The Annals of Probability
2017, Vol. 45, No. 5, 3075–3144
DOI: 10.1214/16-AOP1132

THE SCALING LIMIT OF THE MINIMUM SPANNING TREE OF THE COMPLETE GRAPH

BY LOUIGI ADDARIO-BERRY[1,2,*], NICOLAS BROUTIN[3,†], CHRISTINA GOLDSCHMIDT[2,4,‡] AND GRÉGORY MIERMONT[5,§]

McGill University, Inria Rocquencourt-Paris*†, University of Oxford*‡ and École Normale Supérieure de Lyon/Institut universitaire de France*§

Consider the minimum spanning tree (MST) of the complete graph with n vertices, when edges are assigned independent random weights. Endow this tree with the graph distance renormalized by $n^{1/3}$ and with the uniform measure on its vertices. We show that the resulting space converges in distribution as $n \to \infty$ to a random compact measured metric space in the Gromov–Hausdorff–Prokhorov topology. We additionally show that the limit is a random binary $\mathbb{R}$-tree and has Minkowski dimension 3 almost surely. In particular, its law is mutually singular with that of the Brownian continuum random tree or any rescaled version thereof. Our approach relies on a coupling between the MST problem and the Erdős–Rényi random graph. We exploit the explicit description of the scaling limit of the Erdős–Rényi random graph in the so-called critical window, established in [*Probab. Theory Related Fields* **152** (2012) 367–406], and provide a similar description of the scaling limit for a "critical minimum spanning forest" contained within the MST. In order to accomplish this, we introduce the notion of $\mathbb{R}$-graphs, which generalise $\mathbb{R}$-trees, and are of independent interest.

CONTENTS

Received January 2015; revised April 2016.
[1]Supported by NSERC Discovery grant and by an FQRNT Nouveau Chercheur grant.
[2]Supported by Royal Society International Exchange Award IE111100.
[3]Supported by ANR-09-BLAN-0011.
[4]Supported by EPSRC fellowship EP/N004833/1.
[5]Supported by Grants ANR-08-BLAN-0190, ANR-08-BLAN-0220-01 and ANR-14-CE25-0014.
MSC2010 subject classifications. Primary 60C05; secondary 60F05.
Key words and phrases. Minimum spanning tree, scaling limit, R-tree, R-graph, Gromov–Hausdorff–Prokhorov distance, critical random graphs.

1. Introduction.

1.1. *A brief history of minimum spanning trees.* The minimum spanning tree (MST) problem is one of the first and foundational problems in the field of combinatorial optimisation. In its initial formulation by Borůvka [24], one is given distinct, positive edge weights (or *lengths*) for K_n, the complete graph on vertices labelled by the elements of $\{1, \dots, n\}$. Writing $\{w_e, e \in E(K_n)\}$ for this collection of edge weights, one then seeks the unique connected subgraph T of K_n with vertex set $V(T) = \{1, \dots, n\}$ that minimizes the *total length*

$$\sum_{e \in E(T)} w_e. \tag{1.1}$$

Algorithmically, the MST problem is among the easiest in combinatorial optimisation: procedures for building the MST are both easily described and provably efficient. The most widely known MST algorithms are commonly called *Kruskal's algorithm* and *Prim's algorithm*.[6] Both procedures are important in this work; as their descriptions are short, we provide them immediately.

[6]Both of these names are misnomers or, at the very least, obscure aspects of the subject's development; see Graham and Hell [37] or Schriver [68] for careful historical accounts.

KRUSKAL'S ALGORITHM. Start from a forest of n isolated vertices $\{1, \dots, n\}$. At each step, add the unique edge of smallest weight joining two distinct components of the current forest. Stop when all vertices are connected.

PRIM'S ALGORITHM. Fix a starting vertex i. At each step, consider all edges joining the component currently containing i with its complement, and from among these add the unique edge of smallest weight. Stop when all vertices are connected.

Unfortunately, efficient procedures for constructing MSTs do not automatically yield efficient methods for understanding the typical structure of the resulting objects. To address this, a common approach in combinatorial optimisation is to study a procedure by examining how it behaves when given random input; this is often called *average case* or *probabilistic analysis.*

The probabilistic analysis of MSTs dates back at least as far as Beardwood, Halton and Hammersley [21] who studied the *Euclidean MST* of n points in $\mathbb{R}^d$. Suppose that μ is an absolutely continuous measure on $\mathbb{R}^d$ with bounded support, and let $(P_i, i \geq 1)$ be independent and identically distributed (i.i.d.) samples from μ. For edge $e = \{i, j\}$, take w_e to be the Euclidean distance between P_i and P_j. Then there exists a constant $c = c(\mu)$ such that if X_n is the total length of the minimum spanning tree, then

$$\frac{X_n}{n^{(d-1)/d}} \xrightarrow{\text{a.s.}} c.$$

This *law of large numbers for Euclidean MSTs* spurred a massive amount of research: on more general laws of large numbers [65, 69, 71, 80], on central limit theorems [16, 44, 46, 47, 64, 82], and on the large-n scaling of various other "localizable" functionals of random Euclidean MSTs [45, 60, 61, 63, 72]. (The above references are representative, rather than exhaustive. The books of Penrose [59] and of Yukich [81] are comprehensive compendia of the known results and techniques for such problems.)

From the perspective of Borůvka's original formulation, the most natural probabilistic model for the MST problem may be the following. Weight the edges of the complete graph K_n with i.i.d. random edge weights $\{W_e : e \in E(K_n)\}$ whose common distribution μ is atomless and has support contained in $[0, \infty)$, and let $\mathbb{M}^n$ be the resulting random MST. The conditions on μ ensure that all edge weights are positive and distinct. Frieze [34] showed that if the common distribution function F is differentiable at 0^+ and $F'(0^+) > 0$, then the total weight X_n satisfies

$$F'(0^+) \cdot \mathbb{E}[X_n] \to \zeta(3), \tag{1.2}$$

whenever the edge weights have finite mean. It is also known that $F'(0^+) \cdot X_n \xrightarrow{\text{p}} \zeta(3)$ without any moment assumptions for the edge weights [34, 70]. Results analogous to (1.2) have been established for other graphs, including the hypercube [62], high-degree expanders and graphs of large girth [22], and others [33, 35].

Returning to the complete graph K_n, Aldous [8] proved a distributional convergence result corresponding to (1.2) in a very general setting where the edge weight distribution is allowed to vary with n, extending earlier, related results [20, 75]. Janson [39] showed that for i.i.d. Uniform$[0,1]$ edge weights on the complete graph, $n^{1/2}(X_n - \zeta(3))$ is asymptotically normally distributed, and gave an expression for the variance that was later shown [42] to equal $6\zeta(4) - 4\zeta(3)$.

If one is interested in the *graph theoretic* structure of the tree $\mathbb{M}^n$ rather than in information about its edge weights, the choice of distribution μ is irrelevant. To see this, observe that the behaviour of both Kruskal's algorithm and Prim's algorithm is fully determined once we order the edges in increasing order of weight, and for any distribution μ as above, ordering the edges by weight yields a uniformly random permutation of the edges. We are thus free to choose whichever distribution μ is most convenient, or simply to choose a uniformly random permutation of the edges. Taking μ to be uniform on $[0,1]$ yields a particularly fruitful connection to the now-classical *Erdős–Rényi random graph process*. This connection has proved fundamental to the detailed understanding of the global structure of $\mathbb{M}^n$ and is at the heart of the present paper, so we now explain it.

Let the edge weights $\{W_e : e \in E(K_n)\}$ be i.i.d. Uniform$[0,1]$ random variables. The Erdős–Rényi graph process $(\mathbb{G}(n,p), 0 \le p \le 1)$ is the increasing graph process obtained by letting $\mathbb{G}(n,p)$ have vertices $\{1,\dots,n\}$ and edges $\{e \in E(K_n) : W_e \le p\}$.[7] For fixed p, each edge of K_n is independently present with probability p. Observing the process as p increases from zero to one, the edges of K_n are added one at a time in exchangeable random order. This provides a natural coupling with the behaviour of Kruskal's algorithm for the same weights, in which edges are *considered* one at a time in exchangeable random order, and added precisely if they join two distinct components. More precisely, for $0 < p < 1$ write $\mathbb{M}(n,p)$ for the subgraph of the MST $\mathbb{M}^n$ with edge set $\{e \in E(\mathbb{M}^n) : W_e \le p\}$. Then for every $0 < p < 1$, the connected components of $\mathbb{M}(n,p)$ and of $\mathbb{G}(n,p)$ have the same vertex sets.

In their foundational paper on the subject [29], Erdős and Rényi described the *percolation phase transition* for their eponymous graph process. They showed that for $p = c/n$ with c fixed, if $c < 1$ (the *subcritical* case) then $\mathbb{G}(n,p)$ has largest component of size $O(\log n)$ in probability, whereas if $c > 1$ (the *supercritical* case) then the largest component of $\mathbb{G}(n,p)$ has size $(1 + o_p(1))\gamma(c)n$, where $\gamma(c)$ is the survival probability of a Poisson(c) branching process. They also showed that for $c > 1$, all components aside from the largest have size $O(\log n)$ in probability.

In view of the above coupling between the graph process and Kruskal's algorithm, the results of the preceding paragraph strongly suggest that "most of" the global structure of the MST $\mathbb{M}^n$ should already be present in the largest component

[7] Later, it will be convenient to allow $p \in \mathbb{R}$, and we note that the definition of $\mathbb{G}(n,p)$ still makes sense in this case.

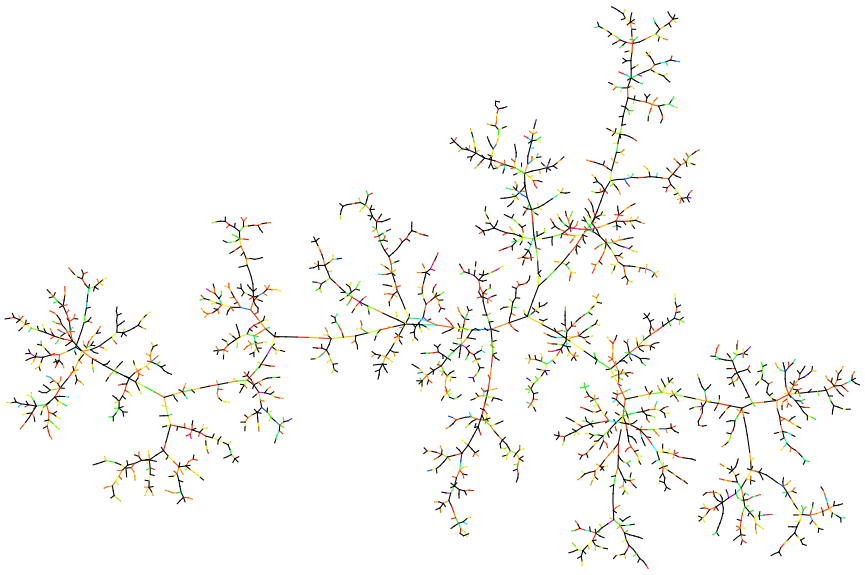

FIG. 1. *A simulation of the minimum spanning tree on K_n for $n = 3000$. Black edges have weights less than $1/n$; for coloured edges, weights increase as colours vary from from red to purple.*

of $\mathbb{M}(n, c/n)$, for any $c > 1$. In order to understand $\mathbb{M}^n$, then a natural approach is to delve into the structure of the forest $\mathbb{M}(n, p)$ for $p \sim 1/n$ (the *near-critical* regime) and, more specifically, to study how the components of this forest attach to one another as p increases through the critical window. (This is the reason for the colour-coding in Figure 1.) In this paper, we use such a strategy to show that after suitable rescaling of distances and of mass, the tree $\mathbb{M}^n$, viewed as a measured metric space, converges in distribution to a random compact measured metric space $\mathscr{M}$ of total mass measure one, which is a *random* $\mathbb{R}$*-tree* in the sense of [30, 49].

The space $\mathscr{M}$ is the scaling limit of the minimum spanning tree on the complete graph. It is binary and its mass measure is concentrated on the leaves of $\mathscr{M}$. The space $\mathscr{M}$ shares all these features with the first and most famous random $\mathbb{R}$-tree, the Brownian continuum random tree or CRT [9–11, 49]. However, $\mathscr{M}$ is not the CRT; we rule out this possibility by showing that $\mathscr{M}$ almost surely has Minkowski dimension 3. Since the CRT has both Minkowski dimension 2 [28] and Hausdorff dimension 2 [28, 38], this shows that the law of $\mathscr{M}$ is mutually singular with that of the CRT, or any rescaled version thereof.

The remainder of the Introduction is structured as follows. First, in Section 1.2, below, we provide the precise statement of our results. Second, in Section 1.3 we provide an overview of our proof techniques. Finally, in Section 1.4, we situate our results with respect to the large body of work by the probability and statistical physics communities on the convergence of minimum spanning trees, and briefly address the question of universality.

1.2. *The main results of this paper.* Before stating our results, a brief word on the spaces in which we work is necessary. We formally introduce these spaces in Section 2, and here only provide a brief summary. First, let $\mathcal{M}$ be the set of measured isometry-equivalence classes of compact measured metric spaces, and let $\mathrm{d_{GHP}}$ denote the Gromov–Hausdorff–Prokhorov distance on $\mathcal{M}$; the pair $(\mathcal{M}, \mathrm{d_{GHP}})$ forms a Polish space.

We wish to think of $\mathbb{M}^n$ as an element of $(\mathcal{M}, \mathrm{d_{GHP}})$. In order to do this, we introduce a measured metric space M^n obtained from $\mathbb{M}^n$ by rescaling distances by $n^{-1/3}$ and assigning mass $1/n$ to each vertex. The main contribution of this paper is the following theorem.

THEOREM 1.1. *There exists a random, compact measured metric space $\mathscr{M}$ such that, as $n \to \infty$,*

$$M^n \xrightarrow{d} \mathscr{M}$$

in the space $(\mathcal{M}, \mathrm{d_{GHP}})$. The limit $\mathscr{M}$ is a random $\mathbb{R}$-tree. It is almost surely binary, and its mass measure is concentrated on the leaves of $\mathscr{M}$. Furthermore, almost surely, the Minkowski dimension of $\mathscr{M}$ exists and is equal to 3.

A consequence of the last statement is that $\mathscr{M}$ is not a rescaled version of the Brownian CRT $\mathscr{T}$, in the sense that for any nonnegative random variable A, the laws of $\mathscr{M}$ and the space $\mathscr{T}$, in which all distances are multiplied by A, are mutually singular. Indeed, the Brownian tree has Minkowski dimension 2 almost surely. The assertions of Theorem 1.1 are contained within the union of Theorems 4.9 and 5.1 and Corollary 5.3, below.

In a preprint [2], the first author of this paper shows that the *unscaled* tree $\mathbb{M}^n$, when rooted at vertex 1, converges in the local weak sense to a random infinite tree, and that this limit almost surely has cubic volume growth. The results of [2] form a natural complement to Theorem 1.1.

As mentioned earlier, we approach the study of M^n and of its scaling limit $\mathscr{M}$ via a detailed description of the graph $\mathbb{G}(n, p)$ and of the forest $\mathbb{M}(n, p)$, for p near $1/n$. As is by this point well known, it turns out that the right scaling for the "critical window" is given by taking $p = 1/n + \lambda/n^{4/3}$, for $\lambda \in \mathbb{R}$, and for such p, the largest components of $\mathbb{G}(n, p)$ typically have size of order $n^{2/3}$ and possess a bounded number of cycles [12, 52]. Adopting this parametrisation, for $\lambda \in \mathbb{R}$ write

$$(\mathbb{G}_\lambda^{n,i}, i \geq 1)$$

for the components of $\mathbb{G}(n, 1/n + \lambda/n^{4/3})$ listed in decreasing order of size (among components of equal size, list components in increasing order of smallest vertex label, say). For each $i \geq 1$, we then write $G_\lambda^{n,i}$ for the measured metric space obtained from $\mathbb{G}_\lambda^{n,i}$ by rescaling distances by $n^{-1/3}$ and giving each vertex mass $n^{-2/3}$, and let

$$G_\lambda^n = (G_\lambda^{n,i}, i \geq 1).$$

We likewise define a sequence $(\mathbb{M}_\lambda^{n,i}, i \geq 1)$ of graphs, and a sequence $M_\lambda^n = (M_\lambda^{n,i}, i \geq 1)$ of measured metric spaces, starting from $\mathbb{M}(n, 1/n+\lambda/n^{4/3})$ instead of $\mathbb{G}(n, 1/n + \lambda/n^{4/3})$.

In order to compare sequences $\mathbf{X} = (\mathrm{X}_i, i \geq 1)$ of elements of $\mathcal{M}$ (i.e., elements of $\mathcal{M}^{\mathbb{N}}$), we let $\mathbb{L}_p$, for $p \geq 1$, be the set of sequences $\mathbf{X} \in \mathcal{M}^{\mathbb{N}}$ with

$$\sum_{i\geq 1} \mathrm{diam}(\mathrm{X}_i)^p + \sum_{i\geq 1} \mu_i(X_i)^p < \infty,$$

and for two such sequences $\mathbf{X} = (\mathrm{X}_i, i \geq 1)$ and $\mathbf{X}' = (\mathrm{X}'_i, i \geq 1)$, we let

$$\mathrm{dist}_{\mathrm{GHP}}^p(\mathbf{X}, \mathbf{X}') = \left(\sum_{i\geq 1} \mathrm{d}_{\mathrm{GHP}}(\mathrm{X}_i, \mathrm{X}'_i)^p\right)^{1/p}.$$

The resulting metric space $(\mathbb{L}_p, \mathrm{dist}_{\mathrm{GHP}}^p)$ is a Polish space. Convergence in this space in particular implies convergence of the first coordinates with respect to the product topology, but is stronger as it additionally yields information about the tails of the sequences.

The second main result of this paper is the following (see Theorems 4.4 and 4.9 below).

THEOREM 1.2. *Fix $\lambda \in \mathbb{R}$. Then there exists a random sequence $\mathscr{M}_\lambda = (\mathscr{M}_\lambda^i, i \geq 1)$ of compact measured metric spaces, such that as $n \to \infty$,*

$$M_\lambda^n \xrightarrow{d} \mathscr{M}_\lambda \tag{1.3}$$

in the space $(\mathbb{L}_4, \mathrm{dist}_{\mathrm{GHP}}^4)$. Furthermore, let $\hat{\mathscr{M}}_\lambda^1$ be the first term $\mathscr{M}_\lambda^1$ of the limit sequence $\mathscr{M}_\lambda$, with its measure renormalized to be a probability. Then as $\lambda \to \infty$, $\hat{\mathscr{M}}_\lambda^1$ converges in distribution to $\mathscr{M}$ in the space $(\mathcal{M}, \mathrm{d}_{\mathrm{GHP}})$.

1.3. *An overview of the proof.* Theorem 1 of [4] states that for each $\lambda \in \mathbb{R}$, there is a random sequence

$$\mathscr{G}_\lambda = (\mathscr{G}_\lambda^i, i \geq 1)$$

of compact measured metric spaces, such that

$$G_\lambda^n \xrightarrow{d} \mathscr{G}_\lambda, \tag{1.4}$$

in the space $(\mathbb{L}_4, \mathrm{dist}_{\mathrm{GHP}}^4)$. (Theorem 1 of [4] is, in fact, slightly weaker than this because the metric spaces there are considered without their accompanying measures, but it is easily strengthened; see Section 4.) The limiting spaces are similar to $\mathbb{R}$-trees; we call them $\mathbb{R}$-*graphs*. The introduction of $\mathbb{R}$-graphs, and a description of some of their basic properties, is one of the conceptual contributions of this paper. We expect this generalisation of the theory of $\mathbb{R}$-trees to find further applications.

In Section 2, we provide an an overview of the theory of $\mathbb{R}$-graphs. In particular, in Section 2.3 we develop a decomposition of $\mathbb{R}$-graphs analogous to the classical "core and kernel" decomposition for finite connected graphs (see, e.g., [41]). The more technical aspects of the proofs are deferred to Section 6. The main results of [3] provide precise distributional descriptions of the cores and kernels of the components of $\mathscr{G}_\lambda$, which may now be phrased as statements about the laws of random $\mathbb{R}$-graphs.

It turns out that, having understood the distribution of G_λ^n, we can access the distribution of M_λ^n by using a minimum spanning tree algorithm called *cycle breaking*. This algorithm finds the minimum weight spanning tree of a graph by listing edges in *decreasing* order of weight, then considering each edge in turn and removing it if its removal leaves the graph connected.

The continuum analogue of cycle breaking involves removing points from an $\mathbb{R}$-graph, an operation we call *cutting*. Ensuring that this operation is well defined and commutes with the appropriate limits is somewhat involved. We describe cycle breaking for graphs, and cutting for $\mathbb{R}$-graphs, in Section 3. Once again, proofs of some of the more challenging assertions are postponed, this time to Section 7.

In Section 4, using the convergence in (1.4) and an analysis of the cycle breaking algorithm, we will establish Theorem 1.2. The sequence $\mathscr{M}_\lambda$ is constructed from $\mathscr{G}_\lambda$ by cutting at appropriately sampled random points.

For fixed n, the process $(M_\lambda^{n,1}, \lambda \in \mathbb{R})$ is eventually constant, and we note that $\mathbb{M}^n = \lim_{\lambda\to\infty} \mathbb{M}_\lambda^{n,1}$ In order to establish that M^n converges in distribution in the space $(\mathcal{M}, \mathrm{d_{GHP}})$ as $n \to \infty$, we rely on two ingredients. First, the convergence in (1.3) implies that the first component $M_\lambda^{n,1}$ converges in distribution as $n \to \infty$ to a limit $\mathscr{M}_\lambda^1$ in the space $(\mathcal{M}, \mathrm{d_{GHP}})$.

Second, the results in [5] entail Lemma 4.5, which in particular implies that for any $\epsilon > 0$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\mathrm{d_{GH}}\big(M_\lambda^{n,1}, M^n\big) \geq \epsilon\big) = 0. \tag{1.5}$$

This is enough to prove a version of our main result for the metric spaces without their measures. In Proposition 4.8, below, we strengthen this statement. Let $\hat{M}_\lambda^{n,1}$ be the measured metric space obtained from $M_\lambda^{n,1}$ by rescaling so that the total mass is one [in $M_\lambda^{n,1}$ we gave each vertex mass $n^{-2/3}$; now we give each vertex mass $|V(\mathbb{M}_\lambda^{n,1})|^{-1}$]. We show that for any $\epsilon > 0$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\mathrm{d_{GHP}}\big(\hat{M}_\lambda^{n,1}, M^n\big) \geq \epsilon\big) = 0. \tag{1.6}$$

Since $(\mathcal{M}, \mathrm{d_{GHP}})$ is a complete, separable space, the so-called principle of accompanying laws entails that

$$M^n \xrightarrow{d} \mathscr{M}$$

in the space $(\mathcal{M}, d_{GHP})$ for some limiting random measured metric space $\mathscr{M}$ which is thus the scaling limit of the minimum spanning tree on the complete graph. Furthermore, still as a consequence of the principle of accompanying laws, $\mathscr{M}$ is also the limit in distribution of $\mathscr{M}_\lambda^1$ as $\lambda \to \infty$ in the space $(\mathcal{M}, d_{GHP})$.

Section 5 is focussed on the properties of $\mathscr{M}$. Many of the results in this section follow easily from the results on convergence established in Section 4.

First, for fixed $\lambda \in \mathbb{R}$, it follows from the distributional construction of $\mathscr{M}_\lambda$ given in Section 4.2 that the trees of the forest $\mathscr{M}_\lambda$ are almost surely binary. Since $\mathscr{M}$ is compact and (if the measure is ignored) is an increasing limit of $\mathscr{M}_\lambda^1$ as $\lambda \to \infty$, we shall deduce that $\mathscr{M}$ is almost surely binary; see Theorem 5.1.

Next, to prove that the mass measure is concentrated on the leaves of $\mathscr{M}$, we use a result of Łuczak [50] on the size of the largest component in the barely supercritical regime. This result in particular implies that for all $\epsilon > 0$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\left| \frac{|V(\mathbb{M}_\lambda^{n,1})|}{2\lambda n^{2/3}} - 1 \right| > \epsilon \right) = 0. \tag{1.7}$$

Since $\mathbb{M}_\infty^{n,1}$ has n vertices, it follows that for any $\lambda \in \mathbb{R}$, the proportion of the mass of $M_\infty^{n,1}$ "already present in $M_\lambda^{n,1}$" is asymptotically negligible. But (1.5) tells us that for λ large, with high probability every point of $M_\infty^{n,1}$ not in $M_\lambda^{n,1}$ has distance $o_{\lambda\to\infty}(1)$ from a point of $M_\lambda^{n,1}$, so has distance $o_{\lambda\to\infty}(1)$ from a leaf of $M_\infty^{n,1}$. Equation (1.7) then implies that essentially all the mass lies close to the leaves of $\mathbb{M}_\infty^{n,1}$. Passing this argument to the limit, it will follow that $\mathscr{M}$ almost surely places all its mass on its leaves.

Finally, the statement on the Minkowski dimension of $\mathscr{M}$, which appears in Proposition 5.2, depends crucially on an explicit description of the components of $\mathscr{G}_\lambda$ from [3], which allows us to estimate the number of balls needed to cover $\mathscr{M}_\lambda^1$. Along with a refined version of (1.5), which yields an estimate of the distance between $\mathscr{M}_\lambda^1$ and $\mathscr{M}$, we are able to obtain bounds on the covering number of $\mathscr{M}$. This is the most technical aspect of Section 5. We give further intuition for our argument just after the statement of Proposition 5.2.

This completes our overview, and we now proceed with a brief discussion of related work, before turning to details.

1.4. *Related work.* In the majority of existing work on convergence of MSTs, the limiting object is a noncompact infinite tree or forest. As detailed above, the study bifurcates into the "geometric" case in which the points lie in a Euclidean space $\mathbb{R}^d$, and the "mean-field" case where the underlying graph is K_n with i.i.d. edge weights. In both cases, a standard approach is to pass directly to an infinite underlying graph or point set, and define the minimum spanning tree (or forest) directly on such a point set.

It is not *a priori* obvious how to define the minimum spanning tree, or forest, of an infinite graph, as neither of the algorithms described above are necessarily well-defined (there may be no smallest weight edge leaving a given vertex or

component). However, it is known [13] that given an infinite locally finite graph $G = (V, E)$ and distinct edge weights $\mathbf{w} = \{w_e, e \in E\}$, the following variant of Prim's algorithm is well-defined and builds a forest, each component of which is an infinite tree.

INVASION PERCOLATION. For each $v \in V$, run Prim's algorithm starting from v and call the resulting set of edges E_v. Then let $\mathrm{MSF}(G, \mathbf{w})$ be the graph with vertices V and edges $\bigcup_{v \in V} E_v$.

The graph $\mathrm{MSF}(G, \mathbf{w})$ is also described by the following rule, which is conceptually based on the coupling between Kruskal's algorithm and the percolation process, described above. For each $r > 0$, let G_r be the subgraph with edges $\{e \in E : w_e < r\}$. Then an edge $e = uv \in E$ with $w_e = r$ is an edge of $\mathrm{MSF}(G, \mathbf{w})$ if and only if u and v are in distinct components of G_r and one of these components is finite.

The latter characterisation again allows the MSF to be studied by coupling with a percolation process. This connection was exploited by Alexander and Molchanov [17] in their proof that the MSF almost surely consists of a single, one-ended tree for the square, triangular and hexagonal lattices with i.i.d. Uniform[0, 1] edge weights and, later, to prove the same result for the MSF of the points of a homogeneous Poisson process in $\mathbb{R}^2$ [15]. Newman [57] has also shown that in lattice models in $\mathbb{R}^d$, the critical percolation probability $\theta(p_c)$ is equal to 0 if and only if Prim's algorithm a.s. only explores a vanishing proportion of the lattice (i.e., if for any fixed $v \in \mathbb{Z}^d$, almost surely $|E_v \cap [-r, r]^d|/r^d \to 0$). Lyons, Peres and Schramm [53] developed the connection with critical percolation. Among several other results, they showed that if G is any Cayley graph for which $\theta(p_c(G)) = 0$, then the component trees in the MSF all have one end almost surely, and that almost surely every component tree of the MSF itself has percolation threshold $p_c = 1$. (See also [74] for subsequent work on a similar model.) For two-dimensional lattice models, more detailed results about the behaviour of the so-called "invasion percolation tree", constructed by running Prim's algorithm once from a fixed vertex, have also recently been obtained [26, 27].

In the mean-field case, one common approach is to study the MST or MSF from the perspective of local weak convergence [14]. This leads one to investigate the minimum spanning forest of Aldous' *Poisson-weighted infinite tree* (PWIT). Such an approach is used implicitly in [54] in studying the first $O(\sqrt{n})$ steps of Prim's algorithm on K_n. Aldous [8] establishes a local weak limit for the subtree of the MST of K_n obtained as follows. Delete the (typically unique) edge whose removal minimizes the size of the component containing vertex 1 in the resulting graph, then keep only the component containing 1. The behaviour of invasion percolation on regular trees and on the PWIT has been studied in [6, 18, 19].

Almost nothing is known about *compact* scaling limits for whole MSTs. In two dimensions, Aizenman, Burchard, Newman and Wilson [7] have shown tightness

for the family of random sets given by considering the subtree of the MST connecting a finite set of points (the family is obtained by varying the set of points), either in the square, triangular or hexagonal lattice, or in a Poisson process. They also studied the properties of subsequential limits for such families, showing, among other results, that any limiting "tree" has Hausdorff dimension strictly between 1 and 2, and that the curves connecting points in such a tree are almost surely Hölder continuous of order α for any $\alpha < 1/2$. Recently, Garban, Pete, and Schramm [36] proved the existence of a scaling limit for the MST of the 2D triangular lattice, and showed that the limit is invariant under scalings, rotations and translations (the limit is *not* expected to be conformally invariant; see [79]). In the mean-field case, however, we are not aware of any previous work on scaling limits for the MST. In short, the scaling limit $\mathscr{M}$ that we identify in this paper appears to be a novel mathematical object, and is one of the first scaling limits to be identified for any problem from combinatorial optimisation.

We expect $\mathscr{M}$ to be a universal object: the MST's of a wide range of "high-dimensional" graphs should also have $\mathscr{M}$ as a scaling limit. By way of analogy, we mention two facts. First, Peres and Revelle [66] have shown the following universality result for *uniform* spanning trees (here informally stated). Let $\{G_n\}$ be a sequence of vertex transitive graphs of size tending to infinity. Suppose that (a) the uniform mixing time of simple random walk on G_n is $o(|G_n^{1/2}|)$, and (b) G_n is sufficiently "high-dimensional", in that the expected number of meetings between two random walks with the same starting point, in the first $|G_n|^{1/2}$ steps, is uniformly bounded. Then after a suitable rescaling of distances, the spanning tree of G_n converges to the CRT in the sense of finite-dimensional distributions. Second, under a related set of conditions, van der Hofstad and Nachmias [76] have very recently proved that the largest component of critical percolation on G_n in the barely supercritical phase has the same scaling as in the Erdős–Rényi graph process (we omit a precise statement of their result as it is rather technical, but mention that their conditions are general enough to address the notable case of percolation on the hypercube). However, a proof of an analogous result for the MST seems, at this time, quite distant. As will be seen below, our proof requires detailed control on the metric and mass structure of all components of the Kruskal process in the critical window and, for the moment, this is not available for any other model. In another direction, the recent paper by Bhamidi, Broutin, Sen and Wang [23] paves the way for such universality results on models of random graphs, including inhomogeneous random graphs and the configuration model, by identifying the scaling limit of critical percolation clusters for these models, hence generalizing the results [4] which serve as a basis for the present work.

2. Metric spaces and types of convergence. The reader may wish to simply skim this section on a first reading, referring back to it as needed. Interested readers can find proofs of the various assertions regarding convergence of metric spaces in, for example, [1, 30, 55].

2.1. *Notions of convergence.*

The Gromov–Hausdorff distance. Given a metric space (X, d), we write $[X, d]$ for the isometry class of (X, d), and use the notation X for either (X, d) or $[X, d]$ when there is no risk of ambiguity. For a metric space (X, d), we write $\operatorname{diam}((X, d)) = \sup_{x,y\in X} d(x, y)$, which may be infinite.

Let $\mathrm{X} = (X, d)$ and $\mathrm{X}' = (X', d')$ be metric spaces. If C is a subset of $X \times X'$, the *distortion* $\operatorname{dis}(C)$ is defined by

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

A *correspondence* C between X and X' is a measurable subset of $X \times X'$ such that for every $x \in X$, there exists $x' \in X'$ with $(x, x') \in C$ and vice versa. Write $C(X, X')$ for the set of correspondences between X and X'. The Gromov–Hausdorff distance $\mathrm{d_{GH}}(\mathrm{X}, \mathrm{X}')$ between the isometry classes of (X, d) and (X', d') is

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

and when X and X′ are compact, there is a correspondence which achieves this infimum. (In fact, since our metric spaces are assumed separable, the requirement that the correspondence be measurable is not strictly necessary.) It can be verified that $\mathrm{d_{GH}}$ is indeed a distance and, writing $\mathring{\mathcal{M}}$ for the set of isometry classes of compact metric spaces, that $(\mathring{\mathcal{M}}, \mathrm{d_{GH}})$ is itself a complete separable metric space. (We reserve the notation $\mathcal{M}$ for compact *measured* metric spaces.)

Let $(X, d, (x_1, \ldots, x_k))$ and $(X', d', (x'_1, \ldots, x'_k))$ be metric spaces, each with an ordered set of k distinguished points (we call such spaces *k-pointed metric spaces*).[8] We say that these two k-pointed metric spaces are *isometry-equivalent* if there exists an isometry $\phi : X \to X'$ such that $\phi(x_i) = x'_i$ for every $i \in \{1, \ldots, k\}$. As before, we write $[X, d, (x_1, \ldots, x_k)]$ for the isometry equivalence class of $(X, d, (x_1, \ldots, x_k))$, and denote either by X when there is little chance of ambiguity.

The *k-pointed Gromov–Hausdorff distance* is defined as

$$\mathrm{d}^k_{\mathrm{GH}}(\mathrm{X}, \mathrm{X}') = \frac{1}{2}\inf\bigl\{\operatorname{dis}(C) : C \in C(X, X') \text{ s.t. } (x_i, x'_i) \in C, 1 \le i \le k\bigr\}.$$

Much as above, the space $(\mathcal{M}^{(k)}, \mathrm{d}^k_{\mathrm{GH}})$ of isometry classes of k-pointed compact metric spaces is itself a complete separable metric space [55].

[8] When $k = 1$, we simply refer to pointed (rather than 1-pointed) metric spaces, and write (X, d, x) rather than $(X, d, (x))$.

The Gromov–Hausdorff–Prokhorov distance. A *compact measured metric space* is a triple (X, d, μ) where (X, d) is a compact metric space and μ is a (nonnegative) finite measure on $(X, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on (X, d). Given a measured metric space (X, d, μ), a metric space (X', d') and a measurable function $\phi : X \to X'$, we write $\phi_*\mu$ for the push-forward of the measure μ to the space (X', d'). Two compact measured metric spaces (X, d, μ) and (X', d', μ') are called *isometry-equivalent* if there exists an isometry $\phi : (X, d) \to (X', d')$ such that $\phi_*\mu = \mu'$. The isometry-equivalence class of (X, d, μ) will be denoted by $[X, d, \mu]$. Again, both will often be denoted by X when there is little risk of ambiguity. If $\mathrm{X} = (X, d, \mu)$, then we write $\mathrm{mass}(\mathrm{X}) = \mu(X)$.

There are several natural distances on compact measured metric spaces that generalize the Gromov–Hausdorff distance; see, for instance, [1, 31, 55, 78]. The presentation we adopt is still different from these references, but closest in spirit to [1] since we are dealing with arbitrary finite measures rather than just probability measures. In particular, it induces the same topology as the compact Gromov–Hausdorff–Prokhorov metric of [1].

If (X, d) and (X', d') are two metric spaces, let $M(X, X')$ be the set of finite nonnegative Borel measures on $X \times X'$. We will denote by p, p' the canonical projections from $X \times X'$ to X and X'.

Let μ and μ' be finite nonnegative Borel measures on X and X', respectively. The *discrepancy* of $\pi \in M(X, X')$ with respect to μ and μ' is the quantity

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

where $\|\nu\|$ is the total variation of the signed measure ν. Since $p_*\pi(X) = \pi(X \times X')$, we have $\|\mu - p_*\pi\| \geq |\mu(X) - \pi(X \times X')|$ by the definition of total variation. By the triangle inequality it follows that $D(\pi; \mu, \mu') \geq |\mu(X) - \mu'(X')|$. If μ and μ' are probability distributions (or have the same mass), a measure $\pi \in M(X, X')$ with $D(\pi; \mu, \mu') = 0$ is a coupling of μ and μ' in the standard sense.

Recall that the Prokhorov distance between two finite nonnegative Borel measures μ and μ' on the *same* metric space (X, d) is given by

$$\inf\{\epsilon > 0 : \mu(F) \leq \mu'(F^\epsilon) + \epsilon \text{ and } \mu'(F) \leq \mu(F^\epsilon) + \epsilon \text{ for every closed } F \subseteq X\}.$$

An alternative distance, which generates the same topology but more easily extends to the setting where μ and μ' are measures on different metric spaces, is given by

$$\inf\{\epsilon > 0 : D(\pi; \mu, \mu') < \epsilon, \pi(\{(x, x') \in X \times X : d(x, x') \geq \epsilon\}) < \epsilon \text{ for some } \pi \in M(X, X)\}.$$

To extend this, we replace the condition on $\{(x, x') \in X \times X : d(x, x') \geq \epsilon\}$ by an analogous condition on the measure of the set of pairs lying outside the correspondence. More precisely, let $\mathrm{X} = (X, d, \mu)$ and $\mathrm{X}' = (X', d', \mu')$ be measured metric

spaces. The Gromov–Hausdorff–Prokhorov distance between X and X′ is defined as

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

the infimum being taken over all $C \in C(\mathrm{X},\mathrm{X}')$ and $\pi \in M(X,X')$. Here and elsewhere, we write $x \vee y = \max(x,y)$ [and, likewise, $x \wedge y = \min(x,y)$].

Just as for $\mathrm{d_{GH}}$, it can be verified that $\mathrm{d_{GHP}}$ is a distance and that writing $\mathcal{M}$ for the set of measured isometry-equivalence classes of compact measured metric spaces, $(\mathcal{M},\mathrm{d_{GHP}})$ is a complete separable metric space (see, e.g., [1]).

Note that $\mathrm{d_{GHP}}((X,d,0),(X',d',0)) = \mathrm{d_{GH}}((X,d),(X',d'))$. In other words, the mapping $[X,d] \mapsto [X,d,0]$ is an isometric embedding of $(\mathring{\mathcal{M}},\mathrm{d_{GH}})$ into $\mathcal{M}$, and we will sometimes abuse notation by writing $[X,d] \in \mathcal{M}$. Note also that

$$\begin{aligned}\mathrm{d_{GH}}(\mathrm{X},\mathrm{X}') \vee \big|\mu(X) - \mu'(X')\big| &\le \mathrm{d_{GHP}}(\mathrm{X},\mathrm{X}')\\ &\le \frac{1}{2}\big(\operatorname{diam}(\mathrm{X}) + \operatorname{diam}(\mathrm{X}')\big) \vee \big(\mu(X) + \mu'(X')\big).\end{aligned}$$

In particular, if Z is the "zero" metric space consisting of a single point with measure 0, then

$$\mathrm{d_{GHP}}(\mathrm{X},\mathrm{Z}) = \frac{\operatorname{diam}(\mathrm{X})}{2} \vee \mu(X) \qquad \text{for every } \mathrm{X} = [X,d,\mu]. \tag{2.1}$$

Finally, we can define an analogue of $\mathrm{d_{GHP}}$ for measured isometry-equivalence class of spaces of the form $(X,d,\mathbf{x},\boldsymbol{\mu})$ where $\mathbf{x} = (x_1,\dots,x_k)$ are points of X and $\boldsymbol{\mu} = (\mu_1,\dots,\mu_l)$ are finite Borel measures on X. If $(X,d,\mathbf{x},\boldsymbol{\mu})$ and $(X',d',\mathbf{x}',\boldsymbol{\mu}')$ are such spaces, whose measured, pointed isometry classes are denoted by X and X′, we let

$$\mathrm{d}^{k,l}_{\mathrm{GHP}}(\mathrm{X},\mathrm{X}') = \inf\left\{\frac{1}{2}\operatorname{dis}(C) \vee \max_{1\le j\le l}\big(D(\pi_j;\mu_j,\mu'_j) \vee \pi_j\big(C^c\big)\big)\right\},$$

where the infimum is over all $C \in C(X,X')$ such that $(x_i,x'_i) \in C$, $1 \le i \le k$ and all $\pi_j \in M(X,X')$, $1 \le j \le l$. Writing $\mathcal{M}^{k,l}$ for the set of measured isometry-equivalence classes of compact metric spaces equipped with k marked points and l finite Borel measures, we again obtain a complete separable metric space $(\mathcal{M}^{k,l},\mathrm{d}^{k,l}_{\mathrm{GHP}})$. We will need the following fact, which is in essence [55], Proposition 10, except that we have to take into account more measures and/or marks. This is a minor modification of the setting of [55], and the proof is similar.

PROPOSITION 2.1. *Suppose that* $\mathrm{X}_n = (X_n,d_n,\mathbf{x}_n,\boldsymbol{\mu}_n)$ *converges to* $\mathrm{X}_\infty = (X_\infty,d_\infty,\mathbf{x}_\infty,\boldsymbol{\mu}_\infty)$ *in* $\mathcal{M}^{k,l}$, *and assume that the first measure* μ^1_n *of* $\boldsymbol{\mu}_n$ *is a probability measure for every* $n \in \mathbb{N}\cup\{\infty\}$. *Let* y_n *be a random variable with distribution* μ^1_n, *and let* $\tilde{\mathbf{x}}_n = (x^1_n,\dots,x^k_n,y_n)$. *Then* $(X_n,d_n,\tilde{\mathbf{x}}_n,\boldsymbol{\mu}_n)$ *converges in distribution to* $(X_\infty,d_\infty,\tilde{\mathbf{x}}_\infty,\boldsymbol{\mu}_\infty)$ *in* $\mathcal{M}^{k+1,l}$.

Sequences of metric spaces. We now consider a natural metric on certain sequences of measured metric spaces. For $p \geq 1$ and $\mathbf{X} = (\mathrm{X}_i, i \geq 1), \mathbf{X}' = (\mathrm{X}'_i, i \geq 1)$ in $\mathcal{M}^{\mathbb{N}}$, we let

$$\mathrm{dist}^p_{\mathrm{GHP}}(\mathbf{X}, \mathbf{X}') = \Bigl(\sum_{i\geq 1} \mathrm{d}_{\mathrm{GHP}}(\mathrm{X}_i, \mathrm{X}'_i)^p\Bigr)^{1/p}.$$

If $\mathbf{X} \in \mathcal{M}^n$ for some $n \in \mathbb{N}$, we consider $\mathbf{X}$ as an element of $\mathcal{M}^{\mathbb{N}}$ by appending to $\mathbf{X}$ an infinite sequence of copies of the "zero" metric space Z. This allows us to use $\mathrm{dist}^p_{\mathrm{GHP}}$ to compare sequences of metric spaces with different numbers of elements, and to compare finite sequences with infinite sequences. In particular, let $\mathbf{Z} = (\mathrm{Z}, \mathrm{Z}, \ldots)$, and recall from Section 1.2 that

$$\mathbb{L}_p = \{\mathbf{X} \in \mathcal{M}^{\mathbb{N}} : \mathrm{dist}^p_{\mathrm{GHP}}(\mathbf{X}, \mathbf{Z}) < \infty\}.$$

By (2.1), $\mathbf{X} \in \mathbb{L}_p$ if and only if the sequences $(\mathrm{diam}(\mathrm{X}_i), i \geq 1)$ and $(\mu_i(X_i), i \geq 1)$ are in $\ell^p(\mathbb{N})$.

2.2. *Some general metric notions.* Let (X, d) be a metric space. For $x \in X$ and $r \geq 0$, we let $B_r(x) = \{y \in X : d(x, y) < r\}$ and $\overline{B}_r(x) = \{y \in X : d(x, y) \leq r\}$. We say (X, d) is *degenerate* if $|X| = 1$. As regards metric spaces, we mostly follow [25] for our terminology.

Paths, length, cycles. Let $\mathcal{C}([a, b], X)$ be the set of continuous functions from $[a, b]$ to X, hereafter called *paths with domain* $[a, b]$ or *paths from a to b*. The image of a path is called an *arc*; it is a *simple arc* if the path is injective. If $f \in \mathcal{C}([a, b], X)$, the length of f is defined by

$$\mathrm{len}(f) = \sup\Biggl\{\sum_{i=1}^{k} d\bigl(f(t_{i-1}), f(t_i)\bigr) : k \geq 1, t_0, \ldots, t_k \in [a, b], t_0 \leq \cdots \leq t_k\Biggr\}.$$

If $\mathrm{len}(f) < \infty$, then the function $\varphi : [a, b] \to [0, \mathrm{len}(f)]$ defined by $\varphi(t) = \mathrm{len}(f|_{[a,t]})$ is nondecreasing and surjective. The function $f \circ \varphi^{-1}$, where φ^{-1} is the right-continuous inverse of φ, is easily seen to be continuous, and we call it the path f *parameterized by arc-length*.

The *intrinsic distance* (or *intrinsic metric*) associated with (X, d) is the function d_l defined by

$$d_l(x, y) = \inf\bigl\{\mathrm{len}(f) : f \in \mathcal{C}\bigl([0, 1], X\bigr), f(0) = x, f(1) = y\bigr\}.$$

The function d_l need not take finite values. When it does, then it defines a new distance on X such that $d \leq d_l$. The metric space (X, d) is called *intrinsic* if $d = d_l$. Similarly, if $Y \subset X$ then the *intrinsic metric on* Y is given by

$$d_l(x, y) = \inf\bigl\{\mathrm{len}(f) : f \in \mathcal{C}\bigl([0, 1], Y\bigr), f(0) = x, f(1) = y\bigr\}.$$

Intrinsic metrics will play a key role when we consider the effect on the metric structure of removing edges in a graph.

Given $x, y \in X$, a *geodesic between x and y* (also called a *shortest path between x and y*) is an isometric embedding $f : [a, b] \to X$ such that $f(a) = x$ and $f(b) = y$ [so that obviously $\text{len}(f) = b - a = d(x, y)$]. In this case, we call the image $\text{Im}(f)$ a *geodesic arc between x and y*.

A metric space (X, d) is called a geodesic space if for any two points x, y there exists a geodesic between x and y. A geodesic space is obviously an intrinsic space. If (X, d) is compact, then the two notions are in fact equivalent. Also note that for every x in a geodesic space and $r > 0$, $\overline{B}_r(x)$ is the closure of $B_r(x)$. All of the limiting metric spaces (X, d) that we consider in this paper are in fact compact geodesic spaces. When working with a graph, it is convenient to move back and forth between two viewpoints: it may be viewed as a finite metric space consisting of the vertex set equipped with the graph distance, or as a geodesic space, by viewing edges as line segments of length one.

A path $f \in \mathcal{C}([a, b], X)$ is a *local geodesic* between x and y if $f(a) = x$, $f(b) = y$, and for any $t \in [a, b]$ there is a neighborhood V of t in $[a, b]$ such that $f|_V$ is a geodesic. It is then straightforward that $b - a = \text{len}(f)$. (Our terminology differs from that of [25], where this would be called a geodesic. We also note that we do not require x and y to be distinct.)

An *embedded cycle* is the image of a continuous injective function $f : \mathbb{S}_1 \to X$, where $\mathbb{S}_1 = \{z \in \mathbb{C} : |z| = 1\}$. The length $\text{len}(f)$ is the length of the path $g : [0, 1] \to X$ defined by $g(t) = f(e^{2i\pi t})$ for $0 \le t \le 1$. It is easy to see that this length depends only on the embedded cycle $c = \text{Im}(f)$ rather than its particular parametrisation. We call it the *length* of the embedded cycle, and write $\text{len}(c)$ for this length. A metric space with no embedded cycle is called *acyclic*, and a metric space with exactly one embedded cycle is called *unicyclic*.

2.3. $\mathbb{R}$*-trees and* $\mathbb{R}$*-graphs*. A metric space $\mathrm{X} = (X, d)$ is an $\mathbb{R}$*-tree* if it is an acyclic geodesic metric space. If (X, d) is an $\mathbb{R}$-tree then for $x \in T$, the *degree* $\deg_X(x)$ of x is the number of connected components of $X \setminus \{x\}$. A *leaf* is a point of degree 1; we let $\mathcal{L}(\mathrm{X})$ be the set of leaves of X.

A metric space (X, d) is an $\mathbb{R}$*-graph* if it is locally an $\mathbb{R}$-tree in the following sense. Note that by definition an $\mathbb{R}$-graph is connected, being a geodesic space.

DEFINITION 2.2. A compact geodesic metric space (X, d) is an $\mathbb{R}$*-graph* if for every $x \in X$, there exists $\epsilon = \epsilon(x) > 0$ such that $(B_\epsilon(x), d|_{B_\epsilon(x)})$ is an $\mathbb{R}$-tree.

Let $\mathrm{X} = (X, d)$ be an $\mathbb{R}$-graph and fix $x \in X$. The degree of x, denoted by $\deg_X(x)$ and with values in $\mathbb{N} \cup \{\infty\}$, is defined to be the degree of x in $B_\epsilon(x)$ for every ϵ small enough so that $(B_\epsilon(x), d)$ is an $\mathbb{R}$-tree, and this definition does not depend on a particular choice of ϵ. If $Y \subset X$ and $x \in Y$, we can likewise define the degree $\deg_Y(x)$ of x in Y as the degree of x in the $\mathbb{R}$-tree $(B_\epsilon(x) \cap Y(x)) \setminus \{x\}$,

where $Y(x)$ is the connected component of Y that contains x, for any ϵ small enough. Obviously, $\deg_Y(x) \le \deg_{Y'}(x)$ whenever $Y \subset Y'$.

Let

$$\mathcal{L}(\mathrm{X}) = \{x \in X : \deg_X(x) = 1\}, \qquad \mathrm{skel}(\mathrm{X}) = \{x \in X : \deg_X(x) \ge 2\}.$$

An element of $\mathcal{L}(\mathrm{X})$ is called a *leaf of* X, and the set skel(X) is called the *skeleton of* X. A point with degree at least 3 is called a *branchpoint of* X. We let $k(\mathrm{X})$ be the set of branchpoints of X.

We provide two simple examples to help understand these definitions. First, suppose that X is a cycle (i.e., is isometric to $\mathbb{S}^1$). Then $\mathrm{skel}(\mathrm{X}) = X$, and there are no branchpoints or leaves. Second, suppose that X is a lollipop (i.e. consists of a cycle and a line segment joined to the cycle at one end). Then there is one branchpoint (the common point of the cycle and the line segment), which has degree 3, and one leaf (the other end of the line segment); the leaf is also the only point not in skel(X).

If X is, in fact, an $\mathbb{R}$-tree, then skel(X) is the set of points whose removal disconnects the space, but this is not true in general. Alternatively, it is easy to see that

$$\mathrm{skel}(\mathrm{X}) = \bigcup_{\substack{x,y \in X \\ c \in \Gamma(x,y)}} c \setminus \{x, y\},$$

where for $x, y \in X$, $\Gamma(x, y)$ denotes the collection of all geodesic arcs between x and y. Since (X, d) is separable, this may be re-written as a countable union, and so there is a unique σ-finite Borel measure ℓ on X with $\ell(\mathrm{Im}(g)) = \mathrm{len}(g)$ for every injective path g, and such that $\ell(X \setminus \mathrm{skel}(\mathrm{X})) = 0$. The measure ℓ is the Hausdorff measure of dimension 1 on X, and we refer to it as the *length measure* on X. If (X, d) is an $\mathbb{R}$-graph then the set $\{x \in X : \deg_X(x) \ge 3\}$ is countable, and hence this set has measure zero under ℓ. To see that this set is countable, assume (X, d) is an $\mathbb{R}$-tree (otherwise consider a finite open cover of X by $\mathbb{R}$-trees). Fix a countable dense set of points $Y \subset X$, and let Z be the set of points of the subtree spanned by Y. It is easily seen that the set $\{x \in Z : \deg_X(x) \ge 3\}$ is countable. But the closure of Z is X so all points of $X \setminus Z$ must have degree 1.

DEFINITION 2.3. Let (X, d) be an $\mathbb{R}$-graph. Its *core*, denoted by core(X), is the union of all the simple arcs having both endpoints in embedded cycles of X. If it is nonempty, then $(\mathrm{core}(\mathrm{X}), d)$ is an $\mathbb{R}$-graph with no leaves.

The last part of this definition is in fact a proposition, which is stated more precisely and proved below as Proposition 6.2. Since the core of X encapsulates all the embedded cycles of X, it is intuitively clear that when we remove core(X) from X, we are left with a family of $\mathbb{R}$-trees. This can be formalized as follows. Fix $x \in X \setminus \mathrm{core}(\mathrm{X})$, and let f be a shortest path from x to core(X), that is, a geodesic

from x to $y \in \mathrm{core(X)}$, where $y \in \mathrm{core(X)}$ is chosen so that $\mathrm{len}(f)$ is minimal [recall that core(X) is a closed subspace of X]. This shortest path is unique, otherwise we would easily be able to construct an embedded cycle c not contained in core(X), contradicting the definition of core(X). Let $\alpha(x)$ be the endpoint of this path not equal to x, which is thus the unique point of core(X) that is closest to x. By convention, we let $\alpha(x) = x$ if $x \in \mathrm{core(X)}$. We call $\alpha(x)$ the *point of attachment* of x.

PROPOSITION 2.4. *The relation* $x \sim y \iff \alpha(x) = \alpha(y)$ *is an equivalence relation on* X. *If* $[x]$ *is the equivalence class of* x, *then* $([x], d)$ *is a compact* $\mathbb{R}$*-tree. The equivalence class* $[x]$ *of a point* $x \in \mathrm{core(X)}$ *is a singleton if and only if* $\deg_X(x) = \deg_{\mathrm{core(X)}}(x)$.

PROOF. The fact that $\sim$ is an equivalence relation is obvious. Fix any equivalence class $[x]$. Note that $[x] \cap \mathrm{core(X)}$ contains only the point $\alpha(x)$, so that $[x]$ is connected and acyclic by definition. Hence, any two points of $[x]$ are joined by a unique simple arc (in $[x]$). This path is moreover a shortest path for the metric d, because a path starting and ending in $[x]$, and visiting $X \setminus [x]$, must pass at least twice through $\alpha(x)$ [if this were not the case, we could find an embedded cycle not contained in core(X)]. The last statement is easy and left to the reader. □

COROLLARY 2.5. *If* (X, d) *is an* $\mathbb{R}$*-graph, then* core(X) *is the maximal closed subset of* X *having only points of degree greater than or equal to* 2.

PROOF. If Y is closed and strictly contains core(X), then we can find $x \in Y$ such that $d(x, \mathrm{core(X)}) = d(x, \alpha(x)) > 0$ is maximal. Then $Y \cap [x]$ is included in the set of points $y \in [x]$ such that the geodesic arc from y to $\alpha(x)$ does not pass through x. This set is an $\mathbb{R}$-tree in which x is a leaf, so $\deg_Y(x) \leq 1$. □

Note that this characterisation is very close to the definition of the core of a (discrete) graph. Another important structural component is conn(X), the set of points of core(X) such that $X \setminus \{x\}$ is connected. Figure 2 summarizes the preceding definitions. The space conn(X) is not connected or closed in general. Clearly, a point of conn(X) must be contained in an embedded cycle of X, but the converse is not necessarily true. A partial converse is as follows.

PROPOSITION 2.6. *Let* $x \in \mathrm{core(X)}$ *have degree* $\deg_X(x) = 2$ *and suppose* x *is contained in an embedded cycle of* X. *Then* $x \in \mathrm{conn(X)}$.

PROOF. Let c be an embedded cycle containing x. Fix $y, y' \in X \setminus \{x\}$, and let ϕ, ϕ' be geodesics from y, y' to their respective closest points $z, z' \in c$. Note that z is distinct from x because otherwise, x would have degree at least 3. Likewise, $z' \neq x$.

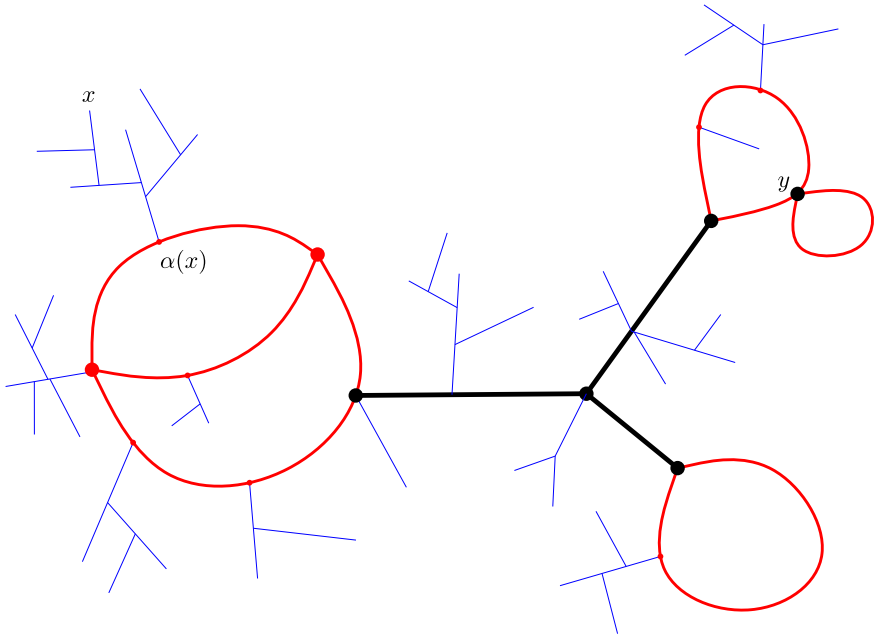

FIG. 2. *An example of an* $\mathbb{R}$*-graph* (X, d)*, emphasizing the structural components.* conn(X) *is in thick red line, and* core(X) $\setminus$ conn(X) *is in very thick black line. The point* y *is in* core(X) *but not* conn(X)*. The subtrees hanging from* core(X) *are in thin blue line. Kernel vertices are represented as large dots. An example of the projection* $\alpha : X \to$ core(X) *is provided.*

Let ϕ'' be a parametrisation of the arc of c between z and z' that does not contain x, then the concatenation of ϕ, ϕ' and the time-reversal of the path ϕ'' is a path from y to y', not passing through x. Hence, $X \setminus \{x\}$ is connected. □

Let us now discuss the structure of core(X). Equivalently, we need to describe $\mathbb{R}$-graphs with no leaves, because such graphs are equal to their cores by Corollary 2.5.

A *graph with edge-lengths* is a triple $(V, E, (l(e), e \in E))$ where (V, E) is a finite connected multigraph, and $l(e) \in (0, \infty)$ for every $e \in E$. With every such object, one can associate an $\mathbb{R}$-graph without leaves, which is the *metric graph* obtained by viewing the edges of (V, E) as segments with respective lengths $l(e)$. Formally, this $\mathbb{R}$-graph is the metric gluing of disjoint copies Y^e of the real segments $[0, l(e)], e \in E$ according to the graph structure of (V, E). We refer the reader to [25] for details on metric gluings and metric graphs.

THEOREM 2.7. *An* $\mathbb{R}$*-graph with no leaves is either a cycle, or is the metric gluing of a finite connected multigraph with edge-lengths in which all vertices have degree at least* 3*. The associated multigraph, without the edge-lengths, is called the kernel of* X*, and denoted by* $\mathrm{ker(X)} = (k(\mathrm{X}), e(\mathrm{X}))$.

The precise definition of ker(X), and the proof of Theorem 2.7, both appear in Section 6.3.

For a connected multigraph $G = (V, E)$, the *surplus* $s(G)$ is $|E| - |V| + 1$. For an $\mathbb{R}$-graph (X, d), we let $s(\mathrm{X}) = s(\ker(\mathrm{X}))$ if ker(X) is nonempty. Otherwise, either (X, d) is an $\mathbb{R}$-tree or core(X) is a cycle. In the former case, we set $s(\mathrm{X}) = 0$; in the latter we set $s(\mathrm{X}) = 1$. Since the degree of every vertex in ker(X) is at least 3, we have $2|e(\mathrm{X})| = \sum_{v \in k(\mathrm{X})} \deg(v) \geq 3|k(\mathrm{X})|$, and so if $s(\mathrm{X}) \geq 1$ we have

$$\big|k(\mathrm{X})\big| \leq 2s(\mathrm{X}) - 2, \tag{2.2}$$

with equality precisely if ker(X) is 3-regular.

3. Cycle-breaking in discrete and continuous graphs.

3.1. *The cycle-breaking algorithm.* Let $G = (V, E)$ be a finite connected multigraph. Let $\mathrm{conn}(G)$ be the set of of all edges $e \in E$ such that $G \setminus e = (V, E \setminus \{e\})$ is connected.

If $s(G) > 0$, then G contains at least one cycle and $\mathrm{conn}(G)$ is nonempty. In this case, let e be a uniform random edge in $\mathrm{conn}(G)$, and let $K(G, \cdot)$ be the law of the multigraph $G \setminus e$. If $s(G) = 0$, then $K(G, \cdot)$ is the Dirac mass at G. By definition, K is a Markov kernel from the set of graphs with surplus s to the set of graphs with surplus $(s - 1) \vee 0$. Writing K^n for the n-fold application of the kernel K, we have that $K^n(G, \cdot)$ does not depend on n for $n \geq s(G)$. We define the kernel $K^\infty(G, \cdot)$ to be equal to this common value: a graph has law $K^\infty(G, \cdot)$ if it is obtained from G by repeatedly removing uniform nondisconnecting edges.

PROPOSITION 3.1. *The probability distribution $K^\infty(G, \cdot)$ is the law of the minimum spanning tree of G, when the edges E are given exchangeable, distinct random edge-weights.*

PROOF. We prove by induction on the surplus of G the stronger statement that $K^\infty(G, \cdot)$ is the law of the minimum spanning tree of G, when the weights of $\mathrm{conn}(G)$ are given exchangeable, distinct random edge-weights [the edge weights outside $\mathrm{conn}(G)$ may have arbitrary values]. For $s(G) = 0$, the result is obvious.

Assume the result holds for every graph of surplus s, and let G have $s(G) = s + 1$. Let e be the edge of $\mathrm{conn}(G)$ with maximal weight, and condition on e and its weight. Then note that the weights of the edges in $\mathrm{conn}(G) \setminus \{e\}$ are still in exchangeable random order, and the same is true of the edges of $\mathrm{conn}(G \setminus e)$. By the induction hypothesis, $K^s(G \setminus e, \cdot)$ is the law of the minimum spanning tree of $G \setminus e$. But e is not in the minimum spanning tree of G, because by definition we can find a path between its endpoints that uses only edges having strictly smaller weights. Hence, $K^s(G \setminus e, \cdot)$ is the law of the minimum spanning tree of G. On the other hand, by exchangeability, the edge e of $\mathrm{conn}(G)$ with largest weight is uniform in $\mathrm{conn}(G)$, so the unconditional law of a random variable with law $K^s(G \setminus e, \cdot)$ is $K^{s+1}(G, \cdot)$. $\square$

3.2. *Cutting the cycles of an* $\mathbb{R}$*-graph.* There is a continuum analogue of the cycle-breaking algorithm in the context of $\mathbb{R}$-graphs, which we now explain. Recall that $\mathrm{conn}(\mathbb{X})$ is the set of points x of the $\mathbb{R}$-graph $\mathbb{X} = (X, d)$ such that $x \in \mathrm{core}(\mathbb{X})$ and $X \setminus \{x\}$ is connected. For $x \in \mathrm{conn}(\mathbb{X})$, we let (X_x, d_x) be the space $\mathbb{X}$ "cut at x". Formally, it is the metric completion of $(X \setminus \{x\}, d_{X\setminus\{x\}})$, where $d_{X\setminus\{x\}}$ is the intrinsic distance: $d_{X\setminus\{x\}}(y, z)$ is the minimal length of a path from y to z that does not visit x.

DEFINITION 3.2. A point $x \in X$ in a measured $\mathbb{R}$-graph $\mathbb{X} = (X, d, \mu)$ is a *regular point* if $x \in \mathrm{conn}(\mathbb{X})$, and moreover $\mu(\{x\}) = 0$ and $\deg_X(x) = 2$. A marked space $(X, d, x, \mu) \in \mathcal{M}^{1,1}$, where (X, d) is an $\mathbb{R}$-graph and x is a regular point, is called *safely pointed*. We say that a pointed $\mathbb{R}$-graph (X, d, x) is safely pointed if $(X, d, x, 0)$ is safely pointed.

Before continuing, we briefly motivate the above definition. In Section 7, we will consider the effect of cutting cycles in each element of a convergent sequence of $\mathbb{R}$-graphs, and we wish to ensure that the cutting operation behaves continuously with respect to taking limits of such sequences. For a point $x \in \mathrm{conn}(\mathbb{X})$, enforcing that $\deg_X(x) = 2$ is equivalent to requiring that $x \notin k(\mathbb{X})$, which is enough to yield continuity for the graph structure of the kernels. We insist that $\mu(\{x\}) = 0$ so that μ induces a measure (still denoted by μ) on the space $\mathbb{X}_x$ with the same total mass. We will give a precise description of the space $\mathbb{X}_x = (X_x, d_x, \mu)$ in Section 7.1: in particular, it is a measured $\mathbb{R}$-graph with $s(\mathbb{X}_x) = s(\mathbb{X}) - 1$.

Note that if $s(\mathbb{X}) > 0$ and if

$$L = \ell\big(\cdot \cap \mathrm{conn}(\mathbb{X})\big)$$

is the length measure restricted to $\mathrm{conn}(\mathbb{X})$, then L-almost every point is regular. Also, L is a finite measure by Theorem 2.7. Therefore, it makes sense to let $\mathcal{K}(\mathbb{X}, \cdot)$ be the law of $\mathbb{X}_x$, where x is a random point of X with law $L/L(\mathrm{conn}(\mathbb{X}))$. If $s(\mathbb{X}) = 0$, we let $\mathcal{K}(\mathbb{X}, \cdot) = \delta_{\{\mathbb{X}\}}$. Again, $\mathcal{K}$ is a Markov kernel from the set of measured $\mathbb{R}$-graphs with surplus s to the set of measured $\mathbb{R}$-graphs of surplus $(s - 1) \vee 0$, and $\mathcal{K}^n(\mathbb{X}, \cdot) = \mathcal{K}^{s(\mathbb{X})}(\mathbb{X}, \cdot)$ for every $n \geq s(\mathbb{X})$: we denote this by $\mathcal{K}^\infty(\mathbb{X}, \cdot)$.

In Section 7, we will give details of the proofs of the aforementioned properties, as well as of the following crucial result. For $r \in (0, 1)$ we let $\mathcal{A}_r$ be the set of measured $\mathbb{R}$-graphs with $s(\mathbb{X}) \leq 1/r$ and whose core, seen as a graph with edge-lengths $(k(\mathbb{X}), e(\mathbb{X}), (\ell(e), e \in e(\mathbb{X})))$, is such that

$$\min_{e\in e(\mathbb{X})} \ell(e) \geq r \quad \text{and} \quad \sum_{e\in e(\mathbb{X})} \ell(e) \leq 1/r$$

(if $s(\mathbb{X}) = 1$, this should be understood as the fact that $\mathrm{core}(\mathbb{X})$ is a cycle with length in $[r, 1/r]$).

THEOREM 3.3. *Fix $r \in (0,1)$. Let (X^n, d^n, μ^n) be a sequence of measured $\mathbb{R}$-graphs in $\mathcal{A}_r$, converging as $n \to \infty$ to $(X, d, \mu) \in \mathcal{A}_r$ in $(\mathcal{M}, \mathrm{d_{GHP}})$. Then $\mathcal{K}^\infty(\mathrm{X}^n, \cdot)$ converges weakly to $\mathcal{K}^\infty(\mathrm{X}, \cdot)$.*

3.3. *A relation between the discrete and continuum procedures.* We can view any finite connected multigraph $G = (V, E)$ as a metric space (V, d), where $d(u, v)$ is the least number of edges in any chain from u to v. We may also consider the metric graph $(m(G), d_{m(G)})$ associated with G by treating edges as segments of length 1 (this is sometimes known as the *cable system* for the graph G [77]). Then $(m(G), d_{m(G)})$ is an $\mathbb{R}$-graph. Note that $\mathrm{d_{GH}}((V, d), (m(G), d_{m(G)})) < 1$ and, in fact, $(m(G), d_{m(G)})$ contains an isometric copy of (V, d). Also, temporarily writing H for the *graph-theoretic* core of G, that is, the maximal subgraph of G of minimum degree two, it is straightforwardly checked that $\mathrm{core}(m(G))$ is isometric to $(m(H), d_{m(H)})$.

Conversely, let (X, d) be an $\mathbb{R}$-graph, and let S_X be the set of points in X with degree at least three. We say that (X, d) *has integer lengths* if all local geodesics between points in S_X have lengths in $\mathbb{Z}_+$. If G is a finite graph and the edges of G are all viewed as having length 1, then cutting G at a point chosen according to its length measure yields an $\mathbb{R}$-graph with integer lengths. This is essentially the reason why such objects will be useful to us.

Let

$$v(\mathrm{X}) = \{x \in X : d(x, S_X) \in \mathbb{Z}_+\},$$

and note that if (X, d) is compact and has integer lengths then necessarily $|S_X| < \infty$ and $|v(\mathrm{X})| < \infty$. The removal of all points in $v(\mathrm{X})$ separates X into a finite collection of paths, each of which is either an open path of length one between two points of $v(\mathrm{X})$, or a half-open path of length strictly less than one between a point of $v(\mathrm{X})$ and a leaf. Create an edge between the endpoints of each such *open* path, and call the collection of such edges $e(\mathrm{X})$. Then let

$$g(\mathrm{X}) = \big(v(\mathrm{X}), e(\mathrm{X})\big);$$

we call the multigraph $g(\mathrm{X})$ the *graph corresponding to* X (see Figure 3).

Now, fix an $\mathbb{R}$-graph (X, d) which has integer lengths and surplus $s(\mathrm{X})$. Let $x_1, \ldots, x_{s(\mathrm{X})}$ be the points sampled by the successive applications of $\mathcal{K}$ to X: given $x_1, \ldots, x_i$, the point x_{i+1} is chosen according to $L/L(\mathrm{X})$ on $\mathrm{conn}(\mathrm{X}_{x_1,\ldots,x_i})$, where $\mathrm{X}_{x_1,\ldots,x_i}$ is the space X cut successively at $x_1, x_2, \ldots, x_i$. Note that x_i can also naturally be seen as a point of X for $1 \le i \le s(\mathrm{X})$. Since the length measure of $v(\mathrm{X})$ is 0, almost surely $x_i \ne v(\mathrm{X})$ for all $1 \le i \le s(\mathrm{X})$. Thus, each point x_i, $1 \le i \le s(\mathrm{X})$, falls in a path component of $\mathrm{core}(\mathrm{X}) \setminus v(\mathrm{X})$ which itself corresponds uniquely to an edge in $e_i \in e(\mathrm{X})$. Note that the edges e_i, $1 \le i \le s(\mathrm{X})$, are distinct by construction. Then let $g_0(\mathrm{X}) = g(\mathrm{X})$, and for $1 \le i \le s(\mathrm{X})$, write

$$g_i(\mathrm{X}) = \big(v(\mathrm{X}), e(\mathrm{X}) \setminus \{e_1, \ldots, e_i\}\big).$$

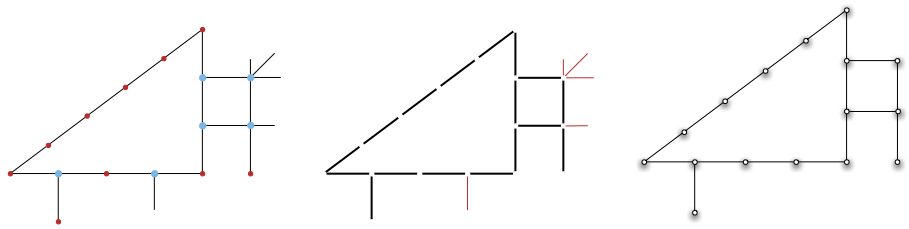

FIG. 3. *Left*: *an* $\mathbb{R}$*-graph with integer lengths. The points of degree at least three are larger and blue, and the remaining points of* $v(\mathrm{X})$ *are smaller and red. Centre*: *the collection of paths after the points of* $v(\mathrm{X})$ *are removed. The paths with noninteger lengths are thinner and are red. Right*: *the graph* $g(\mathrm{X})$.

By construction, the graph $g_i(\mathrm{X})$ is connected and has surplus precisely $s(\mathrm{X}) - i$, and in particular $g_{s(\mathrm{X})}(\mathrm{X})$ is a spanning tree of $g(\mathrm{X})$. Let $\mathrm{cut}(\mathrm{X})$ be the random $\mathbb{R}$-graph resulting from the application of $\mathcal{K}^\infty$, that is obtained by cutting X at the points $x_1, \ldots, x_{s(\mathrm{X})}$ in our setting.

PROPOSITION 3.4. *We have* $\mathrm{d_{GH}}(\mathrm{cut}(\mathrm{X}), g_{s(\mathrm{X})}(\mathrm{X})) < 1$.

PROOF. First, notice that $g_{s(\mathrm{X})}(\mathrm{X})$ and $g(\mathrm{cut}(\mathrm{X}))$ are isomorphic as graphs, so isometric as metric spaces. Also, as noted in greater generality at the start of the subsection, we automatically have $\mathrm{d_{GH}}(\mathrm{cut}(\mathrm{X}), g(\mathrm{cut}(\mathrm{X}))) < 1$. □

PROPOSITION 3.5. *The graph* $g(\mathrm{cut}(\mathrm{X}))$ *is identical in distribution to the minimum-weight spanning tree of* $g(\mathrm{X})$ *when the edges of* $e \in e(\mathrm{X})$ *are given exchangeable*, *distinct random edge weights*.

PROOF. When performing the discrete cycle-breaking on $g(\mathrm{X})$, the set of edges removed from $g(\mathrm{X})$ is identical in distribution to the set $\{e_1, \ldots, e_{s(\mathrm{X})}\}$ of edges that are removed from $g(\mathrm{X})$ to create $g_{s(\mathrm{X})}(\mathrm{X})$, so $g_{s(\mathrm{X})}(\mathrm{X})$ has the same distribution as the minimum spanning tree by Proposition 3.1. Furthermore, as noted in the proof of the preceding proposition, $g_{s(\mathrm{X})}(\mathrm{X})$ and $g(\mathrm{cut}(\mathrm{X}))$ are isomorphic. □

3.4. *Gluing points in* $\mathbb{R}$*-graphs*. We end this section by mentioning the operation of gluing, which in a vague sense is dual to the cutting operation. If (X, d, μ) is an $\mathbb{R}$-graph and x, y are two distinct points of X, we let $\mathrm{X}^{x,y}$ be the quotient metric space [25] of (X, d) by the smallest equivalence relation for which x and y are equivalent. This space is endowed with the push-forward of μ by the canonical projection p. It is not difficult to see that $\mathrm{X}^{x,y}$ is again an $\mathbb{R}$-graph, and that the class of the point $z = p(x) = p(y)$ has degree $\deg_{X^{x,y}}(z) = \deg_X(x) + \deg_X(y)$. Similarly, if $\mathcal{R}$ is a finite set of unordered pairs $\{x_i, y_i\}$ with $x_i \neq y_i$ in X, then one can identify x_i and y_i for each i, resulting in an $\mathbb{R}$-graph $\mathrm{X}^{\mathcal{R}}$.

4. Convergence of the MST. We are now ready to state and prove the main results of this paper. We begin by recalling from the Introduction that we write $\mathbb{M}^n$ for the MST of the complete graph on n vertices and M^n for the measured metric space obtained from $\mathbb{M}^n$ by rescaling the graph distance by $n^{-1/3}$ and assigning mass $1/n$ to each vertex.

4.1. *The scaling limit of the Erdős–Rényi random graph.* Recall that $\mathbb{G}(n, p)$ is the Erdős–Rényi random graph. For $\lambda \in \mathbb{R}$, we write

$$\mathbb{G}_\lambda^n = \big(\mathbb{G}_\lambda^{n,i}, i \geq 1\big)$$

for the components of $\mathbb{G}(n, 1/n + \lambda/n^{4/3})$ listed in decreasing order of size (among components of equal size, list components in increasing order of smallest vertex label, say). For each $i \geq 1$, we then write $G_\lambda^{n,i}$ for the measured metric space obtained from $\mathbb{G}_\lambda^{n,i}$ by rescaling the graph distance by $n^{-1/3}$ and giving each vertex mass $n^{-2/3}$, and let

$$G_\lambda^n = \big(G_\lambda^{n,i}, i \geq 1\big).$$

In a moment, we will state a scaling limit result for G_λ^n; before we can do so, however, we must introduce the limit sequence of measured metric spaces $\mathscr{G}_\lambda = (\mathscr{G}_\lambda^i, i \geq 1)$. We will do this somewhat briefly, and refer the interested reader to [3, 4] for more details and distributional properties.

First, consider the stochastic process $(W_\lambda(t), t \geq 0)$ defined by

$$W_\lambda(t) := W(t) + \lambda t - \frac{t^2}{2},$$

where $(W(t), t \geq 0)$ is a standard Brownian motion. Consider the excursions of W_λ above its running minimum; in other words, the excursions of

$$B_\lambda(t) := W_\lambda(t) - \min_{0 \leq s \leq t} W_\lambda(s)$$

above 0. We list these in decreasing order of length as $(\varepsilon^1, \varepsilon^2, \ldots)$ where, for $i \geq 1$, σ^i is the length of ε^i. (We suppress the λ-dependence in the notation for readability.) For definiteness, we shift the origin of each excursion to 0, so that $\varepsilon^i : [0, \sigma^i] \to \mathbb{R}_+$ is a continuous function such that $\varepsilon^i(0) = e^i(\sigma^i) = 0$ and $\varepsilon^i(x) > 0$ otherwise.

Now for $i \geq 1$ and for $x, x' \in [0, \sigma^i]$, define a pseudo-distance via

$$\hat{d}^i(x, x') = 2\varepsilon^i(x) + 2\varepsilon^i(x') - 4 \inf_{x \wedge x' \leq t \leq x \vee x'} \varepsilon^i(t).$$

Declare that $x \sim x'$ if $\hat{d}^i(x, x') = 0$, so that $\sim$ is an equivalence relation on $[0, \sigma^i]$. Now let $\mathcal{T}^i = [0, \sigma^i]/\sim$ and denote by $\tau^i : [0, \sigma^i] \to \mathcal{T}^i$ the canonical projection. Then $\hat{d}^i$ induces a distance on $\mathcal{T}^i$, still denoted by $\hat{d}^i$, and it is standard (see, e.g.,

[49]) that $(\mathcal{T}^i, \hat{d}^i)$ is a compact $\mathbb{R}$-tree. Write $\hat{\mu}^i$ for the push-forward of Lebesgue measure on $[0, \sigma^i]$ by τ^i, so that $(\mathcal{T}^i, \hat{d}^i, \hat{\mu}^i)$ is a measured $\mathbb{R}$-tree of total mass σ^i.

We now decorate the process B_λ with the points of an independent homogeneous Poisson process in the plane. We can think of the points which fall under the different excursions separately. In particular, to the excursion ε^i, we associate a finite collection $\mathcal{P}^i = \{(x^{i,j}, y^{i,j}), 1 \le j \le s^i\}$ of points of $[0, \sigma^i] \times [0, \infty)$ which are the Poisson points shifted in the same way as the excursion ε^i. (For definiteness, we list the points of $\mathcal{P}^i$ in increasing order of first co-ordinate.) Conditional on $\varepsilon^1, \varepsilon^2, \ldots$, the collections $\mathcal{P}^1, \mathcal{P}^2, \ldots$ of points are independent. Moreover, by construction, given the excursion ε^i, we have $s^i \sim \text{Poisson}(\int_0^{\sigma^i} \varepsilon^i(t)\,\mathrm{d}t)$. Let $z^{i,j} = \inf\{t \ge x^{i,j} : \varepsilon^i(t) = y^{i,j}\}$ and note that, by the continuity of ε^i, $z^{i,j} < \sigma^i$ almost surely. Let

$$\mathcal{R}^i = \big\{\{\tau^i(x^{i,j}), \tau^i(z^{i,j})\}, 1 \le j \le s^i\big\}.$$

Then $\mathcal{R}^i$ is a collection of unordered pairs of points in the $\mathbb{R}$-tree $\mathcal{T}^i$. We wish to glue these points together in order to obtain an $\mathbb{R}$-graph, as in Section 3.4. We define a new equivalence relation $\sim$ by declaring $x \sim x'$ in $\mathcal{T}^i$ if $\{x, x'\} \in \mathcal{R}^i$. Then let $\mathcal{X}^i$ be $\mathcal{T}^i/\!\sim$, let d^i be the quotient metric [25], and let μ^i be the push-forward of $\hat{\mu}^i$ to $\mathcal{X}^i$. Then set $\mathcal{G}^i_\lambda = (\mathcal{X}^i, d^i, \mu^i)$ and $\mathcal{G}_\lambda = (\mathcal{G}^i_\lambda, i \ge 1)$. We note that for each $i \ge 1$, the measure $\hat{\mu}^i$ is almost surely concentrated on the leaves of $\mathcal{T}^i$ since the law of $\mathcal{T}^i$ is absolutely continuous with respect to that of the Brownian CRT. As a consequence, μ^i is almost surely concentrated on the leaves of $\mathcal{X}^i$.

Given an $\mathbb{R}$-graph X, write $r(\mathrm{X})$ for the minimal length of a core edge in X. Then $r(\mathrm{X}) = \inf\{d(u, v) : u, v \in k(\mathrm{X})\}$ whenever $\ker(\mathrm{X})$ is nonempty. We use the convention that $r(\mathrm{X}) = \infty$ if $\mathrm{core}(\mathrm{X}) = \varnothing$ and $r(\mathrm{X}) = \ell(c)$ if X has a unique embedded cycle c. Recall also that $s(\mathrm{X})$ denotes the surplus of X.

THEOREM 4.1. *Fix $\lambda \in \mathbb{R}$. Then as $n \to \infty$, we have the following joint convergence*:

$$G^n_\lambda \xrightarrow{d} \mathcal{G}_\lambda,$$

$$\big(s(G^{n,i}_\lambda), i \ge 1\big) \xrightarrow{d} \big(s(\mathcal{G}^i_\lambda), i \ge 1\big) \quad \textit{and}$$

$$\big(r(G^{n,i}_\lambda), i \ge 1\big) \xrightarrow{d} \big(r(\mathcal{G}^i_\lambda), i \ge 1\big).$$

The first convergence takes place in the space $(\mathbb{L}_4, \mathrm{dist}^4_{\mathrm{GHP}})$. *The others are in the sense of finite-dimensional distributions.*

Recall the definition of mass(X) for a measured metric space X, from Section 2.1. Let $\ell_2^\downarrow = \{x = (x_1, x_2, \ldots) : x_1 \ge x_2 \ge \cdots \ge 0, \sum_{i=1}^\infty x_i^2 < \infty\}$. Corol-

lary 2 of [12] gives the following joint convergence:

$$\begin{aligned}(\text{mass}(G_\lambda^{n,i}), i \geq 1) &\xrightarrow{d} (\text{mass}(\mathscr{G}_\lambda^i), i \geq 1) \quad \text{and} \\ (s(G_\lambda^{n,i}), i \geq 1) &\xrightarrow{d} (s(\mathscr{G}_\lambda^i), i \geq 1),\end{aligned} \tag{4.1}$$

where the first convergence is in $(\ell_2^\downarrow, \|\cdot\|_2)$ and the second is in the sense of finite-dimensional distributions. [Of course, $\text{mass}(\mathscr{G}_\lambda^i) = \sigma^i$ and $s(\mathscr{G}_\lambda^i) = s^i$.] Theorem 1 of [4] extends this to give that, jointly,

$$(G_\lambda^{n,i}, i \geq 1) \xrightarrow{d} (\mathscr{G}_\lambda^i, i \geq 1) \tag{4.2}$$

in the sense of $\text{dist}_{\text{GH}}^4$, where for $\mathbf{X}, \mathbf{Y} \in \mathring{\mathcal{M}}^{\mathbb{N}}$, we let

$$\text{dist}_{\text{GH}}^4(\mathbf{X}, \mathbf{Y}) = \left(\sum_{i=1}^{\infty} \text{d}_{\text{GH}}(\text{X}^i, \text{Y}^i)^4 \right)^{1/4}.$$

We need to improve this convergence from $\text{dist}_{\text{GH}}^4$ to $\text{dist}_{\text{GHP}}^4$. First, we show that we can get GHP convergence componentwise. We do this in two lemmas.

LEMMA 4.2. *Suppose that $(\mathcal{T}, d, \mu)$ and $(\mathcal{T}', d', \mu')$ are measured $\mathbb{R}$-trees, that $\{(x_i, y_i), 1 \leq i \leq k\}$ are pairs of points in $\mathcal{T}$ and that $\{(x_i', y_i'), 1 \leq i \leq k\}$ are pairs of points in $\mathcal{T}'$. Then if $(\hat{\mathcal{T}}, \hat{d}, \hat{\mu})$ and $(\hat{\mathcal{T}}', \hat{d}', \hat{\mu}')$ are the measured metric spaces obtained by identifying x_i and y_i in $\mathcal{T}$ and x_i' and y_i' in $\mathcal{T}'$, for all $1 \leq i \leq k$, we have*

$$\text{d}_{\text{GHP}}\big((\hat{\mathcal{T}}, \hat{d}, \hat{\mu}), (\hat{\mathcal{T}}', \hat{d}', \hat{\mu}')\big) \leq (k+1)\text{d}_{\text{GHP}}^{2k,1}\big((\mathcal{T}, d, \mathbf{x}, \mu), (\mathcal{T}', d', \mathbf{x}', \mu')\big),$$

where $\mathbf{x} = (x_1, \ldots, x_k, y_1, \ldots, y_k)$, and similarly for $\mathbf{x}'$.

PROOF. Let C and π be a correspondence and a measure which realise the Gromov–Hausdorff–Prokhorov distance between $(\mathcal{T}, d, \mathbf{x}, \mu)$ and $(\mathcal{T}', d', \mathbf{x}', \mu')$; write δ for this distance. Note that by definition, $(x_i, x_i') \in C$ and $(y_i, y_i') \in C$ for $1 \leq i \leq k$. Now make the vertex identifications in order to obtain $\hat{\mathcal{T}}$ and $\hat{\mathcal{T}}'$; let $p : \mathcal{T} \to \hat{\mathcal{T}}$ and $p' : \mathcal{T}' \to \hat{\mathcal{T}}'$ be the corresponding canonical projections. Then

$$\hat{C} = \{(p(x), p'(x')) : (x, x') \in \mathcal{R}'\}$$

is a correspondence between $\hat{\mathcal{T}}$ and $\hat{\mathcal{T}}'$. Let $\hat{\pi}$ be the push-forward of the measure π by (p, p'). Then $D(\hat{\pi}; \hat{\mu}, \hat{\mu}') \leq \delta$ and $\hat{\pi}(\hat{C}^c) \leq \delta$. Moreover, by Lemma 21 of [4], we have $\text{dis}(\hat{C}) \leq (k+1)\delta$. The claimed result follows. □

LEMMA 4.3. *Fix $i \geq 1$. Then as $n \to \infty$,*

$$G_\lambda^{n,i} \xrightarrow{d} \mathscr{G}_\lambda^i$$

in $(\mathcal{M}, \text{d}_{\text{GHP}})$.

PROOF. This proof is a fairly straightforward modification of the proof of Theorem 22 in [4], so we will only sketch the argument. Consider the component $\mathbb{G}_\lambda^{n,i}$. Since we have fixed λ and i, let us drop them from the notation and simply write $\mathbb{G}^n$ for the component, and similarly for other objects. Let $\Sigma^n = n^{-2/3}|V(\mathbb{G}^n)|$ and write $S^n \in \mathbb{Z}_+$ for the surplus of $\mathbb{G}^n$. We observe that $\Sigma^n = \mathrm{mass}(G_\lambda^{n,i})$ and $S^n = s(G_\lambda^{n,i})$.

List the vertices of $\mathbb{G}^n$ in depth-first order, starting from the vertex of smallest label, as $v_0, v_1, \ldots, v_{n^{2/3}\Sigma^n - 1}$, and write T^n for the depth-first tree of $\mathbb{G}^n$, both as defined in Section 2 of [4]. We note that T^n is rooted at v_0 and is a spanning subtree of $\mathbb{G}^n$. Let $H^n(k)$ be the graph distance of vertex v_k from v_0 in T^n, and endow T^n with a measure by letting each vertex of T^n have mass $n^{-2/3}$.

Next, let the pairs $\{i_1, j_1\}, \{i_2, j_2\}, \ldots, \{i_{S^n}, j_{S^n}\}$ give the indices of the surplus edges required to obtain $\mathbb{G}^n$ from T^n, listed in increasing order of first co-ordinate. In other words, to build $\mathbb{G}^n$ from T^n, we add an edge between v_{i_k} and v_{j_k} for each $1 \le k \le S^n$ (and re-multiply distances by $n^{1/3}$). Recall that to get G^n from $\mathbb{G}^n$, we rescale the graph distance by $n^{-1/3}$ and assign mass $n^{-2/3}$ to each vertex. It is straightforward that G^n is at GHP distance at most $n^{-1/3}S^n$ from the metric space $\hat{G}^n$ obtained from T^n by *identifying* vertices v_{i_k} and v_{j_k} for all $1 \le k \le S^n$.

From the proof of Theorem 22 of [4], we have jointly

$$\bigl(\Sigma^n, S^n\bigr) \xrightarrow{d} (\sigma, s),$$

$$\bigl(n^{-1/3}H^n\bigl(\lfloor n^{2/3}t\rfloor\bigr), 0 \le t < \Sigma^n\bigr) \xrightarrow{d} \bigl(2\varepsilon(t), 0 \le t < \sigma\bigr),$$

$$\bigl\{\bigl\{n^{-2/3}i_k, n^{-2/3}j_k\bigr\}, 0 \le k \le S^n\bigr\} \xrightarrow{d} \bigl\{\bigl\{x^k, z^k\bigr\}, 1 \le k \le s\bigr\}.$$

By Skorokhod's representation theorem, we may work on a probability space where these convergences hold almost surely. Consider the $\mathbb{R}$-tree $(\mathcal{T}, d_\mathcal{T})$ encoded by 2ε and recall that τ is the canonical projection $[0, \sigma] \to \mathcal{T}$. We extend τ to a function on $[0, \infty)$ by letting $\tau(t) = \tau(t \wedge \sigma)$. Let $\eta^n : [0, \infty) \to \{v_0, v_1, \ldots, v_{n^{2/3}\Sigma^n - 1}\}$ be the function defined by $\eta^n(t) = v_{\lfloor n^{2/3}t\rfloor \wedge (n^{2/3}\Sigma^n - 1)}$. Set

$$C^n = \bigl\{\bigl(\eta^n(t), \tau\bigl(t'\bigr)\bigr) : t, t' \in \bigl[0, \Sigma^n \vee \sigma\bigr], \bigl|t - t'\bigr| \le \delta_n\bigr\},$$

where δ_n converges to 0 slowly enough, that is,

$$\delta_n \ge \max_{1 \le k \le s} \bigl|n^{-2/3}i_k - x^k\bigr| \vee \bigl|n^{-2/3}j_k - z^k\bigr|.$$

Then C^n is a correspondence between T^n and $\mathcal{T}$ that contains (v_{i_k}, x^k) and (v_{j_k}, z^k) for every $k \in \{1, 2, \ldots, s\}$, and with distortion going to 0. Next, let π^n be the push-forward of Lebesgue measure on $[0, \Sigma^n \wedge \sigma]$ under the mapping (η^n, τ). Then the discrepancy of π^n with respect to the uniform measure μ^n on T^n and the image μ of Lebesgue measure by τ on $\mathcal{T}$ is $|\Sigma^n - \sigma|$, and $\pi^n((C^n)^c) = 0$.

For all large enough n, $S^n = s$, so let us assume henceforth that this holds. Then, writing $\mathbf{v} = (v_{i_1}, \dots, v_{i_s}, v_{j_1}, \dots, v_{j_s})$ and $\mathbf{x} = (x^1, \dots, x^k, z^1, \dots, z^k)$, we have

$$\mathrm{d}_{\mathrm{GHP}}^{2s,1}\big((T^n, \mathbf{v}, \mu_n), (\mathcal{T}, \mathbf{x}, \mu)\big) \le \left(\frac{1}{2}\operatorname{dis}(C^n)\right) \vee |\Sigma^n - \sigma| \to 0$$

almost surely, as $n \to \infty$. By Lemma 4.2, we thus have $\mathrm{d}_{\mathrm{GHP}}(\hat{G}^n, \mathscr{G}) \to 0$ almost surely, as $n \to \infty$. Since $\mathrm{d}_{\mathrm{GHP}}(G^n, \hat{G}^n) \le n^{-1/3} S_n \to 0$, it follows that $\mathrm{d}_{\mathrm{GHP}}(G^n, \mathscr{G}) \to 0$ as well. $\square$

PROOF OF THEOREM 4.1. By (4.1), (4.2), Lemma 4.3 and Skorokhod's representation theorem, we may work in a probability space in which the convergence in (4.1) and in (4.2) occur almost surely, and in which for every $i \ge 1$ we almost surely have

$$\mathrm{d}_{\mathrm{GHP}}\big(G_\lambda^{n,i}, \mathscr{G}_\lambda^i\big) \to 0 \tag{4.3}$$

as $n \to \infty$. Now, for each $i \ge 1$,

$$\mathrm{d}_{\mathrm{GHP}}\big(G_\lambda^{n,i}, \mathscr{G}_\lambda^i\big) \le 2\max\big\{\operatorname{diam}\big(G_\lambda^{n,i}\big), \operatorname{diam}\big(\mathscr{G}_\lambda^i\big), \operatorname{mass}\big(G_\lambda^{n,i}\big), \operatorname{mass}\big(\mathscr{G}_\lambda^i\big)\big\}.$$

The proof of Theorem 24 from [4] shows that almost surely

$$\lim_{N\to\infty} \sum_{i=N}^{\infty} \operatorname{diam}\big(\mathscr{G}_\lambda^i\big)^4 = 0,$$

and (4.2) then implies that almost surely

$$\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{i=N}^{\infty} \operatorname{diam}\big(G_\lambda^{n,i}\big)^4 = 0.$$

The $\ell_2^{\downarrow}$ convergence of the masses entails that almost surely

$$\lim_{N\to\infty} \sum_{i=N}^{\infty} \operatorname{mass}\big(\mathscr{G}_\lambda^i\big)^4 = 0$$

and (4.1) then implies that almost surely

$$\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{i=N}^{\infty} \operatorname{mass}\big(G_\lambda^{n,i}\big)^4 = 0.$$

Hence, on this probability space, we have

$$\begin{aligned}&\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{i=N}^{\infty} \mathrm{d}_{\mathrm{GHP}}\big(G_\lambda^{n,i}, \mathscr{G}_\lambda^i\big)^4 \\ &\qquad \le 16 \lim_{N\to\infty} \limsup_{n\to\infty} \sum_{i=N}^{\infty} \big(\operatorname{diam}\big(G_\lambda^{n,i}\big)^4 + \operatorname{diam}\big(\mathscr{G}_\lambda^i\big)^4\end{aligned}$$

$$\begin{aligned} &+ \operatorname{mass}\bigl(G_\lambda^{n,i}\bigr)^4 + \operatorname{mass}\bigl(\mathscr{G}_\lambda^i\bigr)^4\bigr) \\ &= 0 \end{aligned}$$

almost surely. Combined with (4.3), this implies that in this space, almost surely

$$\lim_{n\to\infty} \operatorname{dist}_{\mathrm{GHP}}^4\bigl(G_\lambda^n, \mathscr{G}_\lambda\bigr) = 0.$$

The convergence of $(s(G_\lambda^{n,i}), i \geq 1)$ to $(s(\mathscr{G}_\lambda^i), i \geq 1)$ follows from (4.1).

If i is such that $s(\mathscr{G}_\lambda^i) = 1$ then, by (4.1), we almost surely have $s(G_\lambda^{n,i}) = 1$ for all n sufficiently large. In this case, $r(G_\lambda^{n,i})$ and $r(\mathscr{G}_\lambda^i)$ are the lengths of the unique cycles in $G_\lambda^{n,i}$ and in $\mathscr{G}_\lambda^i$, respectively. Now, $G_\lambda^{n,i} \to \mathscr{G}_\lambda^i$ almost surely in $(\mathring{\mathcal{M}}, \mathrm{d_{GH}})$, and it follows easily that in this space, $r(G_\lambda^{n,i}) \to r(\mathscr{G}_\lambda^i)$ almost surely, for i such that $s(\mathscr{G}_\lambda^i) = 1$.

Finally, by Theorem 4 of [52], $\min(r(G_\lambda^{n,i}) : s(G_\lambda^{n,i}) \geq 2)$ is bounded away from zero in probability. So by Skorokhod's representation theorem, we may assume our space is such that almost surely

$$\liminf_{n\to\infty} \min\bigl(r\bigl(G_\lambda^{n,i}\bigr) : s\bigl(G_\lambda^{n,i}\bigr) \geq 2\bigr) > 0.$$

In particular, it follows from the above that, for any $i \geq 1$ with $s(\mathscr{G}_\lambda^i) \geq 2$, there is almost surely $r > 0$ such that $\mathscr{G}_\lambda^i \in \mathcal{A}^r$ and $G_\lambda^{n,i} \in \mathcal{A}^r$ for all n sufficiently large. Corollary 6.6(i) then yields that in this space, $r(G_\lambda^{n,i}) \to r(\mathscr{G}_\lambda^i)$ almost surely.

Together, the two preceding paragraphs establish the final claimed convergence. For completeness, we note that this final convergence may also be deduced without recourse to the results of [52]; here is a brief sketch, using the notation of the previous lemma. It is easily checked that the points of the kernels of $G_\lambda^{n,i}$ and $\mathscr{G}_\lambda^i$ correspond to the identified vertices (v_{i_k}, v_{j_k}) and (x^k, z^k), and those vertices of degree at least 3 in the subtrees of $T^n, \mathcal{T}$ spanned by the points (v_{i_k}, v_{j_k}), $1 \leq k \leq s$ and (x^k, z^k), $1 \leq k \leq s$, respectively. These trees are combinatorially finite trees (i.e., they are finite trees with edge-lengths), so the convergence of the marked trees $(T^n, \mathbf{v})$ to $(\mathcal{T}, \mathbf{x})$ entails in fact the convergence of the same trees marked not only by $\mathbf{v}, \mathbf{x}$ but also by the points of degree 3 on their skeletons. Write $\mathbf{v}', \mathbf{x}'$ for these enlarged collections of points. Then one concludes by noting that $r(G_\lambda^{n,i})$ [resp., $r(\mathscr{G}_\lambda^i)$] is the minimum quotient distance, after the identifications (v_{i_k}, v_{j_k}) [resp., (x^k, z^k)] between any two distinct elements of $\mathbf{v}'$ (resp., $\mathbf{x}'$). This entails that $r(G_\lambda^{n,i})$ converges almost surely to $r(\mathscr{G}_\lambda^i)$ for each $i \geq 1$. □

The above description of the sequence $\mathscr{G}_\lambda$ of random $\mathbb{R}$-graphs does not make the distribution of the cores and kernels of the components explicit. [Clearly, the kernel of $\mathscr{G}_\lambda^i$ is only nonempty if $s(\mathscr{G}_\lambda^i) \geq 2$ and its core is only nonempty if $s(\mathscr{G}_\lambda^i) \geq 1$.] Such an explicit distributional description was provided in [3], and will be partially detailed below in Section 5.

4.2. *Convergence of the minimum spanning forest.* Recall that $\mathbb{M}(n, p)$ is the minimum spanning forest of $\mathbb{G}(n, p)$ and that we write

$$\mathbb{M}_\lambda^n = (\mathbb{M}_\lambda^{n,i}, i \geq 1)$$

for the components of $\mathbb{M}(n, 1/n + \lambda/n^{4/3})$ listed in decreasing order of size. For each $i \geq 1$, we write $M_\lambda^{n,i}$ for the measured metric space obtained from $\mathbb{M}_\lambda^{n,i}$ by rescaling the graph distance by $n^{-1/3}$ and giving each vertex mass $n^{-2/3}$. We let

$$M_\lambda^n = (M_\lambda^{n,i}, i \geq 1).$$

Recall the cutting procedure introduced in Section 3.2, and that for an $\mathbb{R}$-graph X, we write cut(X) for a random variable with distribution $\mathcal{K}^\infty(\mathrm{X}, \cdot)$. For $i \geq 1$, if $s(\mathscr{G}_\lambda^i) = 0$, let $\mathscr{M}_\lambda^i = \mathscr{G}_\lambda^i$. Otherwise, let $\mathscr{M}_\lambda^i = \mathrm{cut}(\mathscr{G}_\lambda^i)$, where the cutting mechanism is run independently for each i. We note for later use that the mass measure on $\mathscr{M}_\lambda^i$ is almost surely concentrated on the leaves of $\mathscr{M}_\lambda^i$, since this property holds for $\mathscr{G}_\lambda^i$, and $\mathscr{G}_\lambda^i$ may be obtained from $\mathscr{M}_\lambda^i$ by making an almost surely finite number of identifications.

THEOREM 4.4. *Fix $\lambda \in \mathbb{R}$. Then as $n \to \infty$,*

$$M_\lambda^n \xrightarrow{d} \mathscr{M}_\lambda$$

in the space $(\mathbb{L}_4, \mathrm{dist}_{\mathrm{GHP}}^4)$.

PROOF. Write

$$I = \sup\{i \geq 1 : s(\mathscr{G}_\lambda^i) > 1\},$$

with the convention that $I = 0$ when $\{i \geq 1 : s(\mathscr{G}_\lambda^i) > 1\} = \varnothing$. Likewise, for $n \geq 1$ let $I_n = \{i \geq 1 : s(G_\lambda^{n,i}) > 1\}$. We work in a probability space in which the convergence statements of Theorem 4.1 are all almost sure. In this probability space, by Theorem 5.19 of [41] we have that I is almost surely finite and that $I_n \to I$ almost surely.

By Theorem 4.1, almost surely $r(G_\lambda^{n,i})$ is bounded away from zero for all $i \geq 1$. It follows from Theorem 3.3 that almost surely for every $i \geq 1$ we have

$$\mathrm{d_{GHP}}(\mathrm{cut}(G_\lambda^{n,i}), \mathrm{cut}(\mathscr{G}_\lambda^i)) \to 0.$$

Propositions 3.4 and 3.5 then imply that we may work in a probability space in which almost surely, for every $i \geq 1$,

$$\mathrm{d_{GHP}}(M_\lambda^{n,i}, \mathscr{M}_\lambda^i) \to 0. \tag{4.4}$$

Now, for each $i \geq 1$, we have

$$\mathrm{d_{GHP}}(M_\lambda^{n,i}, \mathscr{M}_\lambda^i) \leq 2\max(\mathrm{diam}(M_\lambda^{n,i}), \mathrm{diam}(\mathscr{M}_\lambda^i), \mathrm{mass}(M_\lambda^{n,i}), \mathrm{mass}(\mathscr{M}_\lambda^i)).$$

Moreover, for each $i \geq I$ the right-hand side is bounded above by

$$4\max\bigl(\operatorname{diam}(G_\lambda^{n,i}), \operatorname{diam}(\mathscr{G}_\lambda^i), \operatorname{mass}(G_\lambda^{n,i}), \operatorname{mass}(\mathscr{G}_\lambda^i)\bigr).$$

Since I is almost surely finite, as in the proof of Theorem 4.1 we thus have that, almost surely,

$$\begin{aligned}
&\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{i=N}^{\infty} \mathrm{d}_{\mathrm{GHP}}\bigl(M_\lambda^{n,i}, \mathscr{M}_\lambda^i\bigr)^4 \\
&\quad \leq 64 \lim_{N\to\infty} \limsup_{n\to\infty} \sum_{i=N}^{\infty} \bigl(\operatorname{diam}\bigl(G_\lambda^{n,i}\bigr)^4 + \operatorname{diam}\bigl(\mathscr{G}_\lambda^i\bigr)^4 \\
&\qquad + \operatorname{mass}\bigl(G_\lambda^{n,i}\bigr)^4 + \operatorname{mass}\bigl(\mathscr{G}_\lambda^i\bigr)^4\bigr) \\
&\quad = 0,
\end{aligned}$$

which combined with (4.4) shows that in this space, almost surely

$$\lim_{n\to\infty} \operatorname{dist}_{\mathrm{GHP}}^4\bigl(M_\lambda^n, \mathscr{M}_\lambda\bigr) = 0. \qquad \square$$

4.3. *The largest tree in the minimum spanning forest.* In this section, we study the largest component $\mathbb{M}_\lambda^{n,1}$ of the minimum spanning forest $\mathbb{M}_\lambda$, as well as its analogue $\mathbb{G}_\lambda^{n,1}$ for the random graph. It will be useful to consider the random variable Λ^n which is the smallest number $\lambda \in \mathbb{R}$ such that $\mathbb{G}_\lambda^{n,1}$ is a subgraph of $\mathbb{G}_{\lambda'}^{n,1}$ for every $\lambda' > \lambda$. In other words, in the race of components, Λ^n is the last instant where a new component takes the lead. It follows from Theorem 7 of [50] that $(\Lambda^n, n \geq 1)$ is tight, that is,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\bigl(\Lambda^n > \lambda\bigr) = 0. \tag{4.5}$$

This result is stated in [50] for random graphs with a fixed number of edges, rather than with a fixed edge probability, but it is standard that results for the former have equivalents for the latter; see [41] for more details.

In the following, if $x \mapsto f(x)$ is a real function, we write $f(x) = \mathrm{oe}(x)$ if there exist positive, finite constants c, c', δ, A such that

$$\bigl|f(x)\bigr| \leq c\exp\bigl(-c'x^\delta\bigr) \qquad \text{for every } x > A.$$

In the following lemma, we write $\mathrm{d}_{\mathrm{H}}(M_\lambda^{n,1}, M^n)$ for the Hausdorff distance between $M_\lambda^{n,1}$ and M^n, seen as subspaces of M^n. Recall that edge lengths in all these objects have been normalized by $n^{-1/3}$. Obviously, $\mathrm{d}_{\mathrm{GH}}(M_\lambda^{n,1}, M^n) \leq \mathrm{d}_{\mathrm{H}}(M_\lambda^{n,1}, M^n)$.

LEMMA 4.5. *For any $\epsilon \in (0,1)$ and λ_0 large enough, we have*

$$\limsup_{n\to\infty} \mathbb{P}\biggl(\mathrm{d}_{\mathrm{H}}\bigl(M_\lambda^{n,1}, M^n\bigr) \geq \frac{1}{\lambda^{1-\epsilon}} \Bigm| \Lambda^n \leq \lambda_0\biggr) = \mathrm{oe}(\lambda).$$

In the course of the proof of Lemma 4.5, we will need the following estimate on the length of the longest path outside the largest component of a random graph within the critical window.

LEMMA 4.6. *For all* $0 < \epsilon < 1$, *there exists* λ_0 *such that for all* $\lambda \geq \lambda_0$ *and all* n *sufficiently large, the probability that a connected component of* $\mathbb{G}_\lambda^n$ *aside from* $\mathbb{G}_\lambda^{n,1}$ *contains a simple path of length at least* $n^{1/3}/\lambda^{1-\epsilon}$ *is at most* $e^{-\lambda^{\epsilon/2}}$.

The proof of Lemma 4.6 follows precisely the same steps as the proof of Lemma 3(b) of [5], which is essentially the special case $\epsilon = 1/2$.[9] Since no new idea is involved, we omit the details.

PROOF OF LEMMA 4.5. Fix $f_0 > 0$ and for $i \geq 0$, let $f_i = (5/4)^i \cdot f_0$. Let $t = t(n)$ be the smallest i for which $f_i \geq n^{1/3}/\log n$. Lemma 4 of [5] (proved via Prim's algorithm) states that

$$\mathbb{E}[\operatorname{diam}(\mathbb{M}^n) - \operatorname{diam}(\mathbb{M}_{f_t}^{n,1})] = O\big(n^{1/6}(\log n)^{7/2}\big);$$

this is established by proving the following stronger bound, which will be useful in the sequel:

$$\mathbb{P}\big(\mathrm{d_H}(M_{f_t}^{n,1}, M^n) > n^{-1/6}(\log n)^{7/2}\big) \leq \frac{1}{n}. \tag{4.6}$$

Let B_i be the event that some component of $\mathbb{G}_{f_i}^n$ aside from $\mathbb{G}_{f_i}^{n,1}$ contains a simple path with more than $n^{1/3}/f_i^{1-\epsilon}$ edges and let

$$I^n = \max\{i \leq t : B_i \text{ occurs}\}.$$

Lemma 4.6 entails that, for f_0 sufficiently large, for all n, and all $0 \leq i \leq t-1$,

$$\mathbb{P}\big(i \leq I^n \leq t\big) \leq \sum_{\ell \geq i} e^{-f_i^{\epsilon/2}} \leq 2e^{-f_i^{\epsilon/2}},$$

where the last inequality holds for all i sufficiently large. For fixed $i < t$, if $\Lambda^n \leq f_i$ then for all $\lambda \in [f_i, f_t]$ we have

$$\mathrm{d_H}(M_\lambda^{n,1}, M_{f_t}^{n,1}) \leq \mathrm{d_H}(M_{f_i}^{n,1}, M_{f_t}^{n,1}).$$

If, moreover, $I^n \leq i$, then we have

$$\mathrm{d_H}(M_{f_i}^{n,1}, M_{f_t}^{n,1}) \leq \sum_{j=i+1}^{t} f_j^{\epsilon-1} \leq \frac{f_i^{\epsilon-1}}{1-(4/5)^{1-\epsilon}} < \frac{10}{f_i^{1-\epsilon}}, \tag{4.7}$$

[9] In [5], it was sufficient for the purpose of the authors to produce a path length bound of $n^{1/3}/\lambda^{1/2}$, but their proof does imply the present stronger result. For the careful reader, the key point is that the last estimate in Theorem 19 of [5] is a specialisation of a more general bound, Theorem 11(iii) of [51]. Using the more general bound in the proof is the only modification required to yield the above result.

the latter inequality holding for $\epsilon < 1/2$.

Finally, fix $\lambda \in \mathbb{R}$ and let $i_0 = i_0(\lambda)$ be such that $\lambda \in [f_{i_0}, f_{i_0+1})$. Since $f_t \to \infty$ as $n \to \infty$, we certainly have $i_0 < t$ for all n large enough. Furthermore,

$$\begin{aligned}
&\mathbb{P}\Big(\mathrm{d_H}(M_\lambda^{n,1}, M^n) \geq \frac{1}{\lambda^{1-\epsilon}} \Big| \Lambda^n \leq \lambda_0\Big) \\
&\quad \leq \mathbb{P}\Big(\mathrm{d_H}(M_\lambda^{n,1}, M_{f_t}^{n,1}) \geq \frac{1}{2}\frac{1}{\lambda^{1-\epsilon}} \Big| \Lambda^n \leq \lambda_0\Big) \\
&\qquad + \mathbb{P}\Big(\mathrm{d_H}(M_{f_t}^{n,1}, M^n) \geq \frac{1}{2}\frac{1}{\lambda^{1-\epsilon}} \Big| \Lambda^n > \lambda_0\Big) \\
&\quad \leq \frac{1}{\mathbb{P}(\Lambda^n \leq \lambda_0)}\left(\mathbb{P}\left(\mathrm{d_H}(M_{f_{i_0}}^{n,1}, M_{f_t}^{n,1}) > \frac{10}{f_{i_0}^{1-\epsilon/2}}\right) + \frac{1}{n}\right),
\end{aligned}$$

for all λ large enough and all n such that $2\lambda \leq n^{1/6}(\log n)^{-7/2}$, by (4.6). It then follows from (4.7) and the tightness of $(\Lambda^n, n \geq 1)$ that there exists a constant $C \in (0, \infty)$ such that for all λ_0 large enough,

$$\begin{aligned}
\mathbb{P}\Big(\mathrm{d_H}(M_\lambda^{n,1}, M^n) \geq \frac{1}{\lambda^{1-\epsilon}} \Big| \Lambda^n \leq \lambda_0\Big) &\leq \frac{\mathbb{P}(i_0(\lambda) \leq I^n \leq t) + \frac{1}{n}}{\mathbb{P}(\Lambda^n \leq \lambda_0)} \\
&\leq C\left(e^{-f_{i_0(\lambda)}^{\epsilon/2}} + \frac{1}{n}\right).
\end{aligned}$$

Letting n tend to infinity proves the lemma. □

We are now in a position to easily prove a partial version of our main result. In what follows, we write $\mathring{M}^n$, $\mathring{M}_\lambda^{n,1}$ and $\mathring{\mathscr{M}}_\lambda^1$ for the metric spaces obtained from M^n, $M_\lambda^{n,1}$ and $\mathscr{M}_\lambda^1$ by ignoring their measures.

LEMMA 4.7. *There exists a random compact metric space $\mathring{\mathscr{M}}$ such that, as $n \to \infty$,*

$$\mathring{M}^n \xrightarrow{d} \mathring{\mathscr{M}} \qquad \text{in } (\mathring{\mathcal{M}}, \mathrm{d_{GH}}).$$

Moreover, as $\lambda \to \infty$,

$$\mathring{\mathscr{M}}_\lambda^1 \xrightarrow{d} \mathring{\mathscr{M}} \qquad \text{in } (\mathring{\mathcal{M}}, \mathrm{d_{GH}}).$$

PROOF. Recall that the metric space $(\mathring{\mathcal{M}}, \mathrm{d_{GH}})$ is complete and separable. Theorem 4.4 entails that

$$\mathring{M}_\lambda^{n,1} \xrightarrow{d} \mathring{\mathscr{M}}_\lambda^1$$

as $n \to \infty$ in $(\mathring{\mathcal{M}}, \mathrm{d_{GH}})$. The stated results then follow from this, Lemma 4.5 and the principle of accompanying laws (see Theorem 3.1.14 of [73] or Theorem 9.1.13 in the second edition). □

Let $\hat{M}_\lambda^{n,1}$ be the measured metric space obtained from $M_\lambda^{n,1}$ by rescaling so that the total mass is one [in $M_\lambda^{n,1}$ we gave each vertex mass $n^{-2/3}$; now we give each vertex mass $|V(\mathbb{M}_\lambda^{n,1})|^{-1}$].

PROPOSITION 4.8. *For any* $\epsilon > 0$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\mathrm{d_{GHP}}\big(\hat{M}_\lambda^{n,1}, M^n\big) \geq \epsilon\big) = 0.$$

The proof of this proposition contains several of the paper's novel conceptual contributions. Before turning to this, we use it to prove our main result. Let $\hat{\mathscr{M}}_\lambda^1$ be the measured metric space obtained from $\mathscr{M}_\lambda^1$ by renormalising the measure to be a probability.

THEOREM 4.9. *There exists a random compact measured metric space* $\mathscr{M}$ *of total mass* 1 *such that as* $n \to \infty$,

$$M^n \xrightarrow{d} \mathscr{M}$$

in the space $(\mathcal{M}, \mathrm{d_{GHP}})$. *Moreover, as* $\lambda \to \infty$,

$$\hat{\mathscr{M}}_\lambda^1 \xrightarrow{d} \mathscr{M}$$

in the space $(\mathcal{M}, \mathrm{d_{GHP}})$. *Finally, writing* $\mathscr{M} = (X, d, \mu)$, *we have* $(X, d) \stackrel{d}{=} \mathring{\mathscr{M}}$ *in* $(\mathcal{M}, \mathrm{d_{GH}})$, *where* $\mathring{\mathscr{M}}$ *is as in Lemma* 4.7.

PROOF. Recall that the metric space $(\mathcal{M}, \mathrm{d_{GHP}})$ is a complete and separable. Theorem 4.4 entails that

$$\hat{M}_\lambda^{n,1} \xrightarrow{d} \hat{\mathscr{M}}_\lambda^1$$

as $n \to \infty$ in $(\mathcal{M}, \mathrm{d_{GHP}})$. The stated results then follow from this, Proposition 4.8 and the principle of accompanying laws (see Theorem 3.1.14 of [73] or Theorem 9.1.13 in the second edition). □

Before proceeding to the proof of Proposition 4.8 we make a final remark about the limit spaces, which we will use in the proof that $\mathscr{M}$ is almost surely binary. Observe that, analogous to the fact that $M_\lambda^{n,1}$ is a subspace of M^n, we can view $\mathscr{M}_\lambda^1$ as a subspace of $\mathscr{M}$. (We emphasize that this does not follow from Theorem 4.9.) To this end, we briefly introduce the marked Gromov–Hausdorff topology of [55], Section 6.4. Let $\mathcal{M}_*$ be the set of ordered pairs of the form (X, Y), where $\mathrm{X} = (X, d)$ is a compact metric space and $Y \subset X$ is a compact subset of X (such pairs are considered up to isometries of X). A sequence (X_n, Y_n) of such pairs converges to a limit (X, Y) if there exist correspondences $C_n \in C(X_n, X)$ whose restrictions to $Y_n \times Y$ are correspondences between Y_n and Y, and such that $\mathrm{dis}(C_n) \to 0$. (In

particular, this implies that Y_n converges to Y for the Gromov–Haudorff distance, when these spaces are equipped with the restriction of the distances on X_n, X.) Moreover, a set $\mathcal{A} \subset \mathcal{M}_*$ is relatively compact if and only if $\{\mathrm{X} : (\mathrm{X}, Y) \in \mathcal{A}\}$ is relatively compact for the Gromov–Hausdorff topology.

Recall the definition of the tight sequence of random variables $(\Lambda^n, n \geq 1)$ at the beginning of this section. By taking subsequences, we may assume that we have the joint convergence in distribution

$$(((\mathring{M}^n, \mathring{M}^{n,1}_\lambda), \lambda \in \mathbb{Z}), \Lambda^n) \xrightarrow{d} (((\tilde{\mathscr{M}}, \tilde{\mathscr{M}}^1_\lambda), \lambda \in \mathbb{Z}), \Lambda),$$

for the product topology on $\mathcal{M}_*^{\mathbb{Z}} \times \mathbb{R}$.[10] This coupling of course has the properties that $\tilde{\mathscr{M}} \overset{d}{=} \mathring{\mathscr{M}}$ and that $\tilde{\mathscr{M}}^1_\lambda \overset{d}{=} \mathring{\mathscr{M}}^1_\lambda$ for every $\lambda \in \mathbb{Z}$. Combining this with Lemma 4.5 we easily obtain the following.

PROPOSITION 4.10. *There exists a probability space on which one may define a triple*

$$(\tilde{\mathscr{M}}, (\tilde{\mathscr{M}}^1_\lambda, \lambda \in \mathbb{Z}), \Lambda)$$

with the following properties: (i) Λ *is an a.s. finite random variable*; (ii) $\tilde{\mathscr{M}} \overset{d}{=} \mathring{\mathscr{M}}$, $\tilde{\mathscr{M}}^1_\lambda \overset{d}{=} \mathring{\mathscr{M}}^1_\lambda$ *and* $(\tilde{\mathscr{M}}, \tilde{\mathscr{M}}^1_\lambda) \in \mathcal{M}_*$ *for every* $\lambda \in \mathbb{Z}$; *and* (iii) *for every* $\epsilon \in (0, 1)$ *and* $\lambda_0 > 0$ *large enough*,

$$\mathbb{P}\big(\mathrm{d_H}(\tilde{\mathscr{M}}, \tilde{\mathscr{M}}^1_\lambda) > \lambda^{\epsilon-1} | \Lambda \leq \lambda_0\big) = \mathrm{oe}(\lambda).$$

In particular, $(\tilde{\mathscr{M}}, \tilde{\mathscr{M}}^1_\lambda) \xrightarrow{d} (\tilde{\mathscr{M}}, \tilde{\mathscr{M}})$ *as* $\lambda \to \infty$ *for the marked Gromov–Hausdorff topology.*

4.4. *Proof of Proposition* 4.8. In order to prove the proposition, we need some notation and a rather substantial auxiliary lemma. Let $\mathbb{F}^n_\lambda$ be the subgraph of $\mathbb{M}^n$ with edge set $E(\mathbb{M}^n) \setminus E(\mathbb{M}^{n,1}_\lambda)$. Then $\mathbb{F}^n_\lambda$ is a forest which we view as rooted by taking the root of a component to be the unique vertex in that component which was an element of $\mathbb{M}^{n,1}_\lambda$. For $v \in V(\mathbb{M}^{n,1}_\lambda)$, let $S^n_\lambda(v)$ be the number of nodes in the component $\mathbb{F}^n_\lambda(v)$ of $\mathbb{F}^n_\lambda$ rooted at v. The fact that the random variables $(S^n_\lambda(v), v \in V(\mathbb{M}^{n,1}_\lambda))$ are exchangeable conditional on $V(\mathbb{M}^{n,1}_\lambda)$ will play a key role in what follows.

LEMMA 4.11. *For any* $\delta > 0$,

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \mathbb{P}\Big(\max_{v \in V(\mathbb{M}^{n,1}_\lambda)} S^n_\lambda(v) > \delta n \Big) = 0. \tag{4.8}$$

[10] This is a slight abuse of notation, in the sense that the limiting spaces $\tilde{\mathscr{M}}$ on the right-hand side should, in principle, depend on λ, but obviously these spaces are almost surely all isometric.

PROOF. Let U_λ^n be the event that vertices 1 and 2 lie in the same component of $\mathbb{F}_\lambda^n$. Note that, conditional on $\max_{v\in V(\mathbb{M}_\lambda^{n,1})} S_\lambda^n(v) > \delta n$, the event U_λ^n occurs with probability at least $\delta^2/2$ for sufficiently large n. So, in order to prove the lemma it suffices to show that

$$\lim_{\lambda\to\infty} \limsup_n \mathbb{P}(U_\lambda^n) = 0. \tag{4.9}$$

To prove (4.9), we consider the following modification of Prim's algorithm. We build the MST conditional on the collection $\mathbb{M}_\lambda^n$ of trees. We start from the component containing vertex 1 in $\mathbb{M}_\lambda^n$. To this component, we add the lowest weight edge connecting it to a new vertex. This vertex lies in a new component of $\mathbb{M}_\lambda^n$, which we add in its entirety, before again seeking the lowest-weight edge leaving the tree we have so far constructed. We continue in this way until we have constructed the whole MST. (Observe that the components we add along the way may, of course, be singletons.) Note that if we think of Prim's algorithm as a discrete-time process, with time given by the number of vertices added so far, then this is simply a time-changed version which looks only at times when we add edges of weight *strictly greater than* $1/n + \lambda/n^{4/3}$. This is because when Prim's algorithm first touches a component of $\mathbb{M}_\lambda^n$, it necessarily adds all of its edges before adding any edges of weight exceeding $1/n + \lambda/n^{4/3}$. For $i \geq 0$, write C_i for the tree constructed by the modified algorithm up to step i, where C_0 is the component of vertex 1 in $\mathbb{M}_\lambda^n$, and let e_i be the edge added at step i. The advantage of the modified approach is that, for each $i \geq 1$, we can calculate the probability that the endpoint of e_i which does not lie in C_{i-1} touches $\mathbb{M}_\lambda^{n,1}$, given that it has not at steps $0, 1, \ldots, i-1$.

Recall that, at each stage of Prim's algorithm, we add the edge of minimal weight leaving the current tree. We are thinking of this tree as a collection of components of $\mathbb{M}_\lambda^n$ connected by edges of weight strictly greater than $1/n + \lambda/n^{4/3}$. In general, different sections of the tree built so far are subject to different conditionings depending on the weights of the connecting edges and the order in which they were added. In particular, the endpoint of e_i contained in C_{i-1} is more likely to be in a section with a lower weight-conditioning. However, the *other* endpoint of e_i is equally likely to be any of the vertices of $\{1, 2, \ldots, n\} \setminus C_{i-1}$ because all that we know about them is that they lie in (given) components of $\mathbb{M}_\lambda^n$.

Formally, let $k = n - 1 - |E(\mathbb{M}_\lambda^n)|$. Let C_0 be the component containing 1 in $\mathbb{M}_\lambda^n$. Recursively, for $1 \leq i \leq k$, let:

- e_i be the smallest-weight edge leaving C_{i-1} and
- C_i be the component containing 1 in the graph with edge-set $E(\mathbb{M}_\lambda^n) \cup \{e_1, \ldots, e_i\}$.

The graph with edge-set $E(\mathbb{M}_\lambda^n) \cup \{e_1, \ldots, e_k\}$ is precisely $\mathbb{M}^n$. Let I_1 be the first index for which $V(\mathbb{M}_\lambda^{n,1}) \subset V(C_{I_1})$, so that I_1 is the time at which the component containing 1 attaches to $\mathbb{M}_\lambda^{n,1}$. For each $1 \leq i \leq k$, the endpoint of e_i not in C_{i-1}

is uniformly distributed among all vertices of $\{1,\dots,n\}\setminus C_{i-1}$. So, conditionally given $\mathbb{M}_\lambda^n$, $e_1,\dots,e_{i-1}$ and on $\{I_1\geq i\}$, the probability that I_1 takes the value i is $|V(\mathbb{M}_\lambda^{n,1})|/(n-|V(C_{i-1})|)$. Therefore,

$$\mathbb{P}\big(I_1>i|\mathbb{M}_\lambda^n\big)\leq\left(1-\frac{|V(\mathbb{M}_\lambda^{n,1})|}{n}\right)^i. \tag{4.10}$$

By Theorem 2 of [56] (see also Lemma 3 of [50]), for all $\delta>0$,

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\left|\frac{|V(\mathbb{M}_\lambda^{n,1})|}{2\lambda n^{2/3}}-1\right|>\delta\right)=0. \tag{4.11}$$

Using (4.10) and (4.11), it follows that for any $\delta>0$, there exists $B>0$ such that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\mathbb{P}\big(I_1>Bn^{1/3}/\lambda\big)<\delta. \tag{4.12}$$

Next, let Z be a uniformly random element of $\{1,\dots,n\}\setminus V(\mathbb{M}_\lambda^{n,1})$, and let L_λ be the size of the component of $\mathbb{M}_\lambda^n$ that contains Z. Theorem A1 of [40] shows that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[\frac{\sum_{i=2}^\infty|V(\mathbb{M}_\lambda^{n,i})|^2}{n^{4/3}/\lambda}\right]<\infty,$$

which implies that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[\frac{L_\lambda}{n^{1/3}/\lambda}\right]<\infty.$$

For each $i\geq1$, given that $i<I_1$, the difference $|V(C_i)|-|V(C_{i-1})|$ is stochastically dominated by L_λ, so that

$$\mathbb{E}\big[\big|V(C_{I_1-1})\big|\mathbb{1}_{\{I_1\leq i\}}\big]\leq i\mathbb{E}[L_\lambda].$$

By (4.12) and Markov's inequality, there exists $B'>0$ such that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\mathbb{P}\big(\big|V(C_{I_1-1})\big|>B'n^{2/3}/\lambda^2\big)<\delta. \tag{4.13}$$

The graph C_* with edge-set $E(C_{I_1-1})\cup\{e_{I_1}\}$ forms part of the component containing 1 in $\mathbb{F}_\lambda^n$; indeed, the endpoint of e_{I_1} not contained in C_{I_1-1} is the root of this component. Write v_1 for this root vertex. Now consider freezing the construction of the MST via the modified version of Prim's algorithm at time I_1 and constructing the rest of the MST using the modified version of Prim's algorithm starting now from vertex 2. Let $\ell=n-1-|E(\mathbb{M}_\lambda^n)|-I_1$. Let D_0 be the component containing 2 in the graph with edge-set $E(\mathbb{M}_\lambda^n)\cup\{e_1,\dots,e_{I_1}\}$. Recursively, for $1\leq j\leq\ell$, let:

- f_j be the smallest-weight edge leaving D_{j-1} and
- D_j be the component containing 2 in the graph with edge-set $E(\mathbb{M}_\lambda^n)\cup\{e_1,\dots,e_{I_1}\}\cup\{f_1,\dots,f_j\}$.

Let I_2 be the first index for which f_{I_2} has an endpoint in $V(\mathbb{M}_\lambda^{n,1}) \setminus \{v_1\}$, and let J_1 be the first index for which f_{J_1} has an endpoint in $V(C_*)$.

Recall that U_λ^n is the event that 1 and 2 lie in the same component of $\mathbb{F}_\lambda^n$. If U_λ^n occurs, then we necessarily have $J_1 < I_2$. To prove (4.9), it therefore suffices to show that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}(J_1 < I_2) = 0. \tag{4.14}$$

In order to do so, we first describe how the construction of C_{I_1} conditions the locations of attachment of the edges f_i. As in the Introduction, for $e \in E(K_n)$, W_e is the weight of edge e, and unconditionally these weights are i.i.d. Uniform[0, 1] random variables.

Write $A_0 = V(C_0)$, and for $1 \le i \le I_1$, let $A_i = V(C_i) \setminus V(C_{i-1})$. In particular, $A_{I_1} = V(\mathbb{M}_\lambda^{n,1})$. After C_{I_1} is constructed, for each $0 \le i \le I_1$, the conditioning on edges incident to A_i is as follows:

(a) Every edge between $V(C_{i-1})$ and A_i has weight at least W_{e_i}.

(b) For each $i < j \le I_1$, every edge between A_i and $[n] \setminus V(C_j)$ has weight at least $\max\{W_{e_j}, i < k \le j\}$.

In particular, (b) implies all edges from A_i to $[n] \setminus V(C_{I_1})$ are conditioned to have weight at least $\max\{W_{e_j}, i < k \le I_1\}$. This entails that components which are added later have lower weight-conditioning. In particular, there is no conditioning on edges from $A_{I_1} = V(\mathbb{M}_\lambda^{n,1})$ to $[n] \setminus V(C_{I_1})$ (except the initial conditioning, that all such edges have weight at least $1/n + \lambda/n^{4/3}$, which comes from conditioning on $\mathbb{M}_\lambda^n$).

It follows that under the conditioning imposed by the construction of C_{I_1}, it is *not* the case that for $1 \le j < \ell$, the endpoint of f_{j+1} *outside* D_j is uniformly distributed among $\{1, \dots, n\} \setminus D_j$. However, the conditioning precisely biases these endpoints away from the sets A_i with $i < I_1$ [but not away from $A_{I_1} = V(\mathbb{M}_\lambda^{n,1})$]. As a consequence, for each $1 \le j \le \ell$, conditional on the edge set $E(\mathbb{M}_\lambda^n) \cup \{e_1, \dots, e_{I_1}\} \cup \{f_1, \dots, f_j\}$ and on the event $\{J_1 \ge j\} \cap \{I_2 \ge j\}$, the probability that $j = I_2$ is at least $(|V(\mathbb{M}_\lambda^{n,1})| - 1)/(n - |V(D_{j-1})|)$ and the probability that $j = J_1$ is at most $|V(C_*)|/(n - |V(D_{j-1})|)$. Hence,

$$\mathbb{P}\big(I_2 > i \,|\, \mathbb{M}_\lambda^n\big) \le \left(1 - \frac{|V(\mathbb{M}_\lambda^{n,1})| - 1}{n}\right)^i$$

and so, by (4.11), we obtain that for any $\delta > 0$ there exists $B'' > 0$ such that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(I_2 > B'' n^{1/3}/\lambda\big) < \delta. \tag{4.15}$$

Moreover,

$$\mathbb{P}\big(J_1 > i \,|\, \mathbb{M}_\lambda^n\big) \ge \left(1 - \frac{|V(C_*)|}{n - |V(D_{J_1-1})|}\right)^i.$$

Note that $|V(C_*)| = |V(C_{I_1-1})| + 1$. Also, just as for the components C_i, given that $i < J_1$, the difference $|V(D_i)| - |V(D_{i-1})|$ is stochastically dominated by L_λ and so we obtain the analogue of (4.13): there exists B''' such that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(|V(D_{J_1-1})| > B'''n^{2/3}/\lambda^2\big) < \delta.$$

Hence, from this and (4.13), we see that there exists B'''' such that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(J_1 \le \lambda^2 n^{1/3}/B''''\big) < \delta. \tag{4.16}$$

Together, (4.15) and (4.16) establish (4.14) and complete the proof. □

Armed with this lemma, we now turn to the proof of Proposition 4.8.

PROOF OF PROPOSITION 4.8. Fix $\epsilon > 0$ and let N_ϵ^n be the minimal number of open balls of radius $\epsilon/4$ needed to cover the (finite) space M^n. This automatically yields a covering of $M_\lambda^{n,1}$ by N_ϵ^n open balls of radius $\epsilon/4$ since $M_\lambda^{n,1}$ is included in M^n. From this covering, we can easily construct a new covering $B_\lambda^{n,1}, \dots, B_\lambda^{n,N_\epsilon^n}$ of $M_\lambda^{n,1}$ by sets of diameter at most $\epsilon/2$ which are pairwise disjoint. Let

$$\tilde{B}_\lambda^{n,i} = \bigcup_{v\in B_\lambda^{n,i}} \mathbb{F}_\lambda^n(v), \qquad 1 \le i \le N_\epsilon^n,$$

and let $C_\lambda^n = \bigcup_{i=1}^{N_\epsilon^n} (B_\lambda^{n,i} \times \tilde{B}_\lambda^{n,i})$, which defines a correspondence between $M_\lambda^{n,1}$ and M^n. Moreover, its distortion is clearly at most $2\mathrm{d_H}(M_\lambda^{n,1}, M^n) + \epsilon$. Therefore, by Lemma 4.5,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\mathrm{dis}(C_\lambda^n) > 2\epsilon\big) = 0. \tag{4.17}$$

Next, write $V_\lambda^n = |V(\mathbb{M}_\lambda^{n,1})|$ and take an arbitrary relabelling of the elements of $V(\mathbb{M}_\lambda^{n,1})$ by $\{1, 2, \dots, V_n^\lambda\}$. Since, conditionally on V_λ^n, $(S_\lambda^n(1), \dots, S_\lambda^n(V_\lambda^n))$ are exchangeable, and $V_\lambda^n \xrightarrow{p} \infty$ as $n \to \infty$, Theorem 16.23 of Kallenberg [43] entails that for any $\delta > 0$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\max_{1\le i\le V_\lambda^n} \left|\sum_{j=1}^{i} \frac{S_\lambda^n(j)}{n} - \frac{i}{V_\lambda^n}\right| > \delta\right) = 0 \tag{4.18}$$

as soon as we have that for all $\delta > 0$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\max_{1\le i\le V_\lambda^n} S_\lambda^n(i) > \delta n\Big) = 0,$$

which is precisely the content of Lemma 4.11.

Now define a measure π^n on $M_\lambda^{n,1} \times M^n$ by

$$\pi^n\big(\{(u,v)\}\big) = \frac{1}{V_\lambda^n |\tilde{B}_\lambda^{n,i}|} \wedge \frac{1}{n|B_\lambda^{n,i}|}, \qquad (u,v) \in B_\lambda^{n,i} \times \tilde{B}_\lambda^{n,i}, 1 \le i \le N_\epsilon^n.$$

Note that $\pi_\lambda^n((C_\lambda^n)^c) = 0$ by definition. Moreover, the marginals of π^n are given by

$$\pi_{(1)}^n\big(\{u\}\big) = \frac{1}{V_\lambda^n} \wedge \frac{|\tilde{B}_\lambda^{n,i}|}{n|B_\lambda^{n,i}|}, \qquad u \in B_\lambda^{n,i}, 1 \le i \le N_\epsilon^n,$$

and

$$\pi_{(2)}^n\big(\{v\}\big) = \frac{1}{n} \wedge \frac{|B_\lambda^{n,i}|}{V_\lambda^n |\tilde{B}_\lambda^{n,i}|}, \qquad v \in \tilde{B}_\lambda^{n,i}, 1 \le i \le N_n^\epsilon.$$

Therefore, the discrepancy $D(\pi_\lambda^n)$ of π_λ^n with respect to the uniform measures on M^n and $M_\lambda^{n,1}$ is at most

$$N_\epsilon^n \max_{1 \le i \le N_\epsilon^n} \left| \frac{|\tilde{B}_\lambda^{n,i}|}{n} - \frac{|B_\lambda^{n,i}|}{V_\lambda^n} \right|$$

which (by relabelling the elements of $M_\lambda^{n,1}$ so that the vertices in each $B_\lambda^{n,i}$ have consecutive labels and using exchangeability) is bounded above by

$$2N_\epsilon^n \max_{1 \le i \le V_\lambda^n} \left| \sum_{j=1}^{i} \frac{S_\lambda^n(j)}{n} - \frac{i}{V_\lambda^n} \right|.$$

Then

$$\begin{aligned}
&\mathbb{P}\big(\mathrm{d_{GHP}}(\hat{M}_\lambda^{n,1}, M^n) \ge \epsilon\big) \\
&\quad \le \mathbb{P}\big(\mathrm{dis}(C_\lambda^n) > 2\epsilon\big) + \mathbb{P}\big(D(\pi_\lambda^n) > \epsilon\big) \\
&\quad \le \mathbb{P}\big(\mathrm{dis}(C_\lambda^n) > 2\epsilon\big) + \mathbb{P}\left(N_\epsilon^n \max_{1 \le i \le V_\lambda^n} \left| \sum_{j=1}^{i} \frac{S_\lambda^n(j)}{n} - \frac{i}{V_\lambda^n} \right| > \frac{\epsilon}{2} \right) \\
&\quad \le \mathbb{P}\big(\mathrm{dis}(C_\lambda^n) > 2\epsilon\big) + \mathbb{P}\left(\max_{1 \le i \le V_\lambda^n} \left| \sum_{j=1}^{i} \frac{S_\lambda^n(j)}{n} - \frac{i}{V_\lambda^n} \right| > \frac{\epsilon}{2K} \right) + \mathbb{P}\big(N_\epsilon^n > K\big).
\end{aligned}$$

But now recall that N_ϵ^n is the minimal number of open balls of radius $\epsilon/4$ needed to cover M^n. Let N_ϵ be the same quantity for $\mathring{\mathscr{M}}$. Then by Lemma 4.7, $\mathring{M}^n \xrightarrow{d} \mathring{\mathscr{M}}$, which easily implies that $\limsup_{n\to\infty} \mathbb{P}(N_\epsilon^n > K) \le \mathbb{P}(N_\epsilon > K)$. In particular, by (4.17) and (4.18)

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\mathrm{d_{GHP}}(\hat{M}_\lambda^{n,1}, M^n) \ge \epsilon\big) \le \mathbb{P}(N_\epsilon > K)$$

and the right-hand side converges to 0 as $K \to \infty$. $\square$

5. Properties of the scaling limit. In this section, we give some properties of the limiting metric space $\mathscr{M}$. We start with some general properties that $\mathscr{M}$ shares with the Brownian CRT of Aldous [9–11].

THEOREM 5.1. *$\mathscr{M}$ is a measured $\mathbb{R}$-tree which is almost surely binary and whose mass measure is concentrated on its leaves.*

PROOF. By the second distributional convergence in Theorem 4.9, we may (and will) work in a space in which almost surely $\lim_{\lambda\to\infty} \mathrm{d}_{\mathrm{GHP}}(\hat{\mathscr{M}}_\lambda^1, \mathscr{M}) = 0$. Since it is the Gromov–Hausdorff limit of the sequence of $\mathbb{R}$-trees $\mathscr{M}_\lambda^1$, $\mathscr{M}$ is itself an $\mathbb{R}$-tree (see for instance [30]). For fixed $\lambda \in \mathbb{R}$, each component of $\mathscr{M}_\lambda$ is obtained from $\mathscr{G}_\lambda$, the scaling limit of G_λ^n, using the cutting process. From the construction of $\mathscr{G}_\lambda$ detailed in Section 4.1, it is clear that $\mathscr{G}_\lambda^1$ almost surely does not contain points of degree more than three, and so $\mathscr{M}_\lambda^1$ is almost surely binary.

Next, let us work with the coupling $(\tilde{\mathscr{M}}, (\tilde{\mathscr{M}}_\lambda^1, \lambda \in \mathbb{Z}))$, of Proposition 4.10. We can assume, using the last statement of this proposition and the Skorokhod representation theorem, that $(\tilde{\mathscr{M}}, \tilde{\mathscr{M}}_\lambda^1) \to (\tilde{\mathscr{M}}, \tilde{\mathscr{M}})$ a.s. in $\mathcal{M}_*$. Now suppose that $\tilde{\mathscr{M}}$ has a point x_0 of degree at least 4 with positive probability. On this event, we can find four points x_1, x_2, x_3, x_4 of the skeleton of $\tilde{\mathscr{M}}$, each having degree 2, and such that the geodesic paths from x_i to x_0 have strictly positive lengths and meet only at x_0. But for λ large enough, $x_0, x_1, \ldots, x_4$ all belong to $\tilde{\mathscr{M}}_\lambda^1$, as well as the geodesic paths from $x_1, \ldots, x_4$ to x_0. This contradicts the fact that $\mathscr{M}_\lambda^1$ is binary. Hence, $\mathscr{M}$ is binary almost surely.

Let x and x^λ be sampled according to the probability measures on $\mathscr{M}$ and on $\hat{\mathscr{M}}_\lambda^1$, respectively. For the remainder of the proof, we abuse notation by writing $(\mathscr{M}, x)$ and $(\hat{\mathscr{M}}_\lambda^1, x^\lambda)$ for the marked spaces (random elements of $\mathcal{M}^{1,1}$) obtained by marking at the points x and x^λ. Then we may, in fact, work in a space in which almost surely

$$\lim_{\lambda\to\infty} \mathrm{d}_{\mathrm{GHP}}^{1,1}\big((\mathscr{M}, x), (\hat{\mathscr{M}}_\lambda^1, x^\lambda)\big) = 0.$$

As noted earlier, the mass measure on $\hat{\mathscr{M}}_\lambda^1$ is almost surely concentrated on the leaves of $\hat{\mathscr{M}}_\lambda^1$, and it follows that for each fixed λ, x^λ is almost surely a leaf. Let

$$\Delta(x) = \sup_f f^{-1}(x) \wedge \big(t - f^{-1}(x)\big),$$

where the supremum is over geodesics $f : [0, t] \to \mathscr{M}$ with $x \in \mathrm{Im}(f)$. In particular, $\Delta(x) = 0$ precisely if x is a leaf. For each fixed λ, since x^λ is almost surely a leaf, it is straightforward to verify that almost surely

$$\mathrm{d}_{\mathrm{GHP}}^{1,1}\big((\mathscr{M}, x), (\hat{\mathscr{M}}_\lambda^1, x^\lambda)\big) \geq \Delta(x)/2.$$

But then taking $\lambda \to \infty$ along any countable sequence shows that $\Delta(x) = 0$ almost surely. □

To distinguish $\mathscr{M}$ from Aldous' CRT, we look at a natural notion of fractal dimension, the *Minkowski (or box-counting) dimension* [32]. Given a compact metric space X and $r > 0$, let $N(\mathrm{X}, r)$ be the minimal number of open balls of radius r needed to cover X.

We define the lower and upper Minkowski dimensions by

$$\underline{\dim}_{\mathrm{M}}(\mathrm{X}) = \liminf_{r\downarrow 0} \frac{\log N(\mathrm{X}, r)}{\log(1/r)} \quad \text{and} \quad \overline{\dim}_{\mathrm{M}}(\mathrm{X}) = \limsup_{r\downarrow 0} \frac{\log N(\mathrm{X}, r)}{\log(1/r)}.$$

If $\underline{\dim}_{\mathrm{M}}(\mathrm{X}) = \overline{\dim}_{\mathrm{M}}(\mathrm{X})$, then this value is called the Minkowski dimension and is denoted $\dim_{\mathrm{M}}(\mathrm{X})$.

PROPOSITION 5.2. *The Minkowski dimension of $\mathscr{M}$ exists and is equal to* 3 *almost surely.*

Since the Brownian CRT $\mathscr{T}$ satisfies $\dim_{\mathrm{M}}(\mathscr{T}) = 2$ almost surely ([28], Corollary 5.3), we obtain the following result, which gives a negative answer to a conjecture of Aldous [8].

COROLLARY 5.3. *For any random variable $A > 0$, the laws of $\mathscr{M}$ and of $A\mathscr{T}$, the metric space $\mathscr{T}$ with distances rescaled by A, are mutually singular.*

Before turning to the proof of Proposition 5.2, let us give some intuition for why it is true. We wish find how many open balls we need to cover $\mathscr{M}$. Since we can couple our trees in such a way that $\mathscr{M}_\lambda^1$ is a subspace of $\mathscr{M}$, we may obtain a lower bound by covering $\mathscr{M}_\lambda^1$. Cutting can only increase distances, so we obtain a further lower bound by covering $\mathscr{G}_\lambda^1$. Finally, we again reduce the number of balls needed by covering $\mathrm{core}(\mathscr{G}_\lambda^1)$.

We show below that $\mathrm{core}(\mathscr{G}_\lambda^1)$ typically has $\Theta(\lambda^3)$ edges, each of expected length $\Theta(1/\lambda)$. So to cover $\mathrm{core}(\mathscr{G}_\lambda^1)$ by balls of radius $1/\lambda$ requires $\Theta(\lambda^3)$ balls. Taking $\lambda \to \infty$, this shows that 3 is a lower bound on the Minkowski dimension of $\mathscr{M}$. On the other hand, cutting can at most double the number of balls required for covering since we cut at points of degree 2. Moreover, it turns out that the points in subtrees pendant to $\mathrm{core}(\mathscr{G}_\lambda^1)$ are typically at distance $\Theta(1/\lambda)$ from $\mathrm{core}(\mathscr{G}_\lambda^1)$, so by e.g. doubling the radius of the balls in the covering, we may also cover most points in the pendant subtrees. By a refined version of this argument, a corresponding upper bound for the covering number of $\mathscr{M}$ may be obtained.

To make these ideas rigorous, we rely on an explicit description of the components of $\mathscr{G}_\lambda$ given in [3]. It turns out that, for any $\lambda \in \mathbb{R}$ and $i \geq 1$, the distribution of $\mathscr{G}_\lambda^i$ depends on λ and i only through $\mathrm{mass}(\mathscr{G}_\lambda^i)$ and $\ker(\mathscr{G}_\lambda^i)$. Moreover, we only need a description of $\mathscr{G}_\lambda^1$ in the sequel, and so we focus on this case.

Note that, given $s(\mathscr{G}_\lambda^1) = k \geq 2$, the kernel $\ker(\mathscr{G}_\lambda^1)$ is a 3-regular multigraph with $3k - 3$ edges, and hence $2(k-1)$ vertices. Fix $\lambda \in \mathbb{R}$, $\sigma \in (0, \infty)$, $k \geq 2$,

and K a 3-regular multigraph with $3k-3$ edges. In the following two procedures, we condition on $\mathrm{mass}(\mathscr{G}_\lambda^1)=\sigma$ and $\ker(\mathscr{G}_\lambda^1)=K$. Label the edges of K by $\{1,2,\dots,3k-3\}$ arbitrarily.

CONSTRUCTION 1 ($\mathrm{core}(\mathscr{G}_\lambda^1)$). Independently sample random variables

$$\Gamma_k \sim \mathrm{Gamma}\big((3k-2)/2,\,1/2\big) \quad \text{and}$$
$$(Y_1,Y_2,\dots,Y_{3k-3}) \sim \mathrm{Dirichlet}(1,1,\dots,1).$$

Attach a line-segment of length $Y_j\sqrt{\sigma\Gamma_k}$ in the place of edge j in K, for $1\le j\le 3k-3$.

Before providing the second construction, we should recall some of the basic properties of the CRT $\mathscr{T}$, referring the reader to, for example, [48] for more details. If $\varepsilon=(\varepsilon(s),0\le s\le 1)$ is a standard normalized Brownian excursion then $\mathscr{T}$ is the quotient space of $[0,1]$ endowed with the pseudo-distance $d_\varepsilon(s,t)=2(\varepsilon(s)+\varepsilon(t)-2\inf_{s\wedge t\le u\le s\vee t}\varepsilon(u))$, by the relation $\{d_\varepsilon=0\}$. It is seen as a measured metric space by endowing it with the mass measure which is the image of Lebesgue measure on $[0,1]$ by the canonical projection $p:[0,1]\to\mathscr{T}$. It is also naturally rooted at the point $p(0)$. Likewise, the CRT with mass σ, denoted by $\mathscr{T}_\sigma$, is coded in a similar fashion by (twice) a Brownian excursion conditioned to have duration σ. By scaling properties of Brownian excursion, this is the same as multiplying distances by $\sqrt{\sigma}$ in $\mathscr{T}$, and multiplying the mass measure by σ.

CONSTRUCTION 2 ($\mathscr{G}_\lambda^1$). Sample

$$(X_1,X_2,\dots,X_{3k-3}) \sim \mathrm{Dirichlet}(1/2,1/2,\dots,1/2)$$

and, given $(X_1,\dots,X_{3k-3})$, let $(\mathscr{T}^{(1)},\dots,\mathscr{T}^{(3k-3)})$ be independent CRT's with masses given by $(\sigma X_1,\dots,\sigma X_{3k-3})$, respectively. For $1\le i\le 3k-3$, let (x_i,x_i') be two independent points in $\mathscr{T}^{(i)}$, chosen according to the normalized mass measure. Take the metric gluing of $(\mathscr{T}^{(i)},1\le i\le 3k-3)$ induced by the graph structure of K, by viewing x_i,x_i' as the extremities of the edge i.

PROPOSITION 5.4 ([3]). *The metric space obtained by Construction 1 (resp. Construction 2) has same distribution as* $\mathrm{core}(\mathscr{G}_\lambda^1)$ *(resp.* $\mathscr{G}_\lambda^1$*), given* $\mathrm{mass}(\mathscr{G}_\lambda^1)=\sigma$ *and* $\ker(\mathscr{G}_\lambda^1)=K$.

The proof of Proposition 5.2 builds on this result and requires a couple of lemmas. Recall the notation X_x from Section 3.2.

LEMMA 5.5. *Let* $\mathrm{X}=(X,d,x)$ *be a safely pointed* $\mathbb{R}$*-graph and fix* $r>0$. *Then* $N(\mathrm{X},r)\le N(\mathrm{X}_x,r)\le N(\mathrm{X},r)+2$.

This lemma will be proved in Section 7.1, where we give a more precise description of X_x. The next lemma is a concentration result for the mass and surplus of $\mathscr{G}_\lambda^1$. This should be seen as a continuum analogue of similar results in [50, 56]. We stress that these bounds are far from being sharp, and could be much improved by a more careful analysis. In the rest of this section, if $(Y(\lambda), \lambda \geq 0)$ is a family of positive random variables and $(f(\lambda), \lambda \geq 0)$ is a positive function, we write $Y(\lambda) \asymp f(\lambda)$ if for all $a > 1$,

$$\mathbb{P}\big(Y(\lambda) \notin \big[f(\lambda)/a, af(\lambda)\big]\big) = \mathrm{oe}(\lambda).$$

Note that this only constrains the above probability for large λ.

LEMMA 5.6. *It is the case that*

$$\mathrm{mass}\big(\mathscr{G}_\lambda^1\big) \asymp 2\lambda \quad \textit{and} \quad s\big(\mathscr{G}_\lambda^1\big) \asymp \frac{2\lambda^3}{3}.$$

PROOF. We use the construction of $\mathscr{G}_\lambda$ described in Section 4.1. Recall that $(W(t), t \geq 0)$ is a standard Brownian motion, that $W_\lambda(t) = W(t) + \lambda t - t^2/2$, and that $B_\lambda(t) = W_\lambda(t) - \min_{0\leq s\leq t} W_\lambda(s)$. Note that, letting

$$A_\lambda = \big\{|W(t)| \leq (2\lambda) \vee t \text{ for all } t \geq 0\big\}, \tag{5.1}$$

we have $\mathbb{P}(A_\lambda^c) = \mathrm{oe}(\lambda)$. Considering first $t \leq 2\lambda$, by symmetry, the reflection principle and scaling we have that

$$\begin{aligned}\mathbb{P}\Big(\sup_{0\leq t\leq 2\lambda} |W(t)| > \lambda\Big) &\leq 2\mathbb{P}\Big(\sup_{0\leq t\leq 2\lambda} W(t) > \lambda\Big)\\ &= 2\mathbb{P}\big(|W(2\lambda)| > \lambda\big)\\ &= 2\mathbb{P}\big(|W(2)| > \sqrt{\lambda}\big),\end{aligned}$$

and this is $\mathrm{oe}(\lambda)$ since $W(2)$ is Gaussian. Turning to $t > 2\lambda$, note that letting $W' = (W(u + 2\lambda) - W(2\lambda), u \geq 0)$, then W' is a standard Brownian motion by the Markov property. Hence, on the event $\{\sup_{0\leq t\leq 2\lambda} |W(t)| \leq \lambda\}$, the probability that $|W(t)| > t$ for some $t \geq 2\lambda$ is at most

$$\mathbb{P}\big(\exists u \geq 0 : |W'(u)| \geq u + \lambda\big) \leq 2\mathbb{P}\Big(\max_{u\geq 0}\big(W'(u) - u\big) \geq \lambda\Big).$$

We deduce that $\mathbb{P}(A_\lambda^c) = \mathrm{oe}(\lambda)$ from the fact that $\max_{u\geq 0}(W'(u) - u)$ has an exponential distribution; see, for example, [67].

On A_λ,

$$-\frac{t^2}{2} + \lambda t - \big((2\lambda) \vee t\big) \leq W_\lambda(t) \leq -\frac{t^2}{2} + \lambda t + (2\lambda) \vee t, \qquad t \geq 0,$$

from which it is elementary to obtain that if $\lambda \geq 4$, the following properties hold:

(i) The excursion ε of B_λ that straddles the time λ has length in $[2\lambda - 8, 2\lambda + 8]$.

(ii) All other excursions of B_λ have length at most 6.

(iii) The area of ε is in $[2\lambda^3/3 - 4\lambda^2, 2\lambda^3/3 + 8\lambda^2]$.

Note that (i) and (ii) imply that, for $\lambda \geq 8$, on A_λ, the excursion ε of B_λ is the longest, which we previously called ε^1, and which encodes the component $\mathscr{G}_\lambda^1$ of $\mathscr{G}_\lambda$. This implies that $\mathrm{mass}(\mathscr{G}_\lambda^1) \asymp 2\lambda$, since $\mathrm{mass}(\mathscr{G}_\lambda^1)$ is precisely the length of ε^1. Finally, recall that, given ε^1, $s(\mathscr{G}_\lambda^1)$ has a Poisson distribution with parameter equal to the area of ε^1. Therefore, standard large deviation bounds together with (iii) imply that $s(\mathscr{G}_\lambda^1) \asymp 2\lambda^3/3$. $\square$

PROOF THAT $\underline{\dim}_{\mathrm{M}}(\mathscr{M}) \geq 3$ ALMOST SURELY. In this proof, we always work with the coupling from Proposition 4.10, but for convenience omit the decorations from the notation, for example, writing $\mathscr{M}$ in place of $\mathring{\mathscr{M}}$ or of $\tilde{\mathscr{M}}$. In particular, this allows us to view $\mathscr{M}_\lambda^1$ as a subspace of $\mathscr{M}$ for every $\lambda \in \mathbb{Z}$.

Since $\mathscr{M}_\lambda^1$ is obtained from $\mathscr{G}_\lambda^1$ by performing the cutting operation of Section 3.2, Lemma 5.5 implies that for every $r > 0$,

$$\mathbb{P}\big(N(\mathscr{M}, 1/\lambda) < r\big) \leq \mathbb{P}\big(N\big(\mathscr{M}_\lambda^1, 1/\lambda\big) < r\big) \leq \mathbb{P}\big(N\big(\mathscr{G}_\lambda^1, 1/\lambda\big) < r\big). \tag{5.2}$$

Next, by viewing $\mathrm{core}(\mathscr{G}_\lambda^1)$ as a graph with edge-lengths, we obtain that $N(\mathscr{G}_\lambda^1, 1/\lambda)$ is at least equal to the number $N'(1/\lambda)$ of edges of $\mathrm{core}(\mathscr{G}_\lambda^1)$ that have length at least $2/\lambda$, since the open balls with radius $1/\lambda$ centred at the midpoints of these edges are pairwise disjoint.

Now fix $\sigma > 0$, $k \geq 2$ and a 3-regular multigraph K with $3k-3$ edges, and recall the notation of Construction 1. Given that $\mathrm{mass}(\mathscr{G}_\lambda^1) = \sigma$ and $\ker(\mathscr{G}_\lambda^1) = K$, the edge-lengths of $\mathrm{core}(\mathscr{G}_\lambda^1)$ are given by $Y_i\sqrt{\sigma\Gamma_k}$, $1 \leq i \leq 3k-3$, and we conclude that (still conditionally)

$$N'(1/\lambda) \stackrel{d}{=} \big|\{i \in \{1, \ldots, 3k-3\} : Y_i\sqrt{\sigma\Gamma_k} > 2/\lambda\}\big|.$$

Note that this does not depend on K but only on σ and on k. Now $\Gamma_k \sim \mathrm{Gamma}((3k-2)/2, 1/2)$ can be represented as the sum of $3k-2$ independent random variables with distribution $\mathrm{Gamma}(1/2, 1/2)$, which have mean 1, and by standard large deviation results this implies that

$$\sup_{k \in [\lambda^3/2, \lambda^3]} \mathbb{P}\big(\Gamma_k < \lambda^3\big) = \mathrm{oe}(\lambda).$$

Hence, by first conditioning on $\mathrm{mass}(\mathscr{G}_\lambda^1)$, $s(\mathscr{G}_\lambda^1)$ and using Lemma 5.6, for any given $c > 0$,

$$\begin{aligned}
&\mathbb{P}\big(N'(1/\lambda) < c\lambda^3\big) \\
&\quad\le \sup_{\substack{\sigma\ge\lambda \\ k\in[\lambda^3/2,\lambda^3]}} \mathbb{P}\big(N'(1/\lambda) < c\lambda^3 \,|\, \mathrm{mass}\big(\mathscr{G}_\lambda^1\big) = \sigma, s\big(\mathscr{G}_\lambda^1\big) = k\big) + \mathrm{oe}(\lambda) \\
&\quad\le \sup_{\substack{\sigma\ge\lambda \\ k\in[\lambda^3/2,\lambda^3]}} \mathbb{P}\big(\big|\{i \in \{1,\dots,3k-3\} : Y_i\sqrt{\sigma\Gamma_k} > 2/\lambda\}\big| < c\lambda^3\big) + \mathrm{oe}(\lambda) \qquad (5.3) \\
&\quad\le \sup_{k\in[\lambda^3/2,\lambda^3]} \mathbb{P}\big(\big|\{i \in \{1,\dots,3k-3\} : Y_i > 2/\lambda^3\}\big| < c\lambda^3\big) + \mathrm{oe}(\lambda).
\end{aligned}$$

We now use that $(Y_1,\dots,Y_{3k-3}) \sim \mathrm{Dirichlet}(1,\dots,1)$ is distributed as

$$(\gamma_1,\dots,\gamma_{3k-3})/(\gamma_1+\cdots+\gamma_{3k-3}),$$

where $\gamma_1,\dots,\gamma_{3k-3}$ are independent Exponential(1) random variables. Standard large deviations results for gamma random variables imply that

$$\sup_{k\in[\lambda^3/2,\lambda^3]} \mathbb{P}\big(\gamma_1+\cdots+\gamma_{3k-3} > 4\lambda^3\big) = \mathrm{oe}(\lambda).$$

From this, we obtain

$$\begin{aligned}
&\sup_{k\in[\lambda^3/2,\lambda^3]} \mathbb{P}\big(\big|\{i \in \{1,\dots,3k-3\} : Y_i > 2/\lambda^3\}\big| < c\lambda^3\big) \\
&\quad\le \sup_{k\in[\lambda^3/2,\lambda^3]} \mathbb{P}\big(\big|\{i \in \{1,\dots,3k-3\} : \gamma_i > 8\}\big| < c\lambda^3\big) + \mathrm{oe}(\lambda)
\end{aligned}$$

and this is $\mathrm{oe}(\lambda)$ for $c < e^{-8}$, since $|\{i \in \{1,\dots,3k-3\} : \gamma_i > 8\}|$ is $\mathrm{Bin}(3k-3, e^{-8})$ distributed.

It follows that for such c, $\mathbb{P}(N'(1/\lambda) < c\lambda^3) = \mathrm{oe}(\lambda)$, which with (5.2) implies that

$$\mathbb{P}\big(N(\mathscr{M}, 1/\lambda) < c\lambda^3/2\big) = \mathrm{oe}(\lambda).$$

We obtain by the Borel–Cantelli lemma that $N(\mathscr{M}, 1/\lambda) \ge c\lambda^3/2$ for all $\lambda \in \mathbb{Z}$ sufficiently large. By sandwiching $1/r$ between consecutive integers, this yields that almost surely

$$\underline{\dim}_{\mathrm{M}}(\mathscr{M}) = \liminf_{r\to 0} \frac{\log N(\mathscr{M}, r)}{\log(1/r)} \ge 3. \qquad \square$$

We now prove the upper bound from Proposition 5.2.

PROOF THAT $\overline{\dim}_{\mathrm{M}}(\mathscr{M}) \le 3$ ALMOST SURELY. Recall the definition of Λ from Proposition 4.10. Fix $\lambda_0 > 0$ and an integer $\lambda > \lambda_0$. We work conditionally on the event $\{\Lambda \le \lambda_0\}$. Next, fix $\epsilon > 0$. If $B_1, \dots, B_N$ is a covering of $\mathscr{M}_\lambda^1$ by balls of radius $1/\lambda^{1-\epsilon}$ then, since $\mathscr{M}_\lambda^1 \subset \mathscr{M}$, the centres $x_1, \dots, x_N$ of these balls are elements of $\mathscr{M}$. On the event $\{\mathrm{d_H}(\mathscr{M}_\lambda^1, \mathscr{M}) < 1/\lambda^{1-\epsilon}\}$, whose complement has conditional probability $\mathrm{oe}(\lambda)$ by Proposition 4.10, the balls with centres $x_1, \dots, x_N$ and radius $2/\lambda^{1-\epsilon}$ then form a covering of $\mathscr{M}$. Hence,

$$
\begin{aligned}
&\mathbb{P}\big(N\big(\mathscr{M}, 2/\lambda^{1-\epsilon}\big) > 5\lambda^3 | \Lambda \le \lambda_0\big) \\
&\quad \le \frac{\mathbb{P}(N(\mathscr{M}_\lambda^1, 1/\lambda^{1-\epsilon}) > 5\lambda^3)}{\mathbb{P}(\Lambda \le \lambda_0)} + \mathrm{oe}(\lambda) \\
&\quad \le \frac{\mathbb{P}(N(\mathscr{G}_\lambda^1, 1/\lambda^{1-\epsilon}) + 2s(\mathscr{G}_\lambda^1) > 5\lambda^3)}{\mathbb{P}(\Lambda \le \lambda_0)} + \mathrm{oe}(\lambda) \\
&\quad \le \frac{\mathbb{P}(N(\mathscr{G}_\lambda^1, 1/\lambda^{1-\epsilon}) > 3\lambda^3)}{\mathbb{P}(\Lambda \le \lambda_0)} + \mathrm{oe}(\lambda),
\end{aligned} \tag{5.4}
$$

where in the penultimate step we used Lemma 5.5 and the fact that $\mathscr{M}_\lambda^1$ is obtained from $\mathscr{G}_\lambda^1$ by performing $s(\mathscr{G}_\lambda^1)$ cuts, and in the last step we used the fact that $s(\mathscr{G}_\lambda^1) \asymp 2\lambda^3/3$ from Lemma 5.6.

To estimate $N(\mathscr{G}_\lambda^1, 1/\lambda^{1-\epsilon})$, we now use Construction 2 to obtain a copy of $\mathscr{G}_\lambda^1$ conditioned to satisfy $\mathrm{mass}(\mathscr{G}_\lambda^1) = \sigma$, $s(\mathscr{G}_\lambda^1) = k$ and $\ker(\mathscr{G}_\lambda^1) = K$, where K is a 3-regular multigraph with $3k-3$ edges. Recall that we glue $3k-3$ Brownian CRT's $(\mathscr{T}_{\sigma X_1}^{(1)}, \dots, \mathscr{T}_{\sigma X_{3k-3}}^{(3k-3)})$ along the edges of K. These CRTs are conditionally independent given their masses $\sigma X_1, \dots, \sigma X_{3k-3}$, and $(X_1, \dots, X_{3k-3})$ has Dirichlet$(1/2, \dots, 1/2)$ distribution. (Here, we include the mass in the notation because it will vary later on.) If each of these trees has diameter less than $1/\lambda^{1-\epsilon}$, then clearly we can cover the glued space by $3k-3$ balls of radius $1/\lambda^{1-\epsilon}$, each centred in a distinct tree $\mathscr{T}_{\sigma X_i}^{(i)}$, $1 \le i \le 3k-3$. Therefore, by first conditioning on $\mathrm{mass}(\mathscr{G}_\lambda^1)$ and on $s(\mathscr{G}_\lambda^1)$, and then using Lemma 5.6,

$$
\begin{aligned}
&\mathbb{P}\big(N\big(\mathscr{G}_\lambda^1, 1/\lambda^{1-\epsilon}\big) > 3\lambda^3\big) \\
&\le \sup_{\substack{\sigma \le 3\lambda \\ k \in [\lambda^3/2, \lambda^3]}} \mathbb{P}\big(N\big(\mathscr{G}_\lambda^1, 1/\lambda^{1-\epsilon}\big) > 3\lambda^3 \,|\, \mathrm{mass}\big(\mathscr{G}_\lambda^1\big) = \sigma, s\big(\mathscr{G}_\lambda^1\big) = k\big) + \mathrm{oe}(\lambda) \\
&\le \sup_{\substack{\sigma \le 3\lambda \\ k \in [\lambda^3/2, \lambda^3]}} \mathbb{P}\Big(\max_{1 \le i \le 3k-3} \mathrm{diam}\big(\mathscr{T}_{\sigma X_i}^{(i)}\big) > 1/\lambda^{1-\epsilon}\Big) + \mathrm{oe}(\lambda).
\end{aligned} \tag{5.5}
$$

We can represent $(X_1, \dots, X_{3k-3})$ as $(\gamma_1, \dots, \gamma_{3k-3})/(\gamma_1 + \cdots + \gamma_{3k-3})$, where $\gamma_1, \dots, \gamma_{3k-3}$ are i.i.d. random variables with law Gamma$(1/2, 1)$. Hence,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le i\le 3k-3} X_i > \lambda^{\epsilon-3}\Big)\\
&\quad\le \mathbb{P}\big(\gamma_1 + \cdots + \gamma_{3k-3} < \lambda^{3-\epsilon/2}\big) + \mathbb{P}\Big(\max_{1\le i\le 3k-3} \gamma_i > \lambda^{\epsilon/2}\Big)\\
&\quad\le \mathbb{P}\big(\gamma_1 + \cdots + \gamma_{3k-3} < \lambda^{3-\epsilon/2}\big) + 1 - \big(1 - \mathbb{P}\big(\gamma_1 > \lambda^{\epsilon/2}\big)\big)^{3k-3}.
\end{aligned}$$

Standard large deviations results for gamma random variables then entail that for all $\epsilon > 0$,

$$\sup_{k\in[\lambda^3/2,\lambda^3]} \mathbb{P}\Big(\max_{1\le i\le 3k-3} X_i > \lambda^{\epsilon-3}\Big) = \mathrm{oe}(\lambda),$$

which in turn implies that

$$\begin{aligned}
&\sup_{\substack{\sigma\le 3\lambda\\ k\in[\lambda^3/2,\lambda^3]}} \mathbb{P}\Big(\max_{1\le i\le 3k-3} \operatorname{diam}\big(\mathscr{T}^{(i)}_{\sigma X_i}\big) > \lambda^{\epsilon-1}\Big)\\
&\quad\le \sup_{k\in[\lambda^3/2,\lambda^3]} \mathbb{P}\Big(\max_{1\le i\le 3k-3} X_i > \lambda^{\epsilon-3}\Big) + \mathbb{P}\Big(\max_{1\le i\le 3\lambda^3} \operatorname{diam}\big(\mathscr{T}^{(i)}_{3/\lambda^{2-\epsilon}}\big) > \lambda^{\epsilon-1}\Big),
\end{aligned}$$

where we used that, by scaling, $\operatorname{diam}(\mathscr{T}_\sigma)$ is stochastically increasing in σ, and $(\mathscr{T}^{(i)}_{3/\lambda^{2-\epsilon}})$, $1 \le i \le \lfloor 3\lambda^3 \rfloor$ are independent CRTs, each with mass $3/\lambda^{3-\epsilon}$. Using this bound and Brownian scaling, it follows that

$$\begin{aligned}
&\sup_{\substack{\sigma\le 3\lambda\\ k\in[\lambda^3/2,\lambda^3]}} \mathbb{P}\Big(\max_{1\le i\le 3k-3} \operatorname{diam}\big(\mathscr{T}^{(i)}_{\sigma X_i}\big) > 1/\lambda^{1-\epsilon}\Big)\\
&\quad\le \mathrm{oe}(\lambda) + 1 - \big(1 - \mathbb{P}\big(\operatorname{diam}\big(\mathscr{T} > \lambda^{\epsilon/2}/\sqrt{3}\big)\big)\big)^{3\lambda^3}.
\end{aligned} \tag{5.6}$$

Next, it is well known that the height of $\mathscr{T}$, that is, the maximal distance from the root to another point, is theta-distributed:

$$\mathbb{P}\big(\operatorname{height}(\mathscr{T}) \ge x\big) = \sum_{k\ge 1} (-1)^{k+1} e^{-k^2x^2} \le e^{-x^2}.$$

Since $\operatorname{diam}(\mathscr{T}) \le 2\operatorname{height}(\mathscr{T})$, it follows that

$$\mathbb{P}\big(\operatorname{diam}(\mathscr{T}) \ge x\big) = \mathrm{oe}(x).$$

We obtain that (5.6) is $\mathrm{oe}(\lambda)$, and (5.5) then yields that

$$\mathbb{P}\big(N\big(\mathscr{G}^1_\lambda, 1/\lambda^{1-\epsilon}\big) > 3\lambda^3\big) = \mathrm{oe}(\lambda).$$

By (5.4), we then have

$$\mathbb{P}\big(N\big(\mathscr{M}, 2/\lambda^{1-\epsilon}\big) > 5\lambda^3 \,|\, \Lambda \le \lambda_0\big) = \mathrm{oe}(\lambda).$$

Therefore, the Borel–Cantelli lemma implies that $N(\mathscr{M}, 2/\lambda^{1-\epsilon}) \leq 5\lambda^3$ a.s. for every integer $\lambda > \lambda_0$ large enough. This implies that, conditionally on $\{\Lambda \leq \lambda_0\}$, $\overline{\dim}_{\mathrm{M}}(\mathscr{M}) \leq 3 + \epsilon$ almost surely for every $\epsilon > 0$, by sandwiching $1/r$ between integers in $\limsup_{r\to 0} \log N(\mathscr{M}, r)/\log(1/r)$. Since Λ is almost surely finite and λ_0 was arbitrary, this then holds unconditionally for any $\epsilon > 0$. $\square$

This concludes the proof of Proposition 5.2.

6. The structure of $\mathbb{R}$-graphs. At this point, it remains to prove Theorems 2.7 and 3.3, and Lemma 5.5. We also made use of the forthcoming Corollary 6.6 in the course of proving Theorem 4.1. In this section, we investigate $\mathbb{R}$-graphs, and prove Theorem 2.7 and Corollary 6.6. Section 7 contains the proofs of Theorem 3.3 and of Lemma 5.5.

6.1. *Girth in $\mathbb{R}$-graphs.* In this section, $\mathrm{X} = (X, d)$ is an $\mathbb{R}$-graph. The *girth* of X is defined by

$$\mathrm{gir}(\mathrm{X}) = \inf\{\mathrm{len}(c) : c \text{ is an embedded cycle in } X\}.$$

If (X, d) is an $\mathbb{R}$-graph, then by definition $(B_{\epsilon(x)}(x), d)$ is an $\mathbb{R}$-tree for every $x \in X$ and for some function $\epsilon : X \to (0, \infty)$. The balls $(B_{\epsilon(x)}(x), x \in X)$ form an open cover of X. By extracting a finite sub-cover, we see that there exists $\epsilon > 0$ such that for every $x \in X$, the space $(B_\epsilon(x), d)$ is an $\mathbb{R}$-tree. We let $R(\mathrm{X})$ be the supremum of all numbers $\epsilon > 0$ with this property. It is immediate that $\mathrm{gir}(\mathrm{X}) \geq 2R(X) > 0$. In fact, it is not difficult to show that $\mathrm{gir}(\mathrm{X}) = 4R(\mathrm{X})$ and that $(B_{R(\mathrm{X})}(x), d)$ is an $\mathbb{R}$-tree for all $x \in X$. More precisely, the closed ball $(\overline{B}_{R(\mathrm{X})}(x), d)$ is also a (compact) $\mathbb{R}$-tree, since it is the closure of the corresponding open ball. These facts are not absolutely crucial in the arguments to come, but they make some proofs more elegant, so we will take them for granted and leave their proofs to the reader, who is also referred to Proposition 2.2.15 of [58].

PROPOSITION 6.1. *If $f \in \mathcal{C}([a, b], X)$ is a local geodesic in* X, *then for every $t \in [a, b]$, the restriction of f to $[t - R(\mathrm{X}), t + R(\mathrm{X})] \cap [a, b]$ is a geodesic. In particular, if c is an embedded cycle and $x \in c$, then c contains a geodesic arc of length $2R(\mathrm{X})$ with mid-point x.*

PROOF. The function f is injective on any interval of length at most $2R(\mathrm{X})$, since otherwise we could exhibit an embedded cycle with length at most $2R(\mathrm{X}) = \mathrm{gir}(\mathrm{X})/2$. In particular, f is injective on the interval $[t - R(\mathrm{X}), t + R(\mathrm{X})] \cap [a, b]$, and takes values in the $\mathbb{R}$-tree $(\overline{B}_{R(\mathrm{X})}(f(t)), d)$, so that its image is a geodesic segment, and since f is parameterized by arc-length, its restriction to the above interval is an isometry. This proves the first statement.

For the second statement, note that every injective path $f \in \mathcal{C}([a, b], X)$ parameterized by arc-length is a local geodesic since, for every t, the path f restricted

to $[t - R(\mathrm{X}), t + R(\mathrm{X})] \cap [a, b]$ is an injective path in the $\mathbb{R}$-tree $(\overline{B}_{R(\mathrm{X})}(f(t)), d)$ parameterized by arc-length, and hence is a geodesic. If now $g : \mathbb{S}_1 \to X$ is an injective continuous function inducing the embedded cycle c, it suffices to apply the previous claim to a parametrisation by arc-length mapping 0 to x of the function $t \mapsto g(e^{2\mathrm{i}\pi t})$. $\square$

6.2. *Structure of the core.* In this section, $\mathrm{X} = (X, d)$ is again an $\mathbb{R}$-graph. Recall that core(X) is the union of all arcs with endpoints in embedded cycles.

PROPOSITION 6.2. *The set* core(X) *is a finite union of embedded cycles and simple arcs that are disjoint from the embedded cycles except at their endpoints. Moreover, the space* $(\mathrm{core}(\mathrm{X}), d)$ *is an* $\mathbb{R}$*-graph with no leaves.*

PROOF. Assume, for a contradiction, that the union of all embedded cycles cannot be written as a finite union of embedded cycles. Then we can find an infinite sequence $c_1, c_2, \ldots$ of embedded cycles such that $c_i \setminus (c_1 \cup \cdots \cup c_{i-1})$ is nonempty for every $i \geq 0$, and thus contains at least one point x_i. Up to taking subsequences, one can assume that x_i converges to some point x, and that $d(x, x_i) < R(\mathrm{X})/2$ for every $i \geq 1$. Let γ_i' be a geodesic from x to x_i: this geodesic takes its values in the $\mathbb{R}$-tree $\overline{B}_{R(\mathrm{X})}(x)$. Since $x_i \in c_i$, by Proposition 6.1 we can find two geodesic paths starting from x_i, meeting only at x_i, with length $R(\mathrm{X}) - d(x, x_i)$, and taking values in $c_i \cap \overline{B}_{R(\mathrm{X})}(x)$. At least one of these paths γ_i'' does not pass through x, and so the concatenation γ_i of γ_i' and γ_i'' is an injective path parameterized by arc-length starting from x and with length $R(\mathrm{X})$. So it is, in fact, a geodesic path, since it takes its values in $\overline{B}_{R(\mathrm{X})}(x)$. We let y_i be the endpoint of γ_i, so that $d(x, y_i) = R(\mathrm{X})$ for every $i \geq 1$. Now, we observe that if $i < j$, the paths γ_i and γ_j both start from the same point x, but since γ_i'' takes values in c_i, since γ_j'' passes through $x_j \notin c_i$, and since $d(x, x_i) \vee d(x, x_j) \leq R(\mathrm{X})/2$, these paths are disjoint outside the ball $B_{R(\mathrm{X})/2}(x)$. This implies that $d(y_i, y_j) \geq R(\mathrm{X})$ for every $i < j$, and contradicts the compactness of X.

Therefore, the union X_0 of all embedded cycles is closed and has finitely many connected components. By definition, core(X) is the union of X_0 together with all simple arcs with endpoints in X_0. Obviously, in this definition, we can restrict our attention to simple arcs having only their endpoints in X_0. So let $x, y \in X_0$ with $x \neq y$ be linked by an simple arc A taking its values outside X_0, except at its endpoints. Necessarily, x and y must be in disjoint connected components of X_0, because otherwise there would exist a path from x to y in X_0 whose concatenation with γ would create an embedded cycle not included in X_0. Furthermore, there can exist at most one arc A, or else it would be easy to exhibit an embedded cycle not included in X_0. So we see that core(X) is a finite union of simple arcs and embedded cycles, which is obviously connected, and is thus closed. Any point in core(X) has degree at least 2 by definition.

It remains to check that the intrinsic metric on core(X) is given by d itself. Let $x, y \in \mathrm{core(X)}$ and γ be a geodesic path from x to y. Assume that γ takes some value $z = \gamma(t)$ outside core(X). Let $t_1 = \sup\{s \le t : \gamma(s) \in \mathrm{core(X)}\}$ and $t_2 = \inf\{s \ge t : \gamma(s) \in \mathrm{core(X)}\}$, so that $\gamma((t_1, t_2)) \cap \mathrm{core(X)} = \varnothing$. Since core(X) is connected, we can join $\gamma(t_1)$ and $\gamma(t_2)$ by a simple arc included in core(X), and the union of this arc with $\gamma((t_1, t_2))$ is an embedded cycle not contained in core(X), a contradiction. □

6.3. *The kernel of $\mathbb{R}$-graphs with no leaves.* In this section, X is an $\mathbb{R}$-graph with no leaves. We now start to prove Theorem 2.7 on the structure of such $\mathbb{R}$-graphs. The set $k(\mathrm{X}) = \{x \in X : \deg_X(x) \ge 3\}$ of branchpoints of X forms the vertex set of ker(X).

PROPOSITION 6.3. *The set $k(\mathrm{X})$ is finite, and $\deg_X(x) < \infty$ for every $x \in k(\mathrm{X})$.*

PROOF. By Proposition 6.2, the number of cycles of X is finite. We assume that X is not acyclic, and argue by induction on the maximal number of independent embedded cycles, that is, of embedded cycles $c_1, \dots, c_k$ such that $c_i \setminus (c_1 \cup \cdots \cup c_{k-1}) \neq \varnothing$ for $1 \le i \le k$. Plainly, we may and will assume that for all $i \in \{1, 2, \dots, k\}$, either c_i is disjoint from $c_1 \cup \cdots \cup c_{i-1}$, or $c_i \setminus (c_1 \cup \cdots \cup c_{i-1})$ is a simple arc of the form $\gamma((0, 1))$, where $\gamma : [0, 1] \to X$ satisfies $\gamma(0), \gamma(1) \in c_1 \cup \cdots \cup c_{i-1}$. The result is trivial if X is unicyclic ($k = 1$). Suppose X has k independent embedded cycles $c_1, \dots, c_k$ as above. Consider the smallest connected subset X' of X containing $c_1, \dots, c_{k-1}$: this subset is the union of $c_1, \dots, c_{k-1}$ with some simple arcs having only their endpoints as elements of $c_1 \cup \cdots \cup c_{k-1}$, and is a closed subset of X.

If c_k does not intersect X', then there exists a unique simple arc with one endpoint a in c_k and the other endpoint b in X', and disjoint from $c_k \cup X'$ elsewhere. Then a, b must be elements of $k(\mathrm{X})$: a is the only element of $k(\mathrm{X})$ in c_k, we have $\deg_X(a) = 3$, and $\deg_X(b) = \deg_{X'}(b) + 1$. Therefore, the number of points in $k(\mathrm{X})$ is at most $2 + k(\mathrm{X}')$, where X' is the set X' endowed with the intrinsic metric inherited from X. This is an $\mathbb{R}$-graph without leaves and with (at most) $k - 1$ independent cycles.

If on the other hand $c_k \cap X' \neq \varnothing$, then by assumption we have $X = X' \cup A$, where A is a sub-arc of c_k disjoint from X' except at its endpoints a, b. The latter are elements of $k(\mathrm{X})$, and satisfy $\deg_X(a) \le \deg_{X'}(a) + 2$ and similarly for b (note that a, b may be equal). After we remove $A \setminus \{a, b\}$ from X, we are left with an $\mathbb{R}$-graph X' (in the induced metric) without leaves, and with at most $k - 1$ independent cycles.

The result follows by induction on k. □

If X is unicyclic, then X is in fact identical to its unique embedded cycle c. In this case, $k(\mathrm{X}) = \varnothing$, and we let $e(\mathrm{X}) = \{c\}$. If X has at least two distinct

embedded cycles, then the previous proof entails that $k(\mathrm{X}) \neq \varnothing$, and more precisely that every embedded cycle contains at least one point of $k(\mathrm{X})$. The set $X \setminus k(\mathrm{X})$ has finitely many connected components [in fact, there are precisely $\frac{1}{2}\sum_{x\in k(\mathrm{X})} \deg_X(x)$ components, as the reader is invited to verify], which are simple arcs of the form $\gamma((0,1))$, where $\gamma : [0,1] \to X$ is such that γ is injective on $[0,1)$, such that $\gamma(0), \gamma(1) \in k(\mathrm{X})$, and such that $\gamma((0,1)) \cap k(\mathrm{X}) = \varnothing$. We let $e(\mathrm{X})$ be the set of the closures of these connected components, that is, the arcs $\gamma([0,1])$ with the above notation, which are called the *kernel edges*. The multigraph $\ker(\mathrm{X}) = (k(\mathrm{X}), e(\mathrm{X}))$ is the *kernel* of X, where the vertices incident to $e \in e(\mathrm{X})$ are, of course, the endpoints of e. An orientation of the edge e is the choice of a parametrisation $\gamma : [0,1] \to X$ of the arc e or its reversal $\gamma(1-\cdot)$, considered up to reparametrisations by increasing bijections from $[0,1]$ to $[0,1]$. If e is given an orientation, then its endpoints are distinguished as the source and target vertices, and are denoted by e^-, e^+, respectively. The next proposition then follows from the definition of $k(\mathrm{X})$.

PROPOSITION 6.4. *The kernel of a non-unicyclic $\mathbb{R}$-graph without leaves is a multigraph of minimum degree at least* 3.

Finally, we prove Theorem 2.7. Assume that X is a non-unicyclic $\mathbb{R}$-graph without leaves, and let $\ell(e) : e \in e(\mathrm{X})$ be the lengths of the kernel edges. Note that if $x, y \in k(\mathrm{X})$, then

$$d(x,y) = \inf\left\{\sum_{i=1}^{k} \ell(e_i) : (e_1,\dots,e_k) \text{ a chain from } x \text{ to } y \text{ in } G\right\},$$

where $(e_1,\dots,e_k)$ is a chain from x to y if it is possible to orient $e_1,\dots,e_k \in e(\mathrm{X})$ in such a way that $e_1^- = x$, $e_k^+ = y$ and $e_i^+ = e_{i+1}^-$ for every $i \in \{1,\dots,k-1\}$. Of course, it suffices to restrict the infimum to those chains that are simple, in the sense that they do not visit the same vertex twice. Since there are finitely many simple chains, the above infimum is, in fact, a minimum. Next, if x and y are elements of e and e' respectively, consider an arbitrary orientation of e, e'. Then a shortest path from x to y either stays in e (in this case $e = e'$), or passes through at least one element of $k(\mathrm{X})$ incident to e, and likewise for e'. Therefore,

$$d(x,y) = d_e(x,y) \wedge \min_{s,t\in\{-,+\}} \{d_e(x,e^s) + d(e^s,(e')^t) + d_{e'}((e')^t,y)\},$$

where we let $d_e(a,b)$ be the length of the arc of e between a and b if $a, b \in e$, and ∞ otherwise. It is shown in [25], Section 3, that this formula gives the distance for the metric gluing of the graph with edge-lengths $(k(\mathrm{X}), e(\mathrm{X}), (\ell(e), e \in e(\mathrm{X})))$. This proves Theorem 2.7.

6.4. *Stability of the kernel in the Gromov–Hausdorff topology.* In this section, we show that kernels of $\mathbb{R}$-graphs are stable under small perturbations in the Gromov–Hausdorff metric, under an assumption which says, essentially, that the girth is uniformly bounded away from 0.

Recall from Section 3.2 that $\mathcal{A}_r$ is the set of measured $\mathbb{R}$-graphs X such that

$$\min_{e\in e(\mathrm{X})} \ell(e) \geq r, \qquad \sum_{e\in e(\mathrm{X})} \ell(e) \leq 1/r \quad \text{and} \quad s(\mathrm{X}) \leq 1/r,$$

where it is understood in this definition that the unicyclic $\mathbb{R}$-graphs (those with surplus 1) are such that their unique embedded cycle has length in $[r, 1/r]$. It follows that the sets $\mathcal{A}_r, 0 < r < 1$, are decreasing, with union the set of all measured $\mathbb{R}$-graphs. If $[X, d]$ is an $\mathbb{R}$-graph, we write $[X, d] \in \mathcal{A}_r$ if $[X, d, 0] \in \mathcal{A}_r$. Note that an element $\mathrm{X} \in \mathcal{A}_r$ has $\mathrm{gir}(\mathrm{X}) \geq r$.

A subset A of X is said to be in correspondence with a subset A' of X' via $C \subset X \times X'$ if $C \cap (A \times A')$ is a correspondence between A and A'. Let X and X′ be $\mathbb{R}$-graphs with surplus at least 2. Given $C \in C(\mathrm{X}, \mathrm{X}')$, for $\epsilon > 0$ we say that C is a ϵ-*overlay* (of X and X′) if $\mathrm{dis}(C) < \epsilon$, and there exists a multigraph isomorphism χ between $\ker(\mathrm{X})$ and $\ker(\mathrm{X}')$ such that:

1. For every $v \in k(\mathrm{X})$, $(v, \chi(v)) \in C$.
2. For every $e \in e(\mathrm{X})$, the edges e and $\chi(e)$ are in correspondence via C, and

$$\big|\ell(e) - \ell\big(\chi(e)\big)\big| \leq \epsilon.$$

If $s(\mathrm{X}) = s(\mathrm{X}') = 1$, an ϵ-overlay is a correspondence with distortion at most ϵ, such that the unique embedded cycles c, c' of X and X′ are in correspondence via C, and $|\ell(c) - \ell(c')| \leq \epsilon$. Finally, if $s(\mathrm{X}) = s(\mathrm{X}') = 0$ then an ϵ-overlay is just a correspondence of distortion at most ϵ.

PROPOSITION 6.5. *Fix $r \in (0, 1)$. For every $\epsilon > 0$ there exists $\delta > 0$ such that if $\mathrm{X} = (X, d)$ and $\mathrm{X}' = (X', d')$ are elements of $\mathcal{A}_r$ and $C \in C(\mathrm{X}, \mathrm{X}')$ has $\mathrm{dis}(C) \leq \delta$, then there exists an ϵ-overlay $C' \in C(\mathrm{X}, \mathrm{X}')$ with $C \subset C'$.*

We say that a sequence of finite graphs with edge-lengths $((V_n, E_n, (l_n(e), e \in E_n)), n \geq 1)$ converges to the graph with edge-lengths $(V, E, (l(e), e \in E))$ if (V_n, E_n) and (V, E) are isomorphic for all but finitely many $n \geq 1$, through an isomorphism χ_n such that $l_n(\chi_n(e)) \to l(e)$ as $n \to \infty$ for every $e \in E$. We now state some consequences of Proposition 6.5 which are used in the proof of Theorem 4.1 and in Section 7.3, before proceeding to the proof of Proposition 6.5. Recall the definition of the distances $\mathrm{d}_{\mathrm{GHP}}^{k,l}$ from Section 2.1.

COROLLARY 6.6. *Fix $r \in (0, 1)$. Let $(\mathrm{X}^n = (X^n, d^n, \mu^n), n \geq 1)$ and $\mathrm{X} = (X, d, \mu)$ be elements of $\mathcal{A}_r$. Suppose that $\mathrm{d}_{\mathrm{GHP}}(\mathrm{X}^n, \mathrm{X}) \to 0$, as $n \to \infty$.*

(i) *Then* $\ker(\mathrm{X}^n)$ *converges to* $\ker(\mathrm{X})$ *as a graph with edge-lengths. As a consequence,* $r(\mathrm{X}^n) \to r(\mathrm{X})$*, and writing* L^n *(resp.* L*) for the restriction of the length measure of* X^n *(resp.,* X*) to* $\mathrm{conn}(\mathrm{X}^n)$ *[resp.,* $\mathrm{conn}(\mathrm{X})$*], it holds that*

$$\mathrm{d}_{\mathrm{GHP}}^{0,2}\big((X^n, d^n, \mu^n, L^n), (X, d, \mu, L)\big) \underset{n\to\infty}{\longrightarrow} 0.$$

(ii) *Let* x^n *be a random variable in* X^n *with distribution* $L^n/L^n(\mathrm{conn}(\mathrm{X}^n))$ *and* x *be a random variable in* X *with distribution* $L/L(\mathrm{conn}(\mathrm{X}))$*. Then as* $n \to \infty$*,*

$$\big(X^n, d^n, x^n, \mu^n\big) \xrightarrow{d} (X, d, x, \mu)$$

in the space $(\mathcal{M}^{1,1}, \mathrm{d}_{\mathrm{GHP}}^{1,1})$.

The above results rely on the following lemma. Given metric spaces (X, d) and (X', d'), $C \subset X \times X'$ and $r > 0$, let

$$C_r = \big\{(y, y') \in X \times X' : d(x, y) \vee d'\big(x', y'\big) \le r \text{ for } \big(x, x'\big) \in C\big\}.$$

C_r is the r-enlargement of C with respect to the product distance. Note that if C is a correspondence between X and X', then C_r is also a correspondence for every $r > 0$. Moreover, $\mathrm{dis}(C_r) \le \mathrm{dis}(C) + 4r$. A mapping $\phi : [a, b] \to [a', b']$ is called *bi-Lipschitz* if ϕ is a bijection such that ϕ and ϕ^{-1} are Lipschitz, and we call the quantity

$$K(\phi) = \inf\big\{K > 1 : \forall x, y \in [a, b],\, K^{-1}|x - y| \le |\phi(x) - \phi(y)| \le K|x - y|\big\}$$

the *bi-Lipschitz constant* of ϕ. By convention, we let $K(\phi) = \infty$ if ϕ is not a bijection, or not bi-Lipschitz.

LEMMA 6.7. *Fix* $r \in (0, 1)$ *and let* $(X, d), (X', d') \in \mathcal{A}_r$*. Suppose there exists a correspondence* C *between* X *and* X' *such that* $\mathrm{dis}(C) < r/56$.

Let $x, y \in X$ *be two distinct points in* X*, and let* f *be a local geodesic from* x *to* y*. Let* $x', y' \in X'$ *be such that* $(x, x'), (y, y') \in C$*. Then there exists a local geodesic* f' *from* x' *to* y' *with*

$$\mathrm{len}\big(f'\big) \le \left(1 + \frac{64\,\mathrm{dis}(C)}{r \wedge \mathrm{len}(f)}\right) \cdot \mathrm{len}(f),$$

and a bi-Lipschitz mapping $\phi : [0, \mathrm{len}(f)] \to [0, \mathrm{len}(f')]$ *such that* $(f(t), f'(\phi(t))) \in C_{8\,\mathrm{dis}(C)}$ *for every* $t \in [0, \mathrm{len}(f)]$*, and*

$$K(\phi) \le \left(1 - \frac{64\,\mathrm{dis}(C)}{r \wedge \mathrm{len}(f)}\right)_+^{-1}.$$

Note that the second part of the statement also implies a lower bound on the length of f', namely,

$$\mathrm{len}\big(f'\big) \ge K(\phi)^{-1}\,\mathrm{len}(f) \ge \mathrm{len}(f)\left(1 - \frac{64\,\mathrm{dis}(C)}{r \wedge \mathrm{len}(f)}\right)_+,$$

which is, of course, useless when $r \wedge \mathrm{len}(f) \le 64\,\mathrm{dis}(C)$.

PROOF OF LEMMA 6.7. Let us first assume that $0 < \mathrm{len}(f) \le r/8$, so in particular $d(x,y) \le R(\mathrm{X})$ and f is the geodesic from x to y. We have

$$d(x,y) - \mathrm{dis}(C) \le d'(x',y') \le d(x,y) + \mathrm{dis}(C) \le r/8 + \mathrm{dis}(C) < R(\mathrm{X}'),$$

so that x' and y' are linked by a unique geodesic f'. Set $\phi(t) = d'(x',y')t/d(x,y)$ for $0 \le t \le d(x,y)$. From the preceding chain of inequalities, we obtain that

$$\mathrm{len}(f') \le \mathrm{len}(f) + \mathrm{dis}(C) \quad \text{and} \quad K(\phi) \le \left(1 - \frac{\mathrm{dis}(C)}{d(x,y)}\right)_+^{-1}.$$

Fix $z = f(t) \in \mathrm{Im}(f)$ and let z'' be such that $(z,z'') \in C$. Then $d'(x',z'') \le d(x,z) + \mathrm{dis}(C) < r/4$, so that z'' belongs to the $\mathbb{R}$-tree $B_{R(\mathrm{X}')}(x')$. Let z' be the (unique) point of $\mathrm{Im}(f')$ that is closest to z''. Then a path from x' or y' to z'' must pass through z', from which we have

$$d'(z'',z') = \frac{d'(x',z'') + d'(y',z'') - d'(x',y')}{2} \le \frac{3}{2}\,\mathrm{dis}(C). \tag{6.1}$$

Therefore,

$$t - \frac{5}{2}\,\mathrm{dis}(C) \le d'(x',z'') - d'(z'',z') \le d'(x',z') \le d'(x',z'') \le t + \mathrm{dis}(C),$$

so after a short calculation we get that

$$\left| d'(x',z') - \frac{d'(x',y')}{d(x,y)}t \right| \le \frac{7}{2}\,\mathrm{dis}(C).$$

From this, we obtain

$$\begin{aligned} d'(z', f'(\phi(t))) &= d'\left(f'(d'(x',z')), f'\left(\frac{d'(x',y')}{d(x,y)}t\right)\right) \\ &= \left| d'(x',z') - \frac{d'(x',y')}{d(x,y)}t \right| \le \frac{7}{2}\,\mathrm{dis}(C), \end{aligned}$$

so that, in conjunction with (6.1), we have $(f(t), f'(\phi(t))) \in C_{5\,\mathrm{dis}(C)}$.

We next assume that $\mathrm{len}(f) > r/8$. Fix an integer N such that $r/16 < \mathrm{len}(f)/N \le r/8$ and let $t_i = i\,\mathrm{len}(f)/N$ and $x_i = f(t_i)$ for $0 \le i \le N$. By Proposition 6.1, since f is a local geodesic and $t_{i+1} - t_{i-1} < R(\mathrm{X})$ for every $i \in \{1,\dots,N-1\}$, the restriction $f|_{[t_{i-1},t_{i+1}]}$ must be a shortest path and so

$$d(x_{i-1},x_{i+1}) = \frac{2\,\mathrm{len}(f)}{N} \le r/4, \qquad d(x_i,x_{i+1}) = \frac{\mathrm{len}(f)}{N} \in [r/16, r/8].$$

Letting x_i'' be a point such that $(x_i, x_i'') \in C$ (where we always make the choice $x_0'' = x'$ and $x_N'' = y'$), we have $d'(x_i'', x_{i+1}'') \le d(x_i,x_{i+1}) + \mathrm{dis}(C) < R(\mathrm{X}')$, so

that we can consider the unique geodesic f_i'' between x_i'' and x_{i+1}''. The concatenation of the paths $f_0'', f_1'', \ldots, f_{N-1}''$ is not necessarily a local geodesic, but by excising certain parts of it we will be able to recover a local geodesic between x' and y'. For each $i \in \{1, \ldots, N-1\}$, the sets $\mathrm{Im}(f_{i-1}'')$ and $\mathrm{Im}(f_i'')$ are included in the $\mathbb{R}$-tree $B_{R(\mathrm{X}')}(x_i'')$, and the concatenation of f_{i-1}'' and f_i'' is a path from x_{i-1}'' to x_{i+1}'' which, as such, must contain the image of the geodesic g_i between these points. Let x_i' be the unique point of $\mathrm{Im}(g_i)$ that is closest to x_i'', and let $x_0' = x'$, $x_N' = y'$. Then

$$d'(x_i', x_i'') = \frac{d'(x_{i-1}'', x_i'') + d'(x_{i+1}'', x_i'') - d'(x_{i-1}'', x_{i+1}'')}{2} \le \frac{3}{2}\,\mathrm{dis}(C),$$

so that, for $i \in \{0, 1, \ldots, N\}$,

$$\begin{aligned} d(x_i, x_{i+1}) - \mathrm{dis}(C) &\le d'(x_i'', x_{i+1}'') \le d'(x_i', x_{i+1}') \\ &\le d'(x_i'', x_{i+1}'') + 3\,\mathrm{dis}(C) \le d(x_i, x_{i+1}) + 4\,\mathrm{dis}(C). \end{aligned}$$

If $x_{i+1}' \in \mathrm{Im}(f_{i-1}'')$ then

$$d'(x_{i-1}'', x_{i+1}'') \le d'(x_{i-1}'', x_i'') + \frac{3}{2}\,\mathrm{dis}(C) \le \frac{\mathrm{len}(f)}{N} + \frac{5}{2}\,\mathrm{dis}(C).$$

However, since $(x_{i-1}, x_{i-1}'') , (x_{i+1}, x_{i+1}'') \in C$ and $\mathrm{dis}(C) < r/56 < 2\,\mathrm{len}(f)/(7N)$, we have

$$d'(x_{i-1}'', x_{i+1}'') \ge \frac{2\,\mathrm{len}(f)}{N} - \mathrm{dis}(C) > \frac{\mathrm{len}(f)}{N} + \frac{5}{2}\,\mathrm{dis}(C),$$

so, in fact, $x_{i+1}' \notin \mathrm{Im}(f_{i-1}'')$ and, in particular, x_{i+1}' does not lie on the shortest path between x_{i-1}' and x_i'. From this, it follows that if f' denotes the concatenation of the geodesic f_i' between x_i' and x_{i+1}', for $0 \le i \le N-1$, then f' is a local geodesic between x' and y'. Its length is certainly bounded by the sum of the lengths of the paths f_i'', so that

$$\mathrm{len}(f') \le \sum_{i=1}^{N} d(x_i, x_{i+1}) + 4N\,\mathrm{dis}(C) \le \mathrm{len}(f) + \frac{64\,\mathrm{len}(f)\,\mathrm{dis}(C)}{r},$$

as claimed. Next, we let $\phi_i(t) = d'(x_i', x_{i+1}')t/d(x_i, x_{i+1})$, so that

$$K(\phi_i) \le \left(1 - \frac{4\,\mathrm{dis}(C)}{d(x_i, x_{i+1})}\right)^{-1} \le \left(1 - \frac{64\,\mathrm{dis}(C)}{r}\right)^{-1}.$$

If $\phi : [0, \mathrm{len}(f)] \to [0, \mathrm{len}(f')]$ is the concatenation of the mappings ϕ_i, $0 \le i \le N-1$, then ϕ is bi-Lipschitz with the same upper-bound for $K(\phi)$ as for each $K(\phi_i)$. Finally, we note that $f' \circ \phi$ is the concatenation of the paths $f_i' \circ \phi_i$. If $z = f_i(t)$, we let z'' be such that $(z, z'') \in C$ and let z' be the point in $\mathrm{Im}(f_i')$ that

is closest to z''. Then similar arguments to before entail that $d'(z', z'') \le 3\,\mathrm{dis}(C)$, so that

$$t - 4\,\mathrm{dis}(C) \le d'\bigl(x_i', z'\bigr) \le d'\bigl(x_i', z''\bigr) \le t + \mathrm{dis}(C),$$

which implies that

$$d'\bigl(z', f_i'(\phi_i'(t))\bigr) = \left| d'\bigl(x_i', z'\bigr) - \frac{d'(x_i', x_{i+1}')}{d(x_i, x_{i+1})} t \right| \le 5\,\mathrm{dis}(C),$$

and we conclude that $(f_i(t), f_i'(\phi_i(t))) \in C_{8\,\mathrm{dis}(C)}$, and so that $(f(s), f'(\phi(s))) \in C_{8\,\mathrm{dis}(C)}$ for every $s \in [0, \mathrm{len}(f)]$. $\square$

PROOF OF PROPOSITION 6.5. We will prove this result only when one of X and X′ (and then, in fact, both) has surplus at least 2, leaving the similar and simpler case of surplus 1 to the reader (the case of surplus 0 is trivial). Also, we may assume without loss of generality that $\epsilon < r/4$.

Fix $\epsilon \in (0, r/4)$, and fix any $\delta \in (0, \epsilon r^2/128)$. Also, fix $\mathrm{X}, \mathrm{X}' \in \mathcal{A}_r$ and a correspondence $C \in C(X, X')$ with $\mathrm{dis}(C) < \delta$. List the elements of $k(\mathrm{X})$ as $v_1, \ldots, v_n$, and fix elements $v_1'', \ldots, v_n''$ of X' with $(v_i, v_i'') \in C$ for each $1 \le i \le n$. Since $\mathrm{dis}(C) < \delta$ and $v_1, \ldots, v_n$ are pairwise at distance at least r, $v_1'', \ldots, v_n''$ are pairwise at distance at least $r - 2\delta > r/2$ and, in particular, are all distinct. Next, for every $e \in e(\mathrm{X})$, say with $e^+ = v_i$, $e^- = v_j$, fix a local geodesic f_e between v_i and v_j with $\mathrm{Im}(f_e) = e$ and $f_e(0) = e^-$. By Lemma 6.7, there exists a geodesic f_e'' from v_i'' to v_j'' and a bi-Lipschitz mapping $\phi_e : [0, \ell(e)] \to [0, \mathrm{len}(f_e'')]$ with

$$K(\phi_e) \le \left(1 - \frac{64\delta}{r}\right)^{-1} < 2,$$

and such that $(f_e(t), f_e''(\phi_e(t))) \in C_{8\delta}$ for every $t \in [0, \ell(e)]$. In particular, it follows that $\mathrm{len}(f_e'') > r/2$. Then we claim that for δ small enough, the following two properties hold:

1. For every $e \in e(\mathrm{X})$, the path $(f_e''(t), \epsilon/8 \le t \le \mathrm{len}(f_e'') - \epsilon/8)$ is injective.
2. For $e_1, e_2 \in e(\mathrm{X})$ with $e_1 \ne e_2$, we have

$$\bigl\{f_{e_1}''(t) : \epsilon/8 \le t \le \mathrm{len}\bigl(f_{e_1}''\bigr) - \epsilon/8\bigr\} \cap \bigl\{f_{e_2}''(t) : \epsilon/8 \le t \le \mathrm{len}\bigl(f_{e_2}''\bigr) - \epsilon/8\bigr\} = \varnothing.$$

To establish the first property, suppose that $f_e''(t) = f_e''(t')$ for some $e \in e(\mathrm{X})$ and distinct $t, t' \in [\epsilon/8, \mathrm{len}(f_e'') - \epsilon/8]$. For concreteness, let us assume that $e^- = v_i$ and $e^+ = v_j$. Since f_e'' is a local geodesic, this implies that $|t - t'| \ge R(\mathrm{X}') \ge r/4$. Moreover, since $(f_e(\phi_e^{-1}(t)), f_e''(t)), (f_e(\phi_e^{-1}(t')), f_e''(t')) \in C_{8\delta}$ and since $\delta < \epsilon/128$, we have

$$d\bigl(f_e(\phi_e^{-1}(t)), f_e(\phi_e^{-1}(t'))\bigr) \le d'\bigl(f_e''(t), f_e''(t')\bigr) + 8\delta = 8\delta < \epsilon/16. \tag{6.2}$$

On the other hand, we have

$$\bigl|\phi_e^{-1}(t) - \phi_e^{-1}(t')\bigr| \ge K(\phi_e)^{-1}|t - t'| \ge \frac{1}{2}|t - t'| \ge r/8 > \epsilon/2,$$

and since $\phi_e^{-1}(0) = 0$ and $\phi_e^{-1}(\mathrm{len}(f_e'')) = \ell(e)$,

$$|\phi_e^{-1}(t)| \geq K(\phi_e)^{-1} t > \epsilon/16, \qquad |\phi_e^{-1}(t) - \ell(e)| > \epsilon/16,$$

and similarly for t'. But if $s, s' \in [\epsilon/16, \ell(e) - \epsilon/16]$, then $d(f_e(s), f_e(s')) \geq (\epsilon/8) \wedge |s - s'|$, because a path from $f_e(s)$ to $f_e(s')$ is either a subarc of e, or passes through both vertices v_i and v_j. It follows that $d(f_e(\phi_e^{-1}(t)), f_e(\phi_e^{-1}(t')) \geq \epsilon/8$, in contradiction with (6.2). This yields that property 1 holds.

The argument for property 2 is similar: for every $t_1 \in (\epsilon/8, \mathrm{len}(f_{e_1}'') - \epsilon/8)$ and $t_2 \in (\epsilon/8, \mathrm{len}(f_{e_2}'') - \epsilon/8)$, there exist $x_1 \in e_1$ and $x_2 \in e_2$ such that $(x_1, f_{e_1}''(t_1))$ and $(x_2, f_{e_2}''(t_2))$ are in $C_{8\delta}$. Then the distance from x_1, x_2 to $k(\mathrm{X})$ is at least $\epsilon/16$ so that $d(x_1, x_2) \geq \epsilon/8$. From this, we deduce that $d'(f_{e_1}''(t_1), f_{e_2}''(t_2)) \geq d(x_1, x_2) - 8\delta > 0$.

Next, for every $i \in \{1, \ldots, n\}$, consider the points $f_e''(\epsilon/8)$, $e \in e(\mathrm{X})$ for which $e^- = v_i$, as well as the points $f_e''(\mathrm{len}(f_e'') - \epsilon/8)$ for which $e^+ = v_i$. These points are on the boundary of the ball $\overline{B}_{\epsilon/8}(v_i'')$, which we recall is an $\mathbb{R}$-tree. Let T_i be the subtree of $\overline{B}_{\epsilon/8}(v_i'')$ spanned by these points. Then property 1 above shows that

$$\bigcup_{1 \leq i \leq n} T_i \cup \bigcup_{e \in e(\mathrm{X})} \{f_e''(t), \epsilon/8 \leq t \leq \mathrm{len}(f_e'') - \epsilon/8\}$$

induces a closed subgraph of (X', d') without leaves, and so this subgraph is in fact a subgraph of $\mathrm{core}(\mathrm{X}')$. Furthermore, property 2 implies that the points of degree at least 3 in this subgraph can only belong to $\bigcup_{1 \leq i \leq n} T_i$. Since any such point is then an element of $k(\mathrm{X}')$ and $\mathrm{diam}(T_i) \leq \epsilon/4 < r$, we see that each T_i can contain at most one element of $k(\mathrm{X}')$. On the other hand, each T_i must contain at least one element of $k(\mathrm{X}')$ because T_i has at least three leaves (since v_i has degree at least 3). Thus, each T_i contains exactly one element of $k(\mathrm{X}')$, which we denote by v_i'. Next, for $e \in e(\mathrm{X}')$, if $e^- = v_i$, $e^+ = v_j$, then we let f_e' be the simple path from v_i' to v_j' that has a nonempty intersection with f_e''. It is clear that this path is well-defined and unique. Letting $\chi(v_i) = v_i'$ for $1 \leq i \leq n$, and letting $\chi(e) = \mathrm{Im}(f_e')$ for $e \in e(\mathrm{X})$, we have therefore defined a multigraph homomorphism from $\ker(\mathrm{X})$ to $\ker(\mathrm{X}')$, and this homomorphism is clearly injective. By symmetry of the roles of X and X', we see that $|k(\mathrm{X})| = |k(\mathrm{X}')|$ and $|e(\mathrm{X})| = |e(\mathrm{X}')|$, and so χ must, in fact, be a multigraph isomorphism.

Finally, since $\mathrm{len}(f_e) = \ell(e) \leq 1/r$, we have

$$|\ell(e) - \mathrm{len}(f_e'')| = |\mathrm{len}(f_e) - \mathrm{len}(f_e'')| \leq \frac{64\delta}{r} \mathrm{len}(f_e) \leq \frac{64\delta}{r^2} < \frac{\epsilon}{2},$$

by our choice of δ. But, by construction, $|\mathrm{len}(f_e'') - \ell(\chi(e))| < \epsilon/2$, since the endpoints of $\chi(e)$ each have distance at most $\epsilon/4$ from an endpoint of $f''(e)$. It follows that $|\ell(e) - \ell(\chi(e))| < \epsilon$. Finally, since every point of $\chi(e)$ is within distance $\epsilon/4$ of f_e'' and $e = \mathrm{Im}(f_e)$ and $\mathrm{Im}(f_e'')$ are in correspondence via $C_{8\delta}$, it follows that e and $\chi(e)$ are in correspondence via $C_{8\delta+\epsilon/4}$. Since $\mathrm{dis}(C_{8\delta+\epsilon/4}) < \mathrm{dis}(C) + 16\delta + \epsilon/2 < \epsilon$, this completes the proof. $\square$

PROOF OF COROLLARY 6.6. Again we only consider the case $s(\mathrm{X}) > 1$, the case $s(\mathrm{X}) = 1$ being easier [and the case $s(\mathrm{X}) = 0$ trivial].

Let $(\mathrm{X}_n, n \geq 1)$ and X be as in the statement of Corollary 6.6. Let $(C^n, n \geq 1)$ and $(\pi^n, n \geq 1)$ be sequences of correspondences and of measures, respectively, such that $\mathrm{dis}(C^n), \pi^n((C^n)^c)$ and $D(\pi^n; \mu^n, \mu)$ each converge to 0 as $n \to \infty$. The fact that $\ker(\mathrm{X}^n)$ converges to $\ker(\mathrm{X})$ as a graph with edge-lengths is then an immediate consequence of Proposition 6.5: for each n sufficiently large, simply replace C^n by $C^n \cup \hat{C}^n$, where $\hat{C}^n$ is an ϵ_n-overlay of X and X^n, for some sequence $\epsilon_n \to 0$ [we may assume $\epsilon_n \geq \mathrm{dis}(C^n)$]. We continue to write C^n instead of $C^n \cup \hat{C}^n$, and note that enlarging C^n diminishes $\pi_n((C^n)^c)$.

In particular, we obtain that for all large enough n, there is an isomorphism χ_n from $(k(\mathrm{X}), e(\mathrm{X}))$ to $(k(\mathrm{X}^n), e(\mathrm{X}^n))$ such that $\ell(e) - \ell(\chi_n(e))$ converges to 0 for every $e \in e(\mathrm{X})$. The fact that $r(\mathrm{X}^n) \to r(\mathrm{X})$ is immediate. We now fix a particular orientation of the edges, and view χ_n as an isomorphism of oriented graphs, in the sense that $\chi_n(e^-) = \chi_n(e)^-$.

For each $e \in e(\mathrm{X})$, let f_e be a local geodesic between e^- and e^+ with $f_e(0) = e^-$ and $f_e(\ell(e)) = e^+$ and, for each $n \geq 1$ and $e \in e(\mathrm{X}^n)$, define f_e accordingly. Then for each n sufficiently large, define a mapping Φ_n with domain $\mathrm{dom}(\Phi_n) = \bigcup_{e \in e(\mathrm{X})} f_e([0, \ell(e) - \epsilon_n])$ by setting $\Phi_n(f_e(t)) = f^n_{\chi_n(e)}(t)$ for each $e \in e(\mathrm{X})$ and each $0 \leq t \leq \ell(e) - \epsilon_n$.

By considering a small enlargement of C^n, or, equivalently, by letting ϵ_n tend to zero sufficiently slowly, we may assume without loss of generality that $(x, \Phi_n(x)) \in C^n$ for all $x \in \mathrm{dom}(\Phi_n)$. This comes from the fact that e and $\chi_n(e)$ are in correspondence via C^n; we leave the details of this verification to the reader. It follows that the relation $\{(x, \Phi_n(x)) : x \in \mathrm{dom}(\Phi_n)\}$ is a subset of C^n.

Let $e_c(\mathrm{X})$ be the set of edges $e \in e(\mathrm{X})$ whose removal from $e(\mathrm{X})$ does not disconnect $\ker(\mathrm{X})$. Clearly, $\mathrm{conn}(X) \subseteq k(\mathrm{X}) \cup \bigcup_{e \in e_c(\mathrm{X})} e$, and the measure L is carried by $\bigcup_{e \in e_c(\mathrm{X})} e$ (in fact, it is carried by the subset of points of $\bigcup_{e \in e_c(\mathrm{X})} e$ with degree 2, by Proposition 2.6). Let L' be the restriction of L to the set $\mathrm{dom}(\Phi_n)$, which has total mass $\sum_{e \in e_c(\mathrm{X})}(\ell(e) - \epsilon_n)$. We consider the push-forward ρ_n of L' by the mapping $x \mapsto (x, \Phi_n(x))$ from X to $X \times X^n$. Then the second marginal of ρ_n is the restriction of L^n to $\bigcup_{e \in e_c(\mathrm{X})} \mathrm{Im}(f^n_e)$, so that

$$D(\rho_n; L, L^n) \leq \sum_{e \in e(\mathrm{X})} \bigl(\epsilon_n + \bigl|\ell(e) - \ell\bigl(\chi_n(e)\bigr)\bigr|\bigr).$$

The latter converges to 0 by the convergence of the edge-lengths. It only remains to note that $\rho_n(X \times X^n \setminus C^n) = 0$ by construction. This yields (i).

Finally, (i) implies that $(X^n, d^n, \mu^n, L^n/L^n(\mathrm{conn}(\mathrm{X}^n)))$ converges to $(X, d, \mu, L/L(\mathrm{conn}(\mathrm{X})))$ in the metric $\mathrm{d}^{0,2}_{\mathrm{GHP}}$ introduced in Section 2.1, and (ii) then follows from Proposition 2.1. □

7. Cutting safely pointed $\mathbb{R}$-graphs. In this section, we will consider a simple cutting procedure on $\mathbb{R}$-graphs, and study how this procedure is perturbed by small variations in the Gromov–Hausdorff distance.

7.1. *The cutting procedure.* Let (X, d) be an $\mathbb{R}$-graph, and let $x \in \mathrm{conn(X)}$. We endow the connected set $X \setminus \{x\}$ with the intrinsic distance $d_{X\setminus\{x\}}$: more precisely, $d_{X\setminus\{x\}}(y, z)$ is defined to be the minimal length of an injective path not visiting x. This is indeed a minimum because there are finitely many injective paths between y and z in X, as a simple consequence of Theorem 2.7 applied to core(X). The space $(X \setminus \{x\}, d_{X\setminus\{x\}})$ is not complete, so we let (X_x, d_x) be its metric completion as in Section 3.2. This space is connected, and thus easily seen to be an $\mathbb{R}$-graph. We call it the $\mathbb{R}$-*graph* (X, d) *cut at the point* x.

From now on, we will further assume that $\deg_X(x) = 2$, so that (X, d, x) is safely pointed as in Definition 3.2. In this case, one can provide a more detailed description of (X_x, d_x). A Cauchy sequence $(x_n, n \geq 1)$ in $(X \setminus \{x\}, d_{X\setminus\{x\}})$ is also a Cauchy sequence in (X, d), since $d \leq d_{X\setminus\{x\}}$. If its limit y in (X, d) is distinct from x, then it is easy to see that $d_{X\setminus\{x\}}(x_n, y) \to 0$, by considering a ball $B_\epsilon(y)$ not containing x within which $d = d_{X\setminus\{x\}}$.

So let us assume that $(x_n, n \geq 1)$ converges to x for the distance d. Since x has degree 2, the $\mathbb{R}$-tree $B_{R(\mathrm{X})}(x) \setminus \{x\}$ has exactly two components, say Y_1, Y_2. It is clear that $d_{X\setminus\{x\}}(z_1, z_2) \geq 2R(\mathrm{X})$ for every $z_1 \in Y_1, z_2 \in Y_2$. Since $(x_n, n \geq 1)$ is a Cauchy sequence for $(X \setminus \{x\}, d_{X\setminus\{x\}})$, we conclude that it must eventually take all its values in precisely one of Y_1 and Y_2, let us say Y_1 for definiteness. Note that the restrictions of d and $d_{X\setminus\{x\}}$ to Y_1 are equal, so that if $(x'_n, n \geq 1)$ is another Cauchy sequence in $(X \setminus \{x\}, d_{X\setminus\{x\}})$ which converges in (X, d) to x and takes all but a finite number of values in Y_1, then $d_{X\setminus\{x\}}(x_n, x'_n) = d(x_n, x'_n) \to 0$, and so this sequence is equivalent to $(x_n, n \geq 1)$.

We conclude that the completion of $(X \setminus \{x\}, d_{X\setminus\{x\}})$ adds exactly two points to $X \setminus \{x\}$, corresponding to classes of Cauchy sequences converging to x in (X, d) "from one side" of x. So we can write $X_x = (X \setminus \{x\}) \cup \{x_{(1)}, x_{(2)}\}$ and describe d_x as follows:

- If $y, z \notin \{x_{(1)}, x_{(2)}\}$ then $d_x(y, z)$ is the minimal length of a path from y to z in X not visiting x.
- If $y \neq x_{(2)}$ then $d_x(x_{(1)}, y)$ is the minimal length of an injective path from x to y in X which takes its values in the component Y_1 on some small initial interval $(0, \epsilon)$, and similarly for $d(x_{(2)}, y)$ with $y \neq x_{(1)}$.
- Finally, $d_x(x_{(1)}, x_{(2)})$ is the minimal length of an embedded cycle passing through x.

If (X, d, x, μ) is a pointed measured metric space such that (X, d, x) is a safely pointed $\mathbb{R}$-graph, and $\mu(\{x\}) = 0$, then the space (X_x, d_x) carries a natural measure μ', such that $\mu'(\{x_{(1)}, x_{(2)}\}) = 0$ and, for any open subset $A \subseteq X_x$ not containing $x_{(1)}$ and $x_{(2)}$, $\mu'(A) = \mu(A)$ if on the right-hand side we view A as an open subset of X. Consequently, there is little risk of ambiguity in using the notation μ instead of μ' for this induced measure.

We finish this section by proving Lemma 5.5 on the number of balls required to cover the cut space.

PROOF OF LEMMA 5.5. Let $B_1, B_2, \ldots, B_N$ be a covering of X by open balls of radius $r > 0$, centred at $x_1, \ldots, x_N$, respectively. By definition, any point of X can be joined to the centre of some ball B_i by a geodesic path of length $< r$. If such a path does not pass through x, then it is also a geodesic path in X_x. Now since $B_1, \ldots, B_N$ is a covering of X, this implies that any point y in X can either:

- be joined to some point x_i by a path of length $< r$ that does not pass through x,
- or can be joined to x through a path γ of length $< r$.

In the first case, this means that y belongs to the ball with centre x_i and radius r in X_x. In the second case, depending on whether the initial segment of γ belongs to Y_1 or Y_2, this means that y belongs to the ball with centre $x_{(1)}$ or $x_{(2)}$ with radius r in X_x. This yields a covering of X_x with at most $N + 2$ balls, as desired.

Conversely, it is clear that if N balls are sufficient to cover X_x then the same is true of X, because distances are smaller in X than in X_x (if x is identified with the points $\{x_{(1)}, x_{(2)}\}$). □

7.2. *Stability of the cutting procedure.* The following statement will be used in conjunction with Corollary 6.6(ii). Recall the definition of $\mathcal{A}_r$ and the definition of safely pointed $\mathbb{R}$-graphs, both given in Section 3.2. Then let $\mathcal{A}_r^\bullet$ be the set of (isometry equivalence classes of) safely pointed measured $\mathbb{R}$-graphs (X, d, x, μ) with $(X, d, \mu) \in \mathcal{A}_r$, and say that a pointed $\mathbb{R}$-graph $(X, d, x) \in \mathcal{A}_r^\bullet$ if $(X, d, x, 0) \in \mathcal{A}_r^\bullet$.

THEOREM 7.1. *Fix $r \in (0, 1)$. Let (X^n, d^n, x^n, μ^n), $n \geq 1$ and (X, d, x, μ) be elements of $\mathcal{A}_r^\bullet$. Suppose that*

$$\mathrm{d}_{\mathrm{GHP}}^{1,1}\big((X^n, d^n, x^n, \mu^n), (X, d, x, \mu)\big) \underset{n\to\infty}{\longrightarrow} 0,$$

and that $\mu^n(\{x\}) = \mu(\{x\}) = 0$ for every n. Then

$$\mathrm{d}_{\mathrm{GHP}}\big((X_{x^n}^n, d_{x^n}^n, \mu^n), (X_x, d_x, \mu)\big) \underset{n\to\infty}{\longrightarrow} 0.$$

Our proof of Theorem 7.1 hinges on two lemmas; to state these lemmas we require a few additional definitions. Let $\mathrm{X} = (X, d, x, \mu) \in \mathcal{A}_r^\bullet$ and recall the definition of the projection $\alpha : X \to \mathrm{core}(\mathrm{X})$. For $\epsilon > 0$, write

$$\widetilde{B}_\epsilon(x) = \big\{y \in X : d\big(\alpha(y), x\big) < \epsilon\big\} \quad \text{and} \quad h_\epsilon(\mathrm{X}) = \mathrm{diam}\big(\widetilde{B}_\epsilon(x)\big),$$

so that $B_\epsilon(x) \subseteq \widetilde{B}_\epsilon(x)$. The sets $\widetilde{B}_\epsilon(x)$ decrease to the singleton $\{x\}$ as $\epsilon \downarrow 0$, because $\deg_X(x) = 2$. Consequently, $h_\epsilon(\mathrm{X})$ converges to 0 as $\epsilon \downarrow 0$. For $\epsilon > 0$ sufficiently small, the set $X_{x,\epsilon} = X \setminus \widetilde{B}_\epsilon(x)$, endowed with the intrinsic metric, is an $\mathbb{R}$-graph. In fact, it is easy to see that for $\epsilon < R(X)$, this intrinsic metric is just the restriction of d_x to $X_{x,\epsilon}$.

Let us assume that $\epsilon < d(x, k(\mathrm{X})) \wedge R(\mathrm{X})$. Let $x_{(1),\epsilon}, x_{(2),\epsilon}$ be the two points of core(X) at distance ϵ from x, labelled in such a way that, in the notation of

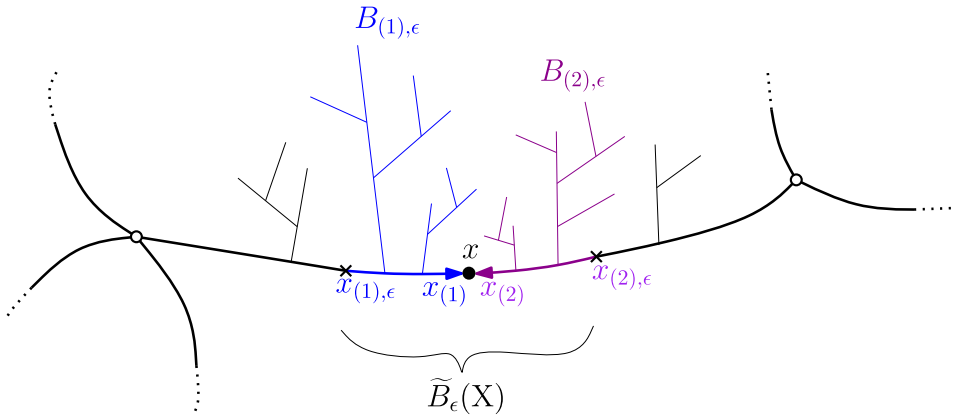

FIG. 4. *Part of an* $\mathbb{R}$*-graph*: core(X) *is in thicker line.*

Section 7.1, $x_{(1),\epsilon}$ is the point closest to $x_{(1)}$ in X_x, or in other words, such that $x_{(1),\epsilon} \in Y_1$. For $i \in \{1,2\}$, let P_i be the geodesic arc between $x_{(i),\epsilon}$ and x in X. We let $B_{(i),\epsilon} = \{w \in \widetilde{B}_\epsilon(x) \setminus \{x\} : \alpha(w) \in P_i\} \cup \{x_{(i)}\}$, which we see as a subset of X_x. See Figure 4 for an illustration.

Now let $\mathrm{X}' = (X', d', x', \mu') \in \mathcal{A}_r^\bullet$. Just as we defined the space $\mathrm{X}_x = (X_x, d_x)$, we define the space $\mathrm{X}'_{x'} = (X'_{x'}, d'_{x'})$, with $X'_{x'} = (X' \setminus \{x'\}) \cup \{x'_{(1)}, x'_{(2)}\}$. We likewise define the sets $\widetilde{B}_\epsilon(x')$ and $B'_{(1),\epsilon}, B'_{(2),\epsilon}$ and the points $x'_{(1),\epsilon}, x'_{(2),\epsilon}$ for $\epsilon < d(x', k(\mathrm{X}')) \wedge R(X')$ as above. We will use the same notation α for the projection $X' \to \mathrm{core}(\mathrm{X}')$.

LEMMA 7.2. *Fix* $\delta > 0$. *If* C *is a* δ*-overlay of* X *and* X′, *then for every* $(y, y') \in C$, *we have* $(\alpha(y), \alpha(y')) \in C_{2\delta}$.

PROOF. Let y'' be such that $(\alpha(y), y'') \in C$. Since C is a δ-overlay, $d'(y'', \mathrm{core}(\mathrm{X}')) < \delta$. In particular, if $\alpha(y'') = \alpha(y')$ then we have $d'(\alpha(y'), y'') < \delta$. Otherwise, a geodesic from y' to y'' must pass through $\alpha(y')$ and $\alpha(y'')$, so that

$$d'(y', \alpha(y')) + d'(\alpha(y'), y'') = d'(y', y'') \le d(y, \alpha(y)) + \delta.$$

On the other hand, since C is an δ-overlay, we know that core(X) and core(X′) are in correspondence via C, which implies that

$$d'(y', \alpha(y')) = d'(y', \mathrm{core}(\mathrm{X}')) > d(y, \mathrm{core}(\mathrm{X})) - \delta = d(y, \alpha(y)) - \delta,$$

so that $d'(\alpha(y'), y'') \le 2\delta$. In all cases, we have $(\alpha(y), \alpha(y')) \in C_{2\delta}$, as claimed. □

LEMMA 7.3. *Fix $r \in (0,1)$. For all $\epsilon > 0$, there exists $\eta > 0$ such that if $\mathrm{d}_{\mathrm{GHP}}^{1,1}(\mathrm{X}, \mathrm{X}') < \eta$ then*

$$\mathrm{d}_{\mathrm{GHP}}(\mathrm{X}_x, \mathrm{X}'_{x'}) \le \mu\big(\widetilde{B}_\epsilon(x)\big) + \mu'\big(\widetilde{B}_\epsilon(x')\big) + 3\max\big(h_\epsilon(x), h_\epsilon(x')\big) + 7\left(2\frac{\mathrm{diam}(\mathrm{X}) \vee \mathrm{diam}(\mathrm{X}')}{r} \vee 1\right)\epsilon.$$

PROOF. Since $\mathrm{d}_{\mathrm{GHP}}^{1,1}(\mathrm{X}, \mathrm{X}') < \eta$, we can find $C_0 \in C(\mathrm{X}, \mathrm{X}')$ with $\mathrm{dis}(C_0) < \eta$ and with $(x, x') \in C_0$, and a measure π with $D(\pi; \mu, \mu') \le \eta$ and $\pi(C_0^c) < \eta$. Fix $\delta > 0$ such that $\delta < \epsilon/10$ and $\delta < r/56$. By choosing $\eta < \delta$ sufficiently small, it follows by Proposition 6.5 that there exists a δ-overlay C of X and X′ with $C_0 \subset C$, so in particular $(x, x') \in C_0$ and $\pi(C^c) < \eta < \delta$. We also remark that $D(\pi; \mu, \mu') \le \delta$.

We next modify C to give a correspondence between $X_{x,\epsilon}$ and $X'_{x',\epsilon}$ by letting

$$C^{(\epsilon)} = \big(C \cap (X_{x,\epsilon} \times X'_{x',\epsilon})\big) \cup A_1 \cup A_2 \cup A'_1 \cup A'_2,$$

where for $i \in \{1, 2\}$, we define

$$A_i = \big\{(y, x'_{(i),\epsilon}) : (y, y') \in C \cap (X_{x,\epsilon} \times B'_{(i),\epsilon})\big\},$$
$$A'_i = \big\{(x_{(i),\epsilon}, y') : (y, y') \in C \cap (B_{(i),\epsilon} \times X'_{x',\epsilon})\big\}.$$

To verify that $C^{(\epsilon)}$ is indeed a correspondence between $X_{x,\epsilon}$ and $X'_{x',\epsilon}$, it suffices to check that there does not exist $y \in X_{x,\epsilon}$ for which $(y, x') \in C$, and similarly that there does not exist $y' \in X'_{x',\epsilon}$ for which $(x, y') \notin C$. In the first case, this is immediate since $d(x, y) \ge \epsilon$, so for all $y' \in X'$ with $(y, y') \in C$ we have $d'(x', y') \ge \epsilon - \delta > 0$. A symmetric argument handles the second case.

We next estimate the distortion of $C^{(\epsilon)}$ when $X_{x,\epsilon}$ and $X'_{x',\epsilon}$ are endowed with the metrics d_x and $d'_{x'}$, respectively. To this end, let $(y, y'), (z, z') \in C^{(\epsilon)}$. We have to distinguish several cases. The simplest case is when (y, y') and (z, z') are, in fact, both in C. In particular, $y, z \in X_{x,\epsilon}$ and $y', z' \in X'_{x',\epsilon}$. Let f be a geodesic from y to z in $X_{x,\epsilon}$, i.e. a local geodesic in $X_{x,\epsilon}$ not passing through x and with minimal length. Let f' be the path from y' to z' associated with f as in Lemma 6.7, which we may apply since $\delta < r/56$. We claim that f' does not pass through x'. Indeed, if it did, then we would be able to find a point $x_0 \in \mathrm{Im}(f)$ such that $(x_0, x') \in C_{8\delta}$. Since also $(x, x') \in C_{8\delta}$, it would follow that

$$d\big(x, \mathrm{Im}(f)\big) \le d(x, x_0) \le d'\big(x', x'\big) + 8\delta < \epsilon,$$

contradicting the fact that f is a path in $X_{x,\epsilon}$. By Lemma 6.7, we deduce that

$$\begin{aligned} d'_{x'}(y',z') &\le d_x(y,z)\left(1+\frac{64\delta}{r\wedge d_x(y,z)}\right) \\ &= d_x(y,z)+64\left(\frac{d_x(y,z)}{r}\vee 1\right)\delta \\ &\le d_x(y,z)+64\left(\frac{2\operatorname{diam}(\mathrm{X})}{r}\vee 1\right)\delta, \end{aligned} \tag{7.1}$$

where at the last step we use that $d_x(y,z)\le \operatorname{diam}(\mathrm{X}_x)\le 2\operatorname{diam}(\mathrm{X})$.

Let us now consider the cases where $(y,y')\notin C$, still assuming that $(z,z')\in C$. There are two possibilities:

1. There exists $y''\in B'_{(i),\epsilon}$ with $(y,y'')\in C$ and $i\in\{1,2\}$, and so $y'=x'_{(i),\epsilon}$.
2. There exists $\overline{y}\in B_{(i),\epsilon}$ with $(\overline{y},y')\in C$ and $i\in\{1,2\}$, and so $y=x_{(i),\epsilon}$.

Let us consider the first case, assuming $i=1$ for definiteness. The argument leading to (7.1) is still valid, with y'' replacing y'. Using $d'_{x'}(y'',x'_{(1),\epsilon})=d'(y'',x'_{(1),\epsilon})\le h_\epsilon(x')$, we obtain

$$d'_{x'}(y',z')\le d_x(y,z)+64\left(\frac{2\operatorname{diam}(\mathrm{X})}{r}\vee 1\right)\delta+h_\epsilon(x').$$

In the second case (still assuming $i=1$ without loss of generality), we have to modify the argument as follows. We consider a geodesic f from $\overline{y}$ to z in (X_x,d_x). We f' be the associated path from y' to z' (again using Lemma 6.7), and claim that $x'\notin\operatorname{Im}(f')$. Otherwise, f would visit a point at distance less than 8δ from x. On the other hand, the point of $\operatorname{Im}(f)$ that is closest to x is $\alpha(\overline{y})$. But by Lemma 7.2, we have

$$d\big(x,\alpha(\overline{y})\big)\ge d'\big(x',\alpha(y')\big)-2\delta\ge\epsilon-2\delta>8\delta.$$

Finally, since $y=x_{(1),\epsilon}$ we obtain that $d_x(\overline{y},z)\le d_x(y,z)+h_\epsilon(x)$, and the argument leading to (7.1) yields

$$d'_{x'}(y',z')\le d_x(y,z)+h_\epsilon(x)+64\left(\frac{2\operatorname{diam}(\mathrm{X})}{r}\vee 1\right)\delta.$$

Arguing similarly when (z,z') is no longer assumed to belong to C, we obtain the following bound for every $(y,y'),(z,z')\in C^{(\epsilon)}$:

$$d'_{x'}(y',z')\le d_x(y,z)+2\big(h_\epsilon(x)\vee h_\epsilon(x')\big)+64\left(\frac{2\operatorname{diam}(\mathrm{X})}{r}\vee 1\right)\delta.$$

Writing $h_\epsilon=h_\epsilon(x)\vee h_\epsilon(x')$, by symmetry we thus conclude that

$$\operatorname{dis}\big(C^{(\epsilon)}\big)\le 2h_\epsilon+64\left(2\frac{\operatorname{diam}(\mathrm{X})\vee\operatorname{diam}(\mathrm{X}')}{r}\vee 1\right)\delta,$$

where the distortion is measured with respect to the metrics d_x and $d'_{x'}$. Now let $\hat{C}^{(\epsilon)}$ be the h_ϵ-enlargement of $C^{(\epsilon)}$ with respect to d_x and $d'_{x'}$. Since $C^{(\epsilon)}$ is a correspondence between $X_{x,\epsilon}$ and $X'_{x',\epsilon}$, and all points of X_x (resp., $X'_{x'}$) have distance at most h_ϵ from $X_{x,\epsilon}$ (resp., $X'_{x',\epsilon}$) under d_x (resp., $d'_{x'}$), we have that $\hat{C}^{(\epsilon)}$ is a correspondence between X_x and $X'_{x'}$, of distortion at most

$$3h_\epsilon + 64\left(2\frac{\operatorname{diam}(\mathrm{X}) \vee \operatorname{diam}(\mathrm{X}')}{r} \vee 1\right)\delta.$$

Finally, since $D(\pi; \mu, \mu') \le \delta$ and $\pi(C^c) \le \delta$, and since $(C \cap (X_{x,\epsilon} \times X'_{x',\epsilon})) \subset C^{(\epsilon)} \subset \hat{C}^{(\epsilon)}$, we have

$$\begin{aligned}\pi\big((\hat{C}^{(\epsilon)})^c\big) &\le \pi\big(C^c\big) + \pi\big(\widetilde{B}_\epsilon(x) \times X'\big) + \pi\big(X \times \widetilde{B}_\epsilon(x')\big)\\ &\le \delta + \mu\big(\widetilde{B}_\epsilon(x)\big) + \mu'\big(\widetilde{B}_\epsilon(x')\big).\end{aligned}$$

Since $65\delta < 6.5\epsilon < 7\epsilon$, the lemma then follows from the two preceding offset equations and the definition of the distance $\mathrm{d_{GHP}}$. □

PROOF OF THEOREM 7.1. Fix $\epsilon > 0$. Under the hypotheses of the theorem, for all n large enough, by Lemma 7.3 we have

$$\begin{aligned}\mathrm{d_{GHP}}\big(\mathrm{X}_x, \mathrm{X}^n_{x^n}\big) &\le \mu\big(\widetilde{B}_\epsilon(x)\big) + \mu'\big(\widetilde{B}_\epsilon(x^n)\big) + 3\max\big(h_\epsilon(x), h_\epsilon(x^n)\big)\\ &\quad + 7\left(2\frac{\operatorname{diam}(\mathrm{X}) \vee \operatorname{diam}(\mathrm{X}^n)}{r} \vee 1\right)\epsilon.\end{aligned}$$

It is easily checked that, for all $\epsilon > 0$,

$$\limsup_{n\to\infty} h_\epsilon(x^n) \le h_{2\epsilon}(x), \qquad \limsup_{n\to\infty} \mu^n\big(\widetilde{B}_\epsilon(x^n)\big) \le \mu\big(\widetilde{B}_{2\epsilon}(x)\big),$$

which both converge to 0 as $\epsilon \to 0$. The result follows. □

7.3. *Randomly cutting $\mathbb{R}$-graphs.* Let $\mathrm{X} = (X, d, x)$ be a safely pointed $\mathbb{R}$-graph, and write L for the length measure restricted to $\operatorname{conn}(\mathrm{X})$. Then $\mathrm{X}_x = (X_x, d_x)$ is an $\mathbb{R}$-graph with $s(\mathrm{X}_x) = s(\mathrm{X}) - 1$. Indeed, if e is the edge of $\ker(\mathrm{X})$ that contains x, then it is easy to see that $\ker(\mathrm{X}_x)$ is the graph obtained from X by first deleting the interior of the edge e, and then taking the kernel of the resulting $\mathbb{R}$-graph. Taking the kernel of a graph does not modify its surplus, and so the surplus diminishes by 1 during this operation, which corresponds to the deletion of the edge e. Moreover, we see that $\mathcal{A}_r$ is stable under this operation, in the sense that if $(X, d) \in \mathcal{A}_r$, then for every x such that (X, d, x) is safely pointed, the space (X_x, d_x) is again in $\mathcal{A}_r$. Indeed, the edges in $\ker(\mathrm{X}_x)$ are either edges of $\ker(\mathrm{X})$, or a concatenation of edges in $\ker(\mathrm{X})$, and so the minimum edge-length can only increase. On the other hand, the total core length and surplus can only decrease.

Let us now consider the following random cutting procedure for $\mathbb{R}$-graphs. If (X, d) is an $\mathbb{R}$-graph which is not an $\mathbb{R}$-tree, then it contains at least one cycle, and

by Proposition 2.6, ℓ-almost every point of any such cycle is in conn(X). Consequently, the measure $L = \ell(\cdot \cap \text{conn}(\mathrm{X}))$ is nonzero, and we can consider a point x chosen at random in conn(X) with distribution $L/L(\text{conn}(\mathrm{X}))$. Then (X, d, x) is a.s. safely pointed by Proposition 2.6. Let $\mathcal{K}(\mathrm{X}, \cdot)$ be the distribution of (X_x, d_x). By convention, if X is an $\mathbb{R}$-tree, we let $\mathcal{K}(\mathrm{X}, \cdot) = \delta_{\{\mathrm{X}\}}$. By combining Corollary 6.6(ii) with Theorem 7.1, we immediately obtain the following statement.

PROPOSITION 7.4. *Fix $r > 0$, and let $(\mathrm{X}^n, n \geq 1)$ and X be elements of $\mathcal{A}_r$ such that $\mathrm{d_{GHP}}(\mathrm{X}^n, \mathrm{X}) \to 0$ as $n \to \infty$. Then $\mathcal{K}(\mathrm{X}^n, \cdot) \xrightarrow{d} \mathcal{K}(\mathrm{X}, \cdot)$ in $(\mathcal{M}, \mathrm{d_{GHP}})$, as $n \to \infty$.*

In particular, $\mathcal{K}$ defines a Markov kernel from $\mathcal{A}_r$ to itself for every r. Since each application of this kernel decreases the surplus by 1 until it reaches 0, it makes sense to define $\mathcal{K}^\infty(\mathrm{X}, \cdot)$ to be the law of $\mathcal{K}^m(\mathrm{X}, \cdot)$ for every $m \geq s(\mathrm{X})$, where $\mathcal{K}^m$ denotes the m-fold composition of $\mathcal{K}$. The next corollary follows immediately from Proposition 7.4 by induction.

COROLLARY 7.5. *Fix $r > 0$, and let $(\mathrm{X}^n, n \geq 1)$ and X be elements of $\mathcal{A}_r$ with $\mathrm{d_{GHP}}(\mathrm{X}^n, \mathrm{X}) \to 0$ as $n \to \infty$. Then $\mathcal{K}^\infty(\mathrm{X}^n, \cdot) \xrightarrow{d} \mathcal{K}^\infty(\mathrm{X}, \cdot)$ in $(\mathcal{M}, \mathrm{d_{GHP}})$, as $n \to \infty$.*

This proves Theorem 3.3.

Acknowledgements. CG is grateful to Université Paris-Sud for making her professeur invité for March 2011 which enabled work on this project to progress, and for the support of the University of Warwick. GM is grateful to PIMS (CNRS UMI 3069) and UBC, Vancouver, for hospitality in the year 2011–2012, and for the support of Université Paris-Sud.

We are very grateful to the referee, whose extremely careful reading and helpful comments led to several improvements in the exposition of this paper.

REFERENCES

[1] ABRAHAM, R., DELMAS, J.-F. and HOSCHEIT, P. (2013). A note on the Gromov–Hausdorff–Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.* **18** no. 14. MR3035742

[2] ADDARIO-BERRY, L. (2013). The local weak limit of the minimum spanning tree of the complete graph. Available at http://problab.ca/louigi/unscaled.pdf and arXiv:1301.1667 [math.PR].

[3] ADDARIO-BERRY, L., BROUTIN, N. and GOLDSCHMIDT, C. (2010). Critical random graphs: Limiting constructions and distributional properties. *Electron. J. Probab.* **15** 741–775. MR2650781

[4] ADDARIO-BERRY, L., BROUTIN, N. and GOLDSCHMIDT, C. (2012). The continuum limit of critical random graphs. *Probab. Theory Related Fields* **152** 367–406. MR2892951

[5] ADDARIO-BERRY, L., BROUTIN, N. and REED, B. (2009). Critical random graphs and the structure of a minimum spanning tree. *Random Structures Algorithms* **35** 323–347. MR2548517

[6] ADDARIO-BERRY, L., GRIFFITHS, S. and KANG, R. J. (2012). Invasion percolation on the Poisson-weighted infinite tree. *Ann. Appl. Probab.* **22** 931–970. MR2977982

[7] AIZENMAN, M., BURCHARD, A., NEWMAN, C. M. and WILSON, D. B. (1999). Scaling limits for minimal and random spanning trees in two dimensions. *Random Structures Algorithms* **15** 319–367. MR1716768

[8] ALDOUS, D. (1990). A random tree model associated with random graphs. *Random Structures Algorithms* **1** 383–402. MR1138431

[9] ALDOUS, D. (1991). The continuum random tree. I. *Ann. Probab.* **19** 1–28. MR1085326

[10] ALDOUS, D. (1991). The continuum random tree. II. An overview. In *Stochastic Analysis* (*Durham*, 1990). *London Mathematical Society Lecture Note Series* **167** 23–70. Cambridge Univ. Press, Cambridge. MR1166406

[11] ALDOUS, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226

[12] ALDOUS, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25** 812–854. MR1434128

[13] ALDOUS, D. and STEELE, J. M. (1992). Asymptotics for Euclidean minimal spanning trees on random points. *Probab. Theory Related Fields* **92** 247–258. MR1161188

[14] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 1–72. Springer, Berlin. MR2023650

[15] ALEXANDER, K. S. (1995). Percolation and minimal spanning forests in infinite graphs. *Ann. Probab.* **23** 87–104. MR1330762

[16] ALEXANDER, K. S. (1996). The RSW theorem for continuum percolation and the CLT for Euclidean minimal spanning trees. *Ann. Appl. Probab.* **6** 466–494. MR1398054

[17] ALEXANDER, K. S. and MOLCHANOV, S. A. (1994). Percolation of level sets for two-dimensional random fields with lattice symmetry. *J. Stat. Phys.* **77** 627–643. MR1301459

[18] ANGEL, O., GOODMAN, J., DEN HOLLANDER, F. and SLADE, G. (2008). Invasion percolation on regular trees. *Ann. Probab.* **36** 420–466. MR2393988

[19] ANGEL, O., GOODMAN, J. and MERLE, M. (2013). Scaling limit of the invasion percolation cluster on a regular tree. *Ann. Probab.* **41** 229–261. MR3059198

[20] AVRAM, F. and BERTSIMAS, D. (1992). The minimum spanning tree constant in geometrical probability and under the independent model: A unified approach. *Ann. Appl. Probab.* **2** 113–130. MR1143395

[21] BEARDWOOD, J., HALTON, J. H. and HAMMERSLEY, J. M. (1959). The shortest path through many points. *Math. Proc. Cambridge Philos. Soc.* **55** 299–327. MR0109316

[22] BEVERIDGE, A., FRIEZE, A. and MCDIARMID, C. (1998). Random minimum length spanning trees in regular graphs. *Combinatorica* **18** 311–333. MR1721947

[23] BHAMIDI, S., BROUTIN, N., SEN, S. and WANG, X. (2014). Scaling limits of random graph models at criticality: Universality and the basin of attraction of the Erdős–Rényi random graph. Available at arXiv:1411.3417 [math.PR].

[24] BORŮVKA, O. (1926). O jistém problému minimálním. *Práce Moravské Přírodovědecké Společnosti Brno* **3** 37–58.

[25] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418

[26] DAMRON, M. and SAPOZHNIKOV, A. (2011). Outlets of 2D invasion percolation and multiple-armed incipient infinite clusters. *Probab. Theory Related Fields* **150** 257–294. MR2800910

[27] DAMRON, M., SAPOZHNIKOV, A. and VÁGVÖLGYI, B. (2009). Relations between invasion percolation and critical percolation in two dimensions. *Ann. Probab.* **37** 2297–2331. MR2573559

[28] DUQUESNE, T. and LE GALL, J.-F. (2005). Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields* **131** 553–603. MR2147221

[29] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** 17–61. MR0125031

[30] EVANS, S. N., PITMAN, J. and WINTER, A. (2006). Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields* **134** 81–126. MR2221786

[31] EVANS, S. N. and WINTER, A. (2006). Subtree prune and regraft: A reversible real tree-valued Markov process. *Ann. Probab.* **34** 918–961. MR2243874

[32] FALCONER, K. (1990). *Fractal Geometry. Mathematical Foundations and Applications*. Wiley, Chichester. MR1102677

[33] FRIEZE, A., RUSZINKÓ, M. and THOMA, L. (2000). A note on random minimum length spanning trees. *Electron. J. Combin.* **7** Research Paper 4, 5. MR1774728

[34] FRIEZE, A. M. (1985). On the value of a random minimum spanning tree problem. *Discrete Appl. Math.* **10** 47–56. MR0770868

[35] FRIEZE, A. M. and MCDIARMID, C. J. H. (1989). On random minimum length spanning trees. *Combinatorica* **9** 363–374. MR1054012

[36] GARBAN, C., PETE, G. and SCHRAMM, O. (2013). The scaling limits of the minimal spanning tree and invasion percolation in the plane. Available at arXiv:1309.0269 [math.PR].

[37] GRAHAM, R. L. and HELL, P. (1985). On the history of the minimum spanning tree problem. *IEEE Ann. Hist. Comput.* **7** 43–57. MR0783327

[38] HAAS, B. and MIERMONT, G. (2004). The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.* **9** 57–97 (electronic). MR2041829

[39] JANSON, S. (1995). The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Structures Algorithms* **7** 337–355. MR1369071

[40] JANSON, S. and LUCZAK, M. J. (2008). Susceptibility in subcritical random graphs. *J. Math. Phys.* **49** 125207, 23. MR2484338

[41] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. Wiley-Interscience, New York. MR1782847

[42] JANSON, S. and WÄSTLUND, J. (2006). Addendum to: "The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph" [Random Structures Algorithms **7** (1995), no. 4, 337–355; MR1369071] by Janson. *Random Structures Algorithms* **28** 511–512. MR2225705

[43] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

[44] KESTEN, H. and LEE, S. (1996). The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.* **6** 495–527. MR1398055

[45] KOZMA, G., LOTKER, Z. and STUPP, G. (2006). The minimal spanning tree and the upper box dimension. *Proc. Amer. Math. Soc.* **134** 1183–1187 (electronic). MR2196055

[46] LEE, S. (1997). The central limit theorem for Euclidean minimal spanning trees. I. *Ann. Appl. Probab.* **7** 996–1020. MR1484795

[47] LEE, S. (1999). The central limit theorem for Euclidean minimal spanning trees. II. *Adv. in Appl. Probab.* **31** 969–984. MR1747451

[48] LE GALL, J.-F. (2005). Random trees and applications. *Probab. Surv.* **2** 245–311. MR2203728

[49] LE GALL, J.-F. (2006). Random real trees. *Ann. Fac. Sci. Toulouse Math.* (6) **15** 35–62. MR2225746

[50] ŁUCZAK, T. (1990). Component behavior near the critical point of the random graph process. *Random Structures Algorithms* **1** 287–310. MR1099794

[51] ŁUCZAK, T. (1998). Random trees and random graphs. *Random Structures Algorithms* **13** 485–500. MR1662797

[52] ŁUCZAK, T., PITTEL, B. and WIERMAN, J. C. (1994). The structure of a random graph at the point of the phase transition. *Trans. Amer. Math. Soc.* **341** 721–748. MR1138950

[53] LYONS, R., PERES, Y. and SCHRAMM, O. (2006). Minimal spanning forests. *Ann. Probab.* **34** 1665–1692. MR2271476

[54] MCDIARMID, C., JOHNSON, T. and STONE, H. S. (1997). On finding a minimum spanning tree in a network with random weights. *Random Structures Algorithms* **10** 187–204. MR1611522

[55] MIERMONT, G. (2009). Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér.* (4) **42** 725–781. MR2571957

[56] NACHMIAS, A. and PERES, Y. (2007). Component sizes of the random graph outside the scaling window. *ALEA Lat. Am. J. Probab. Math. Stat.* **3** 133–142. MR2349805

[57] NEWMAN, C. M. (1997). *Topics in Disordered Systems*. Birkhäuser, Basel. MR1480664

[58] PAPADOPOULOS, A. (2005). *Metric Spaces, Convexity and Nonpositive Curvature*. *IRMA Lectures in Mathematics and Theoretical Physics* **6**. European Mathematical Society (EMS), Zürich. MR2132506

[59] PENROSE, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198

[60] PENROSE, M. D. (1996). The random minimal spanning tree in high dimensions. *Ann. Probab.* **24** 1903–1925. MR1415233

[61] PENROSE, M. D. (1998). Extremes for the minimal spanning tree on normally distributed points. *Adv. in Appl. Probab.* **30** 628–639. MR1663521

[62] PENROSE, M. D. (1998). Random minimal spanning tree and percolation on the N-cube. *Random Structures Algorithms* **12** 63–82. MR1637395

[63] PENROSE, M. D. (1999). A strong law for the longest edge of the minimal spanning tree. *Ann. Probab.* **27** 246–260. MR1681114

[64] PENROSE, M. D. and YUKICH, J. E. (2001). Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.* **11** 1005–1041. MR1878288

[65] PENROSE, M. D. and YUKICH, J. E. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13** 277–303. MR1952000

[66] PERES, Y. and REVELLE, D. (2005). Scaling limits of the uniform spanning tree and loop-erased random walk on finite graphs. Available at arXiv:math/0410430 [math.PR].

[67] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357

[68] SCHRIVER, A. (2005). On the history of combinatorial optimization (till 1960). In *Discrete Optimization* (K. Aardal, G. L. Nemhauser and R. Weismantel, eds.). *Handbooks in Operations Research and Management* **12** 1–68. North Holland, Amsterdam.

[69] STEELE, J. M. (1981). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.* **9** 365–376. MR0626571

[70] STEELE, J. M. (1987). On Frieze's $\zeta(3)$ limit for lengths of minimal spanning trees. *Discrete Appl. Math.* **18** 99–103. MR0905183

[71] STEELE, J. M. (1988). Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.* **16** 1767–1787. MR0958215

[72] STEELE, J. M., SHEPP, L. A. and EDDY, W. F. (1987). On the number of leaves of a Euclidean minimal spanning tree. *J. Appl. Probab.* **24** 809–826. MR0913823

[73] STROOCK, D. W. (1993). *Probability Theory, an Analytic View*. Cambridge Univ. Press, Cambridge. MR1267569

[74] TIMÁR, Á. (2006). Ends in free minimal spanning forests. *Ann. Probab.* **34** 865–869. MR2243871

[75] TIMOFEEV, E. A. (1988). On finding the expected length of a random minimal tree. *Theory Probab. Appl.* **33** 361–365.
[76] VAN DER HOFSTAD, R. and NACHMIAS, A. (2017). Hypercube percolation. *J. Eur. Math. Soc.* (*JEMS*) **19** 725–814. MR3612867
[77] VAROPOULOS, N. T. (1985). Long range estimates for Markov chains. *Bull. Sci. Math.* (2) **109** 225–252. MR0822826
[78] VILLANI, C. (2009). *Optimal Transport, Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454
[79] WILSON, D. B. (2004). Red-green-blue model. *Phys. Rev. E* **69** 037105.
[80] YUKICH, J. E. (1996). Ergodic theorems for some classical problems in combinatorial optimization. *Ann. Appl. Probab.* **6** 1006–1023. MR1410126
[81] YUKICH, J. E. (1998). *Probability Theory of Classical Euclidean Optimization Problems*. *Lecture Notes in Math.* **1675**. Springer, Berlin. MR1632875
[82] YUKICH, J. E. (2000). Asymptotics for weighted minimal spanning trees on random points. *Stochastic Process. Appl.* **85** 123–138. MR1730615

L. ADDARIO-BERRY
DEPARTMENT OF MATHEMATICS AND STATISTICS
MCGILL UNIVERSITY
805 SHERBROOKE W.
MONTRÉAL, QUEBEC H3A 2K6
CANADA
E-MAIL: louigi.addario@mcgill.ca

N. BROUTIN
PROJET RAP
INRIA ROCQUENCOURT-PARIS
78153 LE CHESNAY
FRANCE
E-MAIL: nicolas.broutin@inria.fr

C. GOLDSCHMIDT
DEPARTMENT OF STATISTICS
AND LADY MARGARET HALL
UNIVERSITY OF OXFORD
24-29 ST GILES'
OXFORD
OX1 3LB
UNITED KINGDOM
E-MAIL: goldschm@stats.ox.ac.uk

G. MIERMONT
UMPA
ÉCOLE NORMALE SUPÉRIEURE DE LYON
46 ALLÉE D'ITALIE
69364 LYON CEDEX 07
FRANCE
E-MAIL: gregory.miermont@ens-lyon.fr