotes the the number of surplus edges permitted by the planar structure of a tree T , called its *area* in [2]. We know from Theorem 12 and Lemma 14 of [2] that

$$(1-p)^{-a(\mathbb{T}_m)} \xrightarrow{(d)} e^{\int_0^\sigma \mathbf{e}^{(\sigma)}(t) dt},$$

and that the sequence on the left-hand side is bounded in L^2 . We will prove (5) by first showing the analogous statement for \mathbb{T}_m (this is Lemma 5.9 below) and then using the measure change. We need the following lemma, which makes use of Kesten's tree, that is the tree $\hat{\mathbb{T}}$ consisting of a copy of \mathbb{Z}_+ (the *spine*), at each point of which we graft an independent Galton–Watson tree with $\text{Poisson}(1)$ offspring distribution. We root the resulting infinite tree at 0. (This is the local weak limit of \mathbb{T}_m [9].)

LEMMA 5.8. 1. *Let $Y(m)$ be the number of vertices of \mathbb{T}_m which lie outside the largest subtree descending from a child of the root. Then*

$$Y(m) \xrightarrow[m \rightarrow \infty]{(d)} Y,$$

where Y is the number of vertices of $\hat{\mathbb{T}}$ which have no ancestors on the spine apart from the root.

2. Write $(v_i, i \in \{1, \dots, m\})$ for the vertices of T_m in planar order. For $v \in T_m$, let $Z_v(m)$ be the number of vertices in the subtree rooted at v . Let $Y_v(m)$ be the number of such vertices which lie outside the largest of the subtrees rooted at a child of v . Then the $(Y_{v_i}(m), m \in \mathbb{N}, i \leq m)$ are tight:

$$\lim_{M \rightarrow \infty} \limsup_{m \rightarrow \infty} \sup_{1 \leq i \leq m} \mathbb{P}[Y_{v_i}(m) > M] = 0.$$

PROOF. Let $T_1(m), \dots, T_D(m)$ be the subtrees of T_m rooted at its first generation, with D being the degree of the root, listed in decreasing order of size, and define \hat{T}_i similarly (noting that \hat{T}_1 is infinite). It is well-known that T_m is a Galton-Watson tree with Poisson(1) offspring distribution, conditioned to have m vertices and assigned a uniformly random labelling from $[m]$. Knowing this, it is shown within the proof of Proposition 5.2 in [21] that $(|T_2(m)|, \dots, |T_D(m)|, 0, \dots)$ converges in distribution to $(|\hat{T}_2|, \dots, |\hat{T}_D|, 0, \dots)$. Since the sum of the latter is a.s. finite, this also implies convergence of the sum.

Part 2. follows from the fact that, for all i and m , the conditional distribution of $Y_{v_i}(m)$ given $Z_{v_i}(m)$ is the same as that of $Y(Z_{v_i}(m))$. (This is an aspect of the *Markov branching property* of conditioned Galton-Watson trees, see [10].) Hence, we can write, for $M > 0$

$$\begin{aligned} \mathbb{P}[Y_{v_i}(m) > M] &= \mathbb{E}[\mathbb{P}[Y(Z_{v_i})) > M] \mid Z_{v_i}] \\ &\leq \sup_{k \in \mathbb{N}} \mathbb{P}[Y(k) > M] \end{aligned}$$

Since the distributions of $(Y(k), k \in \mathbb{N})$ form a tight sequence, the above upper bound tends to 0 as M tends to infinity. \square

Now add to the tree T_m each of the $a(T_m)$ permitted surplus edges independently with probability p . Conditionally on $a(T_m)$ this yields a $\text{Bin}(a(T_m), p)$ number of surplus edges, for which we write $R(m)$. Write the tails and heads of these surplus edges as $a_{i,m}$ and $b_{i,m}$ respectively, listed in increasing planar order of $a_{i,m}$, for $i \leq R(m)$. We also write $b_{i,m}^-$ for the parent of $b_{i,m}$ in T_m . Let $W_i(m)$ be the number of descendants of $b_{i,m}$. The following lemma is a version of (5) for T_m .

LEMMA 5.9.

$$\lim_{K \rightarrow \infty} \limsup_{m \rightarrow \infty} \mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) > K \right] = 0.$$

PROOF. Fix $\varepsilon > 0$. In Lemma 19 of [2], it is proved that $R(m)$ converges in distribution as $m \rightarrow \infty$. An identical argument shows that $R(m)$ converges in distribution as $m \rightarrow \infty$ and, in particular, is tight. Therefore, there exists $I > 0$ such that $\mathbb{P}[R(m) > I] < \frac{\varepsilon}{2}$ for all m . Moreover,

$$\begin{aligned} \mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) > K \right] &\leq \mathbb{P}[R(m) > I] + \mathbb{P} \left[R(m) \leq I, \sum_{i=1}^{R(m)} W_i(m) > K \right] \\ &\leq \frac{\varepsilon}{2} + \mathbb{P} \left[\sum_{i=1}^{R(m) \wedge I} W_i(m) > K \right]. \end{aligned}$$

We then split the event where $\sum_{i=1}^{R(m) \wedge I} W_i(m) > K$ in two: either, for all $i \leq R(m) \wedge I$, the vertex $a_{i,m}$ lies in the largest of the subtrees rooted at the children of $b_{i,m}^-$, in which case

we also have $\sum_{i=1}^{R(m) \wedge I} Y_{b_{i,m}^-}(m) > K$, or there exists i for which $a_{i,m}$ is *not* in this largest subtree, which then implies, in particular, that $Y_{b_{i,m}^-}(m) \geq d(a_{i,m}, b_{i,m})$. This leads to

$$\begin{aligned} & \mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) > K \right] \\ & \leq \frac{\varepsilon}{2} + \mathbb{P} \left[\sum_{i=1}^{R(m) \wedge I} Y_{b_{i,m}^-}(m) > K \right] + \mathbb{P} \left[\exists i \leq R(m) \wedge I : Y_{b_{i,m}^-}(m) \geq d(a_{i,m}, b_{i,m}) \right]. \end{aligned}$$

By Lemma 5.8, the $(Y_{b_{i,m}^-}(m), i \leq R(m))$ are tight as $m \rightarrow \infty$, and thus so is the sum of at most I of them:

$$\mathbb{P} \left[\sum_{i=1}^{R(m) \wedge I} Y_{b_{i,m}^-}(m) > K \right] \leq \frac{\varepsilon}{4}$$

for all m , for K large enough. For the final term, we may again adapt the argument from Lemma 19 of [2] to see that for each i , $m^{-1/2}d(a_{i,m}, b_{i,m})$ converges in distribution, where d denotes the graph distance in T_m . In particular there exists $\eta > 0$ such that $\mathbb{P}[d(a_{i,m}, b_{i,m}) \leq m^{1/2}\eta] \leq \frac{\varepsilon}{8I}$. We then have

$$\mathbb{P} \left[i \leq R(m), Y_{b_{i,m}^-}(m) \geq d(a_{i,m}, b_{i,m}) \right] \leq \frac{\varepsilon}{8I} + \mathbb{P} \left[i \leq R(m), Y_{b_{i,m}^-}(m) \geq m^{1/2}\eta \right],$$

and by Lemma 5.8 again, $\mathbb{P}[i \leq R(m), Y_{b_{i,m}^-}(m) \geq m^{1/2}\eta] < \varepsilon/8I$ for all m sufficiently large, so that for such m ,

$$\mathbb{P} \left[\exists i \leq R(m) \wedge I : Y_{b_{i,m}^-}(m) \geq d(a_{i,m}, b_{i,m}) \right] \leq \frac{\varepsilon}{4}.$$

Combining all the terms yields

$$\limsup_{m \rightarrow \infty} \mathbb{P} \left[\sum_{i=1}^{R(m)} W_i(m) > K \right] \leq \varepsilon. \quad \square$$

PROOF OF PROPOSITION 5.7. It remains to show that (5) holds. We use the change of measure to pass from T_m to T_m . Call $A(m, K)$ the event where $\sum_{i=1}^{R(m)} W_i(m) > K$ and $A(m, K)$ the event where $\sum_{i=1}^{R(m)} W_i(m) > K$. Then we have

$$\mathbb{P}[A(m, K)] = \frac{\mathbb{E}[(1-p)^{-a(T_m)} \mathbf{1}_{A(m, K)}]}{\mathbb{E}[(1-p)^{-a(T_m)}]} \leq \frac{\sqrt{\mathbb{E}[(1-p)^{-2a(T_m)}]}}{\mathbb{E}[(1-p)^{-a(T_m)}]} \sqrt{\mathbb{P}[A(m, K)]}.$$

We know that $\mathbb{E}[(1-p)^{-2a(T_m)}]$ is bounded and that $\mathbb{E}[(1-p)^{-a(T_m)}]$ converges to a positive limit. So by Lemma 5.9, we obtain

$$\lim_{K \rightarrow \infty} \limsup_{m \rightarrow \infty} \mathbb{P}[A(m, K)] = 0,$$

as required. \square

5.3. Proof of Theorem 1.3. We first prove that the convergence in Theorem 1.3 occurs in the weaker product topology, namely that for any $k \in \mathbb{N}$,

$$n^{-1/3}(C_1(n), C_2(n), \dots, C_k(n)) \xrightarrow{(d)} (C_1, C_2, \dots, C_k)$$

with respect to $d_{\mathcal{G}}^k$. We will later improve this to a convergence with respect to d .

5.3.1. *Convergence in the product topology.* Let (T_1^n, T_2^n, \dots) be the forward exploration trees of $\vec{G}(n, p)$. We list them in decreasing order of their sizes (Z_1^n, Z_2^n, \dots) , and recall that we write $(\|T_1^n\|, \|T_2^n\|, \dots)$ for their heights. We also let (Y_1^n, Y_2^n, \dots) be the subgraphs of $\vec{G}(n, p)$ induced by the vertex-sets of these trees (which include both surplus and back edges). By [3], we have the following convergence for the ℓ^2 topology on sequences:

$$(7) \quad n^{-2/3}(Z_i^n, i \in \mathbb{N}) \xrightarrow{(d)} (\sigma_i, i \in \mathbb{N}),$$

where $(\sigma_i, i \in \mathbb{N})$ are the excursion lengths of W_λ above its running infimum, sorted decreasingly. Again, using Skorokhod's theorem, we may work on a probability space for which this convergence occurs almost surely. Moreover, conditionally on (Z_1^n, Z_2^n, \dots) , the $(Y_i^n, i \in \mathbb{N})$ are independent, each having the distribution of $X_{Z_i^n}$ as in Section 5.2. Since $Z_i^n p^{2/3} \rightarrow \sigma_i$, we have that the rescaled strongly connected components of Y_i^n converge in distribution to those of \mathcal{M}_{σ_i} , and this holds jointly for any finite set of indices i . Taking into account Proposition 5.3, the following proposition will give the convergence in Theorem 1.3 for the product topology.

PROPOSITION 5.10. *For all $k \in \mathbb{N}$, we have*

$$\lim_{K \rightarrow \infty} \mathbb{P}[\forall i \leq k, \exists j \leq K : \mathcal{C}_i \subseteq \mathcal{D}_j] = 1$$

and,

$$\lim_{K \rightarrow \infty} \liminf_{n \rightarrow \infty} \mathbb{P}[\forall i \leq k, \exists j \leq K : C_i(n) \subseteq Y_j^n] = 1.$$

Informally, Proposition 5.10 states that, with high probability, large strongly connected components of $\vec{G}(n, p)$ and \mathcal{D} will only be found in large trees of the forward depth-first forest, making the ordering of both trees and strongly connected components by their lengths compatible. Its proof relies on two lemmas.

LEMMA 5.11. *As $\sigma \rightarrow 0$, we have*

$$(8) \quad \mathbb{P}[\mathcal{M}_\sigma \text{ has a complex component}] = O(\sigma^3).$$

For all $\varepsilon > 0$, we have as $\sigma \rightarrow 0$

$$(9) \quad \mathbb{P}[\|\mathcal{T}_\sigma\| \geq \varepsilon] = O(\sigma^3).$$

Consequently, for all $\varepsilon > 0$,

$$(10) \quad \mathbb{P}[\mathcal{M}_\sigma \text{ has a component with length greater than } \varepsilon] = O(\sigma^3).$$

LEMMA 5.12. *There exists $C > 0$ such that, for all n large enough and $1 \leq m \leq n^{2/3}$,*

$$(11) \quad \mathbb{P}[X_m \text{ has a complex component}] \leq C \frac{m^3}{n^2}$$

and

$$(12) \quad \mathbb{P}[X_m \text{ has a component which contains a surplus edge}] \leq C \frac{m^3}{n^2}.$$

Moreover, for all $\varepsilon > 0$, there exists $C > 0$ such that, for n large enough, and $1 \leq m \leq n^{2/3}$,

$$(13) \quad \mathbb{P}[\|T_m\| \geq n^{1/3}\varepsilon] \leq C \frac{m^2}{n^{4/3}}.$$

Consequently, for all $\varepsilon > 0$, there exists $C > 0$ such that, for n large enough,

$$(14) \quad \mathbb{P}[X_m \text{ has a component with length greater than } n^{1/3}\varepsilon] \leq C \frac{m^2}{n^{4/3}}.$$

Note that for both of these lemmas, the final statement is a consequence of the previous ones by noticing that any component consisting of a single ancestral cycle has length smaller than the height of the tree (plus one in the discrete case).

PROOF OF LEMMA 5.11. By Proposition 4.4,

$$\mathbb{P}[\mathcal{M}_\sigma \text{ has a complex component}] \leq \mathbb{P}[N_\sigma^a \geq 2 \text{ or } N_\sigma^b \geq 1] = O(\sigma^3).$$

Recalling that the height $\|\mathcal{T}_\sigma\|$ has the same distribution as $\sup 2\tilde{\mathbf{e}}^{(\sigma)}$ and that it has exponential moments [15], we have, for $\sigma < 1$,

$$\begin{aligned} \mathbb{P}[\|\mathcal{T}_\sigma\| > \varepsilon] &= \frac{\mathbb{E}\left[\mathbf{1}_{\{\sup \mathbf{e} \geq \varepsilon/2\sqrt{\sigma}\}} \exp\left(\sigma^{3/2} \int_0^1 \mathbf{e}(x) dx\right)\right]}{\mathbb{E}\left[\exp\left(\sigma^{3/2} \int_0^1 \mathbf{e}(x) dx\right)\right]} \\ &\leq \mathbb{E}[e^{\sup \mathbf{e} - \varepsilon/2\sqrt{\sigma}} e^{\sigma^{3/2} \sup \mathbf{e}}] \leq \mathbb{E}[e^{2\sup \mathbf{e}}] e^{-\varepsilon/2\sqrt{\sigma}} = O(\sigma^3). \end{aligned}$$

Note we have used the bound $\mathbf{1}_{\{\sup \mathbf{e} \geq \varepsilon/2\sqrt{\sigma}\}} \leq e^{\sup \mathbf{e} - \varepsilon/2\sqrt{\sigma}}$. This proves (8) and (9); (10) then follows. \square

For Lemma 5.12, we require some preliminary bounds on the height and area of T_m .

LEMMA 5.13. *There exists a constant $M > 0$ such that, for all n large enough such that $1/(2n) < p < 2/n$ and all $1 \leq m \leq n^{2/3}$,*

$$(15) \quad \mathbb{E}[\|T_m\|^4] \leq Mm^2$$

and

$$(16) \quad \mathbb{E}[(a(T_m))^2] \leq Mm^3.$$

PROOF. Lemma 25 from [2] gives $\mathbb{E}[\|T_m\|^4] \leq M \cdot \max(m^6 n^{-4}, 1) \cdot m^2$ for all n large enough and $m \leq n$, and restricting ourselves to $m \leq n^{2/3}$ yields (15).

For (16), we follow the beginning of the proof of Lemma 25 from [2]. Let $q = \max(m^{-3/2}, p)$. Then (6) and Markov's inequality together yield

$$\mathbb{P}[a(T_m) > xm^{3/2}] \leq \frac{\mathbb{E}[(1-q)^{-a(T_m)}]}{(1-q)^{-xm^{3/2}}} \leq \frac{\mathbb{E}[(1-p)(1-q))^{-a(T_m)}]}{(1-q)^{-xm^{3/2}}} \leq \frac{\mathbb{E}[(1-q)^{-2a(T_m)}]}{(1-q)^{-xm^{3/2}}}.$$

From Lemma 14 in [2], we obtain that $\mathbb{E}[(1-q)^{-2a(T_m)}] \leq K \exp 4\kappa\delta^2$ where $\delta = \max(2m^{3/2}/n, 1)$. Since $qm^{3/2} \geq \delta/4$, we get

$$\mathbb{P}[a(T_m) > xm^{3/2}] \leq K e^{4\kappa\delta^2 - x\delta/4},$$

and for $1 \leq m \leq n^{2/3}$, we have $1 \leq \delta \leq 2$, so that

$$\mathbb{P}[a(T_m) > xm^{3/2}] \leq K e^{64\kappa - x/4}.$$

It follows that

$$\mathbb{E}\left[\frac{(a(T_m))^2}{m^3}\right] \leq K e^{64\kappa} \int_0^\infty e^{-\sqrt{x}/4} dx = 32K e^{64\kappa},$$

which completes the proof. \square

PROOF OF LEMMA 5.12. We take n large enough for (15) and (16) to hold, and $m \leq n^{2/3}$. Notice first that (13) follows from (15) and Markov's inequality:

$$\mathbb{P}[\|T_m\| \geq n^{1/3}\varepsilon] \leq \frac{\mathbb{E}[\|T_m\|^4]}{\varepsilon^4 n^{4/3}} \leq \frac{Mm^2}{\varepsilon^4 n^{4/3}}.$$

We now want to show that the probability that X_m contains a strongly connected component which is complex or features surplus edges is also bounded by $m^3 n^{-2}$. Such a component can only arise if one of the following four events occurs:

$$A_m = \{X_m \text{ has at least two ancestral back edges.}\}$$

$$B_m = \{X_m \text{ has one ancestral back edge, and at least one other back edge which points inside the created cycle.}\}$$

$$C_m = \{X_m \text{ has at least two surplus edges.}\}$$

$$D_m = \{X_m \text{ has one surplus edge } (a, b) \text{ and at least one back edge pointing to an ancestor of } a.\}$$

We will bound the probabilities of each of these events separately.

Conditionally on the tree T_m , the number of ancestral back edges in X_m has distribution $\text{Bin}(S_m, p)$, where S_m is the sum of the heights of all vertices in T_m . By using the well-known stochastic domination of $\text{Bin}(k, p)$ by $\text{Poi}(-k \log(1 - p))$ and the fact that $\mathbb{P}[\text{Poi}(\mu) \geq 2] \leq \mu^2$, we have

$$\mathbb{P}[A_m \mid T_m] \leq (-S_m \log(1 - p))^2 \leq M(S_m p)^2.$$

From now on, the constant M can vary from line to line, but never depends on n or m .

Since $S_m \leq m\|T_m\|$, by using (15) again, for n large enough we end up with

$$\mathbb{P}[A_m] \leq M \frac{m^2}{n^2} \mathbb{E}[\|T_m\|^2] \leq M \frac{m^3}{n^2}.$$

Given that there is exactly one ancestral back edge in X_m , the number of back edges which point back into the cycle created is stochastically dominated by $\text{Bin}(m\|T_m\|, p)$. Hence we have

$$\begin{aligned} \mathbb{P}[B_m \mid T_m] &\leq p S_m (1 - p)^{S_m - 1} (1 - (1 - p)^{m\|T_m\|}) \\ &\leq M p S_m (m\|T_m\| \log(1 - p)) \\ &= M n^{-2} (m\|T_m\|)^2. \end{aligned}$$

This is the same bound as above, thus leading to

$$\mathbb{P}[B_m] \leq M \frac{m^3}{n^2}$$

Since the number of surplus edges has distribution $\text{Bin}(a(T_m), p)$, we get $\mathbb{P}[C_m \mid T_m] \leq M p^2 a(T_m)^2$ and

$$\mathbb{P}[C_m] \leq M n^{-2} \mathbb{E}[a(T_m)^2].$$

A similar argument as for B_m also yields

$$\mathbb{P}[D_m] \leq M n^{-2} m \mathbb{E}[\|T_m\| a(T_m)] \leq M n^{-2} m \sqrt{\mathbb{E}[\|T_m\|^2]} \sqrt{\mathbb{E}[a(T_m)^2]},$$

and an application of (16) concludes the proof. \square

We can now prove the proposition.

PROOF OF PROPOSITION 5.10. Fix $k \in \mathbb{N}$ and $\eta > 0$, and let $\varepsilon > 0$ be small enough that

$$\mathbb{P}[\text{len}(\mathcal{C}_k) > \varepsilon] > 1 - \eta.$$

By Lemmas 5.11 and 5.12, there exists $C > 0$ such that, for all $K \in \mathbb{N}$,

$$\begin{aligned} & \mathbb{P}[\exists i > K : \mathcal{D}_i \text{ contains a component with length greater than } \varepsilon] \\ & \leq \mathbb{P}[\sigma_{K+1} > 1] + C\mathbb{E}\left[\sum_{i>K} \sigma_i^3\right] \end{aligned}$$

and

$$\begin{aligned} & \mathbb{P}[\exists i > K : Y_i^n \text{ contains a component with length greater than } n^{1/3}\varepsilon] \\ & \leq \mathbb{P}[Z_{K+1}^n > n^{2/3}] + C\mathbb{E}\left[\sum_{i>K} \frac{(Z_i^n)^2}{n^{4/3}}\right]. \end{aligned}$$

By Proposition 4.1 and (7), there exists K sufficiently large that both of these are smaller than η . Then

$$\mathbb{P}[\mathcal{C}_1, \dots, \mathcal{C}_k \text{ are in } \mathcal{D}_1, \dots, \mathcal{D}_K, \text{len}(\mathcal{C}_k) > \varepsilon] \geq 1 - 2\eta.$$

From the fact that $(n^{-1/3}Y_1^n, \dots, n^{-1/3}Y_K^n) \xrightarrow{(d)} (\mathcal{D}_1, \dots, \mathcal{D}_N)$, we deduce that, for n greater than some $n_0 \in \mathbb{N}$,

$$\mathbb{P}[C_1(n), \dots, C_k(n) \text{ are in } Y_1^n, \dots, Y_K^n, \text{len}(C_k(n)) > \varepsilon n^{1/3}] \geq 1 - 3\eta$$

and hence

$$\mathbb{P}[C_1(n), \dots, C_k(n) \text{ are in } Y_1^n, \dots, Y_K^n] \geq 1 - 4\eta. \quad \square$$

5.3.2. *Controlling the tail.* The proof of Theorem 1.3 will be completed if we can show that, for all $\varepsilon > 0$,

$$\lim_{k \rightarrow \infty} \mathbb{P}\left[\sum_{i=k+1}^{\infty} d_{\bar{\mathcal{G}}}(C_i, \mathfrak{L}) > \varepsilon\right] = 0$$

and

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left[\sum_{i=k+1}^{\infty} d_{\bar{\mathcal{G}}}(C_i(n), \mathfrak{L}) > n^{1/3}\varepsilon\right] = 0.$$

Fix $\eta > 0$. For $k \in \mathbb{N}$, let $Q(k)$ be the largest integer such that

$$\mathbb{P}[\text{all components in } \mathcal{D}_1, \dots, \mathcal{D}_{Q(k)} \text{ have lengths exceeding } \text{len}(\mathcal{C}_k)] > 1 - \eta.$$

Then by Proposition 5.10 and the convergence of $n^{-1/3}C_k(n)$ to \mathcal{C}_k , it also holds that, for n large enough, all the components of $Y_1^n, \dots, Y_{Q(k)}^n$ have lengths exceeding that of $C_k(n)$ with probability at least $1 - 2\eta$. Thus we have

$$\mathbb{P}\left[\sum_{i=k+1}^{\infty} d_{\bar{\mathcal{G}}}(C_i, \mathfrak{L}) > \varepsilon\right] \leq \eta + \mathbb{P}\left[\left(\sum_{i=Q(k)+1}^{\infty} \sum_{j: \mathcal{C}_j \subset \mathcal{D}_i} d_{\bar{\mathcal{G}}}(C_j, \mathfrak{L})\right) > \varepsilon\right]$$

and similarly

$$\begin{aligned} & \mathbb{P} \left[\sum_{i=k+1}^{\infty} d_{\bar{\mathcal{G}}}(C_i(n), \mathfrak{L}) > n^{1/3} \varepsilon \right] \\ & \leq 2\eta + \mathbb{P} \left[\left(\sum_{i=Q(k)+1}^{\infty} \sum_{j: C_j(n) \subset X_i^n} d_{\bar{\mathcal{G}}}(C_j(n), \mathfrak{L}) \right) > n^{1/3} \varepsilon \right]. \end{aligned}$$

Note that $Q(k) \rightarrow \infty$ as $k \rightarrow \infty$: indeed, it is non-decreasing, and so if it did possess a finite limit Q , then the probability of $\mathcal{D}_1, \dots, \mathcal{D}_{Q+1}$ containing a smallest component of \mathcal{C} would be at least η , a contradiction since there is no smallest component. It is therefore enough to prove that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\sum_{i=N+1}^{\infty} \sum_{j: C_j \subset \mathcal{D}_i} d_{\bar{\mathcal{G}}}(C_j, \mathfrak{L}) > \varepsilon \right] = 0$$

and

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\sum_{i=N+1}^{\infty} \sum_{j: C_j(n) \subset X_i^n} d_{\bar{\mathcal{G}}}(C_j(n), \mathfrak{L}) > n^{1/3} \varepsilon \right] = 0.$$

However, by (8), (11) and (12), for N large enough, all the components contained in $(\mathcal{D}_i, i \geq N+1)$ are single ancestral cycles with probability at least $1 - \eta$, and for n large enough, this also holds for those contained in $(Y_i^n, i \geq N+1)$. Noting that such components have length at most the height of the underlying tree (plus one in the discrete case), and that their number is at most the number of ancestral back edges, we are reduced to proving the following statements:

$$(17) \quad \lim_{K \rightarrow \infty} \mathbb{P} \left[\sum_{i=K+1}^{\infty} N_{\sigma_i}^a \|\mathcal{T}_{\sigma_i}\| > \varepsilon \right] = 0$$

and

$$(18) \quad \lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\sum_{i=K+1}^{\infty} A_i^n \|T_i^n\| > n^{1/3} \varepsilon \right] = 0,$$

where A_i^n is the number of ancestral back edges in Y_i^n . These may be obtained using the following lemma.

LEMMA 5.14. 1. *There exists $C > 0$ such that, for $\sigma < 1$,*

$$\mathbb{E}[N_{\sigma}^a \|\mathcal{T}_{\sigma}\|] \leq C \sigma^2.$$

2. *There exists $C > 0$ such that, for n large enough, and $1 \leq m \leq n^{2/3}$,*

$$\mathbb{E}[A_m \|\mathcal{T}_m\|] \leq C \frac{m^2}{n},$$

where A_m is the number of ancestral back edges in X_m .

PROOF. Part 1. is straightforward: assuming $\|\mathcal{T}_\sigma\|$ and N_σ^a are built from a tilted excursion $\tilde{\mathbf{e}}^{(\sigma)}$, and remembering that $\mathbb{E}[N_\sigma^a \mid \tilde{\mathbf{e}}^{(\sigma)}] = \int_0^\sigma 2\tilde{\mathbf{e}}^{(\sigma)}(t)dt$, we have

$$\begin{aligned} \mathbb{E}[N_\sigma^a \mid \|\mathcal{T}_\sigma\|] &= \mathbb{E} \left[\sup 2\tilde{\mathbf{e}}^{(\sigma)} \int_0^\sigma 2\tilde{\mathbf{e}}^{(\sigma)}(t)dt \right] \\ &\leq 4\sigma \mathbb{E}[\sup(\tilde{\mathbf{e}}^{(\sigma)})^2] \\ &\leq 4\sigma \frac{\mathbb{E} \left[\sup(\sqrt{\sigma}\mathbf{e})^2 \exp \left(\sigma^{3/2} \int_0^1 \mathbf{e}(t)dt \right) \right]}{\mathbb{E} \left[\exp \left(\sigma^{3/2} \int_0^1 \mathbf{e}(t)dt \right) \right]} \\ &\leq 4\sigma^2 \mathbb{E} \left[e^{\int_0^1 \mathbf{e}(t)dt} \sup(\mathbf{e})^2 \right], \end{aligned}$$

the latter expectation being finite (see the proof of Lemma 5.11). For part 2, recall that, conditionally on T_m the distribution of A_m is stochastically dominated by $\text{Bin}(m\|T_m\|, p)$. Thus, we have

$$\mathbb{E}[A_m \mid T_m] \leq p \mathbb{E}[m\|T_m\|^2],$$

and applying Lemma 5.13 concludes the proof. \square

We leave the straightforward adaptation of the arguments used for Proposition 5.10 to prove (17) and (18) to the reader. This completes the proof of Theorem 1.3.

6. Further properties of the scaling limit. We write \mathcal{C} for the list of strongly connected components of \mathcal{D} , and \mathcal{C}_σ for that of \mathcal{M}_σ , in decreasing order of length. Let also $\mathcal{C}_{\text{cplx}}$ be the list of *complex* components of \mathcal{C} , i.e. those that are not cycles, also in decreasing order of length. We have not yet been able to find the exact distribution of \mathcal{C} and \mathcal{C}_σ for $\sigma > 0$: this will be the subject of future research. However, we show here that \mathcal{C}_σ and $\mathcal{C}_{\text{cplx}}$ both have a positive probability of being equal to any appropriate fixed family of directed multigraphs.

For sequences (G_1, \dots, G_k) and (H_1, \dots, H_j) of directed multigraphs, we write $(G_1, \dots, G_k) \equiv (H_1, \dots, H_j)$ if $j = k$ and G_i is isomorphic to H_i for each $i \leq j$. We extend this notation naturally to the case where one or both of the sequences has edge lengths by simply ignoring the edge lengths.

PROPOSITION 6.1. *Let G_1, \dots, G_k be a finite sequence consisting of 3-regular strongly connected directed multigraphs or loops. We have*

$$\mathbb{P}[\mathcal{C}_\sigma \equiv (G_1, \dots, G_k)] > 0.$$

Assuming that G_1, \dots, G_k are all complex, we also have

$$\mathbb{P}[\mathcal{C}_{\text{cplx}} \equiv (G_1, \dots, G_k)] > 0.$$

Let $(e_i, 1 \leq i \leq K)$ be an arbitrary ordering of the edges of (G_1, \dots, G_k) . Then, conditionally on $\mathcal{C}_\sigma \equiv (G_1, \dots, G_k)$ (resp. $\mathcal{C}_{\text{cplx}} \equiv (G_1, \dots, G_k)$), \mathcal{C}_σ (resp. $\mathcal{C}_{\text{cplx}}$) gives lengths $(\ell(e_i), 1 \leq i \leq K)$ to these edges, and their joint distribution has full support in

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

Constructing 3-regular directed multigraphs from trees and back edges. First, we want to show that any of the graphs in which we are interested can be constructed by a procedure which adds back edges to a plane tree. We set this up in a discrete framework. Let t be a discrete plane tree whose vertices have outdegrees in $\{0, 1, 2\}$. We think of this as a directed graph, with edges pointing away from the root. We assume that t has as many leaves as internal vertices of outdegree one, which we call x_1, \dots, x_n and y_1, \dots, y_n respectively, in the planar order. We assume, moreover, that for each $i \geq 1$, the internal vertex y_i is visited before the leaf x_i in the depth-first exploration. By identifying x_i and y_i for all i , we obtain a directed graph, whose strongly connected components we then extract. Each strongly connected component will have exactly one vertex of degree 2, which we erase, merging its two incident edges. The result is a set of 3-regular strongly connected directed multigraphs. The next lemma asserts that any appropriate collection of such multigraphs can be obtained by this procedure, and Figure 7 provides an example.

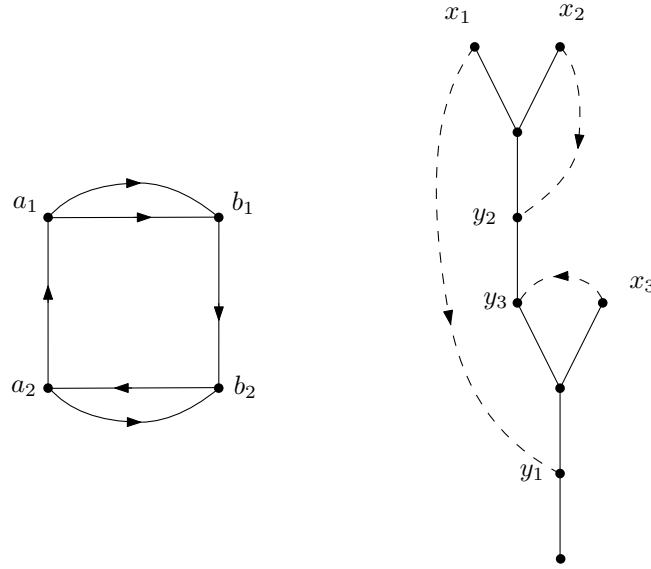


FIG 7. Obtaining a 3-regular connected directed multigraph from a tree with backward identifications. The tree was built using the method presented in the proof of Lemma 6.2.

LEMMA 6.2. *For any (G_1, \dots, G_k) , there exist a discrete plane tree t and pairings (x_i, y_i) such that the above construction results in (G_1, \dots, G_k) .*

PROOF. Notice first that we can focus on the case where $k = 1$. Once this case is treated, the general case can be solved by taking a tree t which contains distinct subtrees corresponding to each G_i .

So let G be a fixed strongly connected 3-regular directed multigraph. Noticing that it cannot have vertices with outdegree 0 or 3, and that the sum of the outdegrees of all the vertices is equal to that of all the indegrees, we deduce that there exists $n \in \mathbb{N}$ such that G has n vertices with indegree 1 and outdegree 2, and n vertices with indegree 2 and outdegree 1. Let a_1, \dots, a_n be the former and b_1, \dots, b_n the latter, for any ordering such that the edge (b_1, a_1) exists.

We will give a method to construct the necessary plane tree as well as the backward links between leaves and edges. At each step, t will contain a certain number of vertices of G , as well as some “open” edges, which have their tails at points in t but are missing their heads.

Start with t initially containing three vertices: a root with outdegree 1, its child (which we arbitrarily call ρ_0) which has outdegree 1 as well, and its next neighbour a_1 , from which originate two open edges. At each step of the algorithm, let z be the leftmost of the deepest vertices of t which have open edges, choose any edge of G starting at z which is not yet featured in t , call u the head of that edge, and do the following:

- If u is not already in t , add it at the end of the leftmost open edge, and add one or two open edges at u corresponding to its outdegree in G . The edge (z, u) is then a tree edge in t .
- If u is already in t but $u \neq a_1$, add a leaf at the end of the leftmost open edge, label that leaf x_j for the smallest available j and let also $u = y_j$. The edge (z, u) is then featured in t as the tree edge (z, x_j) , identifying x_j with u .
- If $u = a_1$, put a leaf at the end of the leftmost open edge, label that leaf x_j for the smallest available j , and let $y_j = \rho_0$. The edge (z, u) is then featured in t as the merging of the tree edges (z, x_j) and (ρ_0, a_1) , identifying x_j with ρ_0 .

Note that this algorithm terminates, and that identifying the pairs (x_i, y_i) in t and removing the root (which is not in its strongly connected component) and ρ_0 (which has degree 2 in the strongly connected component) gives us G .

Moreover, by construction, the successive vertices appearing as z follow the planar ordering of t . This means that at any step, any other vertex of t can be found earlier than z in the contour process, and thus in every pair (x_i, y_i) , the vertex y_i is seen earlier than x_i in the exploration process, and the identifications indeed go backwards. This completes the proof. \square

The marked tree has full support. If T is a discrete plane tree and \mathcal{T} is a discrete plane tree with edge lengths (equivalently an \mathbb{R} -tree with finitely many leaves which are ordered), we write $\mathcal{T} \equiv T$ if the discrete plane structure underlying \mathcal{T} is T . If $\mathcal{T} \equiv T$ then the lengths of the edges of \mathcal{T} , in planar order, form a vector in \mathbb{R}_+^k where k is the number of edges of T .

Let T be a fixed binary rooted discrete plane tree with $n \in \mathbb{N}$ leaves. For an excursion function $f : [0, \sigma] \rightarrow \mathbb{R}_+$, we let $D_T(f)$ be the set of increasing sequences $\mathbf{t} = (t_1, \dots, t_n) \in [0, \sigma]^n$ such that the $\mathcal{T}_f(t_1, \dots, t_n) \equiv T$. This is an open subset of $[0, \sigma]^n$ which can be written explicitly as

$$D_T(f) = \{\mathbf{t} \in [0, \sigma]^n : t_1 < t_2 < \dots < t_n \text{ and } \forall k \in \{3, \dots, n\}, \\ \hat{f}(t_{i(k)}, t_{k-1}) < \hat{f}(t_{k-1}, t_k) < \hat{f}(t_{j(k)}, t_{k-1})\}.$$

Here the indices $i(k)$ and $j(k)$ are defined as follows, and illustrated by Figure 8. Let L_1, \dots, L_n be the leaves of T in planar order (we add $L_0 = \rho$ for the sake of convenience). For $k \in \{3, \dots, n\}$, we then take $i(k) < j(k)$ to be any two integers in $\{0, 1, 2, \dots, k-1\}$ such that, on the path $[[\rho, L_{k-1}]]$, the two points $L_{i(k)} \wedge L_{k-1}$ and $L_{j(k)} \wedge L_{k-1}$ are respectively maximal and minimal such that $L_{i(k)} \wedge L_{k-1} \leq L_{k-1} \wedge L_k \leq L_{j(k)} \wedge L_{k-1}$ for the genealogical/planar order.

LEMMA 6.3. *We have*

$$\mathbb{P}[\mathcal{T}_f^{\text{mk}} = T] = \int_{\mathbf{t} \in D_T(f)} d\mathbf{t} \prod_{k=1}^n \left(\sum_{i=1}^k f(t_i) - \hat{f}(t_{i-1}, t_i) \right) \\ \exp \left(- \int_0^\sigma \left(f(t) - \hat{f}(t_{I(t)}, t) + \sum_{i=1}^{I(t)} f(t_i) - \hat{f}(t_{i-1}, t_i) \right) dt \right),$$

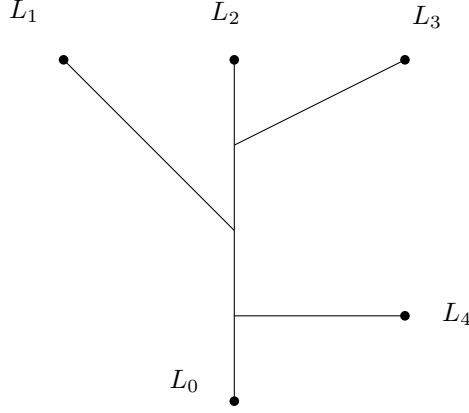


FIG 8. For this tree, $i(3) = 1$, $j(3) = 2$, $i(4) = 0$, and $j(4) = 1$. Given an excursion function f , a sequence $t_1 < t_2 < t_3 < t_4$ will then be in $D_T(f)$ iff $\hat{f}(t_1, t_2) < \hat{f}(t_2, t_3) < f(t_2)$ and $0 < \hat{f}(t_3, t_4) < \hat{f}(t_1, t_4)$.

where $t_0 = 0$ and, for $t \in [0, \sigma]$, $I(t) = \max\{i : t_i < t\}$.

Moreover, if we take $f = 2\tilde{e}^{(\sigma)}$ for $\sigma > 0$, then

$$\mathbb{P}[\mathcal{T}_\sigma^{\text{mk}} \equiv T] > 0,$$

and conditionally on $\mathcal{T}_\sigma^{\text{mk}} \equiv T$, the joint distribution of the edge lengths of $\mathcal{T}_\sigma^{\text{mk}}$ has full support in \mathbb{R}_+^{2n-1} .

PROOF. The first statement comes from Lemma 3.2. For the second statement, we use a comparison with the scaling limit of the undirected random graph. Specifically, Lemma 10 of [1] gives the joint distribution of the tree shape and the edge lengths in the subtree of \mathcal{T}_σ spanned by the root and a random collection of leaves obtained as the projection of a Poisson point process on $[0, \sigma]$ with intensity $\tilde{e}^{(\sigma)}(\cdot)$ onto the tree.² In particular, the probability that this procedure gives the tree shape T and that the lengths of the edges (in planar order) lie in an open set $A \subset \mathbb{R}_+^{2n-1}$ is positive, that is

$$\mathbb{E} \left[\int_{\mathbf{t} \in D_T(2\tilde{e}^{(\sigma)})} d\mathbf{t} \mathbf{1}_{\{(2\tilde{e}^{(\sigma)}(\mathbf{t})) \in A'\}} \prod_{k=1}^n \tilde{e}^{(\sigma)}(t_k) \exp \left(- \int_0^\sigma \tilde{e}^{(\sigma)}(t) dt \right) \right] > 0,$$

where $A' \in \mathbb{R}_+^n$ is the open set such that the heights the leaves of T are in A' iff its edge lengths are in A . This implies that $\mathbb{E}[G] > 0$ where

$$G = \int_{\mathbf{t} \in D_T(2\tilde{e}^{(\sigma)})} d\mathbf{t} \mathbf{1}_{\{(2\tilde{e}^{(\sigma)}(\mathbf{t})) \in A'\}} \prod_{k=1}^n \left(\sum_{i=1}^k 2\tilde{e}^{(\sigma)}(t_i) \right),$$

(since G is larger than the random variable in the expectation above). We then have

$$\mathbb{P}[\mathcal{T}_\sigma^{\text{mk}} \equiv T, \text{lengths in } A] \geq \mathbb{E} \left[G \exp(-\sigma(n+1) \sup \tilde{e}^{(\sigma)}) \right],$$

and this is positive since $\sup \tilde{e}^{(\sigma)}$ is a.s. finite. \square

²The sampled leaves in the undirected graph setting come from a Poisson point process with intensity $\tilde{e}^{(\sigma)}(\cdot)$ rather than the intensity $2\tilde{e}^{(\sigma)}(\cdot)$ we have in our construction for the directed graph. This is because (as seen in [2]) in the setting of the undirected graph the identifications arise as the limit of the surplus edges: the number of potential surplus edges originating at a single vertex is given not by the height of the vertex but rather by the number of vertices sitting on the stack in the depth-first exploration (the so-called *depth-first walk*). The depth-first walk is asymptotically half the size of the height, and so has scaling limit $\tilde{e}^{(\sigma)}$ rather than $2\tilde{e}^{(\sigma)}$.

PROOF OF PROPOSITION 6.1. We first show the result for \mathcal{C}_σ . Let \mathbf{t} and $((x_i, y_i), i \in \{1, \dots, n\})$ be the discrete tree and pairing of leaves and outdegree-1 vertices given by Lemma 6.2. Moreover, let T be obtained from \mathbf{t} by erasing the vertices of degree 2, and merging their adjacent edges. Let

- (e_1, \dots, e_K) be the edges of (G_1, \dots, G_k) , in any order;
- $(e_1, \dots, e_K, e_{K+1}, \dots, e_N)$ be those of \mathbf{t} , in any order completing the previous one;
- (f_1, \dots, f_M) be those of T , in planar order.³

By construction, each edge of (G_1, \dots, G_k) , is an edge of \mathbf{t} , justifying the notation for the edges of \mathbf{t} . Moreover, each edge of T is obtained by merging edges of \mathbf{t} , so there exists a partition of $\{1, \dots, N\}$ with blocks $(S(i), 1 \leq i \leq M)$ such that for each $i \in \{1, \dots, M\}$, f_i is obtained by merging e_j for $j \in S(i)$. For $i \in \{1, \dots, n\}$, let $e_T(y_i)$ be the edge of T containing y_i . Given this information, we call a collection of positive lengths $\ell(e_i)$, and $\ell(f_i)$ such that $\ell(f_i) = \sum_{j \in S(i)} \ell(e_j)$ an *admissible length assignment*.

Recall that, from the construction given in Section 3.2.2, conditionally on $\mathcal{T}_\sigma^{\text{mk}}$ with leaves L_1, \dots, L_p , the marked internal points z_1, \dots, z_p are independent and, for each j , z_j is uniform on $\cup_{k=1}^j [\rho, L_k]$. If $\mathcal{T}_\sigma^{\text{mk}} \equiv T$ then this gives rise to a length assignment ℓ on T , and we have

$$\mathbb{P} \left[z_j \in e_T(y_j), \forall j \in \{1, \dots, n\} \mid \mathcal{T}_\sigma^{\text{mk}} \equiv T \right] \geq \prod_{j=1}^n \frac{\ell(g(y_j))}{\text{len}(\mathcal{T}_\sigma^{\text{mk}})}.$$

Moreover, conditionally on the event $\{z_j \in e_T(y_j), \forall j \in \{1, \dots, n\}, \mathcal{T}_\sigma^{\text{mk}} \equiv T\}$, for any edge f_i of T , the probability that z_j , for j such that $y_j \in f_i$, are in the right order on f_i is $\frac{1}{|S(i)|!}$. If this occurs, then it gives rise to a length assignment ℓ on \mathbf{t} as well, making the whole length assignment admissible. We then have $(\ell(e_j), j \in S(i)) = (D_1(i)\ell(f_i), \dots, D_{|S(i)|}(i)\ell(f_i))$ where $D(i) = (D_1(i), \dots, D_{|S(i)|}(i)) \in \Delta_{|S(i)|}$ has the Dirichlet(1, ..., 1) distribution on the $(|S(i)| - 1)$ -dimensional simplex $\Delta_{|S(i)|}$. These events occur independently for different $i \in \{1, \dots, M\}$.

Let A be an open set in \mathbb{R}_+^K . Take open sets $B \subset \mathbb{R}_+^M$ and $C_i \in \Delta_{|S(i)|}$ for $i \in \{1, \dots, M\}$ such that, for any admissible length assignment, if $(\ell(f_i), i \in \{1, \dots, M\}) \in B$ and, for all i , $(\frac{\ell(e_j)}{\ell(f_i)}, j \in S(i)) \in C_i$, then we have $\ell(e_i), i \in \{1, \dots, K\} \in A$. Then

$$\begin{aligned} & \mathbb{P} \left[C_\sigma \equiv (G_1, \dots, G_k), (\ell(e_i), i \in \{1, \dots, K\}) \in A \right] \\ & \geq \mathbb{E} \left[\mathbf{1}_{\{\mathcal{T}_\sigma^{\text{mk}} \equiv T, (\ell(f_i), i \in \{1, \dots, M\}) \in B\}} \prod_{j=1}^n \frac{\ell(f(y_j))}{\text{len}(\mathcal{T}_\sigma^{\text{mk}})} \prod_{i=1}^M \frac{1}{|S(i)|!} \mathbf{1}_{\{D(i) \in C_i\}} \right]. \end{aligned}$$

By Lemma 6.3, the event $\{\mathcal{T}_\sigma^{\text{mk}} \equiv T, (\ell(f_i), i \in \{1, \dots, M\}) \in B\}$ occurs with positive probability, and since Dirichlet distributions charge the full simplex, we do indeed have that

$$\mathbb{P} \left[C_\sigma \equiv (G_1, \dots, G_k), (\ell(e_i), i \in \{1, \dots, K\}) \in A \right] > 0.$$

We finally turn to the result for $\mathcal{C}_{\text{cplx}}$. Recall that $(\sigma_i, i \geq 1)$ are the ranked excursion lengths of a Brownian motion with parabolic drift and that, conditionally on the lengths,

³Note that in fact we have $M = 2n - 1$, $N = 3n - 1$ and $K = 3(n - k) + k'$, where k' is the number of unicycles amongst (G_1, \dots, G_k) ; however, this fact is not useful here.

$\mathcal{C}_i, i \geq 1$ are independent copies of \mathcal{C}_{σ_i} . Notice that

$$\begin{aligned} \mathbb{P}[\mathcal{C}_{\text{cplx}} \equiv (G_1, \dots, G_k), \text{ lengths in } A] \\ \geq \mathbb{P}[\mathcal{C}_1 \equiv (G_1, \dots, G_k), \text{ lengths in } A, \mathcal{C}_i \text{ has no complex components } \forall i \geq 2]. \end{aligned}$$

From Propositions 4.1 and 4.4, we deduce that $(\mathcal{C}_i, i \geq 2)$ has no complex components with positive probability. An application of the first part of the proposition then completes the proof. \square

APPENDIX A: GROMOV-HAUSDORFF DISTANCES

In this section, we give relevant background on the Gromov-Hausdorff distance and its variants that we use in this paper. For additional details and proofs, we refer to Chapter 7 of [4], Section 6 of [19], and the references therein.

A.1. Definitions. Consider two compact metric spaces X and X' . The *Gromov-Hausdorff distance* $d_{\text{GH}}(X, X')$ between them is defined to be

$$d_{\text{GH}}(X, X') = \inf_{\phi, \phi'} d_{\text{H}, \mathcal{Z}}(\phi(X), \phi'(X')),$$

where the infimum is taken over all possible isometric embeddings ϕ and ϕ' into a common metric space \mathcal{Z} , and $d_{\text{H}, \mathcal{Z}}$ denotes the Hausdorff distance between compact subsets of \mathcal{Z} .

We use two variants of the Gromov-Hausdorff distance which include *marked points* and *probability measures* on X and X' , respectively. First, consider $k \in \mathbb{N}$ and points $x_i \in X$ and $x'_i \in X'$ for $i \in [k]$. We define the *k-pointed Gromov-Hausdorff distance* between $(X, (x_i, i \in [k]))$ and $(X', (x'_i, i \in [k]))$ to be

$$\begin{aligned} (19) \quad d_{\text{GH}}^k((X, (x_i, i \in [k])), (X', (x'_i, i \in [k]))) \\ = \inf_{\phi, \phi'} \left(d_{\text{H}, \mathcal{Z}}(\phi(X), \phi'(X')) \vee \max_{i \in [k]} d_{\mathcal{Z}}(\phi(x_i), \phi'(x'_i)) \right), \end{aligned}$$

where ϕ, ϕ' , and \mathcal{Z} are as before and $d_{\mathcal{Z}}$ is the metric on \mathcal{Z} . We will also need a version of the pointed Gromov-Hausdorff distance which allows for a random (but finite) number of marked points. For compact metric spaces X and X' let $S \subset X$ and $S' \subset X'$ be such that $|S| < \infty$ and $|S'| < \infty$. Then define

$$(20) \quad d_{\text{GH}}^*((X, S), (X', S')) = \begin{cases} d_{\text{GH}}^k((X, S), (X', S')) & \text{if } |S| = |S'| = k \geq 0 \\ \infty & \text{otherwise.} \end{cases}$$

Next, let ν and ν' be Borel probability measures on X and X' respectively. The *Gromov-Hausdorff-Prokhorov distance* between (X, ν) and (X', ν') is defined to be

$$d_{\text{GHP}}((X, \nu), (X', \nu')) = \inf_{\phi, \phi'} (d_{\text{H}, \mathcal{Z}}(\phi(X), \phi'(X')) \vee d_{\text{P}, \mathcal{Z}}(\phi^{-1}(\nu), (\phi')^{-1}(\nu'))),$$

where ϕ, ϕ' , and \mathcal{Z} are as before and $d_{\text{P}, \mathcal{Z}}$ is the Prokhorov metric between probability measures on \mathcal{Z} .

Note that these definitions are flexible: we can add more probability measures, combine marks and measures, and so on. Write $d_{\text{GHP}}^{k, l}$ for the k -pointed and l -measured Gromov-Hausdorff-Prokhorov distance. We will make particular use of the distance d_{GHP}^* defined as follows: for $S \subset X$, $S' \subset X'$, natural numbers L and L' , Borel probability measures $(\nu_i, i \in [L])$ on X and $(\nu'_i, i \in [L'])$ Borel probability measures on X' , let

$$\begin{aligned} d_{\text{GHP}}^*((X, S, (\nu_i, i \in [L])), (X', S', (\nu'_i, i \in [L']))) \\ = \begin{cases} d_{\text{GHP}}^{k, l}((X, S, (\nu_i, i \in [L])), (X', S', (\nu'_i, i \in [L']))) & \text{if } |S| = |S'| = k \text{ and } L = L' = l, \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

These distances all make their respective sets of isometry classes into Polish spaces. The following lemma is a variant of Proposition 10 in [19] and is particularly useful for us.

LEMMA A.1. *Let $((X^n, (x_i^n, i \in [k]), (\nu_i^n, i \in [l])), n \in \mathbb{N})$ be a sequence of random k -pointed and l -measured compact metric spaces which converges in distribution to $(X, (x_i, i \in [k]), (\nu_i, i \in [l]))$. For all $n \in \mathbb{N}$, conditionally on $(X^n, (x_i^n, i \in [k]), (\nu_i^n, i \in [l]))$, let $(y_i^n, i \in [l])$ be independent random variables taking values in X^n with respective distributions $(\nu_i^n, i \in [l])$. Then the sequence of $(k + l)$ -pointed metric spaces, $(X^n, (x_i^n, i \in [k]), (y_i^n, i \in [l]), n \in \mathbb{N})$, converges in distribution as $n \rightarrow \infty$ to $(X, (x_i, i \in [k]), (y_i, i \in [l]))$, where the $(y_i, i \in [l])$ are defined analogously to the $(y_i^n, i \in [l])$.*

A.2. Correspondences and their use. A *correspondence* between X and X' is a relation \mathcal{R} such that, for any $x \in X$, there exists at least one $x' \in X'$ such that $x\mathcal{R}x'$ and, for any $x' \in X'$, there exists at least one $x \in X$ such that $x\mathcal{R}x'$. Correspondences are a very convenient tool with which to study Gromov–Hausdorff distances, by quantifying them through their *distortions*. The distortion of the correspondence \mathcal{R} is defined by

$$\text{dis } \mathcal{R} = \sup \{ |d(x, y) - d(x', y')| : x\mathcal{R}x', y\mathcal{R}y' \}.$$

It is then classical that

$$d_{\text{GH}}(X, X') = \frac{1}{2} \inf_{\mathcal{R}} \text{dis } \mathcal{R},$$

where the infimum is taken over all correspondences \mathcal{R} between X and X' .

The following useful lemma showcases the use of a correspondence to bound a pointed Gromov–Hausdorff distance.

LEMMA A.2. *Let f and g be two excursion functions on $[0, \sigma]$, and let $s_1, \dots, s_k, s'_1, \dots, s'_k$ be points in $[0, \sigma]$. Then*

$$d_{\text{GH}}^k((\mathcal{T}_f, p_f(s_1), \dots, p_f(s_k)), (\mathcal{T}_g, p_g(s'_1), \dots, p_g(s'_k))) \leq 2\|f - g\| + \omega_\delta(f),$$

where $\delta = \sup_{1 \leq i \leq k} |s_i - s'_i|$ and $\omega_\delta(f)$ is the δ -modulus of continuity of f .

PROOF. The relation $\mathcal{R} = \{p_f(s), p_g(s')\} \subset \mathcal{T}_f \times \mathcal{T}_g$ is well-known to be a correspondence, with $\text{dis}(\mathcal{R}) \leq 4\|f - g\|$. As in the proof of Theorem 7.3.25 in [4], we can then build a metric d on the disjoint union $\mathcal{T}_f \cup \mathcal{T}_g$ which extends their intrinsic metrics by letting, for s and s' in $[0, \sigma]$,

$$d(p_f(s), p_f(s')) = \inf_{t \in [0, \sigma]} \left\{ d_f(s, t) + d_g(t, s') + \frac{1}{2} \text{dis}(\mathcal{R}) \right\}.$$

It is then straightforward to verify that, under this embedding, $d_{\text{H}}(\mathcal{T}_f, \mathcal{T}_g) \leq 2\|f - g\|$ and $d(p_f(s_i), p_g(s'_i)) \leq 2\|f - g\| + \omega_\delta(f)$. \square

We end this section with a formulation of the multiply pointed and measured Gromov–Hausdorff–Prokhorov distance in terms of correspondences. Let X and X' be compact metric spaces with marked points $(x_i, i \in [k])$ and $(x'_i, i \in [k])$ as well as Borel probability measures $(\nu_i, i \in [l])$ and $(\nu'_i, i \in [l])$. We let R be the set of all correspondences \mathcal{R} between X and X' such that $x_i\mathcal{R}x'_i$ for $i \in [k]$. For $i \in [l]$, let $\mathcal{C}(\nu_i, \nu'_i)$ be the set of *couplings* of ν_i and ν'_i , namely Borel probability measures on $X \times X'$ which have ν_i and ν'_i as marginals on X and X' respectively. We then have

$$(21) \quad d_{\text{GHP}}^k((X, (x_i, i \in [k]), (\nu_i, i \in [l])), (X', (x'_i, i \in [k]), (\nu'_i, i \in [l]))) \\ = \inf \left\{ \rho > 0 : \exists \mathcal{R} \in R(X, Y), \mu_i \in \mathcal{C}(\nu_i, \nu'_i), i \in [l] \text{ s.t. } \inf_{i \in [l]} \mu_i(\mathcal{R}) \geq 1 - \rho, \text{dis } \mathcal{R} \leq \rho \right\}.$$

Acknowledgments. We would like to thank Nicolas Broutin and Julien Berestycki for helpful discussions. We are very grateful to Éric Brunet for the proof of the second statement in Proposition 4.1. We would like to thank the referees for their careful reading of the paper and insightful comments, which led to many improvements.

Funding. This research was supported by EPSRC Fellowship EP/N004833/1.

REFERENCES

- [1] ADDARIO-BERRY, L., BROUTIN, N. and GOLDSCHMIDT, C. (2010). Critical Random Graphs: Limiting Constructions and Distributional Properties. *Electron. J. Probab.* **15** 741–775. <https://doi.org/10.1214/EJP.v15-772>
- [2] ADDARIO-BERRY, L., BROUTIN, N. and GOLDSCHMIDT, C. (2012). The continuum limit of critical random graphs. *Probability Theory and Related Fields* **152** 367–406.
- [3] ALDOUS, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25** 812–854. <https://doi.org/10.1214/aop/1024404421>
- [4] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A course in metric geometry. Graduate Studies in Mathematics* **33**. American Mathematical Society, Providence, RI. <https://doi.org/10.1090/gsm/033> MR1835418
- [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. *Ann. Probab.*, to appear. Available at [arXiv:2002.04954](https://arxiv.org/abs/2002.04954).
- [6] COULSON, M. (2022). The critical window in random digraphs. *Combinatorics, Probability and Computing* **31** 411–429. <https://doi.org/doi:10.1017/S096354832100033X>
- [7] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** 17–61. MR0125031
- [8] GABOW, H. N. (2000). Path-based depth-first search for strong and biconnected components. *Information Processing Letters* **74** 107–114. [https://doi.org/10.1016/S0020-0190\(00\)00051-X](https://doi.org/10.1016/S0020-0190(00)00051-X)
- [9] GRIMMETT, G. R. (1980/81). Random labelled trees and their branching networks. *J. Austral. Math. Soc. Ser. A* **30** 229–237. MR607933
- [10] HAAS, B. and MIERMONT, G. (2012). Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.* **40** 2589–2666.
- [11] HARRISON, J. M. (1985). *Brownian motion and stochastic flow systems*. John Wiley & Sons, New York; Chichester.
- [12] JANSON, S. (2007). Brownian excursion area, Wright’s constants in graph enumeration, and other Brownian areas. *Probab. Surveys* **4** 80–145. <https://doi.org/10.1214/07-PS104>
- [13] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian motion and stochastic calculus*, second ed. *Graduate Texts in Mathematics* **113**. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4612-0949-2> MR1121940
- [14] KARP, R. M. (1990). The transitive closure of a random digraph. *Random Structures and Algorithms* **1** 73–93.
- [15] KENNEDY, D. P. (1976). The distribution of the maximum Brownian excursion. *Journal of Applied Probability* **13** 371–376. <https://doi.org/10.1017/S0021900200094468>
- [16] LE GALL, J.-F. (2005). Random trees and applications. *Probab. Surveys* **2** 245–311. <https://doi.org/10.1214/154957805100000140>
- [17] ŁUCZAK, T. (1990). The phase transition in the evolution of random digraphs. *Random Structures and Algorithms* **14** 217–223.
- [18] ŁUCZAK, T. and SEIERSTAD, T. G. (2009). The critical behavior of random digraphs. *Random Structures and Algorithms* **35** 271–293. <https://doi.org/10.1002/rsa.20283>
- [19] MIERMONT, G. (2009). Tessellations of random maps of arbitrary genus. *Annales scientifiques de l’École Normale Supérieure Ser. 4*, **42** 725–781. <https://doi.org/10.24033/asens.2108>
- [20] NORRIS, J. R. (1997). *Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511810633>
- [21] PAGNARD, C. (2017). Local limits of Markov branching trees and their volume growth. *Electron. J. Probab.* **22** 53 pp. <https://doi.org/10.1214/17-EJP96>
- [22] REVUZ, D. and YOR, M. (2004). *Continuous Martingales and Brownian Motion*, third ed. *Grundlehren der Mathematischen Wissenschaften*. Springer.
- [23] TARJAN, R. (1972). Depth-First Search and Linear Graph Algorithms. *SIAM Journal on Computing* **1** 146–160. <https://doi.org/10.1137/0201010>