Bernoulli **23**(4A), 2017, 2380–2433
DOI: 10.3150/16-BEJ813

Cutting down **p**-trees and inhomogeneous continuum random trees

NICOLAS BROUTIN[1] and MINMIN WANG[2]

[1]*Inria Paris–Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay, France.*
E-mail: nicolas.broutin@inria.fr
[2]*PRES Sorbonne Universités, Université Pierre et Marie Curie, LPMA (UMR 7599), 4 place Jussieu, 75005 Paris, France. E-mail: wangminmin03@gmail.com*

We study a fragmentation of the **p**-trees of Camarri and Pitman. We give exact correspondences between the **p**-trees and trees which encode the fragmentation. We then use these results to study the fragmentation of the inhomogeneous continuum random trees (scaling limits of **p**-trees) and give distributional correspondences between the initial tree and the tree encoding the fragmentation. The theorems for the inhomogeneous continuum random tree extend previous results by Bertoin and Miermont about the cut tree of the Brownian continuum random tree.

Keywords: cut tree; inhomogeneous continuum random trees; **p**-tree; random cutting

1. Introduction

The study of random cutting of trees has been initiated by Meir and Moon [45] in the following form: given a rooted (graph theoretic) tree, one can proceed to chop the tree into pieces by iterating the following process: choose a uniform random edge; removing it disconnects the tree into two pieces; discard the part which does not contain the root and keep chopping the portion containing the root until it is reduced to a single node. In the present document, we consider the related version where the vertices are chosen at random and removed (until one is left with an empty tree); each such pick is referred to as a *cut*. We will see that this version is actually much more adapted to the problems we consider here, than the edge cutting procedure presented above.

The main focus in [45] and in most of the subsequential papers has been put on the study of some parameters of this cutting down process, and in particular on how many cuts are necessary for the process to finish. This has been studied for a number of different models of deterministic and random trees such as complete binary trees of a given height, random trees arising from the divide-and-conquer paradigm [25,35–37], and the family trees of finite-variance critical Galton–Watson processes conditioned on the total progeny [31,39,47]. The latter model of random trees turns out to be far more interesting, and it provides an a posteriori motivation for the cutting down process. As we will see shortly, the cutting down process provides an interesting way to investigate some of the structural properties of random trees by partial destruction and recombination, or equivalently as partially resampling the tree.

Let us now be more specific: if L_n denotes the number of cuts required to completely cut down a uniform labelled rooted tree (random Cayley tree, or equivalently conditioned Galton–Watson tree with Poisson offspring distribution) on n nodes, then $n^{-1/2}L_n$ converges in distribution to a

1350-7265

Rayleigh distribution which has density $xe^{-x^2/2}$ on $\mathbb{R}_+$. Janson [39] proved that a similar result holds for any Galton–Watson tree with a finite-variance offspring distribution conditioned on the total progeny to be n. This is the parameter point of view. Addario-Berry *et al.* [3] have shown that for the random Cayley trees, L_n actually has the same distribution as the number of nodes on the path between two uniform random nodes. Their argument relies on the following construction, which amounts to a partial resampling of the Cayley tree. If one considers the sequence of subtrees which are discarded as the cutting process goes on, and adds a path linking their roots, then the resulting tree is a uniform Cayley tree, and the two extremities of the path are independent uniform random nodes. So the properties of the parameter L_n follow from a stronger correspondence between the combinatorial objects themselves.

This strong connection between the discrete objects can be carried to the level of their scaling limit, namely Aldous' Brownian continuum random tree (CRT) [4] (see also Le Gall [40]), which is, roughly speaking, a "tree-like" (random) metric space; see Section 2 for a proper definition. Without being too precise for now, the natural cutting procedure on the Brownian CRT involves a Poisson rain of cuts sampled according to the length measure. (However, not all the cuts contribute to the isolation of the root.) As in the partial resampling of the discrete setting, we glue the sequence of discarded subtrees along an interval, thereby obtaining a new metric space. If the length of the interval is well-chosen (as a function of the cutting process), the tree obtained is distributed like the Brownian CRT and the two ends of the interval are independently random leaves. This identifies in distribution the discarded subtrees from the cutting procedure as the components of the forest that one obtains from a spinal decomposition of the Brownian CRT. The distribution of the latter is provided by the classical path decomposition of a Brownian excursion due to Bismut [17] (see also [26]). Note that a similar identity is proven by Abraham and Delmas [2] for general Lévy trees without using a discrete approximation. A related example is that of the subtree prune and re-graft dynamics of Evans *et al.* [29] (See also [27,30]), which truly resamples the object rather than giving a "recursive" decomposition.

The aim of this paper is two-fold. First, we prove exact identities and give reversible transformations of **p**-trees that are similar to the ones for Cayley trees developed in [3]. The model of **p**-trees generalizes Cayley trees in allowing "weights" on the vertices. In particular, this additional structure of weights introduces some inhomogeneity in the vertex degrees. We then lift these results to the scaling limits, the inhomogeneous continuum random trees (ICRT) of Aldous and Pitman [10], which are closely related to the general additive coalescent [10,13]. Unlike the Brownian CRT or the stable trees (special cases of the Lévy trees [41]), a general ICRT is not self-similar, nor does it enjoy a "branching property" as the Lévy trees do. This lack of "recursivity" ruins the natural approaches such as the one used in Abraham and Delmas [1,2] or the ones in Bertoin and Miermont [15] and Dieuleveut [23], which argue by comparing two fragmentations with the same dislocation measure but different indices of self-similarity [14]. This is one of the reasons why we believe that studying these transformations at the level of the ICRT is interesting. Furthermore, a conjecture of Aldous *et al.* [6], page 185, suggests that the results for ICRTs actually explain the one of Abraham and Delmas [2] for Lévy trees by providing a result "conditional on the degree distribution".

Second, rather than only focusing on the isolation of the root we also consider a cutting procedure which eventually isolates almost all the points: looking at the evolution in time of the masses of the connected components induced by the cuts yields a fragmentation process with an initial

unit mass. When the initial tree is the Brownian CRT, the induced fragmentation process is the one considered in Aldous and Pitman [8]. For a general ICRT, however, it differs from the one in Aldous and Pitman [10], simply because it is non Markovian. Following the idea in the recent works of Bertoin and Miermont [15] and Dieuleveut [23] on Galton–Watson trees, we construct the genealogical tree for the fragmentation that will be called the *cut tree*. We then generalize the results concerning the distribution of the cut tree that Bertoin and Miermont [15] obtained in the case of Brownian CRT. The results of the present document are also used in a companion paper [19], where we prove that the transformation from the Brownian CRT to its genealogical tree is reversible.

Plan of the paper. In the next section, we introduce the necessary notation and relevant background. We then present more formally the discrete (**p**-trees) and continuous (ICRT) models we are considering, and in which sense the inhomogeneous continuum random trees are the scaling limit of **p**-trees. In Section 3 we introduce the cutting down procedures and state our main results. The study of cutting down procedure for **p**-trees is the topic of Section 4. The results are lifted to the level of the scaling limits in Section 5.

2. Notation, models and preliminaries

Before we can present our results, we need some notation. This section may safely be skipped by the impatient reader and referred to later on.

2.1. Aldous–Broder algorithm and p-trees

The model of **p**-trees is well-known in combinatorics (see Pitman [48] and the references therein), and is intimately related to the matrix tree theorem and the Aldous–Broder algorithm for generating random trees [12,18]. It has been applied to the study of general birthday problem by Camarri and Pitman [21], and this is why it is also known under the name *birthday trees*. For more information about the Aldous–Broder algorithm, we refer the reader to Lyons and Peres [44], Section 4.4.

For a general graph G, we denote by $\mathfrak{v}(G)$ and $\mathfrak{e}(G)$ its vertex set and edge set, respectively. Let A be a finite set and $\mathbf{p} = (p_u, u \in A)$ be a probability measure on A such that $\min_{u\in A} p_u > 0$; this ensures that A is indeed the support of $\mathbf{p}$. Let $\mathbb{T}_A$ denote the set of rooted trees labelled with (all the) elements of A (connected acyclic graphs on A, with a distinguished vertex). For $t \in \mathbb{T}_A$, we let $r(t)$ denote its root vertex. For $u, v \in A$, we write $\{u, v\}$ to mean that u and v are adjacent in t. We sometimes write $\langle u, v\rangle$ to mean that $\{u, v\}$ is an edge of t, and that u is on the path between r and v (we think of the edges as pointing towards the root). For a tree $t \in \mathbb{T}_A$ (rooted at r, say) and a node $v \in A$, we let t^v denote the tree re-rooted at v. In this discrete setting of graphs, the natural metric on $t \in \mathbb{T}_A$ is the graph distance and the edges might be thought as having unit length.

We usually abuse notation, but we believe it does not affect the clarity or precision of our statements. For instance, we refer to a node u in the vertex set $\mathfrak{v}(t)$ of a tree t using $u \in t$. Depending on the context, we sometimes write $t \setminus \{u\}$ to denote the graph induced by t on the

vertex set $\mathfrak{v}(t) \setminus \{u\}$ (for a tree t, this is in general a forest). In particular, the edges with endpoint u are not present in the graph $t \setminus \{u\}$. The (in-)degree $C_u(t)$ of a vertex $u \in A$ is the number of edges of the form $\langle u, v\rangle$ with $v \in A$. For a rooted tree t, and a node u of t, we write $\mathrm{Sub}(t, u)$ for the subtree of t rooted at u (above u). For $t \in \mathbb{T}_A$ and $\mathbf{V} \subseteq A$, we write $\mathrm{Span}(t; \mathbf{V})$ for the smallest subtree of t (in the sense of vertex set inclusion) containing $\mathbf{V}$ and the root of $r(t)$. So $\mathrm{Span}(t; \mathbf{V})$ is the subtree induced by t on the set

$$\bigcup_{u \in \mathbf{V}} [\![r(t), u]\!],$$

where $[\![u, v]\!]$ denotes collection of nodes on the (unique) path between u and v in t. In the case where $\mathbf{V} = \{v_1, v_2, \ldots, v_k\}$, we usually write $\mathrm{Span}(t; v_1, \ldots, v_k)$ instead of $\mathrm{Span}(t; \{v_1, \ldots, v_k\})$. We also write

$$\mathrm{Span}^*(t; \mathbf{V}) := \mathrm{Span}(t; \mathbf{V}) \setminus \big\{r(t)\big\}$$

for the graph induced by t on the set $\mathbf{V} \setminus \{r(t)\}$. (This is only used as a convenient way to denote the collection of nodes, and the fact that it is in general not connected has no importance.)

As noticed by Aldous [12] and Broder [18], one can generate random trees on A by extracting a tree from the trace of a random walk on A, where the sequence of steps is given by a sequence of i.i.d. vertices sampled according to $\mathbf{p}$.

Definition (Weighted version of Aldous–Broder algorithm). *Let $\mathbf{Y} = (Y_j, j \geq 0)$ be a sequence of independent variables with common distribution $\mathbf{p}$. Let $\mathcal{T}(\mathbf{Y})$ be the graph rooted at Y_0 with the set of edges*

$$\big\{\langle Y_{j-1}, Y_j\rangle : Y_j \notin \{Y_0, \ldots, Y_{j-1}\}, j \geq 1\big\}. \tag{2.1}$$

The sequence $\mathbf{Y}$ defines a random walk on A, which eventually visits every element of A with probability one, since A is the support of $\mathbf{p}$. So the trace $\{\langle Y_{j-1}, Y_j\rangle : j \geq 1\}$ of the random walk on A is a connected graph on A, rooted at Y_0. Aldous–Broder Algorithm extracts the tree $\mathcal{T}(\mathbf{Y})$ from the trace of the random walk. To see that $\mathcal{T}(\mathbf{Y})$ is a tree, observe that the edge $\langle Y_{j-1}, Y_j\rangle$ is added only if Y_j has never appeared before in the sequence. It follows easily that $\mathcal{T}(\mathbf{Y})$ is a connected graph without cycles, hence a tree on A. Let π denote the distribution of $\mathcal{T}(\mathbf{Y})$.

Lemma 2.1 ([12,18,21,28]). *For $t \in \mathbb{T}_A$, we have*

$$\pi(t) = \pi^{(\mathbf{p})}(t) := \prod_{u \in A} p_u^{C_u(t)}. \tag{2.2}$$

Note that π is indeed a probability distribution on $\mathbb{T}_A$, since by Cayley's multinomial formula [22,49], we have

$$\sum_{t \in \mathbb{T}_A} \pi(t) = \sum_{t \in \mathbb{T}_A} \prod_{u \in A} p_u^{C_u(t)} = \Big(\sum_{u \in A} p_u\Big)^{|A|-1} = 1. \tag{2.3}$$

Definition (p-tree). *A random tree on the set A with the distribution π as specified by (2.2) is called a* **p**-tree *on A.*

When talking of a **p**-tree, we often omit the mention of A as it appears to be the support of **p**. Observe that when **p** is the uniform distribution on $\{1, 2, \dots, n\}$, the **p**-tree is a uniform random rooted tree with n vertices (the Cayley tree of size n). So the results we are about to present generalize the exact distributional results in [3]. However, we believe that the point of view we adopt here is a little cleaner, since it permits to make the transformation *exactly* reversible without any extra anchoring nodes (which prevent any kind of duality at the discrete level).

From now on, we consider $n \geq 1$ and let $[n]$ denote the set $\{1, 2, \dots, n\}$. We write $\mathbb{T}_n$ as a shorthand for $\mathbb{T}_{[n]}$, the set of the rooted trees on $[n]$. Let also $\mathbf{p} = (p_i, 1 \leq i \leq n)$ be a probability measure on $[n]$ satisfying $\min_{i\in[n]} p_i > 0$. For a subset $A \subseteq [n]$ such that $\mathbf{p}(A) > 0$, we let $\mathbf{p}|_A(\cdot) = \mathbf{p}(\cdot \cap A)/\mathbf{p}(A)$ denote **p** conditioned on A, and write $\pi|_A := \pi^{(\mathbf{p}|_A)}$. For a probability distribution μ, we write $X \sim \mu$ to mean that μ is the distribution of the random variable X. Recall the notation t^v for the re-rooting of a tree t at node v. The following claim can be easily verified from (2.2).

Lemma 2.2. *Let T be a* **p**-*tree on $[n]$. If V is an independent vertex of distribution* **p**, *then $T^V \sim \pi$.*

2.2. Measured metric spaces and the Gromov–Prokhorov topology

Standard references on measured metric spaces and Gromov–Prokhorov topology are Gromov [33], Chapter $3\frac{1}{2}$, Greven *et al.* [32]. See also Bertoin and Miermont [15], Section 1.4, for the pointed topology.

If (X, d) is a metric space endowed with the Borel σ-algebra, we denote by $\mathcal{M}_f(X)$ the set of finite measures on X and by $\mathcal{M}_1(X)$ the subset of probability measures on X. If $m \in \mathcal{M}_f(X)$, we denote by $\mathrm{supp}(m)$ the support of m on X, that is the smallest closed set A such that $m(A^c) = 0$. If $f : X \to Y$ is a measurable map between two metric spaces, and if $m \in \mathcal{M}_f(X)$, then the push-forward of m is an element of $\mathcal{M}_f(Y)$, denoted by $f_*m \in \mathcal{M}_f(Y)$, and is defined by $(f_*m)(A) = m(f^{-1}(A))$ for each Borel set A of Y. If $m \in \mathcal{M}_f(X)$ and $A \subseteq X$, we denote by $m\restriction_A$ the restriction of m to A: $m\restriction_A(B) = m(A \cap B)$ for any Borel set B. This should not be confused with the measure conditioned on A, which remains a probability measure and is denoted by $m|_A$.

We say a triple (X, d, μ) is a *measured metric space* (or sometimes a *metric measure space*) if (X, d) is a Polish space (separable and complete) and $\mu \in \mathcal{M}_1(X)$. Two measured metric spaces (X, d, μ) and (X', d', μ') are said to be *weakly isometric* if there exists an isometry ϕ between the supports of μ on X and of μ' on X' such that $\phi_*\mu = \mu'$. This defines an equivalence relation between the measured metric spaces, and we denote by $\mathbb{M}$ the set of equivalence classes. Note that if (X, d, μ) and (X', d', μ') are weakly isometric, the metric spaces (X, d) and (X', d') may not be isometric.

We can define a metric on $\mathbb{M}$ by adapting Prokhorov's distance. Consider a metric space (X, d) and for $\varepsilon > 0$, let $A^\varepsilon := \{x \in X : d(x, A) < \varepsilon\}$ denote the ε-neighborhood of A. Recall that given

two (Borel) probability measures $\mu, \nu \in \mathcal{M}_1(X)$, the *Prokhorov distance* d_P between μ and ν is defined by

$$\mathrm{d}_P(\mu,\nu) := \inf\{\varepsilon > 0 : \mu(A) \le \nu(A^\varepsilon) + \varepsilon \text{ and } \nu(A) \le \mu(A^\varepsilon) + \varepsilon, \text{ for all Borel sets } A\}. \quad (2.4)$$

Note that the definition of the Prokhorov distance (2.4) can be easily extended to a pair of finite (Borel) measures on X. Then, for two measured metric spaces (X,d,μ) and (X',d',μ') the *Gromov–Prokhorov (GP) distance* between them is defined to be

$$\mathrm{d}_{\mathrm{GP}}((X,d,\mu),(X',d',\mu')) = \inf_{\phi,\psi} \mathrm{d}_P(\phi_*\mu, \psi_*\mu'), \quad (2.5)$$

where the infimum is taken over all isometric embeddings ϕ and ψ from $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu')$ into a common metric space Z. It is clear that d_{GP} depends only on the equivalence classes containing (X,d,μ) and (X',d',μ'). Moreover, the Gromov–Prokhorov distance turns $\mathbb{M}$ into a Polish space ([32], Theorem 1).

There is another more convenient characterization of the *GP topology* (the topology induced by d_{GP}) that relies on convergence of distance matrices between random points. Let $\mathcal{X} = (X,d,\mu)$ be a measured metric space and let $(\xi_i, i \ge 1)$ be a sequence of i.i.d. points of common distribution μ. We write $\rho^{\mathcal{X}} = (d(\xi_i,\xi_j), 1 \le i, j < \infty)$ for the *distance matrix* associated with this sequence. One easily verifies that the distribution of $\rho^{\mathcal{X}}$ does not depend on the particular element of an equivalent class of $\mathbb{M}$. Moreover, by Gromov's reconstruction theorem [33], Section $3\frac{1}{2}.5$, the distribution of $\rho^{\mathcal{X}}$ characterizes $\mathcal{X}$ as an element of $\mathbb{M}$. In this work, we are interested in random variables taking values in $\mathbb{M}$. If $\mathcal{X} = (X,d,\mu)$ is such a random variable, we insist on the fact that, by an i.i.d. sequence of μ-distributed points, we mean that we first condition on $\mathcal{X}$ then sample an i.i.d. sequence according to μ. We have the following characterization for the convergence in distribution for random variables in $\mathbb{M}$.

Proposition 2.3 (Corollary 8 of [42]). *If $\mathcal{X}$ is some random element taking values in $\mathbb{M}$ and for each $n \ge 1$, $\mathcal{X}_n$ is a random element taking values in $\mathbb{M}$, then $\mathcal{X}_n$ converges to $\mathcal{X}$ in distribution as $n \to \infty$ if and only if $\rho^{\mathcal{X}_n}$ converges to $\rho^{\mathcal{X}}$ in the sense of finite-dimensional distributions.*

Pointed Gromov–Prokhorov topology. The above characterization by distance matrices turns out to be quite handy when we want to keep track of marked points. Let $k \in \mathbb{N}$. If (X,d,μ) is a measured metric space and $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in X^k$ is a k-tuple, then we say $(X,d,\mu,\mathbf{x})$ is a *k-pointed measured metric space*, or simply a pointed measured metric space. Two pointed metric measure spaces $(X,d,\mu,\mathbf{x})$ and $(X',d',\mu',\mathbf{x}')$ are said to be *weakly isometric* if there exists an isometric bijection

$$\phi : \mathrm{supp}(\mu) \cup \{x_1, x_2, \ldots, x_k\} \to \mathrm{supp}(\mu') \cup \{x_1', x_2', \ldots, x_k'\}$$

such that $\phi_*\mu = \mu'$ and $\phi(x_i) = x_i'$, $1 \le i \le k$, where $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ and $\mathbf{x}' = (x_1', x_2', \ldots, x_k')$. We denote by $\mathbb{M}_k^*$ the space of weak isometry equivalence classes of k-pointed measured metric spaces. Again, we emphasize the fact that the underlying metric spaces (X,d) and (X',d') do not have to be isometric.

A sequence $(X_n, d_n, \mu_n, \mathbf{x}_n)_{n\geq 1}$ of k-pointed measured metric spaces is said to converge to some pointed measured metric space $(X, d, \mu, \mathbf{x})$ in the *k-pointed Gromov–Prokhorov (pGP) topology* if for any $m \geq 1$,

$$\big(d_n\big(\xi^*_{n,i}, \xi^*_{n,j}\big), 1 \leq i, j \leq m\big) \xrightarrow[d]{n\to\infty} \big(d\big(\xi^*_i, \xi^*_j\big), 1 \leq i, j \leq m\big),$$

where for each $n \geq 1$ and $1 \leq i \leq k$, $\xi^*_{n,i} = x_{n,i}$ if $\mathbf{x}_n = (x_{n,1}, x_{n,2}, \ldots, x_{n,k})$ and $(\xi^*_{n,i}, i \geq k+1)$ is a sequence of i.i.d. μ_n-distributed points in X_n. Similarly, $\xi^*_i = x_i$ for $1 \leq i \leq k$ and $(\xi^*_i, i \geq k+1)$ is a sequence of i.i.d. μ-distributed points in X. The k-pointed Gromov–Prokhorov topology is also metrizable: it suffices to modify the definition (2.5) by taking into account the marked points. Then the proof of [32], Theorem 1, can be easily adapted to show that $\mathbb{M}^*_k$ is a Polish space. It is also straightforward to get an analog of Proposition 2.3 for $\mathbb{M}^*_k$.

2.3. Compact metric spaces and the Gromov–Hausdorff metric

We follow Burago *et al.* [20] for the definition of Gromov–Hausdorff metric and Miermont [46] for Gromov–Hausdorff–Prokhorov metric and the pointed metric. See also Evans [27], Evans *et al.* [29], Evans and Winter [30], Villani [50].

Gromov–Hausdorff metric. Let (X, d) be a metric space. Recall the notation A^ε for the ε-neighborhood of $A \subset X$. Two compact subsets A and B are compared using the Hausdorff distance d_H defined by

$$\mathrm{d}_H(A, B) := \inf\big\{\varepsilon > 0 : A \subseteq B^\varepsilon \text{ and } B \subseteq A^\varepsilon\big\}.$$

To compare two compact metric spaces (X, d) and (X', d'), we define their Gromov–Hausdorff (GH) distance d_{GH} by

$$\mathrm{d}_{\mathrm{GH}}\big((X, d), \big(X', d'\big)\big) := \inf_{\phi,\psi} \mathrm{d}_H\big(\phi(X), \psi\big(X'\big)\big),$$

where the infimum ranges over all choices of isometric embeddings ϕ and ψ from X and X' into a common metric space Z. Note that, as opposed to the case of the GP topology, two compact metric spaces that are at GH distance zero are isometric.

Gromov–Hausdorff–Prokhorov metric. If (X, d) and (X', d') are two compact metric spaces and if $\mu \in \mathcal{M}_f(X)$ and $\mu' \in \mathcal{M}_f(X')$, one way to compare simultaneously the metric spaces and the measures is to define

$$\mathrm{d}_{\mathrm{GHP}}\big((X, d, \mu), \big(X', d', \mu'\big)\big) := \inf_{\phi,\psi}\big\{\mathrm{d}_H\big(\phi(X), \psi\big(X'\big)\big) \vee \mathrm{d}_P\big(\phi_*\mu, \psi_*\mu'\big)\big\},$$

where the infimum ranges over all choices of isometric embeddings ϕ and ψ from X and X' into a common metric space Z. If we denote by $\mathbb{M}_c$ the set of equivalence classes of compact measured metric spaces under measure-preserving isometries, then $\mathbb{M}_c$ is Polish when endowed with $\mathrm{d}_{\mathrm{GHP}}$ ([30], Theorem 2.5, [46], Proposition 8).

Pointed Gromov–Hausdorff metric. We fix some $k \in \mathbb{N}$. Given two compact metric spaces (X, d_X) and (Y, d_Y), let $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in X^k$ and $\mathbf{y} = (y_1, y_2, \ldots, y_k) \in Y^k$. Then the pointed Gromov–Hausdorff metric between $(X, d_X, \mathbf{x})$ and $(Y, d_Y, \mathbf{y})$ is defined to be

$$\mathrm{d_{pGH}}\big((X, d_X, \mathbf{x}), (Y, d_Y, \mathbf{y})\big) := \inf_{\phi,\psi}\Big\{\mathrm{d}_H\big(\phi(X), \psi(Y)\big) \vee \max_{1\le i\le k} d_Z\big(\phi(x_i), \psi(y_i)\big)\Big\},$$

where the infimum ranges over all choices of isometric embeddings ϕ and ψ from X and X' into a common metric space (Z, d_Z). Let $\mathbb{M}_c^k$ denote the isometry-equivalence classes of those compact metric spaces with k marked points. It is a Polish space when endowed with $\mathrm{d_{pGH}}$ ([46], Proposition 9).

2.4. Real trees

The notion of real tree (also known as $\mathbb{R}$-tree) existed for quite some time (see, for example, Dress *et al.* [24]) before it was introduced in probability for the study of random trees by Evans *et al.* [29]. More precisely, a metric space (X, d, r) is called a *(rooted) real tree* if $r \in X$ and

- for any two points $x, y \in X$, there exists an isometry ϕ_{xy} from $[0, d(x, y)]$ into X such that $\phi_{xy}(0) = x$ and $\phi_{xy}(d(x, y)) = y$. In this case, the image of ϕ_{xy} is denoted by $[\![x, y]\!]$;
- if $q : [0, 1] \to X$ is a continuous injective map such that $q(0) = x$ and $q(1) = y$, then $q([0, 1]) = [\![x, y]\!]$.

As for discrete trees, when it is clear from context which metric we are talking about, we refer to metric spaces by the sets. For instance, $(\mathcal{T}, d)$ is often referred to as $\mathcal{T}$.

A *measured (rooted) real tree* is a real tree (X, d, r) equipped with a finite (Borel) measure $\mu \in \mathcal{M}_f(X)$. We always assume that the metric space (X, d) is complete and separable. We denote by $\mathbb{T}_w$ the set of the weak isometry equivalence classes of measured rooted real trees, equipped with the pointed Gromov–Prokhorov topology. Also, let $\mathbb{T}_w^c$ be the set of the measure-preserving isometry equivalence classes of those measured rooted real trees (X, d, r, μ) such that (X, d) is compact. We endow $\mathbb{T}_w^c$ with the pointed Gromov–Hausdorff–Prokhorov distance. Then both $\mathbb{T}_w$ and $\mathbb{T}_w^c$ are Polish spaces ([29], Theorem 1, [30], Theorem 2.5). However in our proofs, we do not always distinguish between an equivalence class and the elements it contains.

Let (T, d, r) be a real tree. For $u \in T$, the degree of u in T, denoted by $\deg(u, T)$, is the number of connected components of $T \setminus \{u\}$. We also denote by

$$\mathrm{Lf}(T) = \big\{u \in T : \deg(u, T) = 1\big\} \quad \text{and} \quad \mathrm{Br}(T) = \big\{u \in T : \deg(u, T) \ge 3\big\}$$

the set of *leaves* and the set of *branch points* of T, respectively. The skeleton of T is the complementary set of $\mathrm{Lf}(T)$ in T, denoted by $\mathrm{Sk}(T)$. For two points $u, v \in T$, we denote by $u \wedge v$ the closest common ancestor of u and v, that is, the unique point w of $[\![r, u]\!] \cap [\![r, v]\!]$ such that $d(u, v) = d(u, w) + d(w, v)$.

For a rooted real tree (T, d, r), if $x \in T$ then the *subtree of T above x*, denoted by $\mathrm{Sub}(T, x)$, is defined to be

$$\mathrm{Sub}(T, x) := \big\{u \in T : x \in [\![r, u]\!]\big\}.$$

Spanning subtree. Let (T, d, r) be a real tree and let $\mathbf{x} = (x_1, \ldots, x_k)$ be k points of T for some $k \geq 1$. We denote by $\mathrm{Span}(T; \mathbf{x})$ the smallest connected set of T which contains the root r and $\mathbf{x}$, that is, $\mathrm{Span}(T; \mathbf{x}) = \bigcup_{1 \leq i \leq k} [\![r, x_i]\!]$. We consider $\mathrm{Span}(T; \mathbf{x})$ as a real tree rooted at r and refer to it as a *spanning subtree* or a *reduced tree* of T.

If (T, d, r) is a real tree and there exists some $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in T^k$ for some $k \geq 1$ such that $T = \mathrm{Span}(T; \mathbf{x})$, then the metric aspect of T is rather simple to visualize. More precisely, if we write $x_0 = r$ and let $\rho^{\mathbf{x}} = (d(x_i, x_j), 0 \leq i, j \leq k)$, then $\rho^{\mathbf{x}}$ determines (T, d, r) up to an isometry.

Gluing. If (T_i, d_i), $i = 1, 2$ are two real trees with some distinguished points $x_i \in T_i$, $i = 1, 2$, the result of the *gluing* of T_1 and T_2 at (x_1, x_2) is the metric space (T, δ), where $T = T_1 \cup T_2 \setminus \{x_2\}$, δ coincides with d_1 on $T_1 \times T_1$ and

$$\delta(u, v) = \begin{cases} d_1(u, x_1) + d_2(v, x_2), & \text{if } u \in T_1, v \in T_2 \setminus \{x_2\}; \\ d_2(u, x_2) + d_1(v, x_1), & \text{if } u \in T_2 \setminus \{x_2\}, v \in T_1; \\ d_2(u, v), & \text{if } u, v \in T_2 \setminus \{x_2\}. \end{cases}$$

It is easy to verify that (T, δ) is a real tree obtained by identifying x_1 and x_2 as one point. We use the notation $T_1 \circledast_{x_1 = x_2} T_2 := (T, \delta)$ in the following. Moreover, if T_1 is rooted at some point r, we make the convention that $T_1 \circledast_{x_1 = x_2} T_2$ is also rooted at r.

2.5. Inhomogeneous continuum random trees

The class of inhomogeneous continuum random tree (ICRT) is introduced in [21] and [10]. See also [6,7,9] for more information about the ICRT and related problems.

Let $\boldsymbol{\Theta}$ be the set of sequences $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \ldots) \in \mathbb{R}_+^\infty$ such that $\theta_1 \geq \theta_2 \geq \theta_3 \cdots \geq 0$, $\theta_0 \geq 0$, $\sum_{i \geq 0} \theta_i^2 = 1$, and either $\theta_0 > 0$ or $\sum_{i \geq 1} \theta_i = \infty$.

One way to obtain an ICRT relies on Poisson point processes, and is explained below. This construction, which appears in [21] and [10], is the natural extension of Aldous' stick-breaking construction for the Brownian CRT [4].

Poisson point process construction. For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we define a real tree $\mathcal{T}$ in the following way.

- If $\theta_0 > 0$, let $\mathrm{P}_0 = \{(u_j, v_j), j \geq 1\}$ be a Poisson point process on the first octant $\{(x, y) : 0 \leq y \leq x\}$ of intensity measure $\theta_0^2 \, dx \, dy$, ordered in such a way that $u_1 < u_2 < u_3 < \cdots$.
- For every $i \geq 1$ such that $\theta_i > 0$, let $\mathrm{P}_i = \{\xi_{i,j}, j \geq 1\}$ be a homogeneous Poisson process on $\mathbb{R}_+$ of intensity θ_i under $\mathbb{P}$, such that $\xi_{i,1} < \xi_{i,2} < \xi_{i,3} < \cdots$.

All these Poisson processes are supposed to be mutually independent and defined on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We consider the points of all these processes as marks on the half line $\mathbb{R}_+$, among which we distinguish two kinds: the *cutpoints* and the *joinpoints*. The cutpoints are all those u_j, $j \geq 1$ and all the $\xi_{i,j}$, $j \geq 2$, $i \geq 1$ for which $\theta_i > 0$. For each cutpoint x, we associate a joinpoint x^* as follows: $x^* = v_j$ if $x = u_j$ for some $j \geq 1$ and $x^* = \xi_{i,1}$ if $x = \xi_{i,j}$ for some $i \geq 1$ and $j \geq 2$. One easily verifies that the hypotheses on $\boldsymbol{\theta}$ imply that the set of cutpoints is a.s. finite on each compact set of $\mathbb{R}_+$, while the joinpoints are dense a.s. everywhere. (See, for example, [10] for a proof.) In particular, we can arrange the cutpoints in increasing order

as $0 < \eta_1 < \eta_2 < \eta_3 < \cdots$. This splits $\mathbb{R}_+$ into countably many intervals that we now reassemble into a tree. We write η_k^* for the joinpoint associated to the kth cutpoint η_k. Define R_1 to be the metric space $[0, \eta_1]$ rooted at 0. For $k \geq 1$, let

$$R_{k+1} := R_k \underset{\eta_k^* = \eta_k}{\circledast} [\eta_k, \eta_{k+1}].$$

In words, we graft the intervals $[\eta_k, \eta_{k+1}]$ by gluing the left end at the joinpoint η_k^*. Note that we have $\eta_k^* < \eta_k$ a.s., thus $\eta_k^* \in R_k$ and the above gluing operation is well-defined almost surely. It is not difficult to see that R_k has the shape of a k-leafed discrete (rooted) tree. Moreover, it follows from this Poisson construction that $(R_k)_{k\geq 1}$ a consistent family of trees which also verifies the "leaf-tight" condition in the sense of Aldous [5]. Therefore by [5], Theorem 3, the complete metric space $\mathcal{T} := \overline{\bigcup_{k\geq 1} R_k}$ is a real tree and $\frac{1}{k}\sum_{1\leq i\leq k} \delta_{\eta_i}$, the empirical measure on the leaves of R_k, converges almost surely as $k \to \infty$ to a probability measure μ, called the *mass measure* of $\mathcal{T}$. This measure μ is non atomic and concentrated on the set of leaves of $\mathcal{T}$. Moreover, if conditional on $\mathcal{T}$, $(V_k, k \geq 1)$ is a sequence of i.i.d. points of common law μ, then for each $k \geq 1$, the spanning tree $\mathrm{Span}(T; V_1, V_2, \ldots, V_k)$ has the same distribution as R_k. The distribution of $(\mathcal{T}, \mu)$ is said to be the distribution of an *ICRT of parameter* $\boldsymbol{\theta}$, which is a probability distribution on $\mathbb{T}_w$. The push-forward of the Lebesgue measure on $\mathbb{R}_+$ defines a σ-finite measure ℓ on $\mathcal{T}$, which is concentrated on $\mathrm{Sk}(T)$ and called the *length measure* of $\mathcal{T}$. Furthermore, direct computation shows that the distribution of $\ell(R_1)$ is given by

$$\mathbb{P}\big(\ell(R_1) > r\big) = \mathbb{P}(\eta_1 > r) = e^{-(1/2)\theta_0^2 r^2} \prod_{i\geq 1} (1 + \theta_i r) e^{-\theta_i r}, \qquad r > 0. \tag{2.6}$$

(Note that, since $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the infinite product on the right-hand side converges absolutely.)

In the important special case when $\boldsymbol{\theta} = (1, 0, 0, \ldots)$, the above construction coincides with the line-breaking construction of the Brownian CRT in [4], Algorithm 3, that is, $\mathcal{T}$ is the Brownian CRT. This case will be referred to as the Brownian case in the sequel. We notice that whenever there is an index $i \geq 1$ such that $\theta_i > 0$, the point, denoted by β_i, which corresponds to the joinpoint $\xi_{i,1}$ is a branch point of infinite degree. According to [6], Theorem 2, θ_i is a measurable function of $(\mathcal{T}, \beta_i)$, and we refer to it as the *local time* of β_i in what follows.

ICRTs as scaling limits of **p***-trees*. Let $\mathbf{p}_n = (p_{n1}, p_{n2}, \ldots, p_{nn})$ be a probability measure on $[n]$ such that $p_{n1} \geq p_{n2} \geq \cdots \geq p_{nn} > 0$. Define $\sigma_n \geq 0$ by $\sigma_n^2 = \sum_{i=1}^n p_{ni}^2$ and denote by T^n the corresponding $\mathbf{p}_n$-tree. We also denote by d_{T_n} the graph distance of T_n, that is, the distance between two nodes is the number of edges on the unique path between them. Suppose that the sequence $(\mathbf{p}_n, n \geq 1)$ verifies the following hypothesis: there exists some $\boldsymbol{\theta} = (\theta_i, i \geq 0) \in \boldsymbol{\Theta}$ such that

$$\lim_{n\to\infty} \sigma_n = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{p_{ni}}{\sigma_n} = \theta_i \qquad \text{for every } i \geq 1. \tag{H}$$

Then, writing $\sigma_n T^n$ for the rescaled metric space $([n], \sigma_n d_{T^n})$, Camarri and Pitman [21] have shown that

$$\big(\sigma_n T^n, \mathbf{p}_n\big) \xrightarrow[d,\mathrm{GP}]{n\to\infty} (\mathcal{T}, \mu), \tag{2.7}$$

where $\to_{d,\mathrm{GP}}$ denotes the convergence in distribution with respect to the Gromov–Prokhorov topology.

3. Main results

3.1. Cutting down procedures for p-trees and ICRT

Consider a **p**-tree T on $[n]$. We perform a cutting procedure on T by picking each time a vertex according to **p** conditioned on the remaining part; however, it is more convenient for us to retain the portion of the tree that contains a random node V of distribution **p** rather than the root. Thanks to Lemma 2.2, we know that isolating a random point is equivalent (in distribution) to isolating the root. We denote by $L(T)$ the number of cuts until V is finally picked, and let X_i, $1 \le i \le L(T)$, be the sequence of nodes chosen. The following identity in distribution has been already shown in [3] in the special case of the uniform Cayley tree:

$$L(T) \stackrel{d}{=} \mathrm{Card}\{\text{vertices on the path from the root to } V\}. \tag{3.1}$$

Actually, (3.1) is a consequence of Lemma 4.1 below, which in words can be described as follows. In the cutting procedure described above, we connect the rejected parts, which are subtrees above X_i just before the cutting,[1] by drawing an edge between X_i and X_{i+1}, $i = 1, 2, \ldots, L(T) - 1$ (see Figure 1 in Section 4). We obtain another tree on the same vertex set, which contains a path from the first cut X_1 to the node V. We denote by $\mathrm{cut}(T, V)$ this tree which (partially) encodes the isolating process of V and we let X_1 be its root. We prove in Lemma 4.1 that we have

$$\big(\mathrm{cut}(T, V), V\big) \stackrel{d}{=} (T, V). \tag{3.2}$$

Note that by construction, $L(T)$ equals the number of vertices in the path of $\mathrm{cut}(T, V)$ from the root to V. Thus (3.2) easily entails (3.1). This identity between the pairs of trees contains a lot of information about the distributional structure of the **p**-trees, and our aim is to obtain results similar to (3.2) for ICRTs. The method we use relies on the discrete approximation of ICRT by **p**-trees, and a first step consists in defining the appropriate cutting procedure for ICRT.

In the case of **p**-trees, one may also pick the nodes of T in the order in which they appear in a Poisson random measure. We do not develop it here but one should keep in mind that the cutting procedure may be obtained using a Poisson point process on $\mathbb{R}_+ \times T$ with intensity measure $dt \otimes \mathbf{p}$. In particular, this measure has a natural counterpart in the case of ICRTs, and it is according to this measure that the points should be sampled in the continuous case.

So consider now an ICRT $\mathcal{T}$. Recall that for $\boldsymbol{\theta} \neq (1, 0, \ldots)$, for each $\theta_i > 0$ with $i \ge 1$, there exists a unique point, denoted by β_i, which has infinite degree and local time θ_i. Let $\mathcal{L}$ be the measure on $\mathcal{T}$ defined by

$$\mathcal{L}(dx) := \theta_0^2 \ell(dx) + \sum_{i \ge 1} \theta_i \delta_{\beta_i}(dx), \tag{3.3}$$

[1] V is considered as the new root.

which is almost surely σ-finite (Lemma 5.1). Proving that $\mathcal{L}$ is indeed the relevant cutting measure (in a sense made precise in Proposition 5.2) is the topic of Section 7. Conditional on $\mathcal{T}$, let $\mathcal{P}$ be a Poisson point process on $\mathbb{R}_+ \times \mathcal{T}$ of intensity measure $dt \otimes \mathcal{L}(dx)$ and let V be a μ-distributed point in $\mathcal{T}$. We consider the elements of $\mathcal{P}$ as the successive cuts on $\mathcal{T}$ which try to isolate the random point V. For each $t \geq 0$, define

$$\mathcal{P}_t = \big\{x \in \mathcal{T} : \exists s \leq t \text{ such that } (s, x) \in \mathcal{P}\big\},$$

and let $\mathcal{T}_t = \{u \in \mathcal{T} : [\![u, V]\!] \cap \mathcal{P}_t = \varnothing\}$. Note that $\mathcal{T}_t$ is the connected component of $\mathcal{T} \setminus \mathcal{P}_t$ containing V. Clearly, $\mathcal{T}_{t'} \subset \mathcal{T}_t$ if $t' \geq t$. Let $\mathcal{T}_{t-} := \bigcap_{s<t} \mathcal{T}_s$ for $t > 0$. We set $\mathcal{C} := \{t > 0 : \mu(\mathcal{T}_{t-}) > \mu(\mathcal{T}_t)\}$. Those are the arrival times of the cuts which contribute to the isolation of V. Moreover, with probability one, for each $t \in \mathcal{C}$ there exists a unique $x_t \in \mathcal{T}$ such that $(t, x_t) \in \mathcal{P}$. Observe then that $x_t \in \mathcal{T}_{t-}$.

3.2. Tracking one node and the one-node cut tree

We construct a tree which encodes this cutting process in a similar way that the tree $H = \mathrm{cut}(T, V)$ encodes the cutting procedure for discrete trees. First we construct the "backbone", which is the equivalent of the path we add in the discrete case. For $t \geq 0$, we define

$$L_t := \int_0^t \mu(\mathcal{T}_s)\, ds,$$

and L_∞ the limit as $t \to \infty$ (which might be infinite). Now consider the interval $[0, L_\infty]$, together with its Euclidean metric, that we view as a real tree rooted at 0. For $t \in \mathcal{C}$, recall the notation x_t from above. Then, we graft $\mathcal{T}_{t-} \setminus \mathcal{T}_t$, the portion of the tree discarded at time t, on $[0, L_\infty]$ by identifying x_t with the point L_t (in the sense of the gluing introduced in Section 2.5). This creates a rooted real tree and we denote by $\mathrm{cut}(\mathcal{T}, V)$ its completion. Moreover, we can endow $\mathrm{cut}(\mathcal{T}, V)$ with a (possibly defective probability) measure $\hat{\mu}$ by taking the push-forward of μ under the canonical injection ϕ from $\bigcup_{t\in\mathcal{C}}(\mathcal{T}_{t-} \setminus \mathcal{T}_t)$ to $\mathrm{cut}(\mathcal{T}, V)$. We denote by U the endpoint L_∞ of the interval $[0, L_\infty]$. We show in Section 5 the following theorem.

Theorem 3.1. *We have $L_\infty < \infty$ almost surely. Moreover, under* (H) *we have*

$$\big(\sigma_n \,\mathrm{cut}\big(T^n, V^n\big), \mathbf{p}_n, V^n\big) \xrightarrow[d,\mathrm{pGP}]{n\to\infty} \big(\mathrm{cut}(\mathcal{T}, V), \hat{\mu}, U\big),$$

jointly with the convergence in (2.7).

Combining this with (3.2), we show in Section 5 the following theorem.

Theorem 3.2. *Conditional on $\mathcal{T}$, U has distribution $\hat{\mu}$, and the unconditional distribution of $(\mathrm{cut}(\mathcal{T}, V), \hat{\mu})$ is the same as that of $(\mathcal{T}, \mu)$.*

Theorems 3.1 and 3.2 immediately entail the following.

Corollary 3.3. *Suppose that* (H) *holds. Then*

$$\sigma_n L\big(T^n\big) \xrightarrow[d]{n\to\infty} L_\infty,$$

jointly with the convergence in (2.7). *Moreover, the unconditional distribution of* L_∞ *is the same as that of the distance in* $\mathcal{T}$ *between the root and a* μ*-distributed point* V, *as given in* (2.6).

3.3. The complete cutting procedure

In the procedure of the previous section, the fragmentation only takes place on the portions of the tree which contain the random point V. Following Bertoin and Miermont [15], we consider a more general cutting procedure which keeps splitting all the connected components. The aim here is to describe the genealogy of the fragmentation that this cutting procedure produces. For each $t \geq 0$, $\mathcal{P}_t$ induces an equivalence relation $\sim_t$ on $\mathcal{T}$: for $x, y \in \mathcal{T}$ we write $x \sim_t y$ if $[\![x, y]\!] \cap \mathcal{P}_t = \varnothing$. We denote by $\mathcal{T}_x(t)$ the equivalence class containing x. In particular, we have $\mathcal{T}_V(t) = \mathcal{T}_t$. Let $(V_i)_{i\geq 1}$ be a sequence of i.i.d. μ-distributed points in $\mathcal{T}$. For each $t \geq 0$, define $\mu_i(t) = \mu(\mathcal{T}_{V_i}(t))$. We write $\boldsymbol{\mu}^{\downarrow}(t)$ for the sequence $(\mu_i(t), i \geq 1)$ rearranged in decreasing order. In the case where $\mathcal{T}$ is the Brownian CRT, the process $(\boldsymbol{\mu}^{\downarrow}(t))_{t\geq 0}$ is the fragmentation dual to the standard additive coalescent [10]. In the other cases, however, it is not Markov because of the presence of those branch points β_i with fixed local times θ_i.

As in [15], we can define a genealogical tree for this fragmentation process. For each $i \geq 1$ and $t \geq 0$, let

$$L_t^i := \int_0^t \mu_i(s)\, ds,$$

and let $L_\infty^i \in [0, \infty]$ be the limit as $t \to \infty$. For each pair $(i, j) \in \mathbb{N}^2$, let $\tau(i, j) = \tau(j, i) := \inf\{t > 0 : [\![V_i, V_j]\!] \cap \mathcal{P}_t \neq \varnothing\}$, which is almost surely finite by the properties of $\mathcal{T}$ and $\mathcal{P}$. Note that $L_t^i = L_t^j$ for all $t \leq \tau(i, j)$. Then we can construct a sequence of increasing metric spaces $S_1 \subset S_2 \subset \cdots$ such that for each $k \geq 1$, $S_k = \bigcup_{1\leq i\leq k} [\![\rho_*, U_i]\!]$, and for which we have

$$d(\rho_*, U_i) = L_\infty^i, \qquad d(U_i, U_j) = L_\infty^i + L_\infty^j - 2L_{\tau(i,j)}^i, \qquad 1 \leq i < j \leq k, \tag{3.4}$$

where d denotes the common metric of $(S_k, k \geq 1)$. It is not difficult to see that such a sequence of metric spaces exists and is uniquely determined (in the sense of isometry equivalence classes) by (3.4): intuitively, L_∞^i is the distance between ρ_* and the leaf U_i, and $L_{\tau(i,j)}^i = L_{\tau(i,j)}^j$ is the distance between ρ_* and the point where the two segments $[\![\rho_*, U_i]\!]$ and $[\![\rho_*, U_j]\!]$ split. More precisely, let $S_1 = [\![\rho_*, U_1]\!]$ be a segment of length L_∞^1; then for $k \geq 2$, let $i_k = \arg\max\{\tau(i, k) : 1 \leq i \leq k-1\}$ and b_k be the point of $[\![\rho_*, U_{i_k}]\!] \subset S_{k-1}$ at distance $L_{\tau(i_k,k)}^k = L_{\tau(i_k,k)}^{i_k}$ from ρ_*. Then, S_k is obtained by grafting a segment of length $L_\infty^k - L_{\tau(i_k,k)}^k$ on S_{k-1} by identifying one of its extremities with b_k, and calling U_k the other extremity. It follows that, for every $k \geq 1$, $S_k \setminus S_{k-1} =]\!]b_k, U_k]\!]$ is a segment of length $L_\infty^k - L_{\tau(i_k,k)}^k$. One easily verifies that S_k satisfies (3.4).

We let ρ_* be the common root of all the S_k, $k \geq 1$. Define

$$\mathrm{cut}(\mathcal{T}) := \overline{\bigcup_{k\geq 1} S_k},$$

the completion of the metric space $(\bigcup_k S_k, d)$, which is still a real tree, seen as rooted at ρ_*. In the case where $\mathcal{T}$ is the Brownian CRT, the above definition of $\mathrm{cut}(\mathcal{T})$ coincides with the tree defined by Bertoin and Miermont [15].

Similarly, for each $\mathbf{p}_n$-tree T^n, we can define a complete cutting procedure on T^n by first generating a random permutation $(X_{n1}, X_{n2}, \ldots, X_{nn})$ on the vertex set $[n]$ and then removing the X_{ni}s one after the other. Here the permutation $(X_{n1}, X_{n2}, \ldots, X_{nn})$ is constructed by sampling, for $i \geq 1$, X_{ni} according to $\mathbf{p}_n$ conditioned on $[n] \setminus \{X_{nj}, j < i\}$. We define a new genealogy on $[n]$ by making X_{ni} an ancestor of X_{nj} if $i < j$ and X_{nj} and X_{ni} are in the same connected component when X_{ni} is removed. (See Section 4.3 below for more details.) If we denote by $\mathrm{cut}(T^n)$ the corresponding genealogical tree, then the number of vertices in the path of $\mathrm{cut}(T^n)$ between the root X_{n1} and an arbitrary vertex v is precisely equal to the number of cuts necessary to isolate this vertex v. We have

Theorem 3.4. *Suppose that* (H) *holds. Let* ν *be the weak limit of the empirical measures* $\frac{1}{k}\sum_{i=1}^{k} \delta_{U_i}$, *which exists almost surely conditional on* $\mathcal{T}$. *Then, we have*

$$\big(\sigma_n \,\mathrm{cut}\big(T^n\big), \mathbf{p}_n\big) \xrightarrow[d,\mathrm{GP}]{n\to\infty} \big(\mathrm{cut}(\mathcal{T}), \nu\big),$$

jointly with the convergence in (2.7).

From this, we show the following theorem.

Theorem 3.5. *Conditionally on* $\mathcal{T}$, $(U_i, i \geq 1)$ *has the same distribution as a sequence of i.i.d. points of common law* ν. *Furthermore, the pairs* $(\mathrm{cut}(\mathcal{T}), \nu)$ *and* $(\mathcal{T}, \mu)$ *have the same unconditional distribution.*

In general, the convergence of the $\mathbf{p}_n$-trees to the ICRT in (2.7) cannot be improved to Gromov–Hausdorff (GH) topology, see, for instance, [11], Example 28. However, when the sequence $(\mathbf{p}_n)_{n\geq 1}$ is suitably well-behaved, one does have this stronger convergence. (This is the case for example, when $\mathbf{p}_n$ is the uniform distribution on $[n]$, which gives rise to the Brownian CRT, see also [6], Section 4.2.) In such cases, we can reinforce accordingly the above convergences of the cut trees in the Gromov–Hausdorff topology. Note however that a "reasonable" condition on $\mathbf{p}$ that would ensure the Gromov–Hausdorff convergence seems hard to find. Let us mention a related open question in [6], Section 7, which is to determine a practical criterion for the compactness of a general ICRT. Writing $\to_{d,\mathrm{GHP}}$ for the convergence in distribution with respect to the Gromov–Hausdorff–Prokhorov topology (see Section 2), we have the following.

Theorem 3.6. *Suppose that* $\mathcal{T}$ *is almost surely compact and suppose also as* $n \to \infty$,

$$\big(\sigma_n T^n, \mathbf{p}_n\big) \xrightarrow[d,\mathrm{GHP}]{n\to\infty} (\mathcal{T}, \mu). \tag{3.5}$$

Then, jointly with the convergence in (3.5)*, we have*

$$\big(\sigma_n \operatorname{cut}\big(T^n, V^n\big), \mathbf{p}_n\big) \xrightarrow[d,\mathrm{GHP}]{n\to\infty} \big(\operatorname{cut}(\mathcal{T}, V), \hat{\mu}\big),$$

$$\big(\sigma_n \operatorname{cut}\big(T^n\big), \mathbf{p}_n\big) \xrightarrow[d,\mathrm{GHP}]{n\to\infty} (\operatorname{cut}(\mathcal{T}), \nu).$$

3.4. Reversing the cutting procedure

We also consider the transformation that "reverses" the construction of the trees $\operatorname{cut}(\mathcal{T}, V)$ defined above. Here, by "reversing", we mean that, for an ICRT $(\mathcal{H}, d_{\mathcal{H}}, \hat{\mu})$ and a random point U sampled according to its mass measure $\hat{\mu}$, we should construct a tree $\operatorname{shuff}(\mathcal{H}, U)$ such that

$$\big(\mathcal{T}, \operatorname{cut}(\mathcal{T}, V)\big) \overset{d}{=} \big(\operatorname{shuff}(\mathcal{H}, U), \mathcal{H}\big). \tag{3.6}$$

This reverse transformation is the one described in [3] for the Brownian CRT. For $\mathcal{H}$ rooted at $r(\mathcal{H})$, the path between $[\![r(\mathcal{H}), U]\!]$ decomposes $\mathcal{H}$ into countably many subtrees

$$F_x = \{y \in \mathcal{H} : U \wedge y = x\},$$

where $U \wedge y = \arg\max\{d(u, r(\mathcal{H})) : u \in [\![r(\mathcal{H}), U]\!] \cap [\![r(\mathcal{H}), y]\!]\}$ denotes the closest common ancestor of U and y. Informally, the tree $\operatorname{shuff}(\mathcal{H}, U)$ is the metric space one obtains from $\mathcal{H}$ by attaching each F_x at a random point A_x, which is sampled proportionally to $\hat{\mu}$ in the union of the F_y for which $d_{\mathcal{H}}(U, y) < d_{\mathcal{H}}(U, x)$. We postpone the precise definition of $\operatorname{shuff}(\mathcal{H}, U)$ until Section 6.

The question of reversing the complete cut tree $\operatorname{cut}(\mathcal{T})$ is more delicate and is the subject of the companion paper [19]. There we restrict ourselves to the case of a Brownian CRT: for $\mathcal{T}$ and $\mathcal{G}$ Brownian CRT we construct a tree $\operatorname{shuff}(\mathcal{G})$ such that

$$\big(\mathcal{T}, \operatorname{cut}(\mathcal{T})\big) \overset{d}{=} \big(\operatorname{shuff}(\mathcal{G}), \mathcal{G}\big).$$

We believe that the construction there is also valid for more general ICRTs, but the arguments we use there strongly rely on the self-similarity of the Brownian CRT.

Remarks. i. Theorem 3.2 generalizes Theorem 1.5 in [3], which is about the Brownian CRT. The special case of Theorem 3.1 concerning the convergence of uniform Cayley trees to the Brownian CRT is also found there.

ii. When $\mathcal{T}$ is the Brownian CRT, Theorem 3.5 has been proven by Bertoin and Miermont [15]. Their proof relies on the self-similar property of the Aldous–Pitman's fragmentation. They also proved a convergence similar to the one in Theorem 3.4 for the conditioned Galton–Watson trees with finite-variance offspring distributions. Let us point out that their definition of the discrete cut trees differs from ours, and there is no "duality" at the discrete level for their definitions. Very recently, a result related to Theorem 3.4 has been proved for the case of stable trees [23] (with a different notion of discrete cut tree). Note also that the convergences of the cut trees

proved in [15] and [23] are with respect to the Gromov–Prokhorov topology, so are weaker than the convergences of the corresponding conditioned Galton–Watson trees, which hold in the Gromov–Hausdorff–Prokhorov sense. In our case, the identities imply that the convergence of the cut trees is as strong as that of the $\mathbf{p}_n$-trees (Theorem 3.6).

iii. Abraham and Delmas [2] have shown an analog of Theorem 3.2 for the Lévy tree, introduced in [41]. In passing Aldous *et al.* [6] have conjectured that a Lévy tree is a mixture of ICRTs where the parameters $\boldsymbol{\theta}$ are chosen according to the distribution of the jumps in the bridge process of the associated Lévy process. Then the similarity between Theorem 3.2 and the result of Abraham and Delmas may be seen as a piece of evidence supporting this conjecture.

4. Cutting down and rearranging a p-tree

As we have mentioned in the Introduction, our approach to the theorems about continuum random trees involves taking limits in the discrete world. In this section, we prove the discrete results about the decomposition and the rearrangement of **p**-trees that will enable us to obtain similar decomposition and rearrangement procedures for inhomogeneous continuum random trees.

4.1. Isolating one vertex

As a warm up, and in order to present many of the important ideas, we start by isolating a single node. Let T be a **p**-tree and let V be an independent **p**-distributed node. We isolate the vertex V by removing each time a random vertex of T and preserving only the component containing V until the time when V is picked. Let us recall the notation $\mathfrak{v}(G)$ and $\mathfrak{e}(G)$ for the vertex set and edge set of a graph G.

THE 1-CUTTING PROCEDURE AND THE 1-CUT TREE. Initially, we have $T_0 = T$, and an independent vertex V. Then, for $i \geq 1$, we choose a node X_i according to **p** conditioned on $\mathfrak{v}(T_{i-1})$. We define T_i to be the connected component of the forest induced by T_{i-1} on $\mathfrak{v}(T_{i-1}) \setminus \{X_i\}$ which contains V. If $T_i = \varnothing$, or equivalently $X_i = V$, the process stops and we set $L = L(T) = i$. Since at least one vertex is removed at every step, the process stops in time $L \leq n$.

As we destruct the tree T to isolate V by iteratively pruning random nodes, we construct a tree which records the history of the destruction, that we call the 1-cut tree. This 1-cut tree will, in particular, give some information about the number of cuts which were needed to isolate V. Moreover, we will prove that these two trees are *dual* in a sense that we will make precise shortly.

By construction, $(T_i, 0 \leq i < L)$ is a decreasing sequence of nonempty trees which all contain V, and $(X_i, 1 \leq i \leq L)$ is a sequence of distinct vertices of $T = T_0$. For $1 \leq i \leq L$, we let $F_i = T_{i-1} \setminus T_i$ be the graph induced by T_{i-1} on the vertex set $\mathfrak{v}(T_{i-1}) \setminus \mathfrak{v}(T_i)$. It is not difficult to see that F_i is a tree containing X_i, which we see as the root of F_i. Besides, for each $1 \leq i < L$, $X_i \neq V$ and there is a neighbor U_i of X_i on the path between X_i and V in T_{i-1}. Then $U_i \in T_i$ and we see T_i as rooted at U_i.

When the procedure stops, we have a vector $(F_i, 1 \leq i \leq L)$ of subtrees of T which together span all of $[n]$. We re-arrange them into a new tree by connecting their roots $X_1, X_2, \ldots, X_L$ into

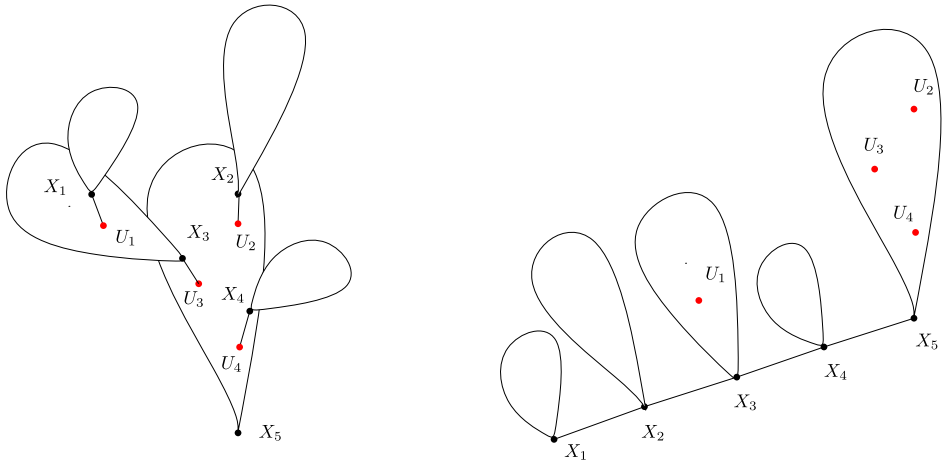

Figure 1. The reorganization of the tree in the one-cutting procedure: on the left the initial tree T, on the right H and the marked nodes $U_1, \dots, U_4$ where to reattach $X_1, \dots, X_4$ in order to recover T.

a path (in this order). We denote by H the resulting tree, calling it the 1-*cut tree*. We also define its root as the node X_1. As we will shortly see, the path $[\![X_1, V]\!]$ plays a special role. Thus we denote it by S and sometimes refer to it as the "backbone" of H.

Note that for $i = 1, \dots, L-1$, we have $U_i \in T_i$. Equivalently, U_i lies in the subtree of H above X_{i+1}. In general, for a tree $t \in \mathbb{T}_n$ and $v \in [n]$, let $x_1, \dots, x_\ell = v$ be the nodes of $\mathrm{Span}(t; v)$. We define $\mathbb{U}(t, v)$ as the collection of nodes $(u_1, \dots, u_{\ell-1})$ such that $u_i \in \mathrm{Sub}(t, x_{i+1})$, for $1 \le i < \ell$. Then by construction, for an $h \in \mathbb{T}_n$, conditional on $H = h$ and $V = v$, we have L equal to the number of the nodes in $\mathrm{Span}(h; v)$ and $(U_1, \dots, U_{L-1}) \in \mathbb{U}(h, v)$ with probability one. For $A \subseteq [n]$, we write $\mathbf{p}(A) := \sum_{i \in A} p_i$. Also recall the measure π from (2.2).

Lemma 4.1. *Let T be a $\mathbf{p}$-tree on $[n]$, and V be an independent $\mathbf{p}$-distributed node. Let $h \in \mathbb{T}_n$, and $v \in [n]$ for which $\mathrm{Span}(h; v)$ is the path made of the nodes $x_1, x_2, \dots, x_{\ell-1}, x_\ell = v$. Let $(u_1, \dots, u_{\ell-1}) \in \mathbb{U}(h, v)$ and $w \in [n]$. Then we have*

$$\begin{aligned}\mathbb{P}\big(H = h; V = v; r(T) = w; U_i = u_i, 1 \le i < \ell\big) \\ = \pi(h) \cdot \prod_{1 \le i < \ell} \frac{p_{u_i}}{\mathbf{p}(\mathrm{Sub}(h, x_{i+1}))} \cdot p_v \cdot p_w.\end{aligned}$$

In particular, $(H, V) \sim \pi \otimes \mathbf{p}$.

As a direct consequence of our construction of H, L is the number of nodes in $\mathrm{Span}(H; V)$, which we write $\#\mathrm{Span}(H; V)$. So Lemma 4.1 entails immediately the following.

Proposition 4.2. *Let T be a* **p***-tree and V be an independent* **p***-distributed node. Then*

$$L \stackrel{d}{=} \#\mathrm{Span}(T; V).$$

Proof of Lemma 4.1. By construction, we have

$$\{H = h; V = v\} \subset \{X_1 = x_1, \ldots, X_{\ell-1} = x_{\ell-1}, X_\ell = v; L = \ell\},$$

and the sequence $(F_i, 1 \le i \le \ell)$ is (f_i), that are obtained when one removes the edges $\{x_i, x_{i+1}\}$, $1 \le i < \ell$ (the edges of the subgraph $\mathrm{Span}(h; v)$). Furthermore, given that $L = \ell$ and $X_i = x_i$, $1 \le i < \ell$, in order to recover the initial tree T it suffices to identify the vertices U_i, $1 \le i < \ell$, for which there used to be an edge $\{X_i, U_i\}$ (which yields the correct adjacencies) and the root of T. Note that U_i is a node of T_i, $1 \le i < \ell$. However, by construction, given that $H = h$ and $V = v$, the set of nodes of T_i is precisely the set of nodes of $\mathrm{Sub}(h, x_{i+1})$.

For $\mathbf{u} = (u_1, \ldots, u_{\ell-1}) \in \mathbb{U}(h, v)$, define $\tau(h, v; \mathbf{u})$ as the tree obtained from h by removing the edges of $\mathrm{Span}(h; v)$, and reconnecting the pieces by adding the edges $\{x_i, u_i\}$. (In particular, the number of edges is unchanged.) We regard $\tau(h, v; \mathbf{u})$ as a tree rooted at $r = x_1$, the root of h. The tree T may be recovered by characterizing T^r, the tree T rerooted at r, and the initial root $r(T)$. We have:

$$\begin{aligned}&\big\{H = h; V = v; r(T) = w; U_i = u_i, 1 \le i < \ell\big\}\\ &\quad = \big\{T^r = \tau(h, v; \mathbf{u}); r(T) = w; X_i = x_i, 1 \le i \le \ell\big\}.\end{aligned}$$

It follows that, for any nodes $u_1, u_2, \ldots, u_{\ell-1}$ as above, we have

$$\begin{aligned}&\mathbb{P}\big(H = h; V = v; r(T) = w; U_i = u_i, 1 \le i < \ell\big)\\ &\quad = \mathbb{P}\big(T = \tau(h, v; \mathbf{u})^w; V = v; X_i = x_i; 1 \le i \le \ell\big)\\ &\quad = \pi\big(\tau(h, v; \mathbf{u})^w\big) \cdot p_v \cdot \prod_{1 \le i \le \ell} \frac{p_{x_i}}{\mathbf{p}(\mathrm{Sub}(h, x_i))}.\end{aligned}$$

Now, by definition, the only nodes that get their (in-)degree modified in the transformation from h to $\tau(h, v; \mathbf{u})$ are u_i, x_{i+1}, $1 \le i < \ell$: every such x_{i+1} gets one less in-edge while u_i gets one more. The re-rooting at w then only modifies the in-degrees of the extremities of the path that is reversed, namely $x_1 = r$ and w. It follows that

$$\pi\big(\tau(h, v; \mathbf{u})^w\big) = \pi(h) \cdot \prod_{1 \le i < \ell} \frac{p_{u_i}}{p_{x_{i+1}}} \cdot \frac{p_w}{p_{x_1}}.$$

Since $\mathbf{p}(\mathrm{Sub}(h, x_1)) = 1$, we have

$$\begin{aligned}&\mathbb{P}\big(H = h; V = v; r(T) = w; U_i = u_i, 1 \le i < \ell\big)\\ &\quad = \pi(h) \cdot \prod_{1 \le i < \ell} \frac{p_{u_i}}{\mathbf{p}(\mathrm{Sub}(h, x_{i+1}))} \cdot p_v \cdot p_w,\end{aligned}$$

which proves the first claim. Summing over all the choices for $\mathbf{u} = (u_1, u_2, \dots, u_{\ell-1}) \in \mathbb{U}(h, v)$, and $w \in [n]$, we obtain

$$\begin{aligned}\mathbb{P}(H = h; V = v) &= \sum_{w\in[n]} \sum_{\mathbf{u}\in\mathbb{U}(h,v)} \pi(h) \cdot \prod_{1\le i<\ell} \frac{p_{u_i}}{\mathbf{p}(\mathrm{Sub}(h, x_{i+1}))} \cdot p_v \cdot p_w \\ &= \pi(h) \cdot p_v \cdot \sum_{\substack{\mathbf{u}=(u_1,\dots,u_{\ell-1}):\\ u_i\in\mathrm{Sub}(h,x_{i+1}),1\le i<\ell}} \frac{p_{u_1}}{\mathbf{p}(\mathrm{Sub}(h, x_2))} \cdots \frac{p_{u_{\ell-1}}}{\mathbf{p}(\mathrm{Sub}(h, x_\ell))} \\ &= \pi(h) \cdot p_v,\end{aligned}$$

which completes the proof. □

THE REVERSE 1-CUTTING PROCEDURE. We have transformed the tree T into the tree H, by somewhat "knitting" a path between the first picked random $\mathbf{p}$-distributed node X_1 and the distinguished node V. This transform is reversible. Indeed, it is possible to "unknit" the path between V and the root of H, and reshuffle the subtrees thereby created in order to obtain a new tree $\tilde{T}$, distributed as T and in which V is an independent $\mathbf{p}$-distributed node. Knowing the (U_i), one could do this exactly, and recover the adjacencies of T (recovering T also requires the information about the root $r(T)$ which has been lost). Defining a reverse transformation reduces to finding the joint distribution of (U_i) and $r(T)$, which is precisely the statement of Lemma 4.1, so that the following reverse construction is now straightforward.

Let $h \in \mathbb{T}_n$, rooted at r and let v be a node in $[n]$. We think of h as $\mathrm{cut}(T, v)$ for some initial tree T. Suppose that $\mathrm{Span}(h; v)$ consists of the vertices $r = x_1, x_2, \dots, x_\ell = v$. Removing the edges of $\mathrm{Span}(h; v)$ from h disconnects it into ℓ connected components which we see as rooted at x_i, $1 \le i \le \ell$. For $w \in \mathrm{Span}^*(h; v) = \mathrm{Span}(h; v) \setminus \{r\}$, sample a node U_w according to the $\mathbf{p}$ conditioned on $\mathrm{Sub}(h, w)$. Let $\mathbf{U} = (U_w, w \in \mathrm{Span}^*(h; v))$ be the obtained vector. Then $\mathbf{U} \in \mathbb{U}(h, v)$. We then define $\mathrm{shuff}(h, v)$ to be the rooted tree which has the adjacencies of $\tau(h, v; \mathbf{U})$, but that is re-rooted at an independent $\mathbf{p}$-distributed node.

It should now be clear that the 1-cutting procedure and the reshuffling operation we have just defined are dual in the following sense.

Proposition 4.3 (1-cutting duality). *Let T be $\mathbf{p}$-tree on $[n]$ and V be an independent $\mathbf{p}$-distributed node. Then,*

$$\big(\mathrm{shuff}(T, V), T, V\big) \stackrel{d}{=} \big(T, \mathrm{cut}(T, V), V\big).$$

In particular, $(\mathrm{shuff}(T, V), V) \sim \pi \otimes \mathbf{p}$.

Note that for the joint distribution in Proposition 4.3, it is necessary to re-root at another independent $\mathbf{p}$-distributed node in order to have the claimed equality. Indeed, T and $\tau(T, V; \mathbf{U})$ share the same root, while T and $\mathrm{cut}(T, V)$ do not (they only have the same root with probability $\sum_{i\ge1} p_i^2 < 1$).

Proof of Proposition 4.3. Let $H = \mathrm{cut}(T, V)$ be the tree resulting from the cutting procedure. Let $L = \#\mathrm{Span}(H; V)$. Recall that U_i denotes a node which used to be a neighbor of X_i in T (see the paragraphs above Lemma 4.1). For $w \in \mathrm{Span}^*(H; V)$, we let $U_w = U_i$ if $w = X_{i+1}$, and let $\mathbf{U}$ be the corresponding vector. Then writing $\hat{r} = r(T)$, with probability one, we have

$$T = \tau(H, V; \mathbf{U})^{\hat{r}}.$$

By Lemma 4.1, $\mathbf{U} \in \mathbb{U}(H, V)$ and conditional on H and V, U_w, $w \in \mathrm{Span}^*(H, V)$ and $\hat{r} = r(T)$ are independent and distributed according to the $\mathbf{p}$ conditioned on $\mathrm{Sub}(H, w)$ and $\mathbf{p}$, respectively. So this coupling indeed gives that $T = \tau(H, V; \mathbf{U})^{\hat{r}}$ is distributed as $\mathrm{shuff}(H, V)$, conditional on H. Since in this coupling $(\mathrm{shuff}(H, V), H, V)$ is equal to (T, H, V) and we have $(H, V) \overset{d}{=} (T, V)$, the proof is complete. □

Remark. Note that the shuffle procedure would allow us to obtain the original tree T exactly if we were to use some information that might be gathered as the cutting procedure goes on. In this discrete case, this is rather clear that one could do this, since the shuffle construction only consists in replacing some edges with others but the vertex set remains the same. This observation will be used in Section 6 to prove a similar statement for the ICRT. There it is much less clear and the result is slightly weaker: it is possible to couple the shuffle in such a way that the tree obtained is measure-isometric to the original one.

4.2. Isolating multiple vertices

We define a cutting procedure analogous to the one described in Section 4.1, but which continues until multiple nodes have been isolated. Again, we let T be a $\mathbf{p}$-tree and, for some $k \geq 1$, let $V_1, V_2, \ldots, V_k$ be k independent vertices chosen according to $\mathbf{p}$ (so not necessarily distinct).

THE k-CUTTING PROCEDURE AND THE k-CUT TREE. We start with $\Gamma_0 = T$. Later on, Γ_i is meant to be the forest induced by T on the nodes that are left. For each time $i \geq 1$, we pick a random vertex X_i according to $\mathbf{p}$ conditioned on $\mathfrak{v}(\Gamma_{i-1})$ and remove it. Then among the connected components of $T \setminus \{X_1, \ldots, X_i\}$, we only keep those containing at least one of $V_1, \ldots, V_k$. We stop at the first time when all k vertices $V_1, \ldots, V_k$ have been chosen, that is at time

$$L^k := \inf\big\{i \geq 1 : \{V_1, \ldots, V_k\} \subseteq \{X_1, \ldots, X_i\}\big\}.$$

For $1 \leq \ell \leq k$ and for $i \geq 0$, we denote by T_i^ℓ the connected component of $T \setminus \{X_1, X_2, \ldots, X_i\}$ containing V_ℓ at time i, or $T_i^\ell = \varnothing$ if $V_\ell \in \{X_1, \ldots, X_i\}$. Then $\Gamma_i := \bigcup_{1 \leq \ell \leq k} T_i^\ell$ is the graph on $[n] \setminus \{X_1, \ldots, X_i\}$ with the edge set $\mathfrak{e}(\Gamma_i) = \bigcup_{1 \leq \ell \leq k} \mathfrak{e}(T_i^\ell)$.

Suppose that at time $i \geq 1$, we have $X_i \in T_{i-1}^\ell$ for some $\ell \in \{1, 2, \ldots, k\}$. If $X_i = V_\ell$, then $T_i^\ell = \varnothing$ and we define $F_i = T_{i-1}^\ell$, re-rooted at $X_i = V_\ell$. Otherwise, let $F_i = \Gamma_{i-1} \setminus \Gamma_i$ be the graph induced by Γ_{i-1} on the vertex set $\mathfrak{v}(\Gamma_{i-1}) \setminus \mathfrak{v}(\Gamma_i)$, rooted at X_i. Moreover, in this case for each $\ell \in [k]$ such that $X_i \in T_{i-1}^\ell$, there exists a neighbor U_i^ℓ of X_i on the path $[\![X_i, V_i]\!]$ in T_{i-1}^ℓ. Then $U_i^\ell \in T_i^\ell$ and we set it to be the root of T_i^ℓ. It is not difficult to see (Figure 2) that F_i is the

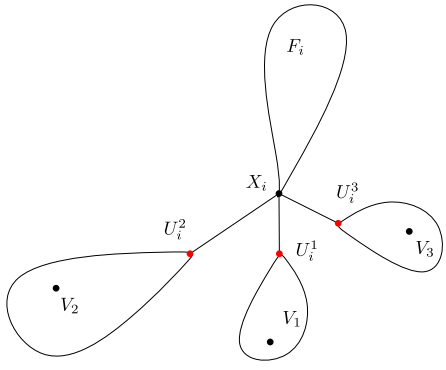

Figure 2. The decomposition of the tree when removing the point X_i from the connected component of Γ_i which contains V_1, V_2 and V_3.

component containing X_i once we have removed the edges $\{X_i, U_i^\ell\}$ from Γ_{i-1} (whenever such an edge exists).

Consider the set of effective cuts which affect the size of T_i^ℓ:

$$\mathcal{E}_\ell^k = \big\{x \in [n] : \text{there exists } i \geq 1, \text{ such that } X_i = x \in T_{i-1}^\ell\big\},$$

and note that $\mathcal{E}_1^k \cup \mathcal{E}_2^k \cup \cdots \cup \mathcal{E}_k^k = \{X_i : 1 \leq i \leq L^k\}$. Let S_k, the *k-cutting skeleton*, be a tree on $\mathcal{E}_1^k \cup \cdots \cup \mathcal{E}_k^k$ that is rooted at X_1, and such that the vertices on the path from X_1 to V_ℓ in S_k are precisely the nodes of $\mathcal{E}_\ell^k$, in the order given by the indices of the cuts. So if we view S_k as a genealogical tree, then in particular, for $1 \leq j, \ell \leq k$, the common ancestors of V_j and V_ℓ are exactly the ones in $\mathcal{E}_j^k \cap \mathcal{E}_\ell^k$. The tree S_k constitutes the *backbone* of a tree on $[n]$ which we now define. For every $x \in S_k$, there is a unique $i = i(x) \geq 1$ such that $x = X_i$. Recall the tree F_i which contains $X_i = x$. We append F_i to S_k at x. Formally, we consider the tree on $[n]$ whose edge set is

$$\mathfrak{e}(S_k) \cup \bigcup_{1 \leq i \leq L^k} \mathfrak{e}(F_i).$$

Furthermore, the tree is considered as rooted at X_1. Then this tree is completely determined by T, $V_1, \ldots, V_k$, and the sequence $\mathbf{X} := (X_i, i = 1, \ldots, L^k)$, and we denote this tree by $\kappa(T; V_1, \ldots, V_k; \mathbf{X})$ when we want to emphasize the dependence in $\mathbf{X}$, or more simply $\mathrm{cut}(T, V_1, \ldots, V_k)$ when no confusion is likely to arise. Clearly, if $H_k = \mathrm{cut}(T, V_1, \ldots, V_k)$, then $S_k = \mathrm{Span}(H_k; V_1, \ldots, V_k)$.

It is convenient to define a *canonical (total) order* $\preceq$ on the vertices of S_k. It will be needed later on in order to define the reverse procedure. For two nodes u, v in S_k, we say that $u \preceq v$ if either $u \in [\![X_1, v]\!]$, or if there exists $\ell \in \{1, \ldots, k\}$ such that $u \in \mathrm{Span}(S_k; V_1, \ldots, V_\ell)$ but $v \notin \mathrm{Span}(S_k; V_1, \ldots, V_\ell)$.

A USEFUL COUPLING. It is useful to see all the trees $\mathrm{cut}(T; V_1, \ldots, V_k)$, $k \geq 1$, on the same probability space. Let V_i, $i \geq 1$, be a sequence of i.i.d. **p**-distributed nodes. Sample another independent i.i.d. sequence Y_i, $i \geq 1$, of common law **p**. For each $k \geq 1$, we define an increasing sequence $\sigma_k = (\sigma_k(j))$ as follows. Let $\sigma_k(1) = 1$. Suppose that we have already defined $X_1^k, \ldots, X_{i-1}^k$. Let Γ_{i-1}^k be the collection of connected components of $T \setminus \{X_1^k, \ldots, X_{i-1}^k\}$ which contain at least one of $V_1, \ldots, V_k$. Let

$$\sigma_k(i) = \inf\bigl\{ j > \sigma_k(i-1) : Y_j \in \Gamma_{i-1}^k \bigr\},$$

and define $X_i^k = Y_{\sigma_k(i)}$. Then, for every k, X_i^k, $1 \leq i \leq L^k = \inf\{i : \Gamma_{i-1}^k = \varnothing\}$, is a sequence of nodes sampled according to **p** conditioned on $\mathfrak{v}(\Gamma_{i-1}^k)$, so that $\mathbf{X}^k := (X_i^k, 1 \leq i \leq L^k)$ can be used for the above k-cutting procedure. Set for $k \geq 1$,

$$H_k := \mathrm{cut}(T, V_1, \ldots, V_k) = \kappa\bigl(T, V_1, \ldots, V_k; \mathbf{X}^k\bigr).$$

By convention let $H_0 = T$ and $\mathrm{Span}(T; \varnothing) = \varnothing$.

Lemma 4.4. *Let* $S_k = \mathrm{Span}(H_k; V_1, \ldots, V_k)$. *Then,* $S_k \subseteq S_{k+1}$ *and*

$$S_k = \mathrm{Span}(S_{k+1}; V_1, \ldots, V_k).$$

Proof. Let T_i^ℓ be the connected component of Γ_i^k which contains V_ℓ. Let $\hat{T}_j^\ell$ be the connected component of $T \setminus \{Y_1, Y_2, \ldots, Y_j\}$ which contains V_ℓ. Then, for $\ell \leq k$, we have

$$\mathcal{E}_\ell^k = \bigl\{x : \exists i \geq 1, x = X_i^k \in T_{i-1}^\ell\bigr\} = \bigl\{y : \exists j \geq 1, y = Y_j \in \hat{T}_{j-1}^\ell\bigr\},$$

so that $\mathcal{E}_\ell^k$ does not depend on k. Then S_k is the tree on $\mathcal{E}_1^k \cup \cdots \cup \mathcal{E}_k^k$ such that the nodes on the path $\mathrm{Span}(S_k; V_\ell)$ are precisely the nodes of $\mathcal{E}_\ell^k$, in the order given by the sequence $\mathbf{X}^k$. It follows that $S_k \subseteq S_{k+1}$ and more precisely that $S_k = \mathrm{Span}(S_{k+1}; V_1, \ldots, V_k)$. □

Remark. The coupling we have just defined justifies an *ordered cutting procedure* which is very similar to the one defined in [3]. Suppose that, for some $j, \ell \in \{1, \ldots, k\}$ we have $x \in \mathcal{E}_j^k \setminus \mathcal{E}_\ell^k$ and $y \in \mathcal{E}_\ell^k \setminus \mathcal{E}_j^k$. Write $(\tilde{X}_i, i \geq 1)$ for the sequence in which we have exchanged the positions of x and y. Then the trees T_i^k, $i \geq \max\{m : X_m = x \text{ or } y\}$ are unaffected if we replace $(X_i, i \geq 1)$ by $(\tilde{X}_i, i \geq 1)$ in the cutting procedure. In particular, if we are only interested in the final tree H_k, we can always suppose that there exist numbers $0 = m_0 < m_1 < m_2 < \cdots < m_k \leq n$ such that, for $1 \leq \ell \leq k$, and if $V_\ell \notin \{V_1, \ldots, V_j\}$, we have

$$\mathcal{E}_\ell^k \setminus \bigcup_{1 \leq j < \ell} \mathcal{E}_j^k = \{X_i : m_{\ell-1} < i \leq m_\ell\}.$$

However, we prefer the coupling over the reordering of the sequence since it does not involve any modification of the distribution of the cutting sequences.

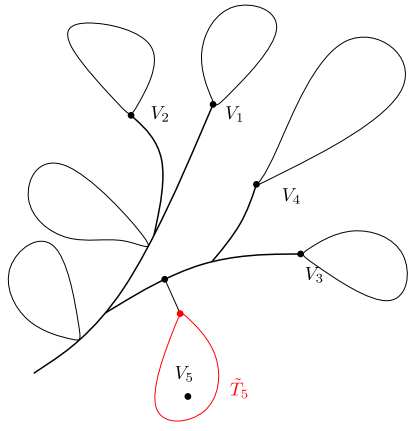

Figure 3. In order to obtain $\mathrm{cut}(T, V_1, \dots, V_k)$ from $\mathrm{cut}(T, V_1, \dots, V_{k-1})$, it suffices to transform the subtree $\tilde{T}_k$ of $\mathrm{cut}(T, V_1, \dots, V_{k-1}) \setminus S_{k-1}$ which contains V_k.

Let $\tilde{T}_k$ be the subtree of $H_{k-1} \setminus \mathrm{Span}(H_{k-1}; V_1, \dots, V_{k-1}) = H_{k-1} \setminus S_{k-1}$ which contains V_k; we agree that $\tilde{T}_k = \varnothing$ if $V_k \in \mathrm{Span}(H_{k-1}; V_1, \dots, V_{k-1})$.

Lemma 4.5. *Let T be a $\mathbf{p}$-tree and let V_k, $k \geq 1$, be a sequence of i.i.d. $\mathbf{p}$-distributed nodes. Then, for each $k \geq 1$, the following statements hold.*

i. *Let $\mathbf{V} \subseteq [n]$ with $\mathbf{V} \neq \varnothing$, then conditional on $V_\ell \in \mathfrak{v}(\tilde{T}_k) = \mathbf{V}$, the pair $(\tilde{T}_k, V_\ell)$ is distributed as $\pi|_{\mathbf{V}} \otimes \mathbf{p}|_{\mathbf{V}}$, and is independent of $(H_{k-1} \setminus \mathbf{V}, V_1, \dots, V_{k-1})$.*

ii. *The joint distribution of $(H_k, V_1, \dots, V_k)$ is given by $\pi \otimes \mathbf{p}^{\otimes k}$.*

Proof. We proceed by induction on $k \geq 1$. Let $\tilde{R}_k$ denote the tree induced by H_k on the vertex set $[n] \setminus \mathfrak{v}(\tilde{T}_k)$. For the base case $k = 1$, the first claim is trivial since $\tilde{T}_1 = T$, and the second is exactly the statement of Lemma 4.1.

Given the two subtrees $\tilde{T}_k$ and $\tilde{R}_k$, it suffices to identify where the tree $\tilde{T}_k$ is grafted on $\tilde{R}_k$ in order to recover the tree H_{k-1}; see Figure 3. By construction, the edge connecting $\tilde{T}_k$ and $\tilde{R}_k$ in H_{k-1} binds the root of $\tilde{T}_k$ to a node of $\mathrm{Span}(\tilde{R}_k; V_1, \dots, V_{k-1})$. Let $t \in \mathbb{T}_{\mathbf{V}}$, $r \in \mathbb{T}_{[n]\setminus\mathbf{V}}$, $v_k \in \mathbf{V}$ and $v_i \in [n] \setminus \mathbf{V}$ for $1 \leq i < k$. Write $\mathbf{v}_{k-1} = \{v_1, \dots, v_{k-1}\}$. For a given node $x \in \mathrm{Span}(r; \mathbf{v}_{k-1})$, let $j_x(r, t)$ (the joint of r and t at x) be the tree obtained from t and r by adding an edge between x and the root of t. By the induction hypothesis, $(H_{k-1}, V_1, \dots, V_{k-1})$ is distributed like a $\mathbf{p}$-tree together with $k-1$ independent $\mathbf{p}$-distributed nodes. Furthermore V_k is independent of $(H_{k-1}, V_1, \dots, V_{k-1})$. It follows that

$$\begin{aligned}
&\mathbb{P}(\tilde{T}_k = t; \tilde{R}_k = r; V_i = v_i, 1 \leq i \leq k) \\
&\quad = \sum_{x \in \mathrm{Span}(r; \mathbf{v}_{k-1})} \mathbb{P}\big(H_{k-1} = j_x(r, t); V_i = v_i, 1 \leq i \leq k\big)
\end{aligned}$$

$$= \sum_{x \in \mathrm{Span}(r;\mathbf{v}_{k-1})} \prod_{i \in \mathbf{V}} p_i^{C_i(t)} \cdot \prod_{j \in [n]\setminus \mathbf{V}} p_j^{C_j(r)} \cdot p_x \cdot \prod_{1 \le i \le k} p_{v_i}$$

$$= \prod_{i \in \mathbf{V}} p_i^{C_i(t)} \cdot p_{v_k} \cdot \prod_{j \in [n]\setminus \mathbf{V}} p_j^{C_j(r)} \cdot \mathbf{p}\big(\mathrm{Span}(r; \mathbf{v}_{k-1})\big) \cdot \prod_{1 \le i < k} p_{v_i}.$$

By summing over t and r and applying Cayley's multinomial formula, we deduce that conditional on $v(\tilde{T}_k) = \mathbf{V} \neq \varnothing$, $(\tilde{T}_k, V_k)$ is independent of $(\tilde{R}_k, V_1, \ldots, V_{k-1})$ and distributed according to $\pi|_{\mathbf{V}} \otimes \mathbf{p}|_{\mathbf{V}}$, which establishes the first claim for k.

Now, conditional on the event $\{V_k \in S_{k-1}\}$, the vertex V_k is distributed according to the $\mathbf{p}$ conditioned on S_{k-1}. In this case, $H_k = H_{k-1}$ so that by the induction hypothesis

$$\text{on } \{V_k \in S_{k-1}\}, \qquad (H_k, V_1, \ldots, V_k) \sim \pi \otimes \mathbf{p}^{k-1} \otimes \mathbf{p}|_{S_{k-1}}. \tag{4.1}$$

On the other hand, if $V_k \notin S_{k-1}$, then $\mathfrak{v}(\tilde{T}_k) \neq \varnothing$ and conditional on $\mathfrak{v}(\tilde{T}_k) = \mathbf{V}$, we have $(\tilde{T}_k, V_k) \sim \pi|_{\mathbf{V}} \otimes \mathbf{p}|_{\mathbf{V}}$. In that case, H_k is obtained from H_{k-1} by replacing $\tilde{T}_k$ by $\mathrm{cut}(\tilde{T}_k, V_k)$. We have already proved that, in this case, $(\tilde{T}_k, V_k)$ is independent of $\tilde{R}_k$, and Lemma 4.1 ensures that the replacement does not alter the distribution. In other words,

$$\text{on } \{V_k \notin S_{k-1}\}, \qquad (H_k, V_1, \ldots, V_k) \sim \pi \otimes \mathbf{p}^{k-1} \otimes \mathbf{p}|_{[n]\setminus S_{k-1}}. \tag{4.2}$$

Since $V_k \sim \mathbf{p}$ is independent of everything else, conditional on S_{k-1}, the event $\{V_k \in S_{k-1}\}$ occurs precisely with probability $\mathbf{p}(S_{k-1})$, so that putting (4.1) and (4.2) together completes the proof of the induction step. □

Corollary 4.6. *Suppose that T is a* **p**-*tree and that $V_1, \ldots, V_k$ are k independent* **p**-*distributed nodes, also independent of T. Then,*

$$S_k \overset{d}{=} \mathrm{Span}(T; V_1, \ldots, V_k).$$

In particular, the total number of cuts needed to isolate $V_1, \ldots, V_k$ in T is distributed as the number of nodes in $\mathrm{Span}(T; V_1, \ldots, V_k)$.

REVERSE k-CUTTING AND DUALITY. As when we were isolating a single node V in Section 4.1, the transformation that yields $H_k = \mathrm{cut}(T, V_1, \ldots, V_k)$ is reversible. To reverse the 1-cutting procedure, we "unknitted" the path between X_1 and V. Similarly, to reverse the k-cutting procedure, we "unknit" the backbone S_k and by doing this obtain a collection of subtrees; then we re-attach these pendant subtrees at random nodes, which are chosen in suitable subtrees in order to obtain a tree distributed like the initial tree T.

For every i, the subtree F_i, rooted at X_i, was initially attached to the set of nodes

$$\mathcal{U}_i := \big\{U_i^j : 1 \le j \le k \text{ such that } T_{i-1}^j \ni X_i\big\}.$$

The corresponding edges have been replaced by some edges which now lie in the backbone S_k; see Figure 4 for an illustration. So, to reverse the cutting procedure knowing the sets $\mathcal{U}_i$, it

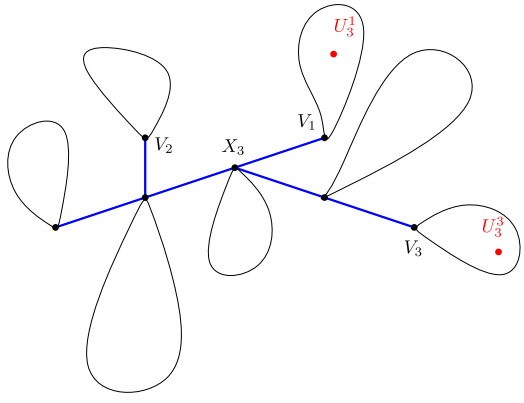

Figure 4. The 3-cut tree and the marked points U_3^1, U_3^3 corresponding to the cut node X_3. The backbone is represented by the subtree in thick blue.

suffices to remove all the edges of S_k, and to re-attach X_i to every node in $\mathcal{U}_i$. In other words, defining a reverse k-cutting transformation knowing only the tree H_k and the distinguished nodes $V_1, \dots, V_k$ reduces to characterizing the distribution of the sets $\mathcal{U}_i$.

Consider a tree $h \in \mathbb{T}_n$, and k nodes $v_1, v_2, \dots, v_k$ not necessarily distinct. Removing the edges of $\mathrm{Span}(h; v_1, \dots, v_k)$ from h disconnects it into connected components f_x, each containing a single vertex x of $\mathrm{Span}(h; v_1, \dots, v_k)$. For a given edge $\langle x, w\rangle$ of $\mathrm{Span}(h; v_1, \dots, v_k)$, let u_w be a node in $\mathrm{Sub}(h, w)$. Let $\mathbf{u}$ be the vector of the u_w, sorted according to the canonical order of w on $\mathrm{Span}(h; v_1, \dots, v_k)$ (see page 2400). For a given tree h and $v_1, \dots, v_k$, we let $\mathbb{U}(h, v_1, \dots, v_k)$ be the set of such vectors $\mathbf{u}$. For $\mathbf{u} \in \mathbb{U}(h, v_1, \dots, v_k)$, define $\tau(h, v_1, \dots, v_k; \mathbf{u})$ as the graph obtained from h by removing every edge $\langle x, w\rangle$ of $\mathrm{Span}(h; v_1, \dots, v_k)$ and replacing it by $\{x, u_w\}$. We regard $\tau(h, v_1, \dots, v_k; \mathbf{u})$ as rooted at the root of h.

Lemma 4.7. *Suppose that $h \in \mathbb{T}_n$, and that $v_1, v_2, \dots, v_k$ are k nodes of $[n]$, not necessarily distinct. Then for every $\mathbf{u} \in \mathbb{U}(h, v_1, \dots, v_k)$, $\tau(h, v_1, \dots, v_k; \mathbf{u})$ is a tree on $[n]$.*

Proof. Write $t := \tau(h, v_1, \dots, v_k; \mathbf{u})$. We proceed by induction on $n \geq 1$. For $n = 1$, $t = h$ is reduced to a single node; so t is a tree.

Suppose now that for any tree t' of size at most $n-1$, any $k \geq 1$, any nodes $v_1, v_2, \dots, v_k \in \mathfrak{v}(t')$, and any $\mathbf{u}' \in \mathbb{U}(t', v_1, \dots, v_k)$, the graph $\tau(t', v_1, \dots, v_k; \mathbf{u}')$ is a tree. Let N be the set of neighbors of the root x_1 of h. For $y \in N$, define $\mathbf{v}_y$ the subset of $\{v_1, \dots, v_k\}$ containing the vertices which lie in $\mathrm{Sub}(h, y)$. If $\mathbf{v}_y \neq \varnothing$, let also $\mathbf{u}_y \in \mathbb{U}(\mathrm{Sub}(h, y), \mathbf{v}_y)$ be obtained from $\mathbf{u}$ by keeping only the vertices u_w for $w \in \mathrm{Span}^*(\mathrm{Sub}(h, y); \mathbf{v}_y)$, still in the canonical order. Then, by construction, the subtrees $\mathrm{Sub}(h, y)$, with $y \in N$ such that $\mathbf{v}_y \neq \varnothing$ are transformed regardless of one another, and the others, for which $\mathbf{v}_y = \varnothing$, are left untouched. So the graph induced by $\tau(h, v_1, \dots, v_k; \mathbf{u})$ on $[n] \setminus \{x_1\}$ consists precisely of $\tau(\mathrm{Sub}(h, y), \mathbf{v}_y; \mathbf{u}_y)$, $y \in N$. By the induction hypothesis, these subgraphs are actually trees. Then $\tau(h, v_1, \dots, v_k; \mathbf{u})$ is simply obtained by adding the node x_1 together with the edges $\{x_1, u_y\}$, for $y \in N$, where $u_y \in \mathrm{Sub}(h, y)$. In other

words, each such edge connects x_1 to a different tree $\tau(\mathrm{Sub}(h, y), \mathbf{v}_y; \mathbf{u}_y)$ so that the resulting graph is also a tree. □

For a given tree h and $v_1, \ldots, v_k \in [n]$ let $\mathbf{U} \in \mathbb{U}(h, v_1, \ldots, v_k)$ be obtained by sampling U_w according to $\mathbf{p}$ conditioned on $\mathrm{Sub}(h, w)$, for every $w \in \mathrm{Span}^*(h; v_1, \ldots, v_k)$. Finally, we define the k-shuffled tree $\mathrm{shuff}(h, v_1, \ldots, v_k)$ to be the tree $\tau(h, v_1, \ldots, v_k; \mathbf{U})$ re-rooted at an independent $\mathbf{p}$-distributed node.

We have the following result, which expresses the fact that the k-cutting and k-shuffling procedures are truly reverses of one another.

Proposition 4.8 (*k*-cutting duality). *Let T be a* $\mathbf{p}$*-tree and let $V_1, \ldots, V_k$ be k independent* $\mathbf{p}$*-distributed nodes, also independent of T. Then, we have the following duality*

$$\big(\mathrm{shuff}(T, V_1, \ldots, V_k), T, V_1, \ldots, V_k\big) \stackrel{d}{=} \big(T, \mathrm{cut}(T, V_1, \ldots, V_k), V_1, \ldots, V_k\big).$$

In particular, $(\mathrm{shuff}(T, V_1, \ldots, V_k), V_1, \ldots, V_k) \sim \pi \otimes \mathbf{p}^{\otimes k}$.

Proof. We consider the coupling we have defined on page 2401: let $H_k = \mathrm{cut}(T, V_1, \ldots, V_k)$ for a $\mathbf{p}$-tree T rooted at $\hat{r} = r(T)$, and for every edge $\langle x, w\rangle$ of $\mathrm{Span}(H_k; V_1, \ldots, V_k)$ we let U_w be the unique node of $\mathrm{Sub}(H_k, w)$ which used to be connected to x in the initial tree T. This defines the vector $\mathbf{U} = (U_w, w \in \mathrm{Span}^*(H_k; V_1, \ldots, V_k))$. We show by induction on $k \geq 1$ that $\tau(H_k, V_1, \ldots, V_k; \mathbf{U})^{\hat{r}} = T$ and that the joint distribution of $(H_k, \hat{r}, V_1, \ldots, V_k, \mathbf{U})$ is that required by the construction above, so that

$$\big(\tau(H_k, V_1, \ldots, V_k; \mathbf{U})^{\hat{r}}, H_k, V_1, \ldots, V_k\big) \stackrel{d}{=} \big(\mathrm{shuff}(H_k, V_1, \ldots, V_k), H_k, V_1, \ldots, V_k\big).$$

Since $(H_k, V_1, \ldots, V_k) \stackrel{d}{=} (T, V_1, \ldots, V_k)$ by Lemma 4.5ii, this would complete the proof.

For $k = 1$, the statement corresponds precisely to the construction of the proof of Proposition 4.3. As before, for $\ell \leq k$, we let $S_\ell = \mathrm{Span}(H_k; V_1, \ldots, V_\ell)$. If $k \geq 2$, let $\tilde{R}_k$ be the connected component of $H_k \setminus S_{k-1}$ which contains V_k, or $\tilde{R}_k = \varnothing$ if $V_k \in S_{k-1}$. In the latter case, $T = \tau(H_k, V_1, \ldots, V_{k-1}, \mathbf{U})^{\hat{r}}$ and the joint distribution of $(H_k, \hat{r}, V_1, \ldots, V_{k-1}, \mathbf{U})$ is correct by the induction hypothesis. Otherwise, let $\mathbf{U}_k$ denote the sub-vector of $\mathbf{U}$ consisting of the components U_w for $w \in \mathrm{Span}^*(\tilde{R}_k, V_k)$, and let $\mathbf{U}_{1,k-1} = (U_w, w \in \mathrm{Span}^*(H_k; V_1, \ldots, V_{k-1}))$. If $\theta \in S_k \setminus S_{k-1}$ is the unique point such that $\tilde{R}_k = \mathrm{Sub}(H_k, \theta)$ (that is, θ is the root of $\tilde{R}_k$), then removing $\tilde{R}_k$ from H_k and replacing it by $\tau(\tilde{R}_k, V_k; \mathbf{U}_k)^{U_\theta}$ yields precisely the tree $H_{k-1} := \mathrm{cut}(T, V_1, \ldots, V_{k-1})$. Also, the distribution of $(\tilde{R}_k, U_\theta, V_k, \mathbf{U}_k)$ is correct, since conditional on the vertex set $\tilde{R}_k$ is distributed as $\pi|_{\mathfrak{v}(\tilde{R}_k)}$ (Lemma 4.5i). Note that this transformation does not modify the distribution of $\mathbf{U}_{1,k-1}$. By the induction hypothesis, $T = \tau(H_{k-1}, V_1, \ldots, V_{k-1}; \mathbf{U}_{1,k-1})^{\hat{r}}$. Since conditionally on $S_{k-1} = \mathrm{Span}(H_k; V_1, \ldots, V_{k-1})$ we have $V_k \in S_{k-1}$ with probability $\mathbf{p}(S_{k-1})$, the proof is complete. □

4.3. The complete cutting and the cut tree

For n a natural number, we may also easily apply the previous procedure until all n nodes have been chosen. In this case, the cutting procedure continues recursively in *all* the connected com-

ponents. The (total) number of cuts is now completely irrelevant (it is a.s. equal to n), and we define the forward transform as follows. Let T be a $\mathbf{p}$-tree and let $(X_i, i \geq 1)$ be a sequence of elements of $[n]$ such that X_i is sampled according to $\mathbf{p}$ conditioned on $[n] \setminus \{X_1, \ldots, X_{i-1}\}$. Let $\Gamma_i = T \setminus \{X_1, \ldots, X_i\}$; we stop precisely at time n, when $\{X_1, \ldots, X_n\} = [n]$ and $\Gamma_n = \varnothing$.

For every $k \in [n]$, define $T_i^{\langle k \rangle}$ as the connected component of Γ_i which contains the vertex k, or $T_i^{\langle k \rangle} = \varnothing$ if $k \in \{X_1, \ldots, X_i\}$. For each $i = 1, \ldots, n$, let $\mathcal{U}_i$ denote the set of neighbors of X_i in Γ_{i-1}. Then we can write $\mathcal{U}_i = \{U_i^{\langle k \rangle} : 1 \leq k \leq n \text{ such that } T_{i-1}^{\langle k \rangle} \ni X_i\}$ where $U_i^{\langle k \rangle}$ is the unique element of $\mathcal{U}_i$ which lies in $T_i^{\langle k \rangle}$. In other words, $U_i^{\langle k \rangle}$ is the neighbor of X_i on the path $[\![X_i, V_k]\!]$. The cuts which affect the connected component containing k are

$$\mathcal{E}_{\langle k \rangle} := \left\{ x \in [n] : \exists i \geq 1, X_i = x \in T_{i-1}^{\langle k \rangle} \right\}.$$

We claim that there exists a tree G such that for every $k \in [n]$, the path $[\![X_1, k]\!]$ in G is precisely made of the nodes in $\mathcal{E}_{\langle k \rangle}$, in the order in which they appear in the sequence $(X_1, X_2, \ldots, X_n)$. In the following, we write $\mathrm{cut}(T) := G$. The following proposition justifies the claim.

Proposition 4.9. *Let T be a $\mathbf{p}$-tree, and let V_k, $k \geq 1$, be i.i.d. $\mathbf{p}$-distributed nodes, independent of T. Then, there exists a random tree $\mathrm{cut}(T)$ on $[n]$ that does not depend on the sequence $(V_k)_{k \geq 1}$, such that, for the coupling of Section 4.2, we have, almost surely as $k \to \infty$,*

$$\mathrm{cut}(T, V_1, \ldots, V_k) \to \mathrm{cut}(T).$$

Proof. Since, for $k \geq 1$, we have $V_1, \ldots, V_k \in S_k$ and $S_k \subseteq S_{k+1}$, the tree S_k converges almost surely to a tree on $[n]$, so that $\lim_{k \to \infty} \mathrm{cut}(T; V_1, \ldots, V_k)$ indeed exists with probability one. In particular, although $\mathrm{cut}(T, V_1, \ldots, V_k)$ certainly depends on $V_1, \ldots, V_k$, the limit only depends on the sequence $(X_i, i \geq 1)$. Indeed, $K := \inf\{k \geq 1 : [n] = \{V_1, \ldots, V_k\}\}$ is a.s. finite, and for every $k \geq K$, one has $\mathrm{cut}(T, V_1, \ldots, V_k) = \mathrm{cut}(T, X_1, \ldots, X_n)$. We then write $\mathrm{cut}(T) := \mathrm{cut}(T, X_1, \ldots, X_n)$. □

Theorem 4.10 (Cut tree). *Let T be a $\mathbf{p}$-tree on $[n]$. Then, we have $\mathrm{cut}(T) \sim \pi$.*

Proof. By Lemma 4.5ii, $H_k = \mathrm{cut}(T, V_1, \ldots, V_k)$ is a π-distributed tree on $[n]$, for every $k \geq 1$. Hence, the claim is straightforward from Proposition 4.9. □

SHUFFLING TREES AND THE REVERSE TRANSFORMATION. Given a tree $g \in \mathbb{T}_n$ that we know is $\mathrm{cut}(t)$ for some tree $t \in \mathbb{T}_n$, and the collections of sets $\mathcal{U}_x$, $x \in [n]$, we cannot recover the initial tree t exactly, for the information about the root has been lost. However, the structure of t as an unrooted tree is easily (in this case, trivially) recovered by connecting every node x to all the nodes in $\mathcal{U}_x$. We now define the reverse operation by sampling the sets $\mathcal{U}_x$ with the correct distribution conditional on g.

Consider a tree $g \in \mathbb{T}_n$, rooted at $r \in [n]$. For each edge $\langle x, w \rangle$ of the tree g, let U_w be a random element sampled according to $\mathbf{p}$ conditioned on $\mathrm{Sub}(g, w)$. Let $\mathbf{U} \in \mathbb{U}(g) := \mathbb{U}(g, 1, 2, \ldots, n)$ be the vector of the U_w, sorted using the canonical order on g with distinguished nodes $1, 2, \ldots, n$.

Let $\tau(g,[n];\mathbf{U})$ denote the graph on $[n]$ whose edges are $\{x,U_w\}$, for which $\langle x,w\rangle$ are edges of g. Then, $\tau(g,[n];\mathbf{U})$ is a tree (Lemma 4.7) and we write $\mathrm{shuff}(g)$ for the random rerooting of $\tau(g,[n];\mathbf{U})$ at an independent **p**-distributed node.

Proposition 4.11. *Let G be a* **p***-tree, and $(V_k,k\geq 1)$ a sequence of i.i.d.* **p***-distributed nodes. Then, there exists a random tree* $\mathrm{shuff}(G)$ *on* $[n]$ *that does not depend on the sequence* $(V_k)_{k\geq 1}$*, such that, for the coupling of Section 4.2, we have, almost surely as* $k\to\infty$*,*

$$\mathrm{shuff}(G,V_1,\ldots,V_k)\to\mathrm{shuff}(G).$$

Proof. For $k\geq 1$, we let $\mathbf{U}_k$ be the subset of $\mathbf{U}$ containing the U_w for which $w\in\mathrm{Span}^*(G;V_1,\ldots,V_k)$, in the canonical order on $\mathrm{Span}^*(G;V_1,\ldots,V_k)$. Then for $k\geq 1$, $\mathbf{U}_k\in\mathbb{U}(G,V_1,\ldots,V_k)$ and since $\mathrm{Span}(G;V_1,\ldots,V_k)$ increases to G, the number of edges of $\tau(G;V_1,\ldots,V_k;\mathbf{U}_k)$ which are constrained by the choices in $\mathbf{U}_k$ increases until the edges are all constrained. It follows that

$$\tau(G,V_1,\ldots,V_k;\mathbf{U}_k)\to\tau(G,1,2,\ldots,n;\mathbf{U})$$

almost surely, as $k\to\infty$. Rerooting all the trees at the same random **p**-distributed node proves the claim. □

We can now state the duality for the complete cutting procedure. It follows readily from the distributional identity in Proposition 4.8

$$\big(T,\mathrm{cut}(T,V_1,\ldots,V_k)\big)\stackrel{d}{=}\big(\mathrm{shuff}(T,V_1,\ldots,V_k),T\big)$$

and the fact that $\mathrm{cut}(T,V_1,\ldots,V_k)\to\mathrm{cut}(T)$ and $\mathrm{shuff}(T,V_1,\ldots,V_k)\to\mathrm{shuff}(T)$ in distribution as $k\to\infty$ (Propositions 4.9 and 4.11).

Proposition 4.12 (Cutting duality). *Let T be a* **p***-tree. Then, we have the following duality in distribution*

$$\big(T,\mathrm{cut}(T)\big)\stackrel{d}{=}\big(\mathrm{shuff}(T),T\big).$$

In particular, $\mathrm{shuff}(T)\sim\pi$.

5. Cutting down an inhomogeneous continuum random tree

Let us recall the notation of Section 2.5. From now on, we fix some $\boldsymbol{\theta}=(\theta_0,\theta_1,\theta_2,\ldots)\in\boldsymbol{\Theta}$. We denote by $I=\{i\geq 1:\theta_i>0\}$ the index set of those θ_i with nonzero values. Let $\mathcal{T}$ be the ICRT of parameter $\boldsymbol{\theta}$ obtained from the Poisson point process construction there. We denote by μ and ℓ its respective mass and length measures. Recall the measure $\mathcal{L}$ defined by

$$\mathcal{L}(dx)=\theta_0^2\ell(dx)+\sum_{i\in I}\theta_i\delta_{\beta_i}(dx),$$

where β_i is the branch point of local time θ_i for $i \in I$. The hypotheses on $\boldsymbol{\theta}$ entail that $\mathcal{L}$ has infinite total mass. On the other hand, we have

Lemma 5.1. *Almost surely, $\mathcal{L}$ is a σ-finite measure concentrated on the skeleton of $\mathcal{T}$. More precisely, if $(V_i, i \geq 1)$ is a sequence of independent points sampled according to μ by first conditioning on $\mathcal{T}$, then for each $k \geq 1$, we have $\mathbb{P}$- almost surely*

$$\mathcal{L}\big(\mathrm{Span}(\mathcal{T}; V_1, V_2, \ldots, V_k)\big) < \infty.$$

Proof. We consider first the case $k = 1$. Recall the Poisson processes $(\mathrm{P}_j, j \geq 0)$ in the Section 2.5 and the notations there. We have seen that $\mathrm{Span}(\mathcal{T}; V_1)$ and R_1 have the same distribution. Then we have

$$\mathcal{L}\big(\mathrm{Span}(\mathcal{T}; V_1)\big) \stackrel{d}{=} \theta_0^2 \eta_1 + \sum_{i \geq 1} \theta_i \delta_{\xi_{i,1}}\big([0, \eta_1]\big).$$

By construction, η_1 is either $\xi_{j,2}$ for some $j \geq 1$ or u_1. This entails that on the event $\{\eta_1 \in \mathrm{P}_j\}$, we have $\eta_1 < \xi_{i,2}$ for all $i \in \mathbb{N} \setminus \{j\}$. Then,

$$\mathbb{E}\Big[\sum_{i \geq 1} \theta_i \delta_{\xi_{i,1}}\big([0, \eta_1]\big)\Big] = \sum_{j \geq 1} \mathbb{E}\Big[\sum_{i \geq 1} \theta_i \cdot \mathbf{1}_{\{\xi_{i,1} \leq \eta_1\}} \mathbf{1}_{\{\eta_1 = \xi_{j,2}\}}\Big] + \mathbb{E}\Big[\sum_{i \geq 1} \theta_i \cdot \mathbf{1}_{\{\xi_{i,1} < \eta_1\}} \mathbf{1}_{\{\eta_1 = u_1\}}\Big].$$

Note that the event $\{\xi_{j,1} \leq \eta_1\} \cap \{\eta_1 = \xi_{j,2}\}$ always occurs. By breaking the first sum on i into $\theta_j + \sum_{i \neq j} \theta_i \mathbf{1}_{\{\xi_{i,1} < \eta_1 < \xi_{i,2}\}}$ and resumming over j, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\sum_{i \geq 1} \theta_i \delta_{\xi_{i,1}}\big([0, \eta_1]\big)\Big] &= \sum_{j \geq 1} \theta_j \mathbb{P}(\eta_1 \in \mathrm{P}_j) + \sum_{j \geq 0} \mathbb{E}\Big[\sum_{i \geq 1, i \neq j} \theta_i \cdot \mathbf{1}_{\{\xi_{i,1} < \eta_1 < \xi_{i,2}\}} \mathbf{1}_{\{\eta_1 \in \mathrm{P}_j\}}\Big] \\
&= \sum_{j \geq 1} \theta_j \mathbb{P}(\eta_1 \in \mathrm{P}_j) + \sum_{j \geq 0} \sum_{i \neq j} \mathbb{E}\big[\theta_i^2 \eta_1 e^{-\theta_i \eta_1} \mathbf{1}_{\{\eta_1 \in \mathrm{P}_j\}}\big] \\
&\leq 1 + \sum_{i \geq 1} \theta_i^2 \cdot \mathbb{E}[\eta_1],
\end{aligned}$$

where we have used the independence of $(\mathrm{P}_j, j \geq 0)$ in the second equality. The distribution of η_1 is given by (2.6). If $\theta_0 > 0$, we have $\mathbb{P}(\eta_1 > r) \leq \exp(-\theta_0^2 r^2/2)$; otherwise, we have $\mathbb{P}(\eta_1 > r) \leq (1 + \theta_1 r) e^{-\theta_1 r}$. In either case, we are able to show that $\mathbb{E}[\eta_1] < \infty$. Therefore,

$$\mathbb{E}\big[\mathcal{L}\big(\mathrm{Span}(\mathcal{T}; V_1)\big)\big] = \theta_0^2 \mathbb{E}[\eta_1] + \mathbb{E}\Big[\sum_{i \geq 1} \theta_i \delta_{\xi_{i,1}}\big([0, \eta_1]\big)\Big] < \infty.$$

In general, the variables $V_1, V_2, \ldots, V_k$ are exchangeable. Therefore

$$\mathbb{E}\big[\mathcal{L}\big(\mathrm{Span}(\mathcal{T}; V_1, V_2, \ldots, V_k)\big)\big] \leq k \mathbb{E}\big[\mathcal{L}\big(\mathrm{Span}(\mathcal{T}; V_1)\big)\big] < \infty,$$

which proves that $\mathcal{L}$ is almost surely finite on the trees spanning finitely many random leaves. Finally, with probability one, $(V_i, i \geq 1)$ is dense in $\mathcal{T}$. Thus $\mathrm{Sk}(\mathcal{T}) = \bigcup_{k\geq 1} [\![r(\mathcal{T}), V_i]\!]$ (see for example [5], Lemma 5). This concludes the proof. □

We recall the Poisson point process $\mathcal{P}$ of intensity measure $dt \otimes \mathcal{L}(dx)$, whose points we have used to define both the one-node-isolation procedure and the complete cutting procedure. As a direct consequence of Lemma 5.1, $\mathcal{P}$ has finitely many atoms on $[0, t] \times \mathrm{Span}(\mathcal{T}; V_1, V_2, \ldots, V_k)$ for all $t > 0$ and $k \geq 1$, almost surely. This fact will be implicitly used in the sequel.

5.1. An overview of the proof

Recall the hypothesis (H) on the sequence of the probability measures $(\mathbf{p}_n, n \geq 1)$:

$$\sigma_n = \left(\sum_{i=1}^{n} p_{ni}^2 \right)^{1/2} \overset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{p_{ni}}{\sigma_n} = \theta_i \qquad \text{for every } i \geq 1. \tag{H}$$

Recall the notation T^n for a $\mathbf{p}_n$-tree, which, from now on, we consider as a measured metric space, equipped with the graph distance and the probability measure $\mathbf{p}_n$. The main results in [21] entail that under hypothesis (H),

$$\left(\sigma_n T^n, \mathbf{p}_n\right) \overset{n\to\infty}{\underset{d,\mathrm{GP}}{\longrightarrow}} (\mathcal{T}, \mu). \tag{5.1}$$

In fact, Camarri and Pitman have shown in [21] a result slightly stronger than (5.1). Let us explain this. For each $n \geq 1$, let $(\xi_k^n)_{k\geq 2}$ be a sequence of i.i.d. $\mathbf{p}_n$-distributed points. Then for $k \geq 2$, write $R_k^n = \mathrm{Span}(T^n; \xi_2^n, \ldots, \xi_k^n)$ for the spanning tree of T^n on the points $\{\xi_2^n, \ldots, \xi_k^n\}$. Similarly, let $(\xi_k)_{k\geq 2}$ be an i.i.d. sequence of points with common law μ conditional on $\mathcal{T}$ and let $R_k = \mathrm{Span}(\mathcal{T}; \xi_2, \ldots, \xi_k)$ for $k \geq 2$. Note that with probability 1, R_k has the shape of a k-leafed tree. Following the terminology in [10] and [21], R_k can be identified with a "(graph) tree with edge-length". More precisely, let $\mathbf{T}_{m,k}$ be the set of trees such that: (1) there are exactly k leaves; (2) there may be extra internal nodes labelled by $\{1, 2, \ldots, m\}$; (3) there may be unlabeled internal nodes; (4) each edge has a strictly positive length. Recall that for $i \geq 1$, β_i denotes the branch point in $\mathcal{T}$ of local time θ_i. By assigning to β_i a label i for $1 \leq i \leq m$, we view R_k as a random variable taking values in $\mathbf{T}_{m,k}$. We treat R_k^n in a similar way: we only retain the labels of those vertices i for $1 \leq i \leq m$. Then by Corollary 15 of [21], for each $k \geq 1$ and $m \geq 0$,

$$\sigma_n R_k^n \overset{n\to\infty}{\underset{d}{\longrightarrow}} R_k \qquad \text{on } \mathbf{T}_{m,k}. \tag{5.2}$$

From this, we easily deduce (5.1) using Proposition 2.3. On the other hand, note that (5.2) also implies that

$$\left(\sigma_n R_k^n, \left(\xi_1^n, \xi_2^n, \ldots, \xi_k^n\right)\right) \overset{n\to\infty}{\underset{d}{\longrightarrow}} \left(R_k, (\xi_1, \xi_2, \ldots, \xi_k)\right), \tag{5.3}$$

with respect to the pointed Gromov–Hausdorff topology. However, even if the trees converge, for the cut trees to converge, one expects that the measures which are used to sample the cuts also converge in a reasonable sense. Recall that the vertex i in T^n has weight p_{ni}. Let

$$\mathcal{L}_n = \sum_{i\in[n]} \frac{p_{ni}}{\sigma_n} \cdot \delta_i = \sigma_n^{-1}\mathbf{p}_n. \tag{5.4}$$

Recall the notation $m\restriction_A$ for the (non-rescaled) restriction of a measure to a subset A. Then (5.2) also allows us to prove the following convergence of the cut-measures.

Proposition 5.2. *Under hypothesis* (H), *we have*

$$\big(\sigma_n R_k^n, \mathcal{L}_n\restriction_{R_k^n}\big) \xrightarrow[d]{n\to\infty} (R_k, \mathcal{L}\restriction_{R_k}) \qquad \forall k \geq 1, \tag{5.5}$$

with respect to the Gromov–Hausdorff–Prokhorov topology.

The proof uses the techniques developed in [6,21] and is postponed until Section 7. We prove in the following subsections that the convergence in Proposition 5.2 is sufficient to entail convergence of the cut trees. To be more precise, we denote by V^n a $\mathbf{p}_n$-distributed node independent of the $\mathbf{p}_n$-tree T^n, and recall that in the construction of $H^n := \mathrm{cut}(T^n, V^n)$, the node V^n ends up at the extremity of the path upon which we graft the discarded subtrees. Recall from the construction of $\mathcal{H} := \mathrm{cut}(\mathcal{T}, V)$ in Section 3 that there is a point U, which is at distance L_∞ from the root. In Section 5.2, we prove Theorem 3.1, that is: if (H) holds, then

$$\big(\sigma_n H^n, \mathbf{p}_n, V^n\big) \xrightarrow[d,\mathrm{pGP}]{n\to\infty} (\mathcal{H}, \hat{\mu}, U), \tag{5.6}$$

jointly with the convergence in (5.5). From there, the proof of Theorem 3.2 is relatively short, and we provide it immediately (taking Theorem 3.1 or equivalently (5.6) for granted).

Proof of Theorem 3.2. Recall that $(\xi_i^n)_{i\geq 2}$ is a sequence of i.i.d. $\mathbf{p}_n$-distributed points. Also let $\xi_1^n = V^n$. Let $(\hat{\xi}_i)_{i\geq 2}$ be a sequence of i.i.d. points of common law $\hat{\mu}$ (by first conditional on $\mathcal{H}$), and let $\hat{\xi}_1 = U$. We let

$$\rho_n = \big(\sigma_n d_{H^n}\big(\xi_i^n, \xi_j^n\big)\big)_{i,j\geq 1} \quad \text{and} \quad \rho_n^* = \big(\sigma_n d_{H^n}\big(\xi_i^n, \xi_j^n\big)\big)_{i,j\geq 2}$$

the distance matrices in $\sigma_n H^n = \sigma_n \,\mathrm{cut}(T^n, V_n)$ associated with the sequences $(\xi_i^n)_{i\geq 1}$ and $(\xi_i^n)_{i\geq 2}$, respectively. According to Lemma 4.1, the distribution of $\xi_1^n = V^n$ is $\mathbf{p}_n$. Therefore, ρ_n is distributed as ρ_n^*. Write similarly

$$\rho = \big(d_{\mathcal{H}}(\hat{\xi}_i, \hat{\xi}_j)\big)_{i,j\geq 1} \quad \text{and} \quad \rho^* = \big(d_{\mathcal{H}}(\hat{\xi}_i, \hat{\xi}_j)\big)_{i,j\geq 2},$$

where $d_{\mathcal{H}}$ denotes the distance of $\mathcal{H} = \mathrm{cut}(\mathcal{T}, V)$. Note that (5.6) entails $\rho_n \to \rho$ in the sense of finite-dimensional distributions. Combined with the previous argument, we deduce that ρ and

ρ^* have the same distribution. However, ρ^* is the distance matrix for an i.i.d. sequence of law $\hat{\mu}$ on $\mathcal{H}$. And the distribution of ρ determines that of U. As a consequence, the law of U is $\hat{\mu}$.

For the unconditional distribution of $(\mathcal{H}, \hat{\mu})$, it suffices to apply the second part of Lemma 4.1, which says that $(H^n, \mathbf{p}_n)$ is distributed like $(T^n, \mathbf{p}_n)$. Then comparing (5.6) with (5.1) shows that the unconditional distribution of $(\mathcal{H}, \hat{\mu})$ is that of $(\mathcal{T}, \mu)$. □

In order to prove Theorem 3.4, the construction of the limit metric space $\mathcal{G} = \mathrm{cut}(\mathcal{T})$ first needs to be justified by resorting to Aldous' theory of continuum random trees [5]. The first step consists in proving that the backbones of $G^n = \mathrm{cut}(T^n)$ converge. For each $n \geq 1$, let $(V_i^n, i \geq 1)$ be a sequence of i.i.d. points of law $\mathbf{p}_n$. Recall that we defined $\mathrm{cut}(\mathcal{T})$ using an increasing family $(S_k)_{k\geq 1}$, defined in (3.4). We show in Section 5.3 the following.

Lemma 5.3. *Suppose that* (H) *holds. Then, we have*

$$\big(\sigma_n d_{G^n}\big(V_i^n, V_j^n\big)\big)_{i,j\geq 1} \xrightarrow[d]{n\to\infty} \big(d_{\mathcal{G}}(U_i, U_j)\big)_{i,j\geq 1} \tag{5.7}$$

in the sense of finite dimensional distributions, jointly with the convergence in (5.1).

Combining this with the identities for the discrete trees in Section 4, we can now prove Theorems 3.4 and 3.5.

Proof of Theorem 3.4. The idea here closely follows the one in [15]. By Theorem 4.10, $(\mathrm{cut}(T^n), \mathbf{p}_n)$ and $(T^n, \mathbf{p}_n)$ have the same distribution for each $n \geq 1$. Recall the notation R_k^n for the spanning tree of T^n. Then it follows from Theorem 4.10 that for each $k \geq 1$,

$$S_k^n := \mathrm{Span}\big(\mathrm{cut}\big(T^n\big); V_1^n, \ldots, V_k^n\big) \overset{d}{=} R_k^n.$$

Now comparing (5.7) with (5.3), we deduce that, for each $k \geq 1$,

$$S_k = \mathrm{Span}\big(\mathrm{cut}(\mathcal{T}); U_1, \ldots, U_k\big) \overset{d}{=} R_k.$$

In particular the family $(S_k)_{k\geq 1}$ is consistent and leaf-tight in the sense of Aldous [5]. This even holds true almost surely conditional on $\mathcal{T}$ and the Poisson point process $\mathcal{P}$. According to Theorem 3 and Lemma 9 of [5], this entails that conditionally on $\mathcal{T}$ and $\mathrm{cut}(\mathcal{T})$, the empirical measure $\frac{1}{k}\sum_{i=1}^k \delta_{U_i}$ converges weakly to some probability measure ν on $\mathrm{cut}(\mathcal{T})$. Moreover, given $\mathcal{T}$ and $\mathrm{cut}(\mathcal{T})$, $(U_i)_{i\geq 1}$ is distributed as an i.i.d. sequence of common law ν.

Now, for the joint convergence, by Lemma 5.3, and using Skorohod's representation theorem, we can find a probability space on which we have almost sure convergence of $(\sigma_n T^n, \mathbf{p}_n)$ in the Gromov–Prokhorov topology, and of the distance matrix $(\sigma_n d_{G_n}(V_i^n, V_j^n))_{1\leq i,j\leq k}$ to $(d_{\mathcal{G}}(U_i, U_j))_{1\leq i,j\leq k}$, for every $k \geq 1$. Since $(V_i^n)_{i\geq 1}$ is a sequence of i.i.d. $\mathbf{p}_n$-distributed points, and by the argument above for the convergence to ν, $(U_i)_{i\geq 1}$ is an i.i.d. sequence of ν-distributed points in $\mathcal{G} = \mathrm{cut}(\mathcal{T})$ conditional on $\mathcal{T}$ and $\mathrm{cut}(\mathcal{T})$, Proposition 2.3 implies that $(G_n, \mathbf{p}_n)$ converges in distribution to $(\mathcal{G}, \nu)$ in the Gromov–Prokhorov topology, which completes the proof. □

Proof of Theorem 3.5. According to Theorem 3 of [5] the distribution of $(\mathrm{cut}(\mathcal{T}), \nu)$ is characterized by the family $(S_k)_{k\geq 1}$. Since S_k and R_k have the same distribution for $k \geq 1$, it follows that $(\mathrm{cut}(\mathcal{T}), \nu)$ is distributed like $(\mathcal{T}, \mu)$. □

5.2. Convergence of the cut-trees $\mathrm{cut}(T^n, V^n)$: Proof of Theorem 3.1

In this part, we prove Theorem 3.1 taking Proposition 5.2 for granted. Let us first reformulate (5.6) in the terms of the distance matrices, which is what we actually show in the following. For each $n \in \mathbb{N}$, recall that $(\xi_i^n, i \geq 2)$ is a sequence of random points of T^n sampled independently according to the mass measure $\mathbf{p}_n$.

We set $\xi_1^n = V^n$ and let ξ_0^n be the root of $H^n = \mathrm{cut}(T^n, V^n)$. Similarly, let $(\xi_i, i \geq 2)$ be a sequence of i.i.d. μ-distributed points and let $\xi_1 = V$. Recall that the mass measure $\hat{\mu}$ of $\mathcal{H} = \mathrm{cut}(\mathcal{T}, V)$ is defined to be the push-forward of μ by the canonical injection ϕ. We set $\widehat{\xi}_i = \phi(\xi_i)$ for $i \geq 2$, $\widehat{\xi}_1 = U$ and $\widehat{\xi}_0$ to be the root of $\mathcal{H}$.

Then the convergence in (5.6) is equivalent to the following:

$$\big(\sigma_n d_{H^n}(\xi_i^n, \xi_j^n), 0 \leq i < j < \infty\big) \xrightarrow[d]{n\to\infty} \big(d_{\mathcal{H}}(\widehat{\xi}_i, \widehat{\xi}_j), 0 \leq i < j < \infty\big), \tag{5.8}$$

jointly with

$$\big(\sigma_n d_{T^n}(\xi_i^n, \xi_j^n), 2 \leq i < j < \infty\big) \xrightarrow[d]{n\to\infty} \big(d_{\mathcal{T}}(\xi_i, \xi_j), 2 \leq i < j < \infty\big), \tag{5.9}$$

in the sense of finite-dimensional distributions. Notice that (5.9) is a direct consequence of (5.1). In order to express the terms in (5.8) with functionals of the cutting process, we introduce the following notations. For $n \in \mathbb{N}$, let $\mathcal{P}_n$ be a Poisson point process on $\mathbb{R}_+ \times T^n$ with intensity measure $dt \otimes \mathcal{L}_n$, where $\mathcal{L}_n = \mathbf{p}_n/\sigma_n$. For $u, v \in T^n$, recall that $[\![u, v]\!]$ denotes the collection of nodes on the path between u and v. For $t \geq 0$, we denote by T_t^n the set of nodes still connected to V^n at time t:

$$T_t^n := \big\{x \in \mathfrak{v}\big(T^n\big) : [0, t] \times \big[\!\big[V^n, x\big]\!\big] \cap \mathcal{P}_n = \varnothing\big\}.$$

Recall that the subtree $\mathcal{T}_t = \{x \in \mathcal{T} : [\![V, x]\!] \cap \mathcal{P}_t = \varnothing\}$ is the remaining part of $\mathcal{T}$ at time t. We then define

$$L_t^n := \mathrm{Card}\big\{s \leq t : \mathbf{p}_n\big(T_s^n\big) < \mathbf{p}_n\big(T_{s-}^n\big)\big\} = \mathrm{Card}\big\{(s, x) \in \mathcal{P}_n : s \leq t, x \in T_{s-}^n\big\}. \tag{5.10}$$

This is the number of cuts that affect the connected component containing V^n before time t. In particular, $L_\infty^n := \lim_{t\to\infty} L_t^n$ has the same distribution as $L(T^n)$ in the notation of Section 4. Indeed, this follows from the coupling on page 2401 and the fact that if $\mathcal{P}_n = \{(t_i, x_i) : i \geq 1\}$ such that $t_1 \leq t_2 \leq \cdots$ then (x_i) is an i.i.d. $\mathbf{p}_n$-sequence. Let us recall that L_t, the continuous analogue of L_t^n, is defined by $L_t = \int_0^t \mu(\mathcal{T}_s)\, ds$ in Section 3. For $n \in \mathbb{N}$ and $x \in T^n$, we define the pair $(\tau_n(x), \varsigma_n(x))$ to be the element of $\mathcal{P}_n$ separating x from V^n, that is,

$$\tau_n(x) := \inf\big\{t > 0 : [0, t] \times \big[\!\big[V^n, x\big]\!\big] \cap \mathcal{P}_n \neq \varnothing\big\},$$

with the convention that $\inf \varnothing = \infty$. In words, $\varsigma_n(x)$ is the first cut that appeared on $[\![V^n, x]\!]$. For $x \in \mathcal{T}$, $(\tau(x), \varsigma(x))$ is defined similarly. We notice that almost surely $\tau(\xi_j) < \infty$ for each $j \geq 2$, since $\tau(\xi_j)$ is an exponential variable with rate $\mathcal{L}([\![V, \xi_j]\!])$, which is positive almost surely. Furthermore, it follows from our construction of $H^n = \mathrm{cut}(T^n, V^n)$ that for $n \in \mathbb{N}$ and $i, j \geq 2$,

$$\begin{aligned}
d_{H^n}\big(\xi_0^n, \xi_1^n\big) &= L_\infty^n - 1,\\
d_{H^n}\big(\xi_0^n, \xi_j^n\big) &= L^n_{\tau_n(\xi_j^n)} - 1 + d_{T^n}\big(\xi_j^n, \varsigma_n\big(\xi_j^n\big)\big),\\
d_{H^n}\big(\xi_1^n, \xi_j^n\big) &= L_\infty^n - L^n_{\tau_n(\xi_j^n)} + d_{T^n}\big(\xi_j^n, \varsigma_n\big(\xi_j^n\big)\big),
\end{aligned}$$

while for $\mathcal{H}$ and $i, j \geq 2$,

$$\begin{aligned}
d_{\mathcal{H}}(\widehat{\xi}_0, \widehat{\xi}_1) &= L_\infty,\\
d_{\mathcal{H}}(\widehat{\xi}_0, \widehat{\xi}_j) &= L_{\tau(\xi_j)} + d_{\mathcal{T}}\big(\xi_j, \varsigma(\xi_j)\big),\\
d_{\mathcal{H}}(\widehat{\xi}_1, \widehat{\xi}_j) &= L_\infty - L_{\tau(\xi_j)} + d_{\mathcal{T}}\big(\xi_j, \varsigma(\xi_j)\big).
\end{aligned}$$

For $n \in \mathbb{N}$ and $i, j \geq 2$, if we define the event

$$\mathcal{A}_n(i, j) := \big\{\tau_n\big(\xi_i^n\big) = \tau_n\big(\xi_j^n\big)\big\} \overset{\text{a.s.}}{=} \big\{\varsigma_n\big(\xi_i^n\big) = \varsigma_n\big(\xi_j^n\big)\big\}, \tag{5.11}$$

and $\mathcal{A}_n^c(i, j)$ its complement, then on the event $\mathcal{A}_n(i, j)$, we have $d_{H^n}(\xi_i^n, \xi_j^n) = d_{T^n}(\xi_i^n, \xi_j^n)$. Similarly, we define $\mathcal{A}(i, j) := \{\tau(\xi_i) = \tau(\xi_j)\}$, and note that $\mathcal{A}(i, j) = \{\varsigma(\xi_i) = \varsigma(\xi_j)\}$ almost surely. Recall that (5.1) implies that $\sigma_n d_{T^n}(\xi_i^n, \xi_j^n) \to d_{\mathcal{T}}(\xi_i, \xi_j)$. Now, on the event $\mathcal{A}_n^c(i, j)$, we have

$$d_{H^n}\big(\xi_i^n, \xi_j^n\big) = \big|L^n_{\tau_n(\xi_j^n)} - L^n_{\tau_n(\xi_i^n)}\big| + d_{T^n}\big(\xi_j^n, \varsigma_n\big(\xi_j^n\big)\big) + d_{T^n}\big(\xi_i^n, \varsigma_n\big(\xi_i^n\big)\big),$$

if $n \in \mathbb{N}$, and

$$d_{\mathcal{H}}(\widehat{\xi}_i, \widehat{\xi}_j) = \big|L_{\tau(\xi_j)} - L_{\tau(\xi_i)}\big| + d_{\mathcal{T}}\big(\xi_j, \varsigma(\xi_j)\big) + d_{\mathcal{T}}\big(\xi_i, \varsigma(\xi_i)\big),$$

for the limit case. Therefore in order to prove (5.8), it suffices to show the joint convergence of the vector

$$\big(\mathbf{1}_{\mathcal{A}_n(i,j)}, \tau_n\big(\xi_i^n\big), \sigma_n d_{T^n}\big(\xi_j^n, \varsigma_n\big(\xi_j^n\big)\big), \big(\sigma_n L_t^n, t \in \mathbb{R}_+ \cup \{\infty\}\big)\big)$$

to the corresponding quantities for $\mathcal{T}$, for each $i, j \geq 2$. We begin with a lemma.

Lemma 5.4. *Under* (H), *we have the following joint convergences as* $n \to \infty$:

$$\big(\mathbf{p}_n\big(T_t^n\big)\big)_{t\geq 0} \xrightarrow{d} \big(\mu(\mathcal{T}_t)\big)_{t\geq 0}, \tag{5.12}$$

in Skorokhod J_1*-topology, along with*

$$(\mathbf{1}_{\mathcal{A}_n(i,j)}, 2 \leq i, j \leq k) \xrightarrow{d} (\mathbf{1}_{\mathcal{A}(i,j)}, 2 \leq i, j \leq k), \tag{5.13}$$

$$\big(\tau_n(\xi_j^n), 2 \le j \le k\big) \xrightarrow{d} \big(\tau(\xi_j), 2 \le j \le k\big) \quad \textit{and} \tag{5.14}$$

$$\big(\sigma_n d_{T^n}(\xi_j^n, \varsigma_n(\xi_j^n)), 2 \le j \le k\big) \xrightarrow{d} \big(d_{\mathcal{T}}(\xi_j, \varsigma(\xi_j)), 2 \le j \le k\big), \tag{5.15}$$

for each $k \ge 2$, and jointly with the convergence in (5.1).

Proof. Recall Proposition 5.2, which says that, for each $k \ge 2$,

$$\big(\sigma_n R_k^n, \mathcal{L}_n \restriction_{R_k^n}\big) \overset{n\to\infty}{\underset{d}{\longrightarrow}} (R_k, \mathcal{L} \restriction_{R_k}),$$

in Gromov–Hausdorff–Prokhorov topology. For $t \ge 0$, let $\{x_1^n, x_2^n, \dots, x_{m_n}^n\}$ be the marks of $\mathcal{P}_n$ on R_k^n before time t, sorted in increasing order of their arrival times. Accordingly, let $\{x_1, x_2, \dots, x_{m_\infty}\}$ be the restriction of $\mathcal{P}_t$ on R_k. Then it follows from Proposition 5.2 and the properties of Poisson point processes that

$$\big(\sigma_n R_k^n, \{x_1^n, x_2^n, \dots, x_{m_n}^n\}\big) \xrightarrow{d} \big(R_k, \{x_1, x_2, \dots, x_{m_\infty}\}\big), \tag{5.16}$$

in the pointed Gromov–Hausdorff topology, jointly with the convergence in (5.5). In particular, the convergence of (5.16) implies that $m_n = m_\infty$ for n large enough. For each $n \in \mathbb{N}$, the pair $(\tau_n(\xi_i^n), \varsigma_n(\xi_i^n))$ corresponds to the first jump of the point process $\mathcal{P}_n$ restricted to $[\![V_1^n, \xi_i^n]\!]$. We notice that for each pair (i, j) such that $2 \le i, j \le k$, the event $\mathcal{A}_n(i, j)$ occurs if and only if $\tau_n(\xi_i^n \wedge \xi_j^n) \le \min\{\tau_n(\xi_i^n), \tau_n(\xi_j^n)\}$. Similarly, $(\tau(\xi_i), \varsigma(\xi_i))$ is the first point of $\mathcal{P}$ on $\mathbb{R} \times [\![V_1, \xi_i]\!]$, and $\mathcal{A}(i, j)$ occurs if and only if $\tau(\xi_i \wedge \xi_j) \le \min\{\tau(\xi_i), \tau(\xi_j)\}$. Therefore, the joint convergences in (5.13), (5.14) and (5.15) follow from (5.16). On the other hand, we have

$$\mathbf{1}_{\{\xi_i^n \in T_t^n\}} = \mathbf{1}_{\{t < \tau_n(\xi_i^n)\}}, \qquad t \ge 0, n \ge 1.$$

For each fixed $t \ge 0$, this sequence of random variables converges to $\mathbf{1}_{\{t<\tau(\xi_i)\}} = \mathbf{1}_{\{\xi_i \in \mathcal{T}_t\}}$ by (5.16). By the law of large numbers, $k^{-1} \sum_{1 \le i \le k} \mathbf{1}_{\{t<\tau(\xi_j)\}} \to \mu(\mathcal{T}_t)$ almost surely. Following the arguments in [8], Section 2.3 and applying (5.14), we can find a sequence $k_n \to \infty$ slowly enough such that

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \mathbf{1}_{\{t<\tau_n(\xi_i^n)\}} \xrightarrow{d} \mu(\mathcal{T}_t).$$

This entails that, for each t fixed, as $n \to \infty$,

$$\mathbf{p}_n\big(T_t^n\big) \xrightarrow{d} \mu(\mathcal{T}_t). \tag{5.17}$$

Using (5.17) for a sequence of times $(t_m, m \ge 1)$ dense in $\mathbb{R}_+$ and combining with the fact that $t \mapsto \mu(\mathcal{T}_t)$ is decreasing, we obtain the convergence in (5.12), jointly with (5.13), (5.14), (5.15) and (5.1). □

Lemma 5.5. *Jointly with* (5.13), (5.14), (5.15) *and* (5.1), *we have for any* $m \geq 1$ *and* $(t_i, 1 \leq i \leq m) \in \mathbb{R}_+^m$,

$$\left(\int_0^{t_i} \mathbf{p}_n\big(T_s^n\big)\, ds, 1 \leq i \leq m\right) \overset{n\to\infty}{\underset{d}{\longrightarrow}} \left(\int_0^{t_i} \mu(\mathcal{T}_s)\, ds, 1 \leq i \leq m\right).$$

Proof. This is a direct consequence of Lemma 5.4. □

Lemma 5.6. *If we let*

$$M_t^n := \sigma_n L_t^n - \int_0^t \mathbf{p}_n\big(T_s^n\big)\, ds, \qquad n \geq 1,$$

then under the hypothesis that $\sigma_n \to 0$ *as* $n \to \infty$, *the sequence of variables* $(M_t^n, n \geq 1)$ *converges to* 0 *in* L^2 *as* $n \to \infty$. *Moreover, this convergence is uniform on compacts.*

Proof. Let $\mathrm{N}_t^n = \mathrm{Card}\{(s,x) \in \mathcal{P}_n : s \leq t\}$ be the counting process of $\mathcal{P}_n$. Then $(\mathrm{N}_t^n, t \geq 0)$ is a Poisson process of rate $1/\sigma_n$. We write $d\mathrm{N}^n$ for the Stieltjes measure associated with $t \mapsto \mathrm{N}_t^n$. For $t \geq 0$, let

$$\mathscr{M}_t^n := L_t^n - \int_{[0,t]} \mathbf{p}_n\big(T_{s-}^n\big)\, d\mathrm{N}_s^n \quad \text{and} \quad \mathscr{N}_t^n := \sigma_n \int_{[0,t]} \mathbf{p}_n\big(T_{s-}^n\big)\, d\mathrm{N}_s^n - \int_0^t \mathbf{p}_n\big(T_s^n\big)\, ds.$$

We notice that, by the definition of L_t^n,

$$\mathscr{M}_t^n = \sum_{(s,x)\in\mathcal{P}_n : s\leq t} \big(\mathbf{1}_{\{x \in T_{s-}^n\}} - \mathbf{p}_n\big(T_{s-}^n\big)\big).$$

Since $\sigma_n^{-1}\mathbf{p}_n = \mathcal{L}_n$, conditionally on T_{s-}^n, $\mathbf{1}_{\{x\in T_{s-}^n\}}$ is a Bernoulli random variable of mean $\mathbf{p}_n(T_{s-}^n)$. Therefore, we have

$$\mathbb{E}\big[\mathscr{M}_t^n \,|\, \big(\mathrm{N}_s^n\big)_{s\leq t}\big] = 0. \tag{5.18}$$

From this, we can readily show that $\mathscr{M}^n$ is a martingale with respect to the natural filtration of $(\mathrm{N}_t^n)_{t\geq 0}$. On the other hand, classical results on the Poisson process entail that $\mathscr{N}^n$ is also a martingale with respect to the same filtration. Once combined, we see that $\mathrm{M}^n = \sigma_n \mathscr{M}^n + \mathscr{N}^n$ itself is a martingale. Therefore, by Doob's maximal inequality for the L^2-norms of martingales, we obtain for any $t \geq 0$,

$$\mathbb{E}\Big[\sup_{s\leq t}\big(\mathrm{M}_s^n\big)^2\Big] \leq 4\mathbb{E}\big[\big(\mathrm{M}_t^n\big)^2\big] = 4\mathbb{E}\big[\big(\sigma_n \mathscr{M}_t^n\big)^2\big] + 4\mathbb{E}\big[\big(\mathscr{N}_t^n\big)^2\big],$$

as a result of (5.18). Direct computation shows that

$$\mathbb{E}\big[\big(\mathscr{M}_t^n\big)^2\big] = \mathbb{E}\left[\frac{1}{\sigma_n}\int_0^t \big(\mathbf{p}_n\big(T_s^n\big) - \mathbf{p}_n^2\big(T_s^n\big)\big)\, ds\right] \quad \text{and} \quad \mathbb{E}\big[\big(\mathscr{N}_t^n\big)^2\big] = \mathbb{E}\left[\sigma_n \int_0^t \mathbf{p}_n^2\big(T_s^n\big)\, ds\right].$$

As a consequence, for any fixed t,

$$\mathbb{E}\Big[\sup_{s\le t}\big(\mathrm{M}_s^n\big)^2\Big] \le 4\sigma_n \mathbb{E}\Big[\int_0^t \mathbf{p}_n\big(T_s^n\big)\,ds\Big] \le 4\sigma_n t \to 0,$$

as $n \to \infty$. □

In particular, Lemmas 5.5 and 5.6 combined entail that for any fixed $t \ge 0$, $\sigma_n L_t^n \to L_t$ in distribution. However, to obtain the convergence of $\sigma_n L_\infty^n$ to L_∞ in distribution we need a tightness condition. For this, we begin with the following observation. For each $n \in \mathbb{N}$ and $s \ge 0$, let $\zeta^n(s) := \inf\{t > 0 : L_t^n \ge \lfloor s \rfloor\}$ be the right-continuous inverse of L_t^n. Recall that from the construction of $H^n = \mathrm{cut}(T^n, V^n)$, there is a correspondence between the vertex sets of the remaining tree at step $\ell - 1$ and the subtree above X_ℓ in H. Then it follows from Lemma 4.1 that

$$\big(\mathfrak{v}\big(T^n_{\zeta^n(s)}\big), 0 \le s < L_\infty^n\big) \overset{d}{=} \big(\mathfrak{v}\big(\mathrm{Sub}\big(T^n, x_s^n\big)\big), 0 \le s < 1 + d_{T^n}\big(r\big(T^n\big), V^n\big)\big),$$

where x_s^n is the point on the path $[\![r(T^n), V^n]\!]$ at distance $\lfloor s \rfloor$ from $r(T^n)$. In particular, this entails

$$\big(\mathbf{p}_n\big(T^n_{\zeta^n(s)}\big), 0 \le s < L_\infty^n\big) \overset{d}{=} \big(\mathbf{p}_n\big(\mathrm{Sub}\big(T^n, x_s^n\big)\big), 0 \le s < 1 + d_{T^n}\big(r\big(T^n\big), V^n\big)\big). \tag{5.19}$$

The limit of the right-hand side can be identified using the convergence of **p**-trees in (5.1). For this, let V be a random point of $\mathcal{T}$ of distribution μ. For $0 \le s \le d_{\mathcal{T}}(r(\mathcal{T}), V)$, let x_s be the point in $[\![r(\mathcal{T}), V]\!]$ at distance s from $r(\mathcal{T})$, or $x_s = V$ if $s > d_{\mathcal{T}}(r(\mathcal{T}), V)$. Similarly, we set $x_s^n = V^n$ if $s \ge 1 + d_{T^n}(r(T^n), V^n)$.

Lemma 5.7. *Under* (H), *we have*

$$\big(\sigma_n L_\infty^n, \big(\mathbf{p}_n\big(T^n_{\zeta^n(s/\sigma_n)}\big)\big)_{s\ge 0}\big) \overset{n\to\infty}{\underset{d}{\longrightarrow}} \big(d_{\mathcal{T}}\big(r(\mathcal{T}), V\big), \big(\mu\big(\mathrm{Sub}(\mathcal{T}, x_s)\big)\big)_{s\ge 0}\big),$$

where the convergence of the second coordinates is with respect to the Skorokhod J_1-topology.

Proof. Because of (5.19) and the fact $\sigma_n \to 0$, it suffices to prove that

$$\big(\mathbf{p}_n\big(\mathrm{Sub}\big(T^n, x^n_{s/\sigma_n}\big)\big), s \ge 0\big) \overset{n\to\infty}{\underset{d}{\longrightarrow}} \big(\mu\big(\mathrm{Sub}(\mathcal{T}, x_s)\big), s \ge 0\big),$$

with respect to the Skorokhod J_1-topology, jointly with $\sigma_n d_{T^n}(r(T^n), V^n) \to d_{\mathcal{T}}(r(\mathcal{T}), V)$ in distribution. Recall that $(\xi_i^n, i \ge 2)$ is a sequence of i.i.d. points of common law $\mathbf{p}_n$ and that we have set $\xi_0^n = r(T^n)$, $\xi_1^n = V^n$, for $n \in \mathbb{N}$. Note that $(\xi_i^n, i \ge 0)$ is still an i.i.d. sequence. Then it follows from (5.1) that

$$\big(\sigma_n d_{T^n}\big(\xi_i^n, \xi_j^n\big), i, j \ge 0\big) \xrightarrow{d} \big(d_{\mathcal{T}}(\xi_i, \xi_j), i, j \ge 0\big)$$

in the sense of finite-dimensional distributions. Taking $i = 0$ and $j = 1$, we get the convergence

$$\sigma_n d_{T^n}\big(V^n, r\big(T^n\big)\big) \xrightarrow{d} d_{\mathcal{T}}\big(V, r(\mathcal{T})\big).$$

On the other hand, for $i \geq 1$, $\xi_i^n \in \mathrm{Sub}(T^n, x_s^n)$ if and only if $d_{T^n}(\xi_i^n \wedge V^n, r(T^n)) \geq s$. Since for any rooted tree (T, d, r) and $u, v \in T$ we have $2d(r, u \wedge v) = d(r, u) + d(r, v) - d(u, v)$, we deduce that for any $k, m \geq 1$ and $(s_j, 1 \leq j \leq m) \in \mathbb{R}_+^m$,

$$(\mathbf{1}_{\{\xi_i^n \in \mathrm{Sub}(T^n, x^n_{s_j/\sigma_n})\}}, 1 \leq i \leq k, 1 \leq j \leq m) \xrightarrow{d} (\mathbf{1}_{\{\xi_i \in \mathrm{Sub}(\mathcal{T}, x_{s_j})\}}, 1 \leq i \leq k, 1 \leq j \leq m),$$

jointly with $\sigma_n d_{T^n}(V^n, r(T^n)) \xrightarrow{d} d_{\mathcal{T}}(V, r(\mathcal{T}))$. Then the argument used to establish (5.17) shows the convergence of $(\mathbf{p}_n(\mathrm{Sub}(T^n, x^n_{s/\sigma_n})), n \geq 1)$ in the sense of finite-dimensional distributions. The convergence in the Skorokhod topology follows from the monotonicity of the mapping $s \mapsto \mathbf{p}_n(\mathrm{Sub}(T^n, x_s^n))$. □

Combining (5.19) and Lemma 5.7 with a time-change argument, we are able to show the following tightness condition.

Lemma 5.8. *Under* (H), *for every* $\delta > 0$,

$$\lim_{t \to \infty} \limsup_{n \to \infty} \mathbb{P}\big(\sigma_n \big(L_\infty^n - L_t^n\big) \geq \delta\big) = 0, \tag{5.20}$$

Proof. Let us begin with a simple observation on the Skorokhod J_1-topology. Let $\mathbb{D}^\uparrow$ be the set of those functions $x : \mathbb{R}_+ \to [0, 1]$ which are nondecreasing and càdlàg. We endow $\mathbb{D}^\uparrow$ with the Skorokhod J_1-topology. Taking $\varepsilon > 0$ and $x \in \mathbb{D}^\uparrow$, we denote by $\kappa_\varepsilon(x) = \inf\{t > 0 : x(t) > \varepsilon\}$. The following is a well-known fact. A proof can be found in [38], Chapter VI, page 304, Lemma 2.10.

FACT. If $x_n \to x$ in $\mathbb{D}^\uparrow$, $n \to \infty$ and $t \mapsto x(t)$ is strictly increasing, then $\kappa_\varepsilon(x_n) \to \kappa_\varepsilon(x)$ as $n \to \infty$.

If $x = (x(t), t \geq 0)$ is a process with càdlàg paths and $t_0 \in \mathbb{R}_+$, we denote by $\mathrm{R}_{t_0}[x]$ the reversed process of x at t_0:

$$\mathrm{R}_{t_0}[x](t) = x\big((t_0 - t)-\big)$$

if $t < t_0$ and $\mathrm{R}_{t_0}[x](t) = x(0)$ otherwise. For each $n \geq 1$, let $x_n(t) = \mathbf{p}_n(T^n_{\zeta^n(t)})$, $t \geq 0$ and denote by $\Lambda_n = \mathrm{R}_{L_\infty^n}[x_n]$ the reversed process at L_∞^n. Similarly, let $y(t) = \mu(\mathrm{Sub}(\mathcal{T}, x_t))$, $t \geq 0$ and denote by $\Lambda = \mathrm{R}_D[y]$ for $D = d_{\mathcal{T}}(V, r(\mathcal{T}))$. Then almost surely $\Lambda_n \in \mathbb{D}^\uparrow$ for $n \in \mathbb{N}$ and $\Lambda \in \mathbb{D}^\uparrow$. Moreover, Lemma 5.7 says that

$$\big(\Lambda_n(t/\sigma_n), t \geq 0\big) \xrightarrow[d]{n \to \infty} \big(\Lambda(t), t \geq 0\big) \tag{5.21}$$

in $\mathbb{D}^\uparrow$. From the construction of the ICRT in Section 2.5 it is not difficult to show that $t \mapsto \Lambda(t)$ is strictly increasing. Then by the above FACT, we have $\sigma_n \kappa_\varepsilon(\Lambda_n) \to \kappa_\varepsilon(\Lambda)$ in distribution, for each $\varepsilon > 0$. In particular, we have for any fixed $\delta > 0$,

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \mathbb{P}\big(\sigma_n \kappa_\varepsilon(\Lambda_n) \geq \delta\big) \leq \lim_{\varepsilon \to 0} \mathbb{P}\big(\kappa_\varepsilon(\Lambda) \geq \delta\big) = 0, \tag{5.22}$$

since almost surely $\Lambda(t) > 0$ for any $t > 0$.

By Lemma 5.4, the sequence $((\mathbf{p}_n(T_t^n))_{t\geq 0}, n \geq 1)$ is tight in the Skorokhod topology. Combined with the fact that, for each fixed n, $\mathbf{p}_n(T_t^n) \searrow 0$ as $t \to \infty$ almost surely, this entails that for any fixed $\varepsilon > 0$,

$$\lim_{t_0\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{t\geq t_0} \mathbf{p}_n\big(T_t^n\big) \geq \varepsilon\Big) = 0. \tag{5.23}$$

Now note that if $L_t^n = k \in \mathbb{N}$, then $T_t^n = T_{\zeta^n(k)}^n$ a.s. since no change occurs until the time of the next cut. In particular we have

$$\mathbf{p}_n\big(T_t^n\big) = \mathbf{p}_n\big(T_{\zeta^n(L_t^n)}^n\big) \qquad \text{a.s.},$$

from which we deduce that

$$\big\{\mathbf{p}_n\big(T_{t_0}^n\big) < \varepsilon\big\} \subseteq \big\{\kappa_\varepsilon(\Lambda_n) \geq L_\infty^n - L_{t_0}^n\big\} \qquad \text{a.s.}$$

Then we have

$$\big\{\sigma_n\big(L_\infty^n - L_{t_0}^n\big) \geq \delta\big\} \cap \Big\{\sup_{t\geq t_0} \mathbf{p}_n\big(T_t^n\big) < \varepsilon\Big\} \subseteq \big\{\sigma_n\kappa_\varepsilon(\Lambda_n) \geq \delta\big\} \qquad \text{a.s.}$$

Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\big(\sigma_n\big(L_\infty^n - L_{t_0}^n\big) \geq \delta\big) \\
&\quad\leq \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{t\geq t_0} \mathbf{p}_n\big(T_t^n\big) \geq \varepsilon\Big) + \limsup_{n\to\infty} \mathbb{P}\Big(\sigma_n\big(L_\infty^n - L_{t_0}^n\big) \geq \delta \text{ and } \sup_{t\geq t_0} \mathbf{p}_n\big(T_t^n\big) < \varepsilon\Big) \\
&\quad\leq \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{t\geq t_0} \mathbf{p}_n\big(T_t^n\big) \geq \varepsilon\Big) + \limsup_{n\to\infty} \mathbb{P}\big(\sigma_n\kappa_\varepsilon(\Lambda_n) \geq \delta\big).
\end{aligned}$$

In above, if we let first $t_0 \to \infty$ and then $\varepsilon \to 0$, we obtain (5.20) as a combined consequence of (5.22) and (5.23). □

Proposition 5.9. *Under* (H), *we have*

$$\big(\sigma_n L_t^n, t \geq 0\big) \xrightarrow[d]{n\to\infty} (L_t, t \geq 0) \tag{5.24}$$

with respect to the uniform topology, and jointly with the convergences in (5.13), (5.14), (5.15) *and* (5.1). *In particular, this entails that $L_\infty < \infty$ almost surely. Moreover, we have*

$$L_\infty \overset{d}{=} d_{\mathcal{T}}\big(r(\mathcal{T}), V\big), \tag{5.25}$$

where V is a random point of distribution μ. The distribution of $d_{\mathcal{T}}(r(\mathcal{T}), V)$ is given in (2.6).

Proof. We fix a sequence of $(t_m, m \geq 1)$, which is dense in $\mathbb{R}_+$. Combining Lemmas 5.5 and 5.6, we obtain, for all $k \geq 1$,

$$\big(\sigma_n L_{t_m}^n, 1 \leq m \leq k\big) \xrightarrow[d]{n\to\infty} (L_{t_m}, 1 \leq m \leq k), \tag{5.26}$$

jointly with the convergences in (5.13), (5.14), (5.15) and (5.1). We deduce from this and Lemma 5.8 that $L_\infty < \infty$ a.s. and

$$\sigma_n L_\infty^n \xrightarrow[d]{n\to\infty} L_\infty, \tag{5.27}$$

jointly with (5.13), (5.14), (5.15) and (5.5), by Theorem 4.2 of [16], Chapter 1. Combined with the fact that $t \mapsto L_t$ is continuous and increasing, this entails the uniform convergence in (5.24). Finally, the distributional identity (5.25) is a direct consequence of Lemma 5.7 and (5.27). □

Proof of Theorem 3.1. We have seen that $L_\infty < \infty$ almost surely. Therefore the cut tree $(\mathrm{cut}(\mathcal{T}, V), \hat{\mu})$ is well defined almost surely. Comparing the expressions of $d_{H^n}(\xi_i^n, \xi_j^n)$ given at the beginning of this subsection with those of $d_{\mathcal{H}}(\xi_i, \xi_j)$, we obtain from Lemma 5.4 and Proposition 5.9 the convergence in (5.8). This concludes the proof. □

Remark. Before concluding this section, let us say a few more words on the proof of Proposition 5.9. The convergence of $(\sigma_n L_t^n, t \geq 0)$ to $(L_t, t \geq 0)$ on any finite interval follows mainly from the convergence in Proposition 5.2. The proof here can be easily adapted to the other models of random trees, see [15,43]. On the other hand, our proof of the tightness condition (5.20) depends on the specific cuttings on the birthday trees, which has allowed us to deduce the distributional identity (5.19). In general, the convergence of L_∞^n may indeed fail. An obvious example is the classical record problem (see Example 1.4 in [39]), where we have $L_t^n \to L_t$ for any fixed t, while $\mathbb{E}(L_\infty^n) \sim \ln n$ and therefore is not tight in $\mathbb{R}$.

5.3. Convergence of the cut-trees $\mathrm{cut}(T^n)$: Proof of Lemma 5.3

Let us recall the settings of the complete cutting down procedure for $\mathcal{T}$: $(V_i, i \geq 1)$ is an i.i.d. sequence of common law μ; $\mathcal{T}_{V_i}(t)$ is the equivalence class of $\sim_t$ containing V_i, whose mass is denoted by $\mu_i(t)$; and $L_t^i = \int_0^t \mu_i(s)\,ds$. The complete cut-tree $\mathrm{cut}(\mathcal{T})$ is defined as the complete and separable metric space $\overline{\bigcup_k S_k}$. We introduce some corresponding notations for the discrete cuttings on T^n. For each $n \geq 1$, we sample a sequence of i.i.d. points $(V_i^n, i \geq 1)$ on T^n of distribution $\mathbf{p}_n$. Recall $\mathcal{P}_n$ the Poisson point process on $\mathbb{R}_+ \times T^n$ of intensity $dt \otimes \mathcal{L}_n$. We define

$$\begin{aligned}
\mu_{n,i}(t) &:= \mathbf{p}_n\big(\{u \in T^n : [0,t] \times [\![u, V_i^n]\!] \cap \mathcal{P}_n = \varnothing\}\big),\\
L_t^{n,i} &:= \mathrm{Card}\big\{s \leq t : \mu_{n,i}(s) < \mu_{n,i}(s-)\big\}, \qquad t \geq 0, i \geq 1,\\
\tau_n(i,j) &:= \inf\big\{t \geq 0 : [0,t] \times [\![V_i^n, V_j^n]\!] \cap \mathcal{P}_n \neq \varnothing\big\}, \qquad 1 \leq i, j < \infty.
\end{aligned}$$

By the construction of $G^n = \mathrm{cut}(T^n)$, we have

$$\begin{aligned}
d_{G^n}\big(V_i^n, r(G^n)\big) &= L_\infty^{n,i} - 1,\\
d_{G^n}\big(V_i^n, V_j^n\big) &= L_\infty^{n,i} + L_\infty^{n,j} - 2L_{\tau_n(i,j)}^{n,i}, \qquad 1 \leq i, j < \infty,
\end{aligned} \tag{5.28}$$

where $L_\infty^{n,i} := \lim_{t\to\infty} L_t^{n,i}$ is the number of cuts necessary to isolate V_i^n. The proof of Lemma 5.3 is quite similar to that of Theorem 3.1. We outline the main steps but leave out the details.

Sketch of proof of Lemma 5.3. First, we can show with essentially the same proof of Lemma 5.4 that we have the following joint convergences: for each $k \geq 1$,

$$\big((\mu_{n,i}(t), 1 \leq i \leq k), t \geq 0\big) \xrightarrow[d]{n\to\infty} \big((\mu_i(t), 1 \leq i \leq k), t \geq 0\big), \tag{5.29}$$

with respect to Skorokhod J_1-topology, jointly with

$$\big(\tau_n(i,j), 1 \leq i, j \leq k\big) \xrightarrow[d]{n\to\infty} \big(\tau(i,j), 1 \leq i, j \leq k\big), \tag{5.30}$$

jointly with the convergence in (5.1). Then we can proceed, with the same argument as in the proof of Lemma 5.6, to show that for any $k, m \geq 1$ and $(t_j, 1 \leq j \leq m) \in \mathbb{R}_+^m$,

$$\left(\int_0^{t_i} \mu_{n,i}(s)\,ds, 1 \leq j \leq m, 1 \leq i \leq k\right) \xrightarrow[d]{n\to\infty} \left(\int_0^{t_i} \mu_i(s)\,ds, 1 \leq j \leq m, 1 \leq i \leq k\right).$$

Since the $V_i^n, i \geq 1$ are i.i.d. $\mathbf{p}_n$-distributed nodes on T^n, each process $(L_t^{n,i})_{t\geq 0}$ has the same distribution as $(L_t^n)_{t\geq 0}$ defined in (5.10). Then Lemmas 5.6 and 5.8 hold true for each $L^{n,i}, i \geq 1$. We are able to show

$$\big(\sigma_n L_t^{n,i}, 1 \leq i \leq k\big)_{t\geq 0} \xrightarrow[d]{n\to\infty} \big(L_t^i, 1 \leq i \leq k\big)_{t\geq 0}, \tag{5.31}$$

with respect to the uniform topology, jointly with the convergences (5.30) and (5.1). Comparing (5.28) with (3.4), we can easily conclude. □

In general, the convergence in (5.1) does not hold in the Gromov–Hausdorff topology. However, in the case where $\mathcal{T}$ is a.s. compact and the convergence (5.1) does hold in the Gromov–Hausdorff sense, then we are able to show that one indeed has GHP convergence as claimed in Theorem 3.6. In the following proof, we only deal with the case of convergence of $\mathrm{cut}(T^n)$. The result for $\mathrm{cut}(T^n, V^n)$ can be obtained using similar arguments and we omit the details.

Proof of Theorem 3.6. We have already shown in Lemma 5.3 the joint convergence of the spanning subtrees: for each $k \geq 1$,

$$\big(\sigma_n R_k^n, \sigma_n S_k^n\big) \xrightarrow[d,\mathrm{GH}]{n\to\infty} (R_k, S_k). \tag{5.32}$$

We now show that for each $\varepsilon > 0$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\max\big\{\mathrm{d_{GH}}\big(R_k^n, T^n\big), \mathrm{d_{GH}}\big(S_k^n, \mathrm{cut}\big(T^n\big)\big)\big\} \geq \varepsilon/\sigma_n\big) = 0. \tag{5.33}$$

Since the couples $(S_k^n, \mathrm{cut}(T^n))$ and (R_k^n, T^n) have the same distribution, it is enough to prove that for each $\varepsilon > 0$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\sigma_n \mathrm{d_{GH}}\big(R_k^n, T^n\big) \geq \varepsilon\big) = 0. \tag{5.34}$$

Let us explain why this is true when $(\sigma_n T^n, \mathbf{p}_n) \to (\mathcal{T}, \mu)$ in distribution in the sense of GHP. Recall the space $\mathbb{M}_c^k$ of equivalence classes of k-pointed compact metric spaces, equipped with the k-pointed Gromov–Hausdorff metric. For each $k \geq 1$ and $\varepsilon > 0$, we set

$$A(k,\varepsilon) := \big\{(T,d,\mathbf{x}) \in \mathbb{M}_c^k : \mathrm{d_{GH}}\big(T, \mathrm{Span}(T;\mathbf{x})\big) \geq \varepsilon\big\}.$$

It is not difficult to check that $A(k,\varepsilon)$ is a closed set of $\mathbb{M}_c^k$. Let $\mathbb{M}_c$ be the set of equivalence classes of compact measured metric spaces, equipped with the Gromov–Hausdorff–Prokhorov metric and for $(T,\mu) \in \mathbb{M}_c$, set

$$m_k\big(T, A(k,\varepsilon)\big) := \int_{T^k} \mu^{\otimes k}(d\mathbf{x}) \mathbf{1}_{\{[T,\mathbf{x}] \in A(k,\varepsilon)\}}.$$

Then according to the proof of Lemma 13 of [46], the mapping from $\mathbb{M}_c$ to $\mathbb{M}_c^k$: $(T,\mu) \mapsto m_k(T, A(k,\varepsilon))$ is upper-semicontinuous. Applying the Portmanteau theorem for upper-semicontinuous mappings [16], page 17, Problem 7, we obtain

$$\limsup_{n\to\infty} \mathbb{E}\big[m_k\big((\sigma_n T^n, \mathbf{p}_n), A(k,\varepsilon)\big)\big] \leq \mathbb{E}\big[m_k\big((\mathcal{T},\mu), A(k,\varepsilon)\big)\big],$$

or, in other words,

$$\limsup_{n\to\infty} \mathbb{P}\big(\sigma_n \mathrm{d_{GH}}\big(T^n, R_k^n\big) \geq \varepsilon\big) \leq \mathbb{P}\big(\mathrm{d_{GH}}(\mathcal{T}, R_k) \geq \varepsilon\big) \underset{k\to\infty}{\longrightarrow} 0,$$

since $\mathrm{d_{GH}}(R_k, \mathcal{T}) \to 0$ almost surely for $\mathcal{T}$ is compact (see for example [5]). This proves (5.34) and thus (5.33). By [16], Chapter 1, Theorem 4.5, (5.32) combined with (5.33) entails the joint convergence in distribution of $(\sigma_n T^n, \sigma_n \mathrm{cut}(T^n))$ to $(\mathcal{T}, \mathrm{cut}(\mathcal{T}))$ in the Gromov–Hausdorff topology. To strengthen to the Gromov–Hausdorff–Prokhorov convergence, one can adopt the arguments in Section 4.4 of [34] and we omit the details. □

6. Reversing the one-cutting transformation

In this section, we justify the heuristic construction of $\mathrm{shuff}(\mathcal{H}, U)$ given in Section 3 for an ICRT $\mathcal{H}$ and a uniform leaf U. The objective is to define formally the shuffle operation in such a way that the identity (3.6) hold. We rely on weak convergence arguments to justify the construction of $\mathrm{shuff}(\mathcal{H}, U)$ by showing it is the limit of the discrete construction in Section 4.1.

Let $(\mathcal{H}, d_{\mathcal{H}}, \mu_{\mathcal{H}})$ be an ICRT rooted at $r(\mathcal{H})$, and let U be a random point in $\mathcal{H}$ of distribution $\mu_{\mathcal{H}}$. Then $\mathcal{H}$ is the disjoint union of the following subsets:

$$\mathcal{H} = \bigcup_{x \in [\![r(\mathcal{H}), U]\!]} F_x \qquad \text{where } F_x := \big\{u \in \mathcal{H} : [\![r(\mathcal{H}), u]\!] \cap [\![r(\mathcal{H}), U]\!] = [\![r, x]\!]\big\}.$$

It is easy to see that F_x is a subtree of $\mathcal{H}$. It is nonempty ($x \in F_x$), but possibly trivial ($F_x = \{x\}$). Let $\mathbf{B} := \{x \in [\![r(\mathcal{H}), U]\!] : \mu_{\mathcal{H}}(F_x) > 0\} \cup \{U\}$, and for $x \in \mathbf{B}$, let $S_x := \mathrm{Sub}(\mathcal{H}, x) \setminus F_x$, which is the union of those F_y such that $d_{\mathcal{H}}(U, y) < d_{\mathcal{H}}(U, x)$. Then for each $x \in \mathbf{B} \setminus \{U\}$, we associate with x an attachment point A_x, which is independent and sampled according to $\mu_{\mathcal{H}}|_{S_x}$, $\mu_{\mathcal{H}}$ conditioned on S_x. We also set $A_U = U$.

Now let $(\xi_i, i \geq 1)$ be a sequence of i.i.d. points of common law $\mu_{\mathcal{H}}$. The set $\boldsymbol{\mathcal{F}} := \bigcup_{x \in \mathbf{B}} F_x$ has full mass with probability one. Thus almost surely $\xi_i \in \boldsymbol{\mathcal{F}}$ for each $i \geq 0$. We will use $(\xi_i)_{i \geq 1}$ to span the tree $\mathrm{shuff}(\mathcal{H}, U)$ and the point ξ_1 is the future root of $\mathrm{shuff}(\mathcal{H}, U)$. For each ξ_i, we define inductively two sequences $\mathbf{x}_i := (x_i(0), x_i(1), \ldots) \in \mathbf{B}$ and $\mathbf{a}_i := (a_i(0), a_i(1), \ldots)$: we set $a_i(0) = \xi_i$, and, for $l \geq 0$,

$$x_i(l) = a_i(l) \wedge U \quad \text{and} \quad a_i(l+1) = A_{x_i(l)}.$$

By definition of $(A_x, x \in \mathbf{B})$, the distance $d_{\mathcal{H}}(r(\mathcal{H}), x_i(k))$ is increasing in $k \geq 1$. For each $i, j \geq 1$, we define the merging time

$$\mathrm{mg}(i, j) := \inf\big\{k \geq 0 : \exists l \leq k \text{ and } x_i(l) = x_j(k-l)\big\},$$

with the convention $\inf \varnothing = \infty$. Another way to present $\mathrm{mg}(i, j)$ is to consider the graph on $\mathbf{B}$ with the edges $\{x, A_x \wedge U\}$, $x \in \mathbf{B}$, then $\mathrm{mg}(i, j)$ is the graph distance between $\xi_i \wedge U$ and $\xi_j \wedge U$. On the event $\{\mathrm{mg}(i, j) < \infty\}$, there is a path in this graph that has only finitely many edges, and the two walks $\mathbf{x}_i$ and $\mathbf{x}_j$ first meet at a point $y(i, j) \in \mathbf{B}$ (where by first, we mean with minimum distance to the root $r(\mathcal{H})$). In particular, if we set $\mathscr{I}(i, j)$, $\mathscr{I}(j, i)$ to be the respective indices of the element $y(i, j)$ appearing in $\mathbf{x}_i$ and $\mathbf{x}_j$, that is,

$$\mathscr{I}(i, j) = \inf\big\{k \geq 0 : x_i(k) = y(i, j)\big\} \quad \text{and} \quad \mathscr{I}(j, i) = \inf\big\{k \geq 0 : x_j(k) = y(i, j)\big\},$$

with the convention that $\mathscr{I}(i, j) = \mathscr{I}(j, i) = \infty$ if $\mathrm{mg}(i, j) = \infty$, then $\mathrm{mg}(i, j) = \mathscr{I}(i, j) + \mathscr{I}(j, i)$. Write $\mathrm{Ht}(u) = d(u, u \wedge U)$ for the height of u in the F_x which contains it. On the event $\{\mathrm{mg}(i, j) < \infty\}$ we define $\gamma(i, j)$ which is meant to be the new distance between ξ_i and ξ_j as follows:

$$\begin{aligned}\gamma(i, j) := &\sum_{k=0}^{\mathscr{I}(i,j)-1} \mathrm{Ht}\big(a_i(k)\big) \\ &+ \sum_{k=0}^{\mathscr{I}(j,i)-1} \mathrm{Ht}\big(a_j(k)\big) + d_{\mathcal{H}}\big(a_i\big(\mathscr{I}(i, j)\big), a_j\big(\mathscr{I}(j, i)\big)\big),\end{aligned}$$

with the convention if k ranges from 0 to -1, the sum equals zero. See Figure 5 for an example.

The justification of the definition for $\mathrm{shuff}(\mathcal{H}, U)$ relies on weak convergence arguments. To this end, let $\mathbf{p}_n$, $n \geq 1$, be a sequence of probability measures such that (H) holds with $\boldsymbol{\theta}$ the parameter of $\mathcal{H}$. Let H^n be a $\mathbf{p}_n$-tree and U^n a $\mathbf{p}_n$-distributed node. Let $(\xi_i^n)_{i \geq 1}$ be a sequence of i.i.d. $\mathbf{p}_n$-distributed points. Then, the quantities S_x^n, $\mathbf{B}^n$, $\mathbf{x}^n$, $\mathbf{a}^n$, and $\mathrm{mg}^n(i, j)$ are defined

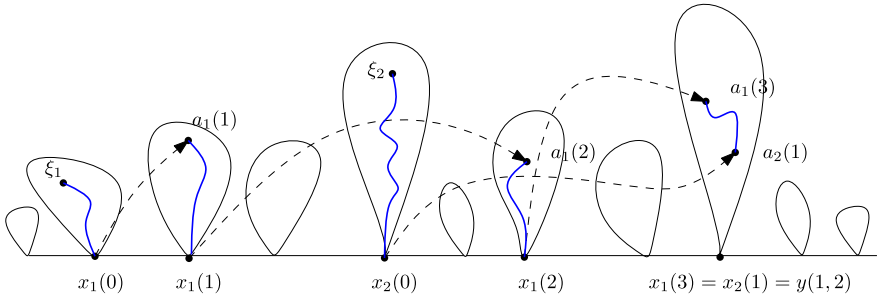

Figure 5. An example with $\mathscr{I}(1,2) = 3$, $\mathscr{I}(2,1) = 1$ and $\mathrm{mg}(1,2) = 4$. The dashed lines indicate the identifications where the root of the relevant subtrees are sent to. The blue lines represent the location of the path between ξ_1 and ξ_2 before the transformation.

for H^n in the same way as S_x, $\mathbf{B}$, $\mathbf{x}$, $\mathbf{a}$, and $\mathrm{mg}(i,j)$ have been defined for $\mathcal{H}$. Let d_{H^n} denote the graph distance on H^n. There is only a slight difference in the definition of the distances

$$\gamma^n(i,j) := \sum_{k=0}^{\mathscr{I}^n(i,j)-1} \big(\mathrm{Ht}\big(a_i^n(k)\big) + 1\big) + \sum_{k=0}^{\mathscr{I}^n(j,i)-1} \big(\mathrm{Ht}\big(a_j^n(k)\big) + 1\big) + d_{H^n}\big(a_i\big(\mathscr{I}^n(i,j)\big), a_j^n\big(\mathscr{I}^n(j,i)\big)\big),$$

to take into account the length of the edges $\{x, A_x^n\}$, for $x \in \mathbf{B}^n$. In that case, the sequence $\mathbf{x}^n$ (resp. $\mathbf{a}^n$) is eventually constant and equal to U^n so that $\mathrm{mg}^n(i,j) < \infty$ with probability one. Furthermore, the unique tree defined by the distance matrix $(\gamma^n(i,j) : i,j \geq 1)$ is easily seen to have the same distribution as the one defined in Section 4.1, since the attaching points are sampled with the same distributions and $(\gamma^n(i,j) : i,j \geq 1)$ coincides with the tree distance after attaching. Recall that we have rerooted $\mathrm{shuff}(H^n, U^n)$ at a random point of law $\mathbf{p}_n$. We may suppose this point is ξ_1^n. Therefore, we have

$$\big(\mathrm{shuff}\big(H^n, U^n\big), H^n\big) \overset{d}{=} \big(H^n, \mathrm{cut}\big(H^n, U^n\big)\big), \tag{6.1}$$

by Lemma 4.1 and Proposition 4.3.

In the case of the ICRT $\mathcal{H}$, it is a priori not clear that $\mathbb{P}(\mathrm{mg}(i,j) < \infty) = 1$. We prove the following theorem.

Theorem 6.1. *For any ICRT $(\mathcal{H}, \mu_{\mathcal{H}})$ and a $\mu_{\mathcal{H}}$-distributed point U, we have the following assertions*:

(a) *almost surely for each $i,j \geq 1$, $\mathrm{mg}(i,j)$ is stochastically dominated by a Poisson random variable with parameter $d_{\mathcal{H}}(\xi_i, \xi_j)E(i,j)$, where $E(i,j)$ is an exponential variable of rate $d_{\mathcal{H}}(U, \xi_i \wedge \xi_j)$; in consequence, we have $\mathrm{mg}(i,j) < \infty$;*

(b) *almost surely the distance matrix* $(\gamma(i,j), 1 \le i, j < \infty)$ *defines a CRT, denoted by* $\mathrm{shuff}(\mathcal{H}, U)$;
(c) $(\mathrm{shuff}(\mathcal{H}, U), \mathcal{H})$ *and* $(\mathcal{H}, \mathrm{cut}(\mathcal{H}, V))$ *have the same distribution.*

The main ingredient in the proof of Theorem 6.1 is the following lemma.

Lemma 6.2. *Under* (H), *for each* $j, l \ge 1$, *we have the following convergences*

$$\begin{aligned}&\big(\sigma_n d_{H^n}\big(r\big(H^n\big), \mathbf{x}_i^n(k)\big), 1 \le i \le j, 0 \le k \le l\big)\\ &\quad\xrightarrow[d]{n\to\infty} \big(d_{\mathcal{H}}\big(r(\mathcal{H}), \mathbf{x}_i(k)\big), 1 \le i \le j, 0 \le k \le l\big),\end{aligned} \tag{6.2}$$

$$\begin{aligned}&\big(\mathbf{p}_n\big(S^n_{x_i^n(k)}\big), 1 \le i \le j, 0 \le k \le l\big)\\ &\quad\xrightarrow[d]{n\to\infty} \big(\mu_{\mathcal{H}}(S_{x_i(k)}), 1 \le i \le j, 0 \le k \le l\big)\end{aligned} \tag{6.3}$$

and

$$\big(\sigma_n H^n, \big(a_i^n(k), 1 \le i \le j, 0 \le k \le l\big)\big) \xrightarrow[d]{n\to\infty} \big(\mathcal{H}, \big(a_i(k), 1 \le i \le j, 0 \le k \le l\big)\big), \tag{6.4}$$

in the weak convergence of the pointed Gromov–Prokhorov topology.

Proof. Fix some $j, l \ge 1$. We argue by induction on k. For $k = 0$, we note that $a_i^n(0) = \xi_i^n$ and $x_i^n(0) = \xi_i^n \wedge U^n$. Then the convergences in (6.4) and (6.2) for $k = 0$ follows easily from (5.1). On the other hand, we can prove (6.3) with the same arguments as employed in Lemma 5.7. Suppose now (6.2), (6.3) and (6.4) hold true for some $k \ge 0$. We notice that $a_i^n(k+1)$ is independently sampled according to $\mathbf{p}_n$ conditioned on $S_{x_i^n(k)}$, we deduce (6.4) for $k+1$ from (5.1). Then the convergence in (6.2) also follows for $k+1$, since $x_i^n(k+1) = a_i^n(k+1) \wedge U^n$. Finally, the very same arguments used in the proof of Lemma 5.7 show that (6.3) holds for $k+1$. □

Proof of Theorem 6.1. *Proof of* (a). By construction, $\mathrm{shuff}(H^n, U^n)$ is the reverse transformation of the one from H^n to $\mathrm{cut}(H^n, U^n)$ in the sense that each attaching "undoes" a cut. In consequence, since $\mathrm{mg}^n(i,j)$ is the number of cuts to undo in order to get ξ_i^n and ξ_j^n in the same connected component, $\mathrm{mg}^n(i,j)$ has the same distribution as the number of the cuts that fell on the path $[\![\xi_i^n, \xi_j^n]\!]$. But the latter is stochastically bounded by a Poisson variable $N_n(i,j)$ of mean $d_{H^n}(\xi_i^n, \xi_j^n) \cdot E_n(i,j)$, where $E_n(i,j)$ is an independent exponential variable of rate $d_{H^n}(U^n, \xi_i^n \wedge \xi_j^n)$. Indeed, each cut is a point of the Poisson point process $\mathcal{P}^n$ and no more cuts fall on $[\![\xi_i^n, \xi_j^n]\!]$ after the time of the first cut on $[\![U^n, \xi_i^n \wedge \xi_j^n]\!]$. But the time of the first cut on $[\![U^n, \xi_i^n \wedge \xi_j^n]\!]$ has the same distribution as $E_n(i,j)$ and is independent of $\mathcal{P}^n$ restricted on $[\![\xi_i^n, \xi_j^n]\!]$. The above argument shows that

$$\mathrm{mg}^n(i,j) = \mathscr{I}^n(i,j) + \mathscr{I}^n(j,i) \le_{\mathrm{st}} N_n(i,j), \qquad i, j \ge 1, n \ge 1, \tag{6.5}$$

where $\le_{\mathrm{st}}$ denotes the stochastic domination order. It follows from (5.1) that, jointly with the convergence in (5.1), we have $N_n(i,j) \to N(i,j)$ in distribution, as $n \to \infty$, where $N(i,j)$ is

a Poisson variable with parameter $d_{\mathcal{H}}(\xi_i, \xi_j) \cdot E(i, j)$. Thus the sequence $(\mathrm{mg}^n(i, j), n \geq 1)$ is tight in $\mathbb{R}_+$.

On the other hand, observe that for $x \in \mathbf{B}$, $\mathbb{P}(A_x \in F_y) = \mu_{\mathcal{H}}(F_y)/\mu_{\mathcal{H}}(S_x)$ if $y \in \mathbf{B}$ and $d_{\mathcal{H}}(U, y) < d_{\mathcal{H}}(U, x)$. In particular, for two distinct points $x, x' \in \mathbf{B}$,

$$\mathbb{P}(\exists y \in \mathbf{B} \text{ such that } A_x \in F_y, A_{x'} \in F_y) = \sum_y \frac{\mu^2_{\mathcal{H}}(F_y)}{\mu_{\mathcal{H}}(S_x)\mu_{\mathcal{H}}(S_{x'})},$$

where the sum is over those $y \in \mathbf{B}$ such that $d_{\mathcal{H}}(U, y) < \min\{d_{\mathcal{H}}(U, x), d_{\mathcal{H}}(U, x')\}$. Similarly, for $n \geq 1$,

$$\mathbb{P}\big(\exists y \in \mathbf{B}^n \text{ such that } A^n_x \in F^n_y, A^n_{x'} \in F^n_y\big) = \sum_y \frac{\mathbf{p}^2_n(F^n_y)}{\mathbf{p}_n(S^n_x)\mathbf{p}_n(S^n_{x'})}.$$

Then it follows from (6.2) and the convergence of the masses in Lemma 5.7 that

$$\begin{aligned}\mathbb{P}\big(\mathscr{I}^n(i, j) = 1; \mathscr{I}^n(j, i) = 1\big) &= \mathbb{P}\big(\exists y \in \mathbf{B}^n \text{ such that } A^n_{x_i(0)} \in F^n_y, A^n_{x_j(0)} \in F^n_y\big) \\ &\overset{n\to\infty}{\longrightarrow} \mathbb{P}\big(\mathscr{I}(i, j) = 1; \mathscr{I}(j, i) = 1\big).\end{aligned}$$

By induction and Lemma 6.2, this can be extended to the following: for any natural numbers $k, l \geq 0$, we have

$$\mathbb{P}\big(\mathscr{I}^n(i, j) = k; \mathscr{I}^n(j, i) = l\big) \overset{n\to\infty}{\longrightarrow} \mathbb{P}\big(\mathscr{I}(i, j) = k; \mathscr{I}(j, i) = l\big).$$

Combined with the tightness of $(\mathrm{mg}^n(i, j), n \geq 1) = (\mathscr{I}^n(i, j) + \mathscr{I}^n(j, i), n \geq 1)$, this entails that

$$\big(\mathscr{I}^n(i, j), \mathscr{I}^n(j, i)\big) \overset{n\to\infty}{\underset{d}{\longrightarrow}} \big(\mathscr{I}(i, j), \mathscr{I}(j, i)\big), \qquad i, j \geq 1, \tag{6.6}$$

jointly with (6.2) and (6.4), using the usual subsequence arguments. In particular, $\mathscr{I}(i, j) + \mathscr{I}(j, i) \leq_{\mathrm{st}} N(i, j) < \infty$ almost surely, which entails that $\mathrm{mg}(i, j) < \infty$ almost surely, for each pair $(i, j) \in \mathbb{N} \times \mathbb{N}$.

Proof of (b). It follows from (6.4), (6.6) and the expression of $\gamma(i, j)$ that

$$\big(\sigma_n \gamma^n(i, j), i, j \geq 1\big) \overset{n\to\infty}{\underset{d}{\longrightarrow}} \big(\gamma(i, j), i, j \geq 1\big), \tag{6.7}$$

in the sense of finite-dimensional distributions, jointly with the Gromov–Prokhorov convergence of $\sigma_n H^n$ to $\mathcal{H}$ in (5.1). However by (6.1), the distribution of $\mathrm{shuff}(H^n, U^n)$ is identical to H^n. Hence, the unconditional distribution of $(\gamma(i, j), 1 \leq i, j < \infty)$ is that of the distance matrix of the ICRT $\mathcal{H}$. We can apply Aldous' CRT theory [5] to conclude that for almost every realization of $\mathcal{H}$, the distance matrix $(\gamma(i, j), i, j \geq 1)$ defines a CRT, denoted by $\mathrm{shuff}(\mathcal{H}, U)$. Moreover, there exists a mass measure $\tilde{\mu}$, such that if $(\tilde{\xi}_i)_{i\geq 1}$ is an i.i.d. sequence of law $\tilde{\mu}$, then

$$\big(d_{\mathrm{shuff}(\mathcal{H},U)}(\tilde{\xi}_i, \tilde{\xi}_j), 1 \leq i, j < \infty\big) \overset{d}{=} \big(\gamma(i, j), 1 \leq i, j < \infty\big).$$

Therefore, we can rewrite (6.7) as

$$\big(\sigma_n \operatorname{shuff}\big(H^n, U^n\big), \sigma_n H^n\big) \xrightarrow[d]{n\to\infty} \big(\operatorname{shuff}(\mathcal{H}, U), \mathcal{H}\big), \tag{6.8}$$

with respect to the Gromov–Prokhorov topology.

Proof of (c). This is an easy consequence of (6.1) and (6.8). Let f, g be two arbitrary bounded functions continuous in the Gromov–Prokhorov topology. Then (6.8) and the continuity of f, g entail that

$$\begin{aligned}\mathbb{E}\big[f\big(\operatorname{shuff}(\mathcal{H}, U)\big) \cdot g(\mathcal{H})\big] &= \lim_{n\to\infty} \mathbb{E}\big[f\big(\sigma_n \operatorname{shuff}\big(H^n, U^n\big)\big) \cdot g\big(\sigma_n H^n\big)\big] \\ &= \lim_{n\to\infty} \mathbb{E}\big[f\big(\sigma_n H^n\big) \cdot g\big(\sigma_n \operatorname{cut}\big(H^n, U^n\big)\big)\big] \\ &= \mathbb{E}\big[f(\mathcal{H}) \cdot g\big(\operatorname{cut}(\mathcal{H}, U)\big)\big],\end{aligned}$$

where we have used (6.1) in the second equality and Theorem 3.1 in the third. Thus, we obtain the identity in distribution in (c). □

7. Convergence of the cutting measures: Proof of Proposition 5.2

Recall the setting at the beginning of Section 5.1. We need to show that for each $k \geq 1$,

$$\big(\sigma_n R_k^n, \mathcal{L}_n \restriction_{R_k^n}\big) \xrightarrow[d]{n\to\infty} \big(R_k(\mathcal{T}), \mathcal{L} \restriction_{R_k}\big) \tag{7.1}$$

in Gromov–Hausdorff–Prokhorov topology. Observe that the Gromov–Hausdorff convergence is clear from (5.2), so that it remains to prove the convergence of the measures.

Case 1. We first prove the claim assuming that $\theta_i > 0$ for every $i \geq 0$. In this case, define

$$m_n := \min\left\{ j : \sum_{i=1}^{j} \left(\frac{p_{ni}}{\sigma_n}\right)^2 \geq \sum_{i\geq 1} \theta_i^2 \right\},$$

and observe that $m_n < \infty$ since $\sum_{i\leq n} (p_{ni}/\sigma_n)^2 = 1 \geq \sum_{i\geq 1} \theta_i^2$. Note also that $m_n \to \infty$. Indeed, for every integer $k \geq 1$, since $p_{ni}/\sigma_n \to \theta_i$, for $i \geq 1$, and $\theta_{k+1} > 0$, we have, for all n large enough,

$$\sum_{i=1}^{k} \left(\frac{p_{ni}}{\sigma_n}\right)^2 < \sum_{i\geq 1} \theta_i^2,$$

so that $m_n > k$ for all n large enough. Furthermore $\lim_{j\to\infty} \theta_j = 0$, and (H) implies that

$$\lim_{n\to\infty} \frac{p_{nm_n}}{\sigma_n} = 0. \tag{7.2}$$

Combining this with the definition of m_n, it follows that, as $n \to \infty$,

$$\sum_{i \le m_n} \left(\frac{p_{ni}}{\sigma_n} \right)^2 \to \sum_{i \ge 1} \theta_i^2. \tag{7.3}$$

If $n, k, M \ge 1$, we set

$$\mathcal{L}_n^* = \sum_{m_n < i \le n} \frac{p_{ni}}{\sigma_n} \delta_i \quad \text{and} \quad \Sigma(n,k,M) = \sum_{M < i \le m_n} \frac{p_{ni}}{\sigma_n} \mathbf{1}_{\{i \in R_k^n\}}.$$

Let ℓ_n denote the (discrete) length measure on T^n, which is induced by the graph distance. Clearly, $\sigma_n \ell_n$ is the length measure of the rescaled tree $\sigma_n T^n$. Recall that d_P stands for the Prokhorov distance.

Lemma 7.1. *Suppose that* (H) *holds. Then, for each* $k \ge 1$, *we have the following assertions*:

(a) *as* $n \to \infty$, *in probability*

$$\mathrm{d}_P \left(\mathcal{L}_n^* \restriction_{R_k^n}, \theta_0^2 \sigma_n \ell_n \restriction_{R_k^n} \right) \to 0; \tag{7.4}$$

(b) *for each* $\varepsilon > 0$, *there exists* $M = M(k, \varepsilon) \in \mathbb{N}$ *such that*

$$\limsup_{n \to \infty} \mathbb{P}\big(\Sigma(n,k,M) \ge \varepsilon \big) \le \varepsilon; \tag{7.5}$$

Before proving Lemma 7.1, let us first explain why this entails Proposition 5.2.

Proof of Proposition 5.2 in the case 1. By Skorokhod representation theorem and a diagonal argument, we can assume that the convergence in (5.2) holds almost surely for all $m \ge 1$. Since the length measure ℓ_n (resp. ℓ) depends continuously on the metric of T^n (resp. the metric of $\mathcal{T}$), according to Proposition 2.23 of [43] this implies that, for each $k \ge 1$,

$$\big(\sigma_n R_k^n, \theta_0^2 \sigma_n \ell_n \restriction_{R_k^n} \big) \to \big(R_k, \theta_0^2 \ell \restriction_{R_k} \big), \tag{7.6}$$

almost surely in the Gromov–Hausdorff–Prokhorov topology. On the other hand, we easily deduce from (5.2) and (H) that, for each fixed $m \ge 1$,

$$\left(\sigma_n R_k^n, \sum_{i=1}^m \frac{p_{ni}}{\sigma_n} \delta_i \restriction_{R_k^n} \right) \to \left(R_k, \sum_{i=1}^m \theta_i \delta_{\mathcal{B}_i} \restriction_{R_k} \right), \tag{7.7}$$

almost surely in the Gromov–Hausdorff–Prokhorov topology. In the following, we write $\mathrm{d}_{\mathrm{P}}^{n,k}$ (resp. $\mathrm{d}_{\mathrm{P}}^{k}$) for the Prokhorov distance on the finite measures on the set R_k^n (resp. R_k). In particular, since the measures below are all restricted to either R_k^n or R_k, we omit the notations $\restriction_{R_k^n}$, $\restriction_{R_k}$ when the meaning is clear from context. We write

$$\mathrm{Kt}_m(\mathcal{L}) := \theta_0^2 \ell + \sum_{i=1}^m \theta_i \delta_{\mathcal{B}_i}$$

for the cut-off measure of $\mathcal{L}$ at level m. By Lemma 5.1, the restriction of $\mathcal{L}$ to R_k^n is a finite measure. Therefore, $\mathrm{Kt}_m(\mathcal{L}) \to \mathcal{L}$ almost surely in $\mathrm{d}_{\mathrm{P}}^k$ as $m \to \infty$.

Now fix some $\varepsilon > 0$. By Lemma 7.1 we can choose some $M = M(k, \varepsilon)$ such that (7.5) holds, as well as

$$\mathbb{P}\big(\mathrm{d}_{\mathrm{P}}^k\big(\mathrm{Kt}_M(\mathcal{L}), \mathcal{L}\big) \geq \varepsilon\big) \leq \varepsilon. \tag{7.8}$$

Define now the approximation

$$\vartheta_{n,M} := \theta_0^2 \sigma_n \ell_n + \sum_{i \leq M} \frac{p_{ni}}{\sigma_n} \delta_i.$$

Then recalling the definition of $\mathcal{L}_n$ in (5.4), and using (7.5) and (7.4), we obtain

$$\limsup_{n\to\infty} \mathbb{P}\big(\mathrm{d}_{\mathrm{P}}^{n,k}(\vartheta_{n,M}, \mathcal{L}_n) \geq \varepsilon\big) \leq \varepsilon. \tag{7.9}$$

We notice that

$$\big(\sigma_n R_k^n, \vartheta_{n,M}\big) \to \big(R_k, \mathrm{Kt}_M(\mathcal{L})\big) \tag{7.10}$$

almost surely in the Gromov–Hausdorff–Prokhorov topology as a combined consequence of (7.6) and (7.7). Finally, by the triangular inequality, we deduce from (7.8), (7.9) and (7.10) that

$$\limsup_{n\to\infty} \mathbb{P}\big(\mathrm{d}_{\mathrm{GHP}}\big(\big(\sigma_n R_k^n, \mathcal{L}_n\big), (R_k, \mathcal{L})\big) \geq 2\varepsilon\big) \leq 2\varepsilon,$$

for any $\varepsilon > 0$, which concludes the proof. □

Proof of Lemma 7.1. We first consider the case $k = 1$. Define

$$D_n := d_{T^n}\big(r\big(T^n\big), V_1^n\big) \quad \text{and} \quad F_n^{\mathcal{L}}(l) := \mathcal{L}_n^*\big(\mathbf{B}\big(r\big(T^n\big), l\big) \cap R_1^n\big), \tag{7.11}$$

where $\mathbf{B}(x, l)$ denotes the ball in T^n centered at x and with radius l. Then the function $F_n^{\mathcal{L}}$ determines the measure $\mathcal{L}_n^* \restriction_{R_1^n}$ in the same way a distributional function determines a finite measure of $\mathbb{R}_+$. Let $(X_j^n, j \geq 0)$ be a sequence of i.i.d. random variables of distribution $\mathbf{p}_n$. We define $\mathfrak{R}_0^n = 0$, and for $m \geq 1$,

$$\mathfrak{R}_m^n = \inf\big\{ j > \mathfrak{R}_{m-1}^n : X_j^n \in \big\{X_1^n, X_2^n, \ldots, X_{j-1}^n\big\}\big\}$$

the mth repeat time of the sequence. For $l \geq 0$, we set

$$F_n(l) := \sum_{j=0}^{l \wedge (\mathfrak{R}_1^n - 1)} \sum_{i > m_n} \frac{p_{ni}}{\sigma_n} \mathbf{1}_{\{X_j^n = i\}}.$$

According to the construction of the birthday tree in [21] and Corollary 3 there, we have

$$\big(D_n, F_n^{\mathcal{L}}(\cdot)\big) \stackrel{d}{=} \big(\mathfrak{R}_1^n - 1, F_n(\cdot)\big). \tag{7.12}$$

Let $q_n \geq 0$ be defined by $q_n^2 = \sum_{i>m_n} p_{ni}^2$. Then (7.3) entails $\lim_{n\to\infty} q_n/\sigma_n = \theta_0$. For $l \geq 0$, we set

$$Z_n(l) := \left| F_n(l) - \frac{q_n^2}{\sigma_n}\big((l+1) \wedge \mathfrak{R}_1^n\big)\right|.$$

We claim that $\sup_{l\geq 0} Z_n(l) \to 0$ in probability as $n \to \infty$. To see this, observe first that

$$Z_n(l) = \left| \sum_{j=0}^{l\wedge(\mathfrak{R}_1^n - 1)} \left(\sum_{i>m_n} \frac{p_{ni}}{\sigma_n} \mathbf{1}_{\{X_j^n = i\}} - \frac{q_n^2}{\sigma_n}\right)\right|,$$

where the terms in the parenthesis are independent, centered, and of variance $\chi_n := \sigma_n^{-2}\sum_{i>m_n} p_{ni}^3 - \sigma_n^{-2} q_n^4 \leq \sigma_n^{-2} q_n^2 p_{nm_n}$. Therefore, Doob's maximal inequality entails that for any fixed number $N > 0$,

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sup_{l\geq 0} Z_n(l)\mathbf{1}_{\{\mathfrak{R}_1^n \leq N/\sigma_n\}}\Big)^2\Big] &\leq \mathbb{E}\left[\left(\sup_{l<\lfloor N/\sigma_n\rfloor} \sum_{j=0}^{l}\left(\sum_{i>m_n} \frac{p_{ni}}{\sigma_n}\mathbf{1}_{\{X_j^n=i\}} - \frac{q_n^2}{\sigma_n}\right)\right)^2\right] \\
&\leq 4N\sigma_n^{-1}\chi_n \\
&\leq 4N \frac{q_n^2}{\sigma_n^2}\frac{p_{nm_n}}{\sigma_n} \to 0,
\end{aligned}$$

by (7.2) and the fact that $q_n/\sigma_n \to \theta_0$. In particular, it follows that

$$\sup_{l\geq 0} Z_n(l)\mathbf{1}_{\{\mathfrak{R}_1^n \leq N/\sigma_n\}} \to 0, \tag{7.13}$$

in probability as $n \to \infty$. On the other hand, the convergence of the $\mathbf{p}_n$-trees in (5.1) implies that the family of distributions of $(\sigma_n D_n, n \geq 1)$ is tight. By (7.12), this entails that

$$\lim_{N\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\mathfrak{R}_1^n > N/\sigma_n\big) = 0. \tag{7.14}$$

Combining this with (7.13) proves the claim.

The generalized distribution function as in (7.11) for the discrete length measure ℓ_n is $l \mapsto l \wedge D_n$. Thus, since $\sup_l Z_n(l) \to 0$ in probability, the identity in (7.12) and $q_n/\sigma_n \to \theta_0$ imply that

$$\mathrm{d}_P\big(\mathcal{L}_n^* \restriction_{R_1^n}, \theta_0^2\sigma_n \ell_n \restriction_{R_1^n}\big) \to 0$$

in probability as $n \to \infty$. This is exactly (7.4) for $k = 1$.

In the general case where $k \geq 1$, we set

$$D_{n,1} := D_n, \qquad D_{n,m} := d_{T^n}\big(b_n(m), V_m^n\big), \qquad m \geq 2,$$

where $b_n(m)$ denotes the branch point of T^n between V_m^n and R_{m-1}^n, i.e., $b_n(m) \in R_{m-1}^n$ such that $[\![r(T^n), V_m^n]\!] \cap R_{m-1}^n = [\![r(T^n), b_n(m)]\!]$. We also define

$$F_{n,1}^{\mathcal{L}}(l) := F_n^{\mathcal{L}} \quad \text{and} \quad F_{n,m}^{\mathcal{L}}(l) := \mathcal{L}_n^*\big(\mathbf{B}\big(b_n(m), l\big) \cap]\!]b_n(m), V_m^n]\!]\big), \qquad m \geq 2.$$

Then conditional on $\mathfrak{R}_k^n$, the vector $(F_{n,1}^{\mathcal{L}}(\cdot), \dots, F_{n,k}^{\mathcal{L}}(\cdot))$ determines the measure $\mathcal{L}_n^* \restriction_{R_k^n}$ for the same reason as before. If we set

$$F_{n,1}(l) := F_n(l) \quad \text{and} \quad F_{n,m}(l) := \sum_{j=\mathfrak{R}_{m-1}^n+1}^{l\wedge(\mathfrak{R}_m^n-1)} \sum_{i>m_n} \frac{p_{ni}}{\sigma_n} \mathbf{1}_{\{X_j^n=i\}}, \qquad m \geq 2,$$

then Corollary 3 of [21] entails the equality in distribution

$$\big(\big(D_{n,m}, F_{n,m}^{\mathcal{L}}(\cdot)\big), 1 \leq m \leq k\big) \stackrel{d}{=} \big(\big(\mathfrak{R}_m^n - \mathfrak{R}_{m-1}^n - 1, F_{n,m}(\cdot)\big), 1 \leq m \leq k\big).$$

Then by the same arguments as before we can show that

$$\max_{1\leq m\leq k} \sup_{l\geq 0} \left| F_{n,m}(l) - \frac{q_n^2}{\sigma_n}\big(l \wedge \big(\mathfrak{R}_m^n - \mathfrak{R}_{m-1}^n - 1\big)\big)\right| \to 0$$

in probability as $n \to \infty$. This then implies (7.4) by the same type of argument as before.

Now let us consider (7.5). The idea is quite similar. For each $M \geq 1$, we set

$$\tilde{Z}_{n,M} := \sum_{j=0}^{\mathfrak{R}_1^n-1} \sum_{M<i\leq m_n} \frac{p_{ni}}{\sigma_n} \mathbf{1}_{\{X_j^n=i\}}.$$

Then

$$\mathbb{E}[\tilde{Z}_{n,M} \mathbf{1}_{\{\mathfrak{R}_1^n \leq N/\sigma_n\}}] \leq N\bigg(\sum_{M<i\leq m_n} \frac{p_{ni}^2}{\sigma_n^2}\bigg).$$

Using (7.3), (H) and the fact that $\sum_i \theta_i^2 < \infty$, we can easily check that for any fixed N,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}[\tilde{Z}_{n,M} \mathbf{1}_{\{\mathfrak{R}_1^n \leq N/\sigma_n\}}] = 0. \tag{7.15}$$

By Markov's inequality, we have

$$\mathbb{P}(\tilde{Z}_{n,M} > \varepsilon) \leq \varepsilon^{-1} \mathbb{E}[\tilde{Z}_{n,M} \mathbf{1}_{\{\mathfrak{R}_1^n \leq N/\sigma_n\}}] + \mathbb{P}\big(\mathfrak{R}_1^n > N/\sigma_n\big).$$

According to (7.14) and (7.15), we can first choose some $N = N(\varepsilon)$ then some $M = M(N(\varepsilon), \varepsilon) = M(\varepsilon)$ such that $\limsup_n \mathbb{P}(\tilde{Z}_{n,M} > \varepsilon) < \varepsilon$. On the other hand, Corollary 3 of [21] says that $\Sigma(n, 1, M)$ is distributed like $\tilde{Z}_{n,M}$. Then we have shown (7.5) for $k = 1$. The general case can be treated in the same way, and we omit the details. □

So far we have completed the proof of Proposition 5.2 in the case where $\boldsymbol{\theta}$ has all strictly positive entries. The other cases are even simpler:

Case 2. Suppose that $\theta_0 = 0$, we take $m_n = n$ and the same argument follows.

Case 3. Suppose that $\boldsymbol{\theta}$ has a finite length I, then it suffices to take $m_n = I$. We can proceed as before.

Acknowledgements

We thank the referees for their careful reading and helpful suggestions. and helpful suggestions. We are grateful to S. Janson and R. Abraham for useful comments on a first version of the paper. We acknowledge partial support from Inria associate team RNA (EQA-021303) and from Agence Nationale de la Recherche grant number ANR-14-CE25-0014 (ANR GRAAL).

References

[1] Abraham, R. and Delmas, J.-F. (2013). Record process on the continuum random tree. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 225–251. MR3083925

[2] Abraham, R. and Delmas, J.-F. (2013). The forest associated with the record process on a Lévy tree. *Stochastic Process. Appl.* **123** 3497–3517. MR3071387

[3] Addario-Berry, L., Broutin, N. and Holmgren, C. (2014). Cutting down trees with a Markov chainsaw. *Ann. Appl. Probab.* **24** 2297–2339. MR3262504

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

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

[6] Aldous, D., Miermont, G. and Pitman, J. (2004). The exploration process of inhomogeneous continuum random trees, and an extension of Jeulin's local time identity. *Probab. Theory Related Fields* **129** 182–218. MR2063375

[7] Aldous, D., Miermont, G. and Pitman, J. (2005). Weak convergence of random p-mappings and the exploration process of inhomogeneous continuum random trees. *Probab. Theory Related Fields* **133** 1–17. MR2197134

[8] Aldous, D. and Pitman, J. (1998). The standard additive coalescent. *Ann. Probab.* **26** 1703–1726. MR1675063

[9] Aldous, D. and Pitman, J. (1999). A family of random trees with random edge lengths. *Random Structures Algorithms* **15** 176–195. MR1704343

[10] Aldous, D. and Pitman, J. (2000). Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Theory Related Fields* **118** 455–482. MR1808372

[11] Aldous, D. and Pitman, J. (2002). Invariance principles for non-uniform random mappings and trees. In *Asymptotic Combinatorics with Application to Mathematical Physics* (*St. Petersburg*, 2001). *NATO Sci. Ser. II Math. Phys. Chem.* **77** 113–147. Kluwer Academic, Dordrecht. MR1999358

[12] Aldous, D.J. (1990). The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.* **3** 450–465. MR1069105

[13] Bertoin, J. (2001). Eternal additive coalescents and certain bridges with exchangeable increments. *Ann. Probab.* **29** 344–360. MR1825153

[14] Bertoin, J. (2002). Self-similar fragmentations. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 319–340. MR1899456

[15] Bertoin, J. and Miermont, G. (2013). The cut-tree of large Galton–Watson trees and the Brownian CRT. *Ann. Appl. Probab.* **23** 1469–1493. MR3098439
[16] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. New York: Wiley. MR1700749
[17] Bismut, J.-M. (1985). Last exit decompositions and regularity at the boundary of transition probabilities. *Z. Wahrsch. Verw. Gebiete* **69** 65–98. MR0775853
[18] Broder, A. (1989). Generating random spanning trees. In *Proc. 30'th IEEE Symp. Found. Comp. Sci.* 442–447. Triangle Park Research: IEEE Computer Society.
[19] Broutin, N. and Wang, M. (2014). Reversing the cut tree of the Brownian CRT. Preprint. Available at arXiv:1408.2924.
[20] Burago, D., Burago, Y. and Ivanov, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Providence, RI: Amer. Math. Soc. MR1835418
[21] Camarri, M. and Pitman, J. (2000). Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.* **5** 18 pp. (electronic). MR1741774
[22] Cayley, A. (1889). A theorem on trees. *Quarterly Journal of Pure and Applied Mathematics* **23** 376–378.
[23] Dieuleveut, D. (2015). The vertex-cut-tree of Galton–Watson trees converging to a stable tree. *Ann. Appl. Probab.* **25** 2215–2262. MR3349006
[24] Dress, A., Moulton, V. and Terhalle, W. (1996). T-theory: An overview. *European J. Combin.* **17** 161–175. MR1379369
[25] Drmota, M., Iksanov, A., Moehle, M. and Roesler, U. (2009). A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures Algorithms* **34** 319–336. MR2504401
[26] Duquesne, T. and Le Gall, J.-F. (2005). Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields* **131** 553–603. MR2147221
[27] Evans, S.N. (2008). *Probability and Real Trees*. *Lecture Notes in Math.* **1920**. Berlin: Springer. MR2351587
[28] Evans, S.N. and Pitman, J. (1998). Construction of Markovian coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **34** 339–383. MR1625867
[29] 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
[30] Evans, S.N. and Winter, A. (2006). Subtree prune and regraft: A reversible real tree-valued Markov process. *Ann. Probab.* **34** 918–961. MR2243874
[31] Fill, J.A., Kapur, N. and Panholzer, A. (2006). Destruction of very simple trees. *Algorithmica* **46** 345–366. MR2291960
[32] Greven, A., Pfaffelhuber, P. and Winter, A. (2009). Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields* **145** 285–322. MR2520129
[33] Gromov, M. (2007). *Metric Structures for Riemannian and Non-Riemannian Spaces*, English ed. *Modern Birkhäuser Classics*. Boston, MA: Birkhäuser. MR2307192
[34] Haas, B. and Miermont, G. (2012). Scaling limits of Markov branching trees with applications to Galton–Watson and random unordered trees. *Ann. Probab.* **40** 2589–2666. MR3050512
[35] Holmgren, C. (2010). Random records and cuttings in binary search trees. *Combin. Probab. Comput.* **19** 391–424. MR2607374
[36] Holmgren, C. (2011). A weakly 1-stable distribution for the number of random records and cuttings in split trees. *Adv. in Appl. Probab.* **43** 151–177. MR2761152
[37] Iksanov, A. and Möhle, M. (2007). A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Commun. Probab.* **12** 28–35. MR2407414

[38] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **288**. Berlin: Springer. MR0959133
[39] Janson, S. (2006). Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29** 139–179. MR2245498
[40] Le Gall, J.-F. (1993). The uniform random tree in a Brownian excursion. *Probab. Theory Related Fields* **96** 369–383. MR1231930
[41] Le Gall, J.-F. and Le Jan, Y. (1998). Branching processes in Lévy processes: The exploration process. *Ann. Probab.* **26** 213–252. MR1617047
[42] Löhr, W. (2013). Equivalence of Gromov–Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces. *Electron. Commun. Probab.* **18** 1–10. MR3037215
[43] Löhr, W., Voisin, G. and Winter, A. (2015). Convergence of bi-measure $\mathbb{R}$-trees and the pruning process. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1342–1368. MR3414450
[44] Lyons, R. and Peres, Y. (2016). *Probability on Trees and Networks*. Cambridge: Cambridge Univ. Press.
[45] Meir, A. and Moon, J.W. (1970). Cutting down random trees. *J. Aust. Math. Soc. A* **11** 313–324. MR0284370
[46] Miermont, G. (2009). Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér.* (4) **42** 725–781. MR2571957
[47] Panholzer, A. (2006). Cutting down very simple trees. *Quaest. Math.* **29** 211–227. MR2233368
[48] Pitman, J. (2001/02). Random mappings, forests, and subsets associated with Abel–Cayley–Hurwitz multinomial expansions. *Sém. Lothar. Combin.* **46** Art. B46h, 45 pp. (electronic). MR1877634
[49] Rényi, A. (1970). On the enumeration of trees. In *Combinatorial Structures and Their Applications* (*Proc. Calgary Internat. Conf., Calgary, Alta.*, 1969) 355–360. New York: Gordon and Breach. MR0263708
[50] Villani, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Berlin: Springer. MR2459454

Received August 2014 and revised July 2015