

# Maximum Agreement Subtrees and Hölder homeomorphisms between Brownian trees

Thomas BUDZINSKI\* & Delphin SÉNIZERGUES†

February 6, 2024

## Abstract

We prove that the size of the largest common subtree between two uniform, independent, leaf-labelled random binary trees of size  $n$  is typically less than  $n^{1/2-\varepsilon}$  for some  $\varepsilon > 0$ . Our proof relies on the coupling between discrete random trees and the Brownian tree and on a recursive decomposition of the Brownian tree due to Aldous. Along the way, we also show that almost surely, there is no  $(1 - \varepsilon)$ -Hölder homeomorphism between two independent copies of the Brownian tree.

## 1 Introduction

**Maximum agreement subtree.** Let  $t, t'$  be two binary trees with  $n$  leaves labelled from 1 to  $n$ . The *maximum agreement subtree* of  $t$  and  $t'$  is the size of the largest subset  $I \subset \{1, \dots, n\}$  such that the subtrees of  $t$  and  $t'$  induced by the labels of  $I$  are the same (as on Figure 1, see also Section 2.1 for precise definitions). This quantity, which will be denoted by  $\text{MAST}(t, t')$ , was introduced by Gordon and Finden [15, 14] in order to measure the compatibility of the outputs of different classifications methods in phylogeny. It is also a generalization of the well studied problem of the longest increasing subsequence of a permutation, and the two problems share a lot of similarities (as noted e.g. in [5]). Since then, it has been studied from algorithmic, extremal and probabilistic points of view. In particular, the quantity  $\text{MAST}(t, t')$  can be computed in polynomial time in  $n$  [29]. On the extremal side, the minimal possible values of  $\text{MAST}(t, t')$  over all pairs  $(t, t')$  of leaf-labelled binary trees of size  $n$  is known to be of order  $\log n$  (the upper bound was proved in [20] and the lower bound in [23]).

**Maximum agreement subtree of random trees.** Another natural question is to understand the *typical* order of magnitude of the maximum agreement subtree, that is, the random variable  $\text{MAST}(T_n, T'_n)$ , where  $T_n$  and  $T'_n$  are random trees of size  $n$ . The most natural model is the one where  $T_n$  and  $T'_n$  are independent and picked uniformly in the set of labelled binary trees of size  $n$ . This model was first investigated by Bryant, McKenzie and Steel [10], who proved by a first moment computation that  $\text{MAST}(T_n, T'_n)$  is  $O(\sqrt{n})$  with high probability. They also provided numerical evidence that  $\text{MAST}(T_n, T'_n)$  should be of order  $n^\beta$  for some  $\beta$  close to  $\frac{1}{2}$ . On the other hand, a polynomial lower bound of order  $n^{1/8}$  was obtained by Bernstein, Ho, Long, Steel, St. John and Sullivant in [7]. This lower bound was recently improved to  $n^{\frac{\sqrt{3}-1}{2}} \approx n^{0.366}$  by Aldous [5] and to  $n^{0.4464}$  by Khezeli [19] in expectation. Finally, we also mention that  $\sqrt{n}$  has been proved to be the right order of magnitude if the trees  $T_n$  and  $T'_n$  are conditioned to have the same shape [25], and that the upper bound in  $\sqrt{n}$  holds robustly for many random trees models arising from branching processes [27].

---

\*ENS de Lyon and CNRS.

†MODAL'X, UPL, Univ. Paris Nanterre, CNRS, F92000 Nanterre France.

thomas.budzinski@ens-lyon.fr  
dsenizer@parisnanterre.fr

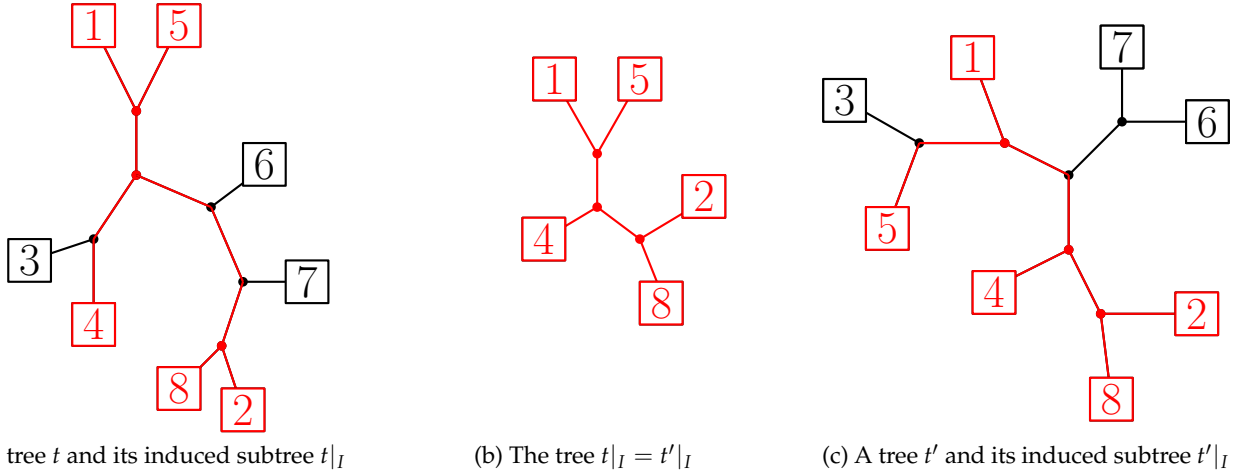


Figure 1: Two labelled binary trees  $t$  and  $t'$  and their largest common subtree, induced by the set  $I = \{1, 2, 4, 5, 8\}$ .

Our main contribution in this paper is to show that the upper bound  $\sqrt{n}$  is actually not optimal in the independent model, which was conjectured by Aldous in [5].

**Theorem 1.** *For all  $n \geq 3$ , let  $T_n$  and  $T'_n$  be two independent uniform labelled binary trees of size  $n$ . There exists a constant  $\varepsilon_1 > 0$  such that we have*

$$\mathbb{P} \left( \text{MAST}(T_n, T'_n) < n^{1/2-\varepsilon_1} \right) \xrightarrow{n \rightarrow +\infty} 1.$$

More explicitly, we find that we can take  $\varepsilon_1 = 10^{-338}$  (see Section 5.1 for a discussion on explicit constants). We have not tried to optimize the constants and this value should be easy to improve, but we do not think that our strategy of proof can give a "reasonable" lower bound (like e.g.  $\varepsilon_1 = 10^{-6}$ ). We also mention that our arguments are sufficient to prove that the probability that  $\text{MAST}(T_n, T'_n)$  exceeds  $n^{1/2-\varepsilon_1}$  is  $O(n^{-a})$  for some  $a > 0$ , and that  $\mathbb{E} [\text{MAST}(T_n, T'_n)] \leq n^{1/2-\varepsilon}$  for some  $\varepsilon > 0$  (see Section 5.2 for a quick discussion).

**Comparison with the Brownian tree.** As recalled before, it is proved in [25] that the MAST of two trees of the same shape is typically of order  $\sqrt{n}$ . Therefore, our strategy relies on the fact that two independent large random trees have "different shapes" at every scale. To formalize this, we make heavy use of the continuous scaling limit of  $T_n$ , which exhibits nice scale invariance properties, and on which more explicit computations can be performed.

More precisely, we denote by  $\mathcal{T}$  the *Brownian tree*, which is the scaling limit of  $T_n$ , seen as a measured metric space, where distances have been normalized by  $\sqrt{n}$  and masses by  $n$ . This compact, continuous random tree with fractal dimension 2 was introduced in [2] and can be built in a natural way from a normalized Brownian excursion (see Section 2.2 for complete definitions). It also has the important property that its branching points all have degree 3. We highlight that comparisons between the discrete trees  $T_n$  and the continuous object  $\mathcal{T}$  already play an important role in the proofs of the lower bounds of [5] and [19].

**Hölder homeomorphisms of Brownian tree.** Since proving Theorem 1 requires to compare the shapes of two independent copies of  $\mathcal{T}$ , we obtain along the way the following result of independent interest.

**Theorem 2.** *Let  $\mathcal{T}$  and  $\mathcal{T}'$  be two independent copies of the Brownian tree. There exists a constant  $\varepsilon_2 > 0$  such that almost surely, there is no  $(1 - \varepsilon_2)$ -Hölder homeomorphism from  $\mathcal{T}$  to  $\mathcal{T}'$ .*

Just like in Theorem 1, we find that we can take  $\varepsilon_2 = 10^{-338}$ , which we did not try to optimize. Although none of Theorems 1 and 2 easily implies the other, they are closely related to each other. Indeed, as can be seen on Figure 1, a common subtree of two trees  $T_n$  and  $T'_n$  gives a "correspondence" between a part of  $T_n$  and a part of  $T'_n$ , which can be extended to a homeomorphism in the continuous limit. This is not a completely new idea, as the arguments of [5] (and the improvements done in [19]) can already be interpreted as a proof of the existence of a homeomorphism from  $\mathcal{T}$  to  $\mathcal{T}'$  with a certain Hölder exponent. As we check in Theorem 22 in the Appendix below, the actual Hölder exponent given by [5] turns out to be  $5 - 2\sqrt{6} \approx 0.1010$ . More generally, statements similar to Theorem 2 on very different objects appear under the name of *Hölder equivalence* in the geometry literature. In geometry, this problem is often of the following form: given a metric space  $X$  that is homeomorphic to  $\mathbb{R}^n$ , what is the optimal Hölder exponent of a homeomorphism from  $\mathbb{R}^n$  to  $X$ ? An immediate upper bound is  $\frac{n}{\dim_H(X)}$ , where  $\dim_H(X)$  is the Hausdorff dimension of  $X$ . We refer to [16] for improved upper bounds in specific contexts such as sub-Riemannian or contact manifolds. However, we are in a very different setting here, as the Brownian trees involved are not manifolds, and so our arguments do not share any commonalities. Another difference between our setting and the one studied by Gromov is that we prove that the Hölder exponent cannot be arbitrarily close to 1 in a context where both sides of the homeomorphism have the same Hausdorff dimension.

We also note that Theorem 2 becomes quite easy if "(1 -  $\varepsilon$ )-Hölder" is replaced by "Lipschitz"<sup>1</sup>. Finally, we remark that our results have a similar flavor to those proved in [6, Theorem 1.2, Theorem 1.7] for a quite different model (largest increasing subsequence of a random separable permutation). More precisely, the proofs in [6] consist of showing that a random tree cannot contain a large subtree satisfying some properties, which improves on the first moment upper-bound and is achieved by comparison with continuous objects. The very recent preprint [9] improves their result (and also provides some lower-bound) using some careful analysis on the Brownian tree and its associated fragmentation process.

**Recursive decomposition of the Brownian tree.** In order to highlight more precisely what Theorem 1 and Theorem 2 have in common, we introduce an important tool in our proofs, which already crucially appears in [5]. This is a randomized recursive decomposition of the Brownian tree  $\mathcal{T}$ , which was introduced by Aldous [4]. The decomposition consists in picking 3 random uniform points  $X_1, X_2, X_3$  in  $\mathcal{T}$ , blowing up  $\mathcal{T}$  into three pieces at the unique branching point that separates  $X_1, X_2, X_3$ , and iterating the decomposition in each of the three pieces (see Section 2.5 for complete definitions). After  $k$  steps, we obtain a (randomized) partition of  $\mathcal{T}$  into  $3^k$  regions, indexed by a set  $\mathbb{T}_3^k$ . We denote those regions by  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3^k}$ . This decomposition enjoys very nice independence properties that we will heavily rely on.

We can now state the key result that we will use to prove both Theorem 1 and Theorem 2. It roughly states that a homeomorphism between two independent realizations  $\mathcal{T}, \mathcal{T}'$  of the Brownian tree cannot be "almost measure-preserving", in the sense that it has to send "most" regions of  $\mathcal{T}$  to regions of  $\mathcal{T}'$  with a much smaller mass. We will denote by  $|A|$  the measure of a subset  $A$  of a Brownian tree.

**Proposition 3.** *Let  $\mathcal{T}$  and  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3^k}$  be a Brownian tree and its recursive decomposition, and let  $\mathcal{T}'$  be an independent copy of  $\mathcal{T}$ . There exist constants  $\xi, \eta > 0$  such that almost surely the following holds for  $k$  large enough: For any homeomorphism  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$ , if we define the set  $U^{k, \Psi} \subset \mathbb{T}_3^k$  of indices as*

$$U^{k, \Psi} := \left\{ \mathbf{i} \in \mathbb{T}_3^k \mid |\Psi(\mathcal{R}[\mathbf{i}])| > e^{-\eta k} |\mathcal{R}[\mathbf{i}]| \right\},$$

*then the subset  $E^{k, \Psi} := \bigcup_{\mathbf{i} \in U^{k, \Psi}} \mathcal{R}[\mathbf{i}]$  of  $\mathcal{T}$  has measure at most  $e^{-\xi k}$ .*

---

<sup>1</sup>For example, take a large branching point  $b_1$  of  $\mathcal{T}$ , and consider an "exceptional scale"  $\delta$  at which  $b_1$  is unusually close to another branching point  $b_2$ . Then  $\Psi(b_1)$  would have to be a large branching point of  $\mathcal{T}'$ , and  $\Psi(b_2)$  would have to be a branching point of  $\mathcal{T}'$  (of scale  $\approx \delta$ ) very close to  $\Psi(b_1)$ , which is unlikely to exist.

Theorem 2 will follow almost immediately from this result. On the other hand, in order to deduce Theorem 1 from Proposition 3, we will rely on the nice coupling existing between the discrete random tree  $T_n$  and the continuous one  $\mathcal{T}$ . We can then argue that a common subtree between  $T_n$  and  $T'_n$  can be extended to a homeomorphism  $\Psi$  from  $\mathcal{T}$  to  $\mathcal{T}'$ . Proposition 3 guarantees that for most of the regions  $\mathcal{R}[i]$ , either  $\mathcal{R}[i]$  or  $\Psi(\mathcal{R}[i])$  is very small, so only few labels can appear in both  $\mathcal{R}[i]$  and in  $\Psi(\mathcal{R}[i])$  when  $T_n$  is coupled to  $\mathcal{T}$  and  $T'_n$  to  $\mathcal{T}'$ . We will conclude by using the classic square root upper bound for each of those regions (Lemma 19).

**Ideas of the proof of Proposition 3.** Finally, let us mention some of the ideas behind the proof of Proposition 3. The proof roughly consists in showing that a certain multiscale exploration of the tree  $\mathcal{T}$  has many "mismatches" with the analog exploration in  $\mathcal{T}'$ , which we believe is the main innovation of the present work. Fix a typical point  $x \in \mathcal{T}$ , and imagine that we try to build a "good" homeomorphism from  $\mathcal{T}$  to  $\mathcal{T}'$ . By looking at smaller and smaller regions of the recursive decomposition around the point  $x$ , we can encode the masses of a sequence of nested neighbourhoods of  $x$  by a sequence  $(f_j(x))_{j \geq 0}$  of i.i.d. numbers, where  $j$  represents decreasing scales. We will argue that it is not possible to find  $x' \in \mathcal{T}'$  such that  $f_j(x')$  is very close to  $f_j(x)$  for most of the scales  $j$ . This will imply that the ratio between the mass of a small region around  $x$  and the mass of its image around  $\Psi(x)$  cannot stay "stationary" as the scale of that region decreases to 0. By a "martingale-like" argument, we will conclude that this ratio must decay quickly, which will yield Proposition 3. This is somewhat reminiscent of some ideas of [6], in the sense that "finding a large substructure is difficult because some positive proportion of the mass is lost at every scale".

**Structure of the paper.** In Section 2, we will introduce all the definitions of discrete and continuous objects that will be needed in the proofs, as well as some basic properties of the Brownian tree and of its recursive decomposition. Section 3 is the central part of the paper and is devoted to the proof of Proposition 3, which represents most of the work. In Section 4, we conclude the proofs of the main theorems. In Section 5, we discuss the quantitative values of  $\varepsilon_1$  and  $\varepsilon_2$  provided by our proof, as well as some open questions. In Appendix A, we construct a Hölder homeomorphism between  $\mathcal{T}$  and  $\mathcal{T}'$ .

**Acknowledgments.** The authors are grateful to Omer Angel for many interesting discussions about various aspects of the model studied in this paper. They would also like to thank Nicolas Curien for useful discussions, and Valentin Féray for noticing the analogy with [6, 9]. They also thank the anonymous referees for their constructive comments and remarks. D.S.'s research has been conducted within the FP2M federation (CNRS FR 2036).

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Definitions and preliminaries</b>	<b>5</b>
2.1	Labelled binary trees and maximum agreement subtree . . . . .	5
2.2	The Brownian tree . . . . .	5
2.3	Dirichlet and Beta distributions . . . . .	6
2.4	Coupling between discrete and continuum random trees . . . . .	7
2.5	The Aldous recursive decomposition . . . . .	9
<b>3</b>	<b>Proof of Proposition 3</b>	<b>13</b>
3.1	Finding mismatches . . . . .	13
3.2	The martingale argument . . . . .	19

<b>4</b>	<b>Proofs of the main results</b>	<b>22</b>
4.1	Hölder homeomorphisms	22
4.2	Maximum agreement subtree bound	23
4.3	The refined square root bound	25
<b>5</b>	<b>Discussion of the results</b>	<b>27</b>
5.1	Explicit constants	27
5.2	Remarks and open questions	28
<b>A</b>	<b>Construction of a Hölder homeomorphism</b>	<b>29</b>

## 2 Definitions and preliminaries

### 2.1 Labelled binary trees and maximum agreement subtree

A *finite tree* is a finite, connected graph with no cycles. A *finite binary tree*  $t$  is a finite tree where all the vertices have degree either 3 or 1. The vertices of degree 3 will be called the *nodes* of  $t$ , and the vertices of degree 1 will be called its *leaves*. A *labelled binary tree of size*  $n \geq 2$  is a finite binary tree with exactly  $n$  leaves (and therefore  $n - 2$  nodes), some of which are labelled by integers so that each label appears at most once. By convention, we also say that a single vertex with no edge is a binary tree of size 1, and the empty tree is a binary tree of size 0.

For  $n \geq 1$ , we denote by  $\mathcal{B}_n$  the set of such trees where all  $n$  leaves are labelled and the labels are  $1, \dots, n$ , up to isomorphism (these are sometimes called *cladograms* in the literature). We highlight that the trees that we consider are not rooted, and they are not plane trees (i.e. there is no clockwise ordering of the neighbours of a fixed vertex).

We recall that for all  $n \geq 3$ , we have

$$\#\mathcal{B}_n = (2n - 5)!! = 1 \cdot 3 \cdots (2n - 7) \cdot (2n - 5).$$

In all the paper, we will denote by  $T_n$  a uniform random variable on  $\mathcal{B}_n$ . We will also denote by  $T'_n$  an independent copy of  $T_n$ . If  $t \in \mathcal{B}_n$  and  $I$  is a subset of  $\{1, \dots, n\}$ , we will denote by  $t|_I$  the *subtree of  $t$  induced by  $I$* . More precisely, this is the labelled binary tree obtained from  $t$  by keeping only the paths joining the labels of  $I$  together, and by contracting the vertices of degree 2 that may appear in the process (see Figure 1). Note that  $t|_I$  does not belong to  $\mathcal{B}_{\#I}$ , unless  $I = \{1, \dots, k\}$ . If  $t, t' \in \mathcal{B}_n$ , we write

$$\text{MAST}(t, t') = \max \{ \#I \mid I \subset \{1, \dots, n\} \text{ such that } t|_I = t'|_I \},$$

where by  $t|_I = t'|_I$  we mean that  $t|_I$  and  $t'|_I$  are isomorphic as leaf-labelled trees.

If  $v_1, v_2$  are nodes of a finite binary tree, a *region  $r$  of  $t$  delimited by  $v_1, v_2$*  is a connected component of the forest obtained by blowing up the nodes  $v_1$  and  $v_2$  into three leaves each. In particular, it is a labelled binary tree, where the leaves coming from  $v_1$  or  $v_2$  are unlabelled. We write  $\#r$  for the number of original leaves of  $t$  that belong to the region  $r$  and call this quantity the *size* of  $r$ . If  $t, t'$  are two labelled binary trees and if  $r \subset t$  and  $r' \subset t'$  are two such regions, we may consider the quantity  $\text{MAST}(r, r')$ , which is the size of the largest subset  $I$  of  $\{1, \dots, n\}$  such that all the elements of  $I$  appear both in  $r$  and in  $r'$ , and  $r|_I = r'|_I$ .

### 2.2 The Brownian tree

We start with the construction of the Brownian tree  $\mathcal{T}$  introduced in [2, 3]. Let  $\mathbf{e} = (e_t)_{0 \leq t \leq 1}$  be a normalized Brownian excursion (that is, a Brownian motion conditioned to stay positive in  $(0, 1)$  and hit 0 at time 1). For  $s, t \in [0, 1]$ , we write  $s \sim_{\mathbf{e}} t$  if

$$e_s = e_t = \min_{[s \wedge t, s \vee t]} e_u,$$

where  $s \wedge t$  and  $s \vee t$  stand respectively for  $\min(s, t)$  and  $\max(s, t)$ . We also write

$$d_e(s, t) = e_s + e_t - 2 \min_{[s \wedge t, s \vee t]} e_u.$$

Then  $d_e$  is a pseudo-distance on  $[0, 1]$ , i.e. it is symmetric and satisfies the triangle inequality. Moreover, we have  $d_e(s, t) = 0$  if and only if  $s \sim_e t$ . Then the quotient  $\mathcal{T} = [0, 1] / \sim_e$  equipped with  $d_e$  is a random compact metric space, which we call the *Brownian tree*. Moreover  $\mathcal{T}$  carries a natural probability measure, which is the pushforward of the Lebesgue measure on  $[0, 1]$  under the canonical map from  $[0, 1]$  to  $\mathcal{T}$ . We will denote by  $|\cdot|$  this probability measure on  $\mathcal{T}$ , which has full support and no atom. We recall that the metric space  $\mathcal{T}$  is almost surely a real tree, i.e. for all  $x, y$  in  $\mathcal{T}$ , there is a unique injective, continuous path from  $x$  to  $y$ , and this path is a geodesic. For  $m > 0$ , we say that a random measured metric space  $(\mathcal{E}, d, \mu)$  is a *Brownian tree of mass  $m$*  if  $(\mathcal{E}, m^{-1/2}d, m^{-1}\mu)$  has the law of  $\mathcal{T}$ . Note in particular that  $\mathcal{T}$  is a Brownian tree with mass 1.

Let  $k \geq 1$  and let  $(\mathcal{E}, d, \mu)$  be a Brownian tree with mass  $m$ . Conditionally on  $(\mathcal{E}, d, \mu)$ , let  $X_1, \dots, X_k$  be  $k$  i.i.d. points of  $\mathcal{E}$ , sampled according to (a normalized version of)  $\mu$ . Note that almost surely, the points  $X_i$  are all leaves, i.e.  $\mathcal{E} \setminus \{X_i\}$  is connected. We call  $(\mathcal{E}, d, \mu, X_1, \dots, X_k)$  a  *$k$ -pointed Brownian tree with mass  $m$* .

In this work, decompositions of  $\mathcal{T}$  into several regions will play a crucial role. For this purpose, we will call *region of  $\mathcal{T}$*  the closure of a connected component of  $\mathcal{T} \setminus F$ , where  $F$  is a subset of  $\mathcal{T}$  of cardinal 0, 1 or 2.

We recall that  $\mathcal{T}$  is almost surely a *binary* real tree, i.e. for all  $x \in \mathcal{T}$ , the space  $\mathcal{T} \setminus \{x\}$  has at most three connected components. A point  $b \in \mathcal{T}$  such that  $\mathcal{T} \setminus \{b\}$  has exactly three connected components will be called a *branching point* of  $\mathcal{T}$ . Moreover, we define the *size* of a branching point  $b$  as the measure of the smallest of the three connected components of  $\mathcal{T} \setminus \{b\}$ , and denote this quantity by  $|b|_{\mathcal{T}}$ . Similarly, if  $\mathcal{R}$  is a region of  $\mathcal{T}$  and  $b \in \mathcal{R}$  is such that  $\mathcal{R} \setminus \{b\}$  has three connected components, the *relative size* of  $b$  in  $\mathcal{R}$  is the measure of the smallest such connected component, and is denoted by  $|b|_{\mathcal{R}}$ .

### 2.3 Dirichlet and Beta distributions

We recall here the definition of the Dirichlet and Beta distributions. For  $(a_1, \dots, a_d) \in (0, \infty)^d$  the *Dirichlet distribution*  $\text{Dir}(a_1, \dots, a_d)$  is the probability measure on the simplex

$$\left\{ (m_1, \dots, m_d) \mid \forall 1 \leq i \leq d, \quad m_i \geq 0 \quad \text{and} \quad \sum_{i=1}^d m_i = 1 \right\}$$

with density

$$\frac{\Gamma(a_1 + \dots + a_d)}{\Gamma(a_1) \dots \Gamma(a_d)} \prod_{i=1}^d m_i^{a_i - 1}$$

with respect to the Lebesgue measure. For  $(W_1, W_2) \sim \text{Dir}(a, b)$ , the first coordinate  $W_1$  is said to have the Beta( $a, b$ ) distribution.

The Dirichlet distribution enjoys two properties that we use throughout the paper. The first one can be derived from the so-called *beta-gamma algebra* results developed in [13]; for the second one we refer to [1, Lemma 17]. Suppose  $(W_1, \dots, W_d) \sim \text{Dir}(a_1, \dots, a_d)$ . Then

- For any  $1 \leq i \leq d - 1$ , we have

$$\left\{ \begin{array}{l} W_1 + \dots + W_i \sim \text{Beta}(a_1 + \dots + a_i, a_{i+1} + \dots + a_d) \\ (W_1 + \dots + W_i)^{-1} \cdot (W_1, \dots, W_i) \sim \text{Dir}(a_1, \dots, a_i) \\ (W_{i+1} + \dots + W_d)^{-1} \cdot (W_{i+1}, \dots, W_d) \sim \text{Dir}(a_{i+1}, \dots, a_d) \end{array} \right. \quad (1)$$

and the three random variables appearing in the last display are independent.



- Let  $I$  be such that  $\mathbb{P}(I = i \mid (W_1, \dots, W_d)) = W_i$ . Then for any  $i \in \{1, \dots, d\}$  we have

$$\mathbb{P}(I = i) = \frac{a_i}{a_1 + \dots + a_d} \quad \text{and} \quad \text{Law}((W_1, \dots, W_d) \mid I = i) = \text{Dir}(a_1, \dots, a_i + 1, \dots, a_d). \quad (2)$$

## 2.4 Coupling between discrete and continuum random trees

We recall the classical coupling between the discrete tree  $T_n$  and the continuous tree  $\mathcal{T}$ . Let  $n \geq 3$  and let  $(e_i)_{1 \leq i \leq 2n-3}$  be an enumeration of the edges of  $T_n$  such that for  $1 \leq i \leq n$ , the edge  $e_i$  is the unique edge incident to the leaf labelled  $i$ .

Let  $(W_i)_{1 \leq i \leq 2n-3}$  be a random vector with distribution  $\text{Dir}(\frac{1}{2}, \dots, \frac{1}{2})$ , independent from  $T_n$ . Conditionally on  $T_n$  and  $(W_i)$ , let  $(\mathcal{T}_i, X_i^1, X_i^2)_{1 \leq i \leq 2n-3}$  be independent bi-pointed Brownian trees with respective masses  $W_i$ . For all  $i$ , let  $v_i^1$  and  $v_i^2$  be the two endpoints of the edge  $e_i$  of  $T_n$ , with the convention that if one of the endpoints is a leaf, then it is  $v_i^1$ . For  $1 \leq i, i' \leq 2n-3$  and  $j, j' \in \{1, 2\}$ , we write  $X_i^j \sim X_{i'}^{j'}$  if  $v_i^j = v_{i'}^{j'}$ , and write

$$\tilde{\mathcal{T}} = \left( \bigcup_{i=1}^{2n-3} \mathcal{T}_i \right) / \sim.$$

As a metric space, this quotient is understood as the "metric gluing" of the  $\mathcal{T}_i$ 's in the sense of [11]. The measure on  $\mathcal{T}$  is straightforwardly obtained from those of the  $\tilde{\mathcal{T}}_i$ , and its total mass is  $\sum_{i=1}^{2n-3} W_i = 1$ .

The following result can be found for  $n = 3$  in [4]. Even though the corresponding result for  $n > 3$  is folklore, we were not able to find a statement in the literature that exactly matches the one that we use here. However, it can be seen as a consequence of the discrete result proven in [26, Exercise 7.4.12 and Exercise 7.4.13] (the proof relies on the Rémy algorithm [28] to build uniform binary trees, and the Dirichlet distribution comes from a Pólya urn argument).

**Theorem 4.** [4, 26] *The  $n$ -pointed measured metric space  $(\tilde{\mathcal{T}}, X_1^1, \dots, X_n^1)$  has the law of an  $n$ -pointed Brownian tree with mass 1.*

See Figure 2 for an illustration of this construction. In particular, the combinatorial structure of the paths joining  $n$  uniform points  $X_1, \dots, X_n$  of  $\mathcal{T}$  is that of  $T_n$ . In the rest of the paper, we will always consider that the continuous tree  $\mathcal{T}$  and the discrete tree  $T_n$  are coupled in this way.

We conclude this section with a lemma comparing the mass of a region in  $\mathcal{T}$  and some corresponding region in  $T_n$ .

**Lemma 5.** *Let  $\varepsilon > 0$  and let  $(\mathcal{T}, X_1, \dots, X_n)$  be an  $n$ -pointed Brownian tree. Then, with probability  $1 - o_n(1)$ , for any region  $\mathcal{R}$  of  $\mathcal{T}$  delimited by at most two branching points, denoting  $R$  the smallest region of  $T_n$  that contains all the leaves with label  $j \in \{1, \dots, n\}$  such that  $X_j \in \mathcal{R}$ , we have the following bound*

$$\#\mathcal{R} \leq n^\varepsilon \vee (n^{1+\varepsilon} \cdot |\mathcal{R}|).$$

*Proof.* Let  $\mathcal{R}$  be a region of  $\mathcal{T}$  delimited by at most two branching points. This way the topological frontier  $\partial\mathcal{R}$  of the region  $\mathcal{R}$  contains at most two elements. With the notation just above, for any  $1 \leq i \leq 2n-3$ , we denote by  $\overset{\circ}{\mathcal{T}}_i$  the interior of the region  $\mathcal{T}_i$ , seen as a subset of  $\mathcal{T}$ . We introduce

$$I = \left\{ i \in \{1, \dots, 2n-3\} \mid \overset{\circ}{\mathcal{T}}_i \cap \mathcal{R} \neq \emptyset \right\}.$$

We note that the set of edges  $\{e_i, i \in I\}$  defines a region  $\tilde{R}$  of  $T_n$  delimited by at most two nodes; we denote  $N := \#\tilde{R}$ . By definition of  $I$ , the region  $\tilde{R}$  contains all the leaves with labels  $j \in \{1, \dots, n\}$

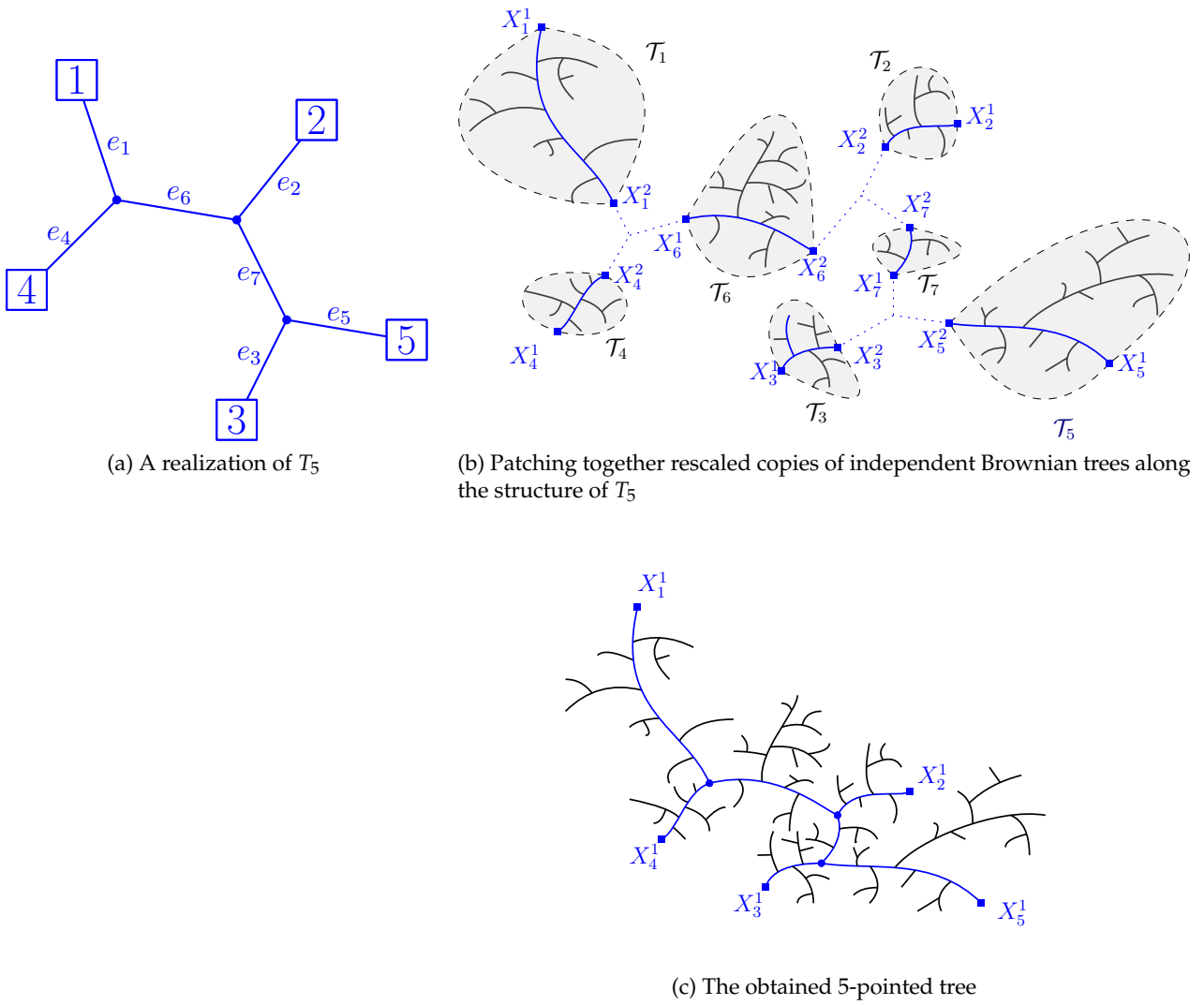


Figure 2: Coupling between the 5-pointed Brownian tree  $(\tilde{\mathcal{T}}, X_1^1, \dots, X_5^1)$  and  $T_5$ . The combinatorial structure of the paths joining the distinguished points, shown in blue, is that of the discrete tree  $T_5$  we started with.



that are such that  $X_j \in \mathcal{R}$ , so  $R \subset \tilde{R}$  and so  $\#R \leq \#\tilde{R} = N$ . If  $N < n^\varepsilon$  we have nothing to prove, so let us assume  $N \geq n^\varepsilon$ . Remark that since  $\tilde{R}$  is a region of  $T_n$ , it has at most two unlabelled leaves, so its number of edges satisfies  $\#I \in \{2N - 3, 2N - 1, 2N + 1\}$ , depending on whether it is delimited by 0 or 1 or 2 nodes.

Now, for any  $i \in I$ , if  $\overset{\circ}{T}_i \cap \mathcal{R} \neq \overset{\circ}{T}_i$  then it means that  $(\partial\mathcal{R}) \cap \overset{\circ}{T}_i \neq \emptyset$ . Since  $\partial\mathcal{R}$  has cardinality at most 2, this can happen for at most two such  $i \in I$ , say  $i_1$  and  $i_2$  (pick them arbitrarily if only 0 or 1 such value exists). This entails  $\mathcal{R} \supset \bigcup_{i \in I \setminus \{i_1, i_2\}} \overset{\circ}{T}_i$ .

Now, let  $(W_i)_{1 \leq i \leq 2n-3}$  be as in the coupling of Theorem 4 (that is,  $W_i$  is the mass of the set of those points  $x \in \overset{\circ}{T}$  such that  $e_i$  is the closest edge of  $T_n$  from  $x$ ). From the above reasoning, we have

$$|\mathcal{R}| \geq \sum_{i \in I \setminus \{i_1, i_2\}} W_i. \quad (3)$$

We now condition on the discrete tree  $T_n$ . We recall that  $(W_i)_{1 \leq i \leq 2n-3}$  is independent of  $T_n$  and has Dirichlet distribution, so according to (1) we have

$$\sum_{i \in I \setminus \{i_1, i_2\}} W_i \sim \text{Beta} \left( \frac{\#I - 2}{2}, \frac{2n - 1 - \#I}{2} \right)$$

conditionally on  $T_n$ . Let us now fix the region  $\tilde{R}$  of  $T_n$ . Using the explicit density of the Beta distribution, we can write, recalling that  $N = \#\tilde{R} \geq n^\varepsilon$  and  $2N - 3 \leq \#I \leq 2N + 1$ :

$$\begin{aligned} \mathbb{P} \left( \sum_{i \in I \setminus \{i_1, i_2\}} W_i \leq n^{-1-\varepsilon} \cdot N \mid T_n \right) &= \frac{\Gamma \left( \frac{2n-3}{2} \right)}{\Gamma \left( \frac{\#I-2}{2} \right) \Gamma \left( \frac{2n-1-\#I}{2} \right)} \int_0^{n^{-1-\varepsilon} \cdot N} x^{\frac{\#I-4}{2}} \cdot (1-x)^{\frac{2n-3-\#I}{2}} dx. \\ &\leq \frac{n^{\#I/2+1}}{(\#I/2)^{(\#I)/2} e^{O(\#I)}} \cdot (n^{-1-\varepsilon} N)^{\frac{\#I-4}{2}} \\ &\leq n^{-\varepsilon n^\varepsilon + O\left(\frac{n^\varepsilon}{\log n}\right)}, \end{aligned}$$

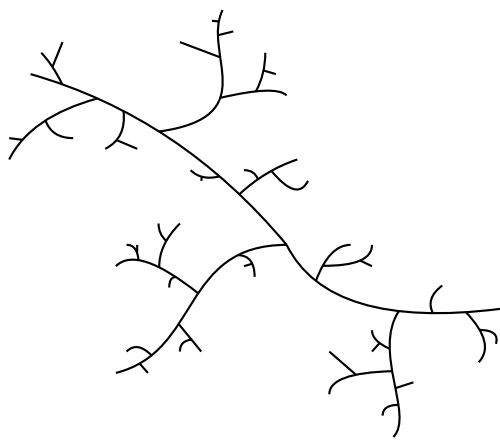
which decays faster than any polynomial in  $n$ . In the above computation we have used the fact that  $\frac{\Gamma(x)}{\Gamma(x-k)} \leq x^k$  for any  $x > 0$  and any integer  $k$  such that  $x - k > 0$ , and in the end our assumption that  $N \geq n^\varepsilon$ . Still conditionally on  $T_n$ , we can perform a union-bound over all the  $O(n^4)$  possibilities for choosing the region  $R$  and the two labels  $i_1, i_2$  corresponding to the removed edges. Combining this with (3), with high probability, for any region  $\mathcal{R}$  and corresponding  $I, i_1, i_2, N$ , we have

$$|\mathcal{R}| \geq \sum_{i \in I \setminus \{i_1, i_2\}} W_i \geq n^{-1-\varepsilon} \cdot N \geq n^{-1-\varepsilon} \cdot \#\tilde{R} \geq n^{-1-\varepsilon} \cdot \#R,$$

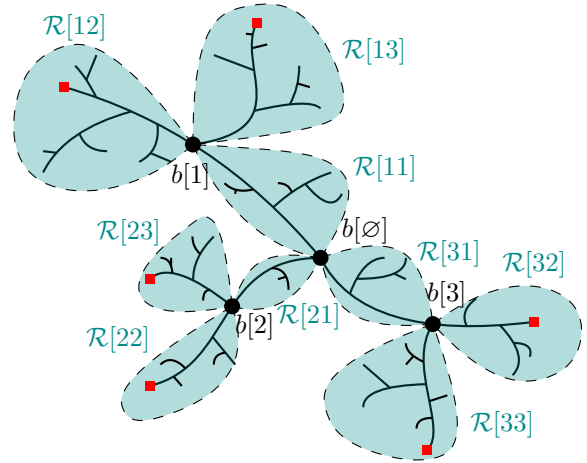
which is what we wanted to prove.  $\square$

## 2.5 The Aldous recursive decomposition

**Decomposing  $\mathcal{T}$  into three regions.** We now introduce a recursive decomposition of the Brownian tree  $\mathcal{T}$ , which consists of repeatedly applying the above decomposition for  $n = 3$ . More precisely, let  $(\mathcal{T}, x[1], x[2], x[3])$  be a 3-pointed Brownian tree with mass 1. Note that almost surely, the points  $x[1]$ ,  $x[2]$  and  $x[3]$  are leaves. In this case, there exists a unique branching point  $b[\emptyset]$  of  $\mathcal{T}$  such that  $x[1]$ ,  $x[2]$  and  $x[3]$  lie in three different connected components of  $\mathcal{T} \setminus \{b[\emptyset]\}$ . For  $i \in \{1, 2, 3\}$ , we call  $\mathcal{R}[i]$  the (closure in  $\mathcal{T}$  of the) connected component containing  $x[i]$ . Then Theorem 4 for  $n = 3$  implies that the vector  $(|\mathcal{R}[1]|, |\mathcal{R}[2]|, |\mathcal{R}[3]|)$  has distribution  $\text{Dir} \left( \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right)$  and that conditionally on this vector, the three regions  $(\mathcal{R}[i], b[\emptyset], x[i])$  are independent bi-pointed Brownian trees with prescribed masses. We will then pick a third point uniformly at random in each  $\mathcal{R}[i]$  and apply this decomposition recursively.

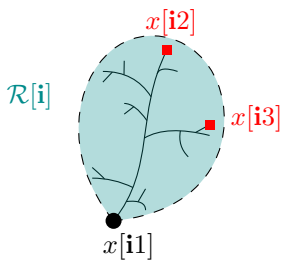


(a) A realization of the Brownian tree  $\mathcal{T}$ .

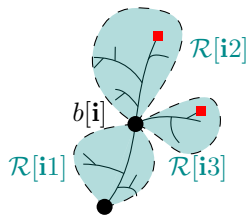


(b) Its recursive decomposition up to depth 2.

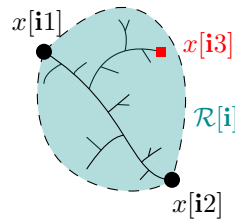
Figure 3: A decomposed Brownian tree  $\mathcal{T}$  and its decomposition  $(\mathcal{R}[i])_{i \in \mathbb{T}_3^2}$  of depth 2.



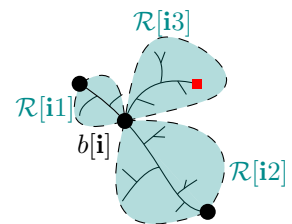
(a) A region delimited by one point.



(b) Its decomposition.



(c) A region delimited by two points.



(d) Its decomposition.

Figure 4: One step of the decomposition for a region  $\mathcal{R}[i]$  delimited by respectively one point and two points. Note that in both cases, the newly created region  $\mathcal{R}[i3]$  is delimited by only one point.

**The complete ternary tree.** More precisely, let

$$\mathbb{T}_3 = \bigcup_{k \geq 0} \{1, 2, 3\}^k$$

be the set of finite words on the alphabet  $\{1, 2, 3\}$ . We will generally use the notation  $\mathbf{i} = i_1 i_2 \dots i_k$  to denote an element of  $\mathbb{T}_3$ . For such a word we call  $k$  the *depth* of  $\mathbf{i}$ , and denote by  $\mathbb{T}_3^k$  the set of elements of  $\mathbb{T}_3$  of depth  $k$ . The set  $\mathbb{T}_3$  can be seen as the complete ternary tree, with root  $\emptyset$  and where the parent of  $i_1 i_2 \dots i_k$  is  $i_1 i_2 \dots i_{k-1}$ . For any  $k \geq 0$  and any word  $\mathbf{i} \in \mathbb{T}_3$  with depth at least  $k$ , we will write  $\mathbf{i}_k = i_1 \dots i_k$ . Finally, we will use concatenation of words of  $\mathbb{T}_3$ : if  $\mathbf{i} = i_1 \dots i_k$  and  $\mathbf{j} = j_1 \dots j_\ell$ , we will write  $\mathbf{ij}$  for  $i_1 \dots i_k j_1 \dots j_\ell$ .

**The recursive decomposition of the Brownian tree.** Our decomposition will associate to each word  $\mathbf{i}$  of  $\mathbb{T}_3$  a region  $\mathcal{R}[\mathbf{i}]$  of  $\mathcal{T}$  in such a way that for all  $\mathbf{i} \in \mathbb{T}_3$ , the regions  $\mathcal{R}[\mathbf{i}1]$ ,  $\mathcal{R}[\mathbf{i}2]$  and  $\mathcal{R}[\mathbf{i}3]$  form a partition of  $\mathcal{R}[\mathbf{i}]$ . More precisely, we write  $\mathcal{R}[\emptyset] = \mathcal{T}$  and we define the branching point  $b[\emptyset]$  and the bi-pointed trees  $\mathcal{R}[1]$ ,  $\mathcal{R}[2]$  and  $\mathcal{R}[3]$  as in the previous paragraph. Moreover, let  $\mathbf{i} \in \mathbb{T}_3$  of depth  $k \geq 1$  and assume that the region  $\mathcal{R}[\mathbf{i}]$  has been defined, and call  $x[\mathbf{i}1]$ ,  $x[\mathbf{i}2]$  its marked points. Conditionally on all the rest, let  $x[\mathbf{i}3]$  be a random point sampled according to the mass measure on  $\mathcal{R}[\mathbf{i}]$ . We denote by  $b[\mathbf{i}]$  the unique branching point of  $\mathcal{R}[\mathbf{i}]$  such that  $x[\mathbf{i}1]$ ,  $x[\mathbf{i}2]$  and  $x[\mathbf{i}3]$  lie in three different connected components of  $\mathcal{R}[\mathbf{i}] \setminus \{b[\mathbf{i}]\}$ , and we call the closures of these three components respectively  $\mathcal{R}[\mathbf{i}1]$ ,  $\mathcal{R}[\mathbf{i}2]$  and  $\mathcal{R}[\mathbf{i}3]$ . Then, for  $a \in \{1, 2, 3\}$ , we set  $x[\mathbf{i}a1] := b[\mathbf{i}]$  and  $x[\mathbf{i}a2] := x[\mathbf{i}a]$  (that is, we mark two points on each of the regions  $\mathcal{R}[\mathbf{i}a]$ , in the same way as in the first step). Finally, we equip the regions  $\mathcal{R}[\mathbf{i}a]$  with the metric and the measure naturally inherited from  $\mathcal{R}[\mathbf{i}]$ . See Figure 3 and Figure 4 for an illustration.

It can easily be checked by induction on the depth that this construction makes sense and that for all  $\mathbf{i} \in \mathbb{T}_3$ , the tree  $\mathcal{R}[\mathbf{i}]$  is a bi-pointed Brownian tree with a randomized mass. Indeed, if  $(\mathcal{R}[\mathbf{i}], x[\mathbf{i}1], x[\mathbf{i}2])$  is a Brownian tree, then almost surely the points  $x[\mathbf{i}1]$ ,  $x[\mathbf{i}2]$  and  $x[\mathbf{i}3]$  defined above are pairwise distinct leaves, so  $b[\mathbf{i}]$  is uniquely characterized and distinct from  $x[\mathbf{i}1]$ ,  $x[\mathbf{i}2]$ ,  $x[\mathbf{i}3]$  and Theorem 4 for  $n = 3$  ensures that the  $\mathcal{R}[\mathbf{i}a]$  are Brownian trees. If  $\mathbf{i}$  has depth  $k$ , we will call  $\mathcal{R}[\mathbf{i}]$  a *region of scale  $k$* . Using Theorem 4 with  $n = 3$  at each scale, we easily have the following independence properties.

**Proposition 6.** *The following statements hold.*

(i) *The vectors*

$$\left( \frac{|\mathcal{R}[\mathbf{i}1]|}{|\mathcal{R}[\mathbf{i}]|}, \frac{|\mathcal{R}[\mathbf{i}2]|}{|\mathcal{R}[\mathbf{i}]|}, \frac{|\mathcal{R}[\mathbf{i}3]|}{|\mathcal{R}[\mathbf{i}]|} \right) \quad (4)$$

*for  $\mathbf{i} \in \mathbb{T}_3$  are i.i.d. with distribution  $\text{Dir}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ .*

(ii) *For all  $k \geq 1$ , conditionally on their masses, the measured metric spaces  $(\mathcal{R}[\mathbf{i}], x[\mathbf{i}1], x[\mathbf{i}2])$  for  $\mathbf{i} \in \mathbb{T}_3^k$  are i.i.d. bi-pointed Brownian trees.*

*Proof.* Let  $\mathcal{F}_k$  be the  $\sigma$ -algebra generated by the masses  $|\mathcal{R}[\mathbf{i}]|$  for  $\mathbf{i}$  of depth at most  $k$ . It is sufficient to show that for all  $k \geq 0$ , conditionally on  $\mathcal{F}_k$ , the vectors (4) for  $\mathbf{i}$  of depth  $k$  are i.i.d. with law  $\text{Dir}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ , and that conditionally on  $\mathcal{F}_k$  and on those vectors, the regions  $\mathcal{R}[\mathbf{i}]$  for  $\mathbf{i}$  of depth  $k + 1$  are independent bi-pointed Brownian trees with prescribed masses. We prove this statement by induction on  $k$ .

For  $k = 0$ , this is just Theorem 4 for  $n = 3$ . If the statement is true for some  $k \geq 0$ , the induction hypothesis guarantees that conditionally on  $\mathcal{F}_k$ , the  $\mathcal{R}[\mathbf{i}]$  for  $\mathbf{i}$  of depth  $k + 1$  are independent bi-pointed Brownian trees with randomized masses. We then apply Theorem 4 for  $n = 3$  to each of these Brownian trees, still conditionally on  $\mathcal{F}_k$ .  $\square$

**Zooming in on a random point.** Finally, let  $x \in \mathcal{T}$  and  $k \geq 0$ . If  $x$  is not one of the points  $x[\mathbf{i}]$ , which is the case for almost every  $x \in \mathcal{T}$ , we will denote by  $\mathbf{i}_k(x)$  the unique index  $\mathbf{i} \in \mathbb{T}_3$  of depth  $k$  such that  $x \in \mathcal{R}[\mathbf{i}]$ , and by  $i_k(x)$  the  $k$ -th letter of the word  $\mathbf{i}_k(x)$ . It will be useful for us to study the recursive decomposition around a random point  $X$  picked uniformly in  $\mathcal{T}$ . In this setting, we will need the following result.

**Lemma 7.** *Let  $(\mathcal{T}, (\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3})$  be a decomposed Brownian tree with a point  $X$  sampled uniformly on  $\mathcal{T}$ , independently of the decomposition. Then the vectors*

$$\left( \frac{|\mathcal{R}[\mathbf{i}_j(X)1]|}{|\mathcal{R}[\mathbf{i}_j(X)]|}, \frac{|\mathcal{R}[\mathbf{i}_j(X)2]|}{|\mathcal{R}[\mathbf{i}_j(X)]|}, \frac{|\mathcal{R}[\mathbf{i}_j(X)3]|}{|\mathcal{R}[\mathbf{i}_j(X)]|}, i_{j+1}(X) \right) \quad (5)$$

for  $j \geq 0$  are i.i.d.. Moreover, they have the distribution of  $(W_1, W_2, W_3, I)$ , where  $(W_1, W_2, W_3) \sim \text{Dir}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$  and  $\mathbb{P}(I = i \mid W_1, W_2, W_3) = W_i$  for  $i \in \{1, 2, 3\}$ . In particular, the index  $I$  is uniform on  $\{1, 2, 3\}$  and the variable  $W_I$  follows the Beta  $(\frac{3}{2}, 1)$  distribution.

*Proof.* The argument is basically the same as the proof of Proposition 6, where  $\mathcal{F}_k$  is replaced by the  $\sigma$ -algebra  $\mathcal{G}_k$  generated by the vectors (5) for  $0 \leq j \leq k-1$ . The only additional element is that since  $X$  is picked independently of the decomposition, conditionally on the Brownian trees  $\mathcal{R}[1], \mathcal{R}[2], \mathcal{R}[3]$ , it has probability  $|\mathcal{R}[a]|$  to be in  $\mathcal{R}[a]$  for  $a \in \{1, 2, 3\}$ . In particular, we have

$$\mathbb{P}(i_1(X) = a \mid |\mathcal{R}[1]|, |\mathcal{R}[2]|, |\mathcal{R}[3]|) = |\mathcal{R}[a]|$$

and conditionally on  $(|\mathcal{R}[1]|, |\mathcal{R}[2]|, |\mathcal{R}[3]|, i_1(X))$ , the region  $\mathcal{R}[i_1(X)]$  is still a (bi-pointed) Brownian tree with prescribed mass. The heredity step is adapted in the same way. Finally, the fact that  $W_I$  has a Beta  $(\frac{3}{2}, 1)$  distribution is a consequence of (1) and (2).  $\square$

**Remark 8.** *Often in the paper, it will be useful to re-interpret a quantity defined as the mass of a certain subset of  $\mathcal{T}$ , typically of the form  $|\cup_{\mathbf{i} \in U^k} \mathcal{R}[\mathbf{i}]|$  with  $U^k \subset \mathbb{T}_3^k$ , as the probability that a random point  $X$  falls into it. This will typically take the following form*

$$\left| \bigcup_{\mathbf{i} \in U^k} \mathcal{R}[\mathbf{i}] \right| = \mathbb{P}\left(\mathbf{i}_k(X) \in U^k \mid \mathcal{T}, (\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3}\right).$$

**Estimates for the size of the regions.** We now present a very rough result controlling the sizes of the regions appearing in the Aldous recursive decomposition of  $\mathcal{T}$ .

**Lemma 9.** *There are constants  $C > c > 0$  such that almost surely, for  $k$  large enough, for all  $\mathbf{i} \in \mathbb{T}_3^k$ , we have*

$$e^{-Ck} \leq |\mathcal{R}[\mathbf{i}]| \leq e^{-ck}.$$

*Proof.* The proof uses classical branching random walk arguments. We first notice that, by the independence properties of our recursive decomposition, the process  $(\log |\mathcal{R}[\mathbf{i}]|)_{\mathbf{i} \in \mathbb{T}_3}$  is a branching random walk. That is, the vectors

$$(\log |\mathcal{R}[\mathbf{i}1]| - \log |\mathcal{R}[\mathbf{i}]|, \log |\mathcal{R}[\mathbf{i}2]| - \log |\mathcal{R}[\mathbf{i}]|, \log |\mathcal{R}[\mathbf{i}3]| - \log |\mathcal{R}[\mathbf{i}]|)$$

for  $\mathbf{i} \in \mathbb{T}_3$  are i.i.d. and have the law of  $(\log W_1, \log W_2, \log W_3)$ , where  $(W_1, W_2, W_3)$  has distribution  $\text{Dir}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ . Therefore, for any  $C > 0$ , using the Chernoff bound we can write, for any  $\lambda > 0$ , for any  $\mathbf{i} \in \mathbb{T}_3^k$ ,

$$\mathbb{P}(\log |\mathcal{R}[\mathbf{i}]| \leq -Ck) \leq e^{-\lambda Ck} \mathbb{E} \left[ e^{-\lambda \log(W_1)} \right]^k. \quad (6)$$

Since Dirichlet distributions have polynomial tails, we can find  $\lambda$  such that the expectation is finite. We can then find  $C$  such that the right-hand side in the last equation is bounded by  $4^{-k}$ . We conclude the proof by a union bound over  $\mathbf{i} \in \mathbb{T}_3^k$  and then using the Borel–Cantelli lemma over  $k$ .

Similarly, for  $c > 0$  and  $\lambda > 0$ , we have

$$\mathbb{P}(\log |\mathcal{R}[\mathbf{i}]| \geq -ck) \leq e^{\lambda ck} \mathbb{E} \left[ e^{\lambda \log(W_1)} \right]^k.$$

Since  $W_1$  does not have an atom at zero, there exists  $\lambda > 0$  such that  $\mathbb{E} \left[ e^{\lambda \log(W_1)} \right] < \frac{1}{5}$ . Once the value of  $\lambda$  is fixed, we can choose  $c > 0$  sufficiently small so that the last right-hand side is bounded by  $4^{-k}$ . This proves the other direction, again by union bound and Borel–Cantelli.  $\square$

### 3 Proof of Proposition 3

This entire section is devoted to proving Proposition 3, which is central to the proof of the main results in the next section.

**Rough idea of the proof.** We start with a rough idea of the proof. Let us fix a point  $x \in \mathcal{T}$  and consider the ratios

$$\left( \frac{|\mathcal{R}[\mathbf{i}_j(x)]|}{|\mathcal{R}[\mathbf{i}_{j-1}(x)]|} \right)_{1 \leq j \leq k} \quad (7)$$

obtained by "zooming scale after scale" around  $x$ . By Lemma 7, these are i.i.d. variables with a fixed, absolutely continuous distribution. Using this, we will argue that if we fix a small region  $r'$  of  $\mathcal{T}'$ , the probability that we can find nested regions of  $\mathcal{T}'$  around  $r'$  that respect the ratios (7) up to a factor  $1 \pm \delta$  at most of the scales  $1 \leq j \leq k$  is of order  $\delta^k$ . By choosing  $\delta$  small enough, we will be able to do a union bound over the possible candidate regions  $r'$  of  $\mathcal{T}'$ . This shows that for any small region  $r'$  of  $\mathcal{T}'$ , a homeomorphism sending  $x$  to a point of  $r'$  will have " $\delta$ -mismatches" between the ratios  $\frac{|\mathcal{R}[\mathbf{i}_j(x)]|}{|\mathcal{R}[\mathbf{i}_{j-1}(x)]|}$  and  $\frac{|\Psi(\mathcal{R}[\mathbf{i}_j(x)])|}{|\Psi(\mathcal{R}[\mathbf{i}_{j-1}(x)])|}$  for "many" scales  $j$  (this is Proposition 13 below, that we prove in Section 3.1). Moreover, the ratio

$$\frac{|\Psi(\mathcal{R}[\mathbf{i}_k(x)])|}{|\mathcal{R}[\mathbf{i}_k(x)]|} \quad (8)$$

can be written as the telescopic product of those mismatches, so the existence of mismatches proves that the ratio (8) will vary "quite often" by a factor  $1 \pm \delta$ . To conclude, we will argue that for most points  $x$ , those mismatches cannot compensate each other. Indeed, if for a region  $\mathcal{R}[\mathbf{i}] = \mathcal{R}[\mathbf{i}1] \cup \mathcal{R}[\mathbf{i}2] \cup \mathcal{R}[\mathbf{i}3]$  we have  $\frac{|\Psi(\mathcal{R}[\mathbf{i}])|}{|\mathcal{R}[\mathbf{i}]|} \ll 1$ , then if a mismatch makes the ratio  $\frac{|\Psi(\mathcal{R}[\mathbf{i}1])|}{|\mathcal{R}[\mathbf{i}1]|}$  closer to 1, it will make the ratios  $\frac{|\Psi(\mathcal{R}[\mathbf{i}2])|}{|\mathcal{R}[\mathbf{i}2]|}$  and  $\frac{|\Psi(\mathcal{R}[\mathbf{i}3])|}{|\mathcal{R}[\mathbf{i}3]|}$  even smaller. Another way to say this is that if  $x$  is picked uniformly at random in  $\mathcal{T}$ , the ratios (8) form a nonnegative martingale in  $k$  (see (19) below), so they converge almost surely. By existence of mismatches, they change by a factor  $1 \pm \delta$  many times, so the limit of the martingale has to be 0. This argument is done in a quantitative way in Section 3.2.

#### 3.1 Finding mismatches

Before giving the precise definition of mismatches, it will be convenient to restrict ourselves to a (not too small) subset of the intermediate scales. More precisely, we first notice that by construction of the Aldous recursive decomposition, if a word  $\mathbf{i}$  has its last letter equal to 3, then the boundary of the region  $\mathcal{R}[\mathbf{i}]$  in  $\mathcal{T}$  is a single point (see Figure 4).

**Definition 10.** Let  $\alpha > 0$  and let  $\mathbf{i} \in \mathbb{T}_3$  with depth  $k$ . We say that a scale  $1 \leq j \leq k-1$  is  $\alpha$ -good for  $\mathbf{i}$  with respect to  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3}$  if  $j$  is odd, if  $i_j = i_{j+1} = 3$ , and if furthermore

$$\min(|\mathcal{R}[\mathbf{i}_j1]|, |\mathcal{R}[\mathbf{i}_j2]|, |\mathcal{R}[\mathbf{i}_j3]|) \geq \alpha |\mathcal{R}[\mathbf{i}_j]|. \quad (9)$$

The reason why we want  $i_j = 3$  is that it implies that there are two small regions  $\mathcal{R}[\mathbf{i}_j2], \mathcal{R}[\mathbf{i}_j3]$  whose boundary is the singleton  $\{b[\mathbf{i}_j]\}$  (see Figure 4), and the ratio between the masses of those regions gives a convenient way to say that  $\mathcal{T}$  and  $\mathcal{T}'$  have "different shapes" at a certain scale. The point of the assumption (9) is that if the branching point  $b[\mathbf{i}]$  splits the region  $\mathcal{R}[\mathbf{i}]$  in a very uneven way, there will be many possible choices for  $\Psi(b[\mathbf{i}])$ , whereas if  $b[\mathbf{i}]$  is "central" in  $\mathcal{R}[\mathbf{i}]$  only few choices will be available.

The first step is to make sure that around most points of the tree  $\mathcal{T}$ , good scales represent a positive proportion of the scales.

**Lemma 11.** Let  $(\mathcal{T}, (\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3})$  be a decomposed Brownian tree. There exists constants  $\alpha, \beta, \gamma > 0$  such that denoting

$$U^k := \left\{ \mathbf{i} \in \mathbb{T}_3^k \mid \mathbf{i} \text{ has at least } \beta k \text{ scales that are } \alpha\text{-good} \right\}$$

and  $E^k := \bigcup_{\mathbf{i} \in \mathbb{T}_3^k \setminus U^k} \mathcal{R}[\mathbf{i}]$ , then almost surely, for  $k$  large enough, we have  $|E^k| < e^{-\gamma k}$ .

*Proof.* Let us prove that we can find  $\alpha, \beta, \gamma$  such that  $\mathbb{E}[|E^k|] < e^{-2\gamma k}$ , so that the conclusion of the lemma can be obtained using the Markov inequality and the Borel–Cantelli lemma. Let  $X$  be a uniform point taken under the mass measure on  $\mathcal{T}$  independently of the decomposition. By Remark 8, we can rewrite  $\mathbb{E}[|E^k|]$  in the following way:

$$\mathbb{E}[|E^k|] = \mathbb{P}(\mathbf{i}_k(X) \text{ has less than } \beta k \text{ scales that are } \alpha\text{-good}).$$

As in the proof of Lemma 7, for  $k \geq 0$ , we denote by  $\mathcal{G}_k$  the  $\sigma$ -algebra generated by the word  $\mathbf{i}_k(X)$  and by the masses  $\mathcal{R}[\mathbf{i}_j(X)a]$  for  $0 \leq j \leq k-1$  and  $a \in \{1, 2, 3\}$ . It is clear that the event that  $j$  is a good scale for  $\mathbf{i}_k(X)$  belongs to  $\mathcal{G}_{j+1}$ .

Moreover, by Lemma 7, for every odd  $j$ , we have

$$\mathbb{P}(j \text{ is } \alpha\text{-good for } X \mid \mathcal{G}_{j-1}) = \frac{1}{3} \cdot \frac{1}{3} \cdot \mathbb{P}(\min(W_1, W_2, W_3) \geq \alpha), \quad (10)$$

where  $(W_1, W_2, W_3) \sim \text{Dir}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ . This probability is deterministic and does not depend on  $j$  so we denote it by  $p$ , and note that  $p > 0$  if  $\alpha > 0$  was chosen small enough. Therefore, the variables

$$\left( \mathbb{1}_{\{\text{the scale } 2j+1 \text{ is } \alpha\text{-good for } X\}} \right)_{0 \leq j < \frac{k}{2}}$$

are i.i.d. Bernoulli variables with positive mean, which is sufficient to prove the lemma by a Chernoff bound.  $\square$

From now on, we will fix  $\alpha, \beta, \gamma > 0$  that satisfy the conclusion of Lemma 11 and we will simply refer to an  $\alpha$ -good scale as a *good scale*.

Now let  $0 < \delta < \alpha$  and let  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$  be a homeomorphism.

**Definition 12.** Let  $k \geq 0$  and  $\mathbf{i} \in \mathbb{T}_3^k$ . We say that a scale  $1 \leq j \leq k-1$  is a  $\delta$ -mismatch for  $\Psi$  at  $\mathbf{i}$  with respect to  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3}$  if it is a good scale and if we have

$$\max_{1 \leq a \leq 3} \left| \frac{|\mathcal{R}[\mathbf{i}_j a]|}{|\mathcal{R}[\mathbf{i}_j]|} - \frac{|\Psi(\mathcal{R}[\mathbf{i}_j a])|}{|\Psi(\mathcal{R}[\mathbf{i}_j])|} \right| > \delta.$$

Since  $\mathcal{T}$ ,  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3}$  and  $\delta$  are fixed throughout the paper, we will often simply write "mismatch for  $\Psi$  at  $\mathbf{i}$ " instead of " $\delta$ -mismatch for  $\Psi$  at  $\mathbf{i}$  with respect to  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3}$ ". Informally, the scale  $j$  being a mismatch for  $\Psi$  at  $\mathbf{i}$  indicates that when we perform one step of the Aldous recursive decomposition in  $\mathcal{T}$ , the decomposition of  $\mathcal{R}[\mathbf{i}_j]$  into three parts and its image by  $\Psi$  in  $\Psi(\mathcal{R}[\mathbf{i}_j])$  do not split the masses with the same proportions.

Our next goal is the following result, which guarantees the existence of mismatches around most points for any homeomorphism. It represents a significant proportion of the proof of Proposition 3.

**Proposition 13.** *There exists  $\delta > 0$  such that almost surely, for  $k$  large enough, for any  $\mathbf{i} \in \mathbb{T}_3^k$  that has at least  $\beta k$  good scales, for any homeomorphism  $\Psi$  from  $\mathcal{T}$  to  $\mathcal{T}'$ , one of the following holds:*

1. either  $|\Psi(\mathcal{R}[\mathbf{i}])| \leq e^{-k} |\mathcal{R}[\mathbf{i}]|$ ,
2. there are at least  $\lfloor \frac{\beta}{2} k \rfloor$  good scales in  $\{1, \dots, k-1\}$  that are  $\delta$ -mismatches for  $\Psi$  at  $\mathbf{i}$ .

*Proof.* Let  $\delta \in (0, \alpha)$ , whose value will be specified later. Let  $k \geq 0$ . Let  $C > 0$  be the constant given by the lower bound in Lemma 9, and let  $B^k$  be the event that for all  $\mathbf{i} \in \mathbb{T}_3$  of depth at most  $k$ , the branching point  $b[\mathbf{i}]$  satisfies  $|b[\mathbf{i}]|_{\mathcal{T}} \geq e^{-Ck}$ . By Lemma 9 (applied to  $k+1$ ), almost surely  $B^k$  occurs for  $k$  large enough.

Now let  $\mathbf{i} \in \mathbb{T}_3^k$ , and let  $J$  be a subset of  $\{1, 2, \dots, k-1\}$  with size  $\lfloor \frac{\beta}{2} k \rfloor$  and with only odd elements. We are interested in the existence of a homeomorphism  $\Psi$  that would satisfy the three following properties:

- (i)  $|b[\mathbf{i}_j]|_{\mathcal{T}} \geq e^{-Ck}$  for all  $0 \leq j \leq k$ ,
- (ii) we have  $|\Psi(\mathcal{R}[\mathbf{i}])| > e^{-k} |\mathcal{R}[\mathbf{i}]|$ ,
- (iii) for all  $j \in J$ , the scale  $j$  is good but is not a  $\delta$ -mismatch for  $\Psi$  at  $\mathbf{i}$ .

We insist that in item (i), we mean  $e^{-Ck}$  and not  $e^{-Cj}$ . We hence define the event  $A_J(\mathbf{i})$  as

$$A_J(\mathbf{i}) = \{ \text{there exists a homeomorphism } \Psi : \mathcal{T} \rightarrow \mathcal{T}' \text{ such that (i), (ii), (iii) hold for } \Psi \text{ and } \mathbf{i} \}.$$

In the rest of the proof, whenever there exists a homeomorphism  $\Psi$  such that (i), (ii), (iii) hold for  $\Psi$  and  $\mathbf{i}$  we will say that " $\Psi$  makes  $A_J(\mathbf{i})$  occur".

Now let us consider what happens on the event where  $B^k$  occurs but none of the  $A_J(\mathbf{i})$  does for any  $\mathbf{i} \in \mathbb{T}_3^k$  and any  $J \subset \{1, 2, \dots, k-1\}$  with  $\#J = \lfloor \frac{\beta}{2} k \rfloor$ . Fix some  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$  and  $\mathbf{i} \in \mathbb{T}_3^k$  that has at least  $\beta k$  good scales. On the event considered, (i) is satisfied so either (ii) fails (and in this case point 1. of the proposition holds); or (ii) holds for  $\Psi$  and  $\mathbf{i}$ , which entails that (iii) fails for all choices of  $J$ . Since we assumed that  $\mathbf{i}$  has more than  $\beta k$  good scales, the fact that we cannot find any  $\lfloor \frac{\beta}{2} k \rfloor$  good scales that are not  $\delta$ -mismatches tells us that at least  $\beta k - \lfloor \frac{\beta}{2} k \rfloor \geq \lfloor \frac{\beta}{2} k \rfloor$  of them are indeed  $\delta$ -mismatches. In view of this, since we already know that  $B^k$  occurs for  $k$  large enough, it suffices to find  $\delta > 0$  such that almost surely, none of the  $A_J(\mathbf{i})$  with  $\mathbf{i} \in \mathbb{T}_3^k$  and  $J \subset \{1, \dots, k\}$  of cardinal  $\lfloor \frac{\beta}{2} k \rfloor$  occurs for  $k$  large enough. For that, we will show using a union bound that the probability of

$$\bigcup_{\mathbf{i} \in \mathbb{T}_3^k} \bigcup_{\substack{J \subset \{1, \dots, k-1\} \\ \#J = \lfloor \frac{\beta}{2} k \rfloor}} A_J(\mathbf{i}) \tag{11}$$

is summable in  $k$ , and conclude with the Borel–Cantelli lemma.

Suppose that there exists some  $\Psi$  that makes the event  $A_J(\mathbf{i})$  occur. We write  $\ell = \lfloor \frac{\beta}{2} k \rfloor$  and denote by  $j_1 < j_2 < \dots < j_\ell$  the elements of  $J$ , and write  $b_h = b[\mathbf{i}_{j_h}]$  for all  $1 \leq h \leq \ell$ . We now try to understand the sequence of points

$$\mathbf{c} = (c_h)_{1 \leq h \leq \ell} = (\Psi(b_h))_{1 \leq h \leq \ell}.$$

We first note that this sequence must satisfy a topological condition: we recall that  $i_{j_h} = i_{j_h+1} = 3$  because the scale  $j_h$  is good. Hence, for all  $1 \leq h \leq \ell$ , the point  $b_h$  is a branching point of  $\mathcal{T}$  with



the points  $b_1, \dots, b_{h-1}$  in the same connected component of  $\mathcal{T} \setminus \{b_h\}$ , and the points  $b_{h+1}, \dots, b_\ell$  in another component. Since  $\Psi$  is a homeomorphism, the same is true for the sequence  $(c_h)$  in  $\mathcal{T}'$ . Moreover, we claim that the sizes of the branching points  $c_h$  cannot be too small. More precisely, for all  $1 \leq h \leq \ell - 1$ , let  $\mathcal{R}'_h(\mathbf{c})$  be the connected component of  $\mathcal{T}' \setminus \{c_h\}$  that contains  $c_{h+1}, \dots, c_\ell$ , with the convention that  $\mathcal{R}'_0(\mathbf{c}) = \mathcal{T}'$ . In other words, we have  $\mathcal{R}'_h(\mathbf{c}) = \Psi(\mathcal{R}[\mathbf{i}_{j_h+1}])$ . Then for all  $1 \leq h \leq \ell$ , we have

$$|c_h|_{\mathcal{R}'_{h-1}(\mathbf{c})} = |c_h|_{\Psi(\mathcal{R}[\mathbf{i}_{j_{h-1}+1}])} \geq |c_h|_{\Psi(\mathcal{R}[\mathbf{i}_{j_h}])} = \min_{a \in \{1,2,3\}} |\Psi(\mathcal{R}[\mathbf{i}_{j_h} a])|.$$

Using successively the fact that  $j_h$  is not a  $\delta$ -mismatch and the fact that the scale  $j_h$  is  $\alpha$ -good, we deduce, for  $1 \leq h \leq \ell - 1$ :

$$|c_h|_{\mathcal{R}'_{h-1}(\mathbf{c})} \geq \left( \frac{\min_{a \in \{1,2,3\}} |\mathcal{R}[\mathbf{i}_{j_h} a]|}{|\mathcal{R}[\mathbf{i}_{j_h}]|} - \delta \right) \cdot |\Psi(\mathcal{R}[\mathbf{i}_{j_h}])| \geq (\alpha - \delta) |\mathcal{R}'_h(\mathbf{c})|. \quad (12)$$

From now on, we assume  $\delta < \frac{\alpha}{4}$ , so that

$$|c_h|_{\mathcal{R}'_{h-1}(\mathbf{c})} \geq \frac{\alpha}{2} |\mathcal{R}'_h(\mathbf{c})|$$

for  $1 \leq h \leq \ell - 1$ . For the same reasons, we also have

$$|c_\ell|_{\mathcal{R}'_{\ell-1}(\mathbf{c})} \geq \frac{\alpha}{2} |\Psi(\mathcal{R}[\mathbf{i}_{j_\ell}])| \geq \frac{\alpha}{2} |\Psi(\mathcal{R}[\mathbf{i}])|$$

so, using (ii) and (i) in the definition of the event  $A_J(\mathbf{i})$ , we have

$$|c_\ell|_{\mathcal{R}'_{\ell-1}(\mathbf{c})} \geq \frac{\alpha}{2} e^{-k} |\mathcal{R}[\mathbf{i}]| \geq \frac{\alpha}{2} e^{-(C+1)k}.$$

Following what precedes, we define a *candidate sequence of length  $\ell$*  as a sequence  $(c_h)_{1 \leq h \leq \ell}$  of branching points of  $\mathcal{T}'$  such that:

- for all  $1 \leq h \leq \ell$ , the points  $c_1, \dots, c_{h-1}$  all lie in one connected component of  $\mathcal{T}' \setminus \{c_h\}$ , and the points  $c_{h+1}, \dots, c_\ell$  lie in another one, denoted if  $h < \ell$  by  $\mathcal{R}'_h(\mathbf{c})$ ;
- for all  $1 \leq h \leq \ell - 1$ , we have  $|c_h|_{\mathcal{R}'_{h-1}(\mathbf{c})} \geq \frac{\alpha}{2} |\mathcal{R}'_h(\mathbf{c})|$ ;
- we have  $|c_\ell|_{\mathcal{R}'_{\ell-1}(\mathbf{c})} \geq \frac{\alpha}{2} e^{-(C+1)k}$ .

Note that crucially, the notion of candidate sequence does not depend on  $\delta$ . By the discussion above, any  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$  that makes the event  $A_J(\mathbf{i})$  occur must send  $(b_h)_{1 \leq h \leq \ell}$  to some candidate sequence of length  $\ell$  in  $\mathcal{T}'$ . We will now provide a bound on the number of such candidate sequences in  $\mathcal{T}'$ . This bound is entirely deterministic, and only uses the fact that  $\mathcal{T}'$  is a real tree with total mass 1 where all branching points have degree 3, which is almost surely the case for a realization of a Brownian tree.

**Lemma 14.** *There exists a constant  $K = K(C, \alpha, \beta)$  such that almost surely, for all  $\ell \geq 2$ , the number of candidate sequences of length  $\ell$  in  $\mathcal{T}'$  is at most  $K^\ell$ .*

*Proof.* If  $\mathbf{c} = (c_h)_{1 \leq h \leq \ell}$  is a candidate sequence, we first define its *sequence of scales*  $(s_h(\mathbf{c}))_{0 \leq h \leq \ell}$  by

$$s_h(\mathbf{c}) = \lfloor -\log |\mathcal{R}'_h(\mathbf{c})| \rfloor$$

for  $1 \leq h \leq \ell - 1$ , with the conventions  $s_0(\mathbf{c}) = 0$  and  $s_\ell(\mathbf{c}) = \lfloor -\log |c_\ell|_{\mathcal{R}'_{\ell-1}(\mathbf{c})} \rfloor$ . Then  $\mathbf{s}(\mathbf{c}) = (s_h(\mathbf{c}))_{1 \leq h \leq \ell}$  is a non-decreasing sequence of integers starting at 0. Moreover, by definition of a candidate sequence, the numbers  $s_h(\mathbf{c})$  are bounded above by

$$-\log \left( \frac{\alpha}{2} e^{-(C+1)k} \right) = -\log \frac{\alpha}{2} + (C+1)k.$$

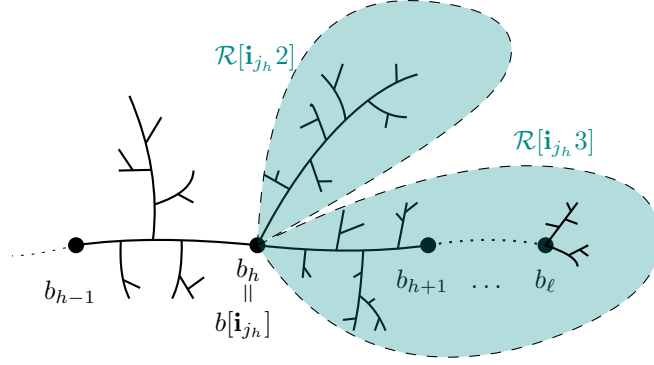


Figure 5: The sequence  $(b_h)_{1 \leq h \leq \ell}$  and their positions relative to  $\mathcal{R}[\mathbf{i}_{j_h} 2]$  and  $\mathcal{R}[\mathbf{i}_{j_h} 3]$ .

Therefore, the number of possible values of the sequence  $(s_h(\mathbf{c}))_{0 \leq h \leq \ell}$  is at most

$$\binom{\log(2/\alpha) + (C+1)k + \ell}{\ell} \leq \frac{2}{\alpha} 2^{(C+2)k}. \quad (13)$$

On the other hand, for any such sequence  $\mathbf{s} = (s_h)_{0 \leq h \leq \ell}$ , let us bound the number of candidate sequences  $\mathbf{c}$  for which the scale sequence  $\mathbf{s}(\mathbf{c})$  is  $\mathbf{s}$ . For this, we start with an easy remark. Let  $\mathcal{R}$  be a region of  $\mathcal{T}$  of mass  $m_0$ . Then we claim the following

**Claim.** *The number of branching points  $c$  in  $\mathcal{R}$  satisfying  $|c|_{\mathcal{R}} \geq m$  is at most  $\frac{m_0}{m}$ .*

Indeed, let  $S_m$  be the set of branching points of size at least  $m$  in  $\mathcal{R}$ . Then  $\mathcal{R}$  can be obtained by gluing the connected components of  $\mathcal{R} \setminus S_m$  along the structure of a finite binary tree. The nodes of this binary tree correspond to the points of  $S_m$ , so this binary tree has  $\#S_m$  nodes and therefore  $\#S_m + 2$  leaves. Those  $\#S_m + 2$  leaves correspond to  $\#S_m + 2$  disjoint parts of  $\mathcal{R}$  with mass at least  $m$  each, so  $(\#S_m + 2)m \leq m_0$  and the claim follows.

Now, let  $\mathbf{s} = (s_h)_{0 \leq h \leq \ell}$  be a non-decreasing sequence of integers with  $s_0 = 0$  and  $s_\ell \leq \log \frac{2}{\alpha} + (C+1)\ell$ . Let us build step by step a candidate sequence  $\mathbf{c}$  satisfying  $\mathbf{s}(\mathbf{c}) = \mathbf{s}$ :

- For  $\mathbf{c}$  to be a candidate sequence,  $c_1$  needs to satisfy

$$|c_1|_{\mathcal{R}'_0(\mathbf{c})} = |c_1|_{\mathcal{T}} \geq \frac{\alpha}{2} \cdot |\mathcal{R}'_1(\mathbf{c})| \geq \frac{\alpha}{2} \cdot e^{-(s_1(\mathbf{c})+1)}$$

by definition of  $s_1(\mathbf{c})$ . Using the last display and the claim, the number of possible choices for  $c_1$  for which  $s_1(\mathbf{c}) = s_1$  is at most  $\frac{2}{\alpha} e^{s_1+1}$ .

- Let  $1 \leq h \leq \ell - 1$  and assume that  $c_1, \dots, c_{h-1}$  have already been chosen. Then  $\mathcal{R}'_{h-2}(\mathbf{c})$  is determined by  $(c_1, \dots, c_{h-1})$  and  $\mathcal{R}'_{h-1}(\mathbf{c})$  is a connected component of  $\mathcal{R}'_{h-2}(\mathbf{c}) \setminus \{c_{h-1}\}$ , so there are only 3 possible choices for  $\mathcal{R}'_{h-1}(\mathbf{c})$ . Moreover, once this region has been chosen, the point  $c_h$  must be a point of  $\mathcal{R}'_{h-1}(\mathbf{c})$  with

$$|c_h|_{\mathcal{R}'_{h-1}(\mathbf{c})} \geq \frac{\alpha}{2} |\mathcal{R}'_h(\mathbf{c})| \geq \frac{\alpha}{2} e^{-(s_h(\mathbf{c})+1)}$$

by definition of  $s_h(\mathbf{c})$ . Using the claim again and the fact that  $|\mathcal{R}'_{h-1}(\mathbf{c})| \leq e^{-s_{h-1}(\mathbf{c})} = e^{-s_{h-1}}$  by construction, the number of possible choices for  $c_h$  that ensure that  $s_h(\mathbf{c}) = s_h$ , given  $c_1, \dots, c_{h-1}$ , is bounded above by

$$3 \cdot \frac{|\mathcal{R}'_{h-1}(\mathbf{c})|}{|c_h|_{\mathcal{R}'_{h-1}(\mathbf{c})}} \leq \frac{6}{\alpha} e^{s_h - s_{h-1} + 1}.$$

- Finally, by the same reasoning, the number of possible choices for  $c_\ell$  given  $c_1, \dots, c_{\ell-1}$  is bounded above by

$$\frac{6}{\alpha} e^{(C+1)\ell - s_{\ell-1}}.$$

Using the above in cascade and reducing the telescopic product we get that, for any  $\mathbf{s}$ , the number of candidate sequences  $\mathbf{c}$  such that  $\mathbf{s}(\mathbf{c}) = \mathbf{s}$  is bounded above by

$$\left(\frac{6}{\alpha}\right)^\ell e^{(C+2)\ell}.$$

Combined with (13), this proves the lemma, with

$$K = \frac{6}{\alpha} \cdot e^{C+2} \cdot 2^{\frac{2C+4}{\beta}}. \quad (14)$$

□

We return to the proof of Proposition 13. From now on, we will work conditionally on the tree  $\mathcal{T}'$  and fix  $J, \mathbf{i}$ . We recall that  $j_1 < \dots < j_\ell$  are the elements of  $J$  and that  $b_h = b[\mathbf{i}_{j_h}]$ . We have seen that if there exists a  $\Psi$  that makes the event  $A_J(\mathbf{i})$  occur, then the sequence  $(\Psi(b_h))_{1 \leq h \leq \ell}$  is a candidate sequence in  $\mathcal{T}'$ . Therefore, we fix such a candidate sequence  $(c_h)_{1 \leq h \leq \ell}$ , and estimate the probability that there exists a  $\Psi$  satisfying (i), (ii), (iii) as well as  $\Psi(b_h) = c_h$  for all  $1 \leq h \leq \ell$ . For this, let  $2 \leq h \leq \ell - 1$ . On the event that such a  $\Psi$  exists, since  $j_h$  is not a  $\delta$ -mismatch for  $\mathbf{i}$ , we have

$$\left( \frac{|\Psi(\mathcal{R}[\mathbf{i}_{j_h} 2])|}{|\Psi(\mathcal{R}[\mathbf{i}_{j_h}])|} - \delta \right) |\mathcal{R}[\mathbf{i}_{j_h}]| \leq |\mathcal{R}[\mathbf{i}_{j_h} 2]| \leq \left( \frac{|\Psi(\mathcal{R}[\mathbf{i}_{j_h} 2])|}{|\Psi(\mathcal{R}[\mathbf{i}_{j_h}])|} + \delta \right) |\mathcal{R}[\mathbf{i}_{j_h}]|, \quad (15)$$

and a similar estimate holds for  $|\mathcal{R}[\mathbf{i}_{j_h} 3]|$ .

On the other hand, by our conventions in the construction of the recursive decomposition and the fact that  $i_{j_h} = i_{j_h+1} = 3$  (since  $j_h$  is a good scale for  $\mathbf{i}$ ), the connected component of  $\mathcal{T} \setminus \{b_h\}$  that contains  $b_\ell$  is  $\mathcal{R}[\mathbf{i}_{j_h} 3]$ , and the component that contains neither  $b_1$  nor  $b_\ell$  is  $\mathcal{R}[\mathbf{i}_{j_h} 2]$ , see Figure 5. Hence  $\Psi(\mathcal{R}[\mathbf{i}_{j_h} 2])$  must be the connected component of  $\mathcal{T}' \setminus \{c_h\}$  that contains neither  $c_1$  nor  $c_\ell$ , and  $\Psi(\mathcal{R}[\mathbf{i}_{j_h} 3])$  must be the connected component of  $\mathcal{T}' \setminus \{c_h\}$  that contains  $c_\ell$ . We denote by  $n_2^h := |\Psi(\mathcal{R}[\mathbf{i}_{j_h} 2])|$  and  $n_3^h := |\Psi(\mathcal{R}[\mathbf{i}_{j_h} 3])|$  the respective masses of these two connected components, and highlight that those masses are completely determined by the sequence  $\mathbf{c}$ . From (15) and the analog equation for  $\mathcal{R}[\mathbf{i}_{j_h} 3]$ , using that  $(x, y) \mapsto \frac{x}{x+y}$  is increasing in  $x$  and decreasing in  $y$ , we get

$$\frac{n_2^h - \delta |\Psi(\mathcal{R}[\mathbf{i}_{j_h}])|}{n_2^h + n_3^h} \leq \frac{|\mathcal{R}[\mathbf{i}_{j_h} 2]|}{|\mathcal{R}[\mathbf{i}_{j_h} 2]| + |\mathcal{R}[\mathbf{i}_{j_h} 3]|} \leq \frac{n_2^h + \delta |\Psi(\mathcal{R}[\mathbf{i}_{j_h}])|}{n_2^h + n_3^h}. \quad (16)$$

Since the scale  $j_h$  is  $\alpha$ -good and we have chosen  $\delta < \frac{\alpha}{2}$ , we have as in (12) that  $n_2^h, n_3^h \geq \frac{\alpha}{2} |\Psi(\mathcal{R}[\mathbf{i}_{j_h}])|$ , so  $n_2^h + n_3^h \geq \alpha |\Psi(\mathcal{R}[\mathbf{i}_{j_h}])|$  and (16) becomes

$$\left| \frac{|\mathcal{R}[\mathbf{i}_{j_h} 2]|}{|\mathcal{R}[\mathbf{i}_{j_h} 2]| + |\mathcal{R}[\mathbf{i}_{j_h} 3]|} - \frac{n_2^h}{n_2^h + n_3^h} \right| \leq \frac{\delta}{\alpha}$$

for all  $2 \leq h \leq \ell - 1$ . On the other hand, by Proposition 6, the variables  $\frac{|\mathcal{R}[\mathbf{i}_j 2]|}{|\mathcal{R}[\mathbf{i}_j 2]| + |\mathcal{R}[\mathbf{i}_j 3]|}$  for  $j \geq 0$  are i.i.d. copies of  $\frac{W_2}{W_2 + W_3}$ , where  $(W_1, W_2, W_3) \sim \text{Dir}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ . It follows that we have

$$\mathbb{P}(\exists \Psi \text{ that makes } A_J(\mathbf{i}) \text{ occur and } \forall 2 \leq h \leq \ell - 1, \Psi(b_h) = c_h \mid \mathcal{T}') \leq \prod_{h=2}^{\ell-1} \mathbb{P}\left(\left| \frac{W_2}{W_2 + W_3} - q_h \right| \leq \frac{\delta}{\alpha}\right),$$

where  $q_h = \frac{n_2^h}{n_2^h + n_3^h} \in [\frac{\alpha}{2}, 1 - \frac{\alpha}{2}]$ . On the other hand, by (1), the law of  $\frac{W_2}{W_2 + W_3}$  is Beta  $(\frac{1}{2}, \frac{1}{2})$  with density  $\frac{1}{\pi} \frac{1}{\sqrt{t(1-t)}} \mathbb{1}_{[0,1]}(t) dt$ . In particular, the density is bounded on  $[\frac{\alpha}{4}, 1 - \frac{\alpha}{4}]$  by  $\frac{1}{\pi \sqrt{\frac{1}{2} \cdot \frac{\alpha}{4}}} \leq \frac{1}{\sqrt{\alpha}}$ .

Therefore, integrating between  $q_h - \frac{\delta}{\alpha}$  and  $q_h + \frac{\delta}{\alpha}$ , assuming that  $\delta < \frac{\alpha^2}{4}$  we have

$$\mathbb{P}(\exists \Psi \text{ that makes } A_J(\mathbf{i}) \text{ occur and } \forall 2 \leq h \leq \ell - 1, \Psi(b_h) = c_h \mid \mathcal{T}') \leq \left( \frac{2\delta}{\alpha\sqrt{\alpha}} \right)^{\ell-2}.$$

We can now take the union bound over all candidate sequences  $(c_h)$ . By Lemma 14, we obtain

$$\mathbb{P}(A_J(\mathbf{i}) \mid \mathcal{T}') \leq K^\ell \left( \frac{2\delta}{\alpha\sqrt{\alpha}} \right)^{\ell-2}.$$

We can now remove the conditioning on  $\mathcal{T}'$  and perform a union bound over the  $3^k$  possible values of  $\mathbf{i}$  and the  $\binom{k-1}{\ell} \leq 2^k$  possible values of the set  $J$ . We find

$$\mathbb{P} \left( \bigcup_{\mathbf{i} \in \mathbb{T}_3^k} \bigcup_{\substack{J \subset \{1, \dots, k-1\} \\ \#J = \ell}} A_J(\mathbf{i}) \right) \leq 6^k \cdot \left( \frac{2\delta K}{\alpha\sqrt{\alpha}} \right)^{\ell-2}. \quad (17)$$

Reminding that  $\ell = \lfloor \frac{\beta k}{2} \rfloor$ , if we choose the constant  $\delta > 0$  sufficiently small, this decays exponentially in  $k$ . Given the discussion before (11), this proves the proposition.  $\square$

### 3.2 The martingale argument

For the next part of the argument, we need to introduce another notion of mismatch that is slightly looser than the one of Definition 12. From now on, we fix a value of  $\delta > 0$  that satisfies the conclusion of Proposition 13. Suppose that  $(\mathcal{T}, (\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3})$  is a decomposed Brownian tree and  $\mathcal{T}'$  is another Brownian tree, and that  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$  is a homeomorphism.

**Definition 15.** Let  $j \geq 1$ , and let  $\mathbf{i} \in \mathbb{T}_3$  of depth at least  $j + 1$ . We say that the scale  $j$  is a weak mismatch for  $\Psi$  at  $\mathbf{i}$  if  $\frac{|\mathcal{R}[\mathbf{i}_j a]|}{|\mathcal{R}[\mathbf{i}_j]|} \geq \alpha$  for all  $a \in \{1, 2, 3\}$  and furthermore

$$\max_{a \in \{1, 2, 3\}} \left| \frac{|\Psi(\mathcal{R}[\mathbf{i}_j a])|}{|\Psi(\mathcal{R}[\mathbf{i}_j])|} - \frac{|\mathcal{R}[\mathbf{i}_j a]|}{|\mathcal{R}[\mathbf{i}_j]|} \right| > \delta.$$

In particular, a mismatch as defined in Definition 12 is also a weak mismatch. The difference between the two definitions is that we have removed the "topological" part of the assumption that  $j$  has to be a good scale, i.e. that  $i_j = i_{j+1} = 3$  (see Definition 10). In this section, we prove the following result.

**Proposition 16.** Let  $\beta$  be as in Lemma 11. There exists a constant  $\eta > 0$  such that for any homeomorphism  $\Psi$  from  $\mathcal{T}$  to  $\mathcal{T}'$ , for any  $k \geq 1$ , the set

$$V^{k, \Psi} := \left\{ \mathbf{i} \in \mathbb{T}_3^k \mid \Psi \text{ has more than } \frac{\beta}{2}k \text{ weak mismatches at } \mathbf{i} \text{ and } \frac{|\Psi(\mathcal{R}[\mathbf{i}])|}{|\mathcal{R}[\mathbf{i}]} \geq e^{-\eta k} \right\}$$

is such that

$$\left| \bigcup_{\mathbf{i} \in V^{k, \Psi}} \mathcal{R}[\mathbf{i}] \right| < e^{-\eta k}.$$

We highlight that the result is in fact deterministic: it holds almost surely for realizations of Brownian trees  $\mathcal{T}, \mathcal{T}'$ , decomposition  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3^k}$  and any homeomorphism  $\Psi$ . In what follows, we will refer to weak mismatches as simply mismatches.

*Proof.* Although the result is deterministic, we will give a probabilistic proof by interpreting the mass of a subset of  $\mathcal{T}$  as the probability that a point  $X$  sampled uniformly in  $\mathcal{T}$  belongs to that subset, as explained in Remark 8. In all the proof, we treat  $\mathcal{T}, \mathcal{T}'$  as deterministic, compact real trees, equipped with a nonatomic mass measure, and we also treat  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3}$  as deterministic. We pick  $X$  in  $\mathcal{T}$  according to its mass measure. For  $j \geq 0$ , we denote by  $\mathcal{H}_j$  the  $\sigma$ -algebra generated by  $\mathbf{i}_j(X)$ , so that  $(\mathcal{H}_j)_{j \geq 0}$  is a filtration. Note that since  $X$  is uniform, we have for  $j \geq 0$  and  $a \in \{1, 2, 3\}$

$$\mathbb{P}(i_{j+1}(X) = a \mid \mathcal{H}_j) = \frac{|\mathcal{R}[\mathbf{i}_j(X)a]|}{|\mathcal{R}[\mathbf{i}_j(X)]|}. \quad (18)$$

We also note that if  $0 \leq j \leq k$ , then the event that scale  $j$  is a mismatch for  $\Psi$  at  $\mathbf{i}_k(X)$  is  $\mathcal{H}_j$ -measurable<sup>2</sup>, since it only depends on  $\mathbf{i}_j(X)$ . Now, for  $j \geq 0$ , we define

$$M_j := \frac{|\Psi(\mathcal{R}[\mathbf{i}_j(X)])|}{|\mathcal{R}[\mathbf{i}_j(X)]|}. \quad (19)$$

A simple computation using (18) shows that the process  $(M_j)_{j \geq 0}$  is an  $(\mathcal{H}_j)$ -martingale.

The idea behind Proposition 16 is that a mismatch gives an opportunity for  $M_{j+1}$  to be significantly different from  $M_j$ . Since the martingale  $M$  is positive, it converges almost surely, and if its value changes often the limit has to be 0. To obtain a quantitative version of this intuition, we will study  $(\log M_j)_{j \geq 0}$ , which is a supermartingale. We will use the fact that the steps corresponding to mismatches tend to bring the value of this process down by more than an additive constant in expectation. Hence, after a large number  $k$  of steps, either we have seen few mismatches or  $\log M$  has gone down by a lot.

More precisely, let us fix a constant  $\mu > 0$  (to be precised later). For  $j \geq 1$ , we introduce

$$Z_j := \log M_j - \log M_{j-1} \quad \text{and} \quad \tilde{Z}_j = Z_j + \mu \cdot \mathbb{1}_{\{\text{scale } j-1 \text{ is a mismatch for } \mathbf{i}_k(X)\}}.$$

We will prove that if  $\mu > 0$  is chosen sufficiently small, then for all  $j \geq 1$ , we have

$$\mathbb{E} \left[ \exp \left( \frac{1}{2} \tilde{Z}_j \right) \mid \mathcal{H}_{j-1} \right] \leq 1. \quad (20)$$

From here, the result follows from using a Chernoff bound. Indeed, using the last display in cascade, we obtain  $\mathbb{E} \left[ \exp \left( \frac{1}{2} \sum_{j=1}^k \tilde{Z}_j \right) \right] \leq 1$  so that we have

$$\mathbb{P} \left( \sum_{j=1}^k \tilde{Z}_j \geq 2\eta k \right) \leq \exp(-\eta k).$$

Writing  $\sum_{j=1}^k \tilde{Z}_j = \log M_k + \mu \cdot \sum_{j=1}^k \mathbb{1}_{\{\text{scale } j-1 \text{ is a mismatch for } \mathbf{i}_k(X)\}}$ , we obtain

$$\mathbb{P} \left( \log M_k \geq (2\eta - \mu\beta/2)k \text{ and } \mathbf{i}_k(X) \text{ has more than } \frac{\beta}{2}k \text{ mismatches} \right) \leq \exp(-\eta k),$$

which is what we want to prove if we set  $\eta = \frac{\mu\beta}{8}$ .

<sup>2</sup>This is the reason why we are looking at the weak mismatches of Definition 15, and not at the mismatches of Definition 12: the assumption  $i_{j+1} = 3$  in the definition of a good scale is not  $\mathcal{H}_j$ -measurable.

So now, we only have to prove (20) for some value  $\mu > 0$ . First, on the event that scale  $j - 1$  is not a mismatch, we have

$$\mathbb{E} \left[ e^{\tilde{Z}_j} \mid \mathcal{H}_{j-1} \right] = \mathbb{E} \left[ e^{Z_j} \mid \mathcal{H}_{j-1} \right] = \mathbb{E} \left[ \frac{M_j}{M_{j-1}} \mid \mathcal{H}_{j-1} \right] = 1$$

by the martingale property. Therefore, by concavity of  $x \mapsto x^{\frac{1}{2}}$  and Jensen's inequality, we have  $\mathbb{E} \left[ e^{\frac{1}{2}\tilde{Z}_j} \mid \mathcal{H}_{j-1} \right] \leq 1^{\frac{1}{2}} = 1$ . Hence we only have to focus on the event where scale  $j - 1$  is a mismatch. For this, we use the following lemma.

**Lemma 17.** *There exists  $\mu = \mu(\alpha, \delta)$  such that the following holds. Let  $(p_1, p_2, p_3)$  and  $(q_1, q_2, q_3)$  be two elements of the simplex  $\{(x_1, x_2, x_3) \mid x_1 + x_2 + x_3 = 1\}$ , and let  $Z$  be a random variable given by*

$$Z = \log q_I - \log p_I, \quad \text{where for all } i \in \{1, 2, 3\}, \quad \mathbb{P}(I = i) = p_i.$$

*If  $\max_{1 \leq i \leq 3} |p_i - q_i| \geq \delta$  and  $p_1, p_2, p_3 \geq \alpha$ , then we have*

$$\mathbb{E} \left[ \exp \left( \frac{1}{2}(Z + \mu) \right) \right] \leq 1. \quad (21)$$

We then apply the lemma to the random variable  $\tilde{Z}_j$  conditionally on  $\mathcal{H}_{j-1}$ , on the event that  $j - 1$  is a mismatch, by taking  $p_i = \frac{|\mathcal{R}[\mathbf{i}_{j-1}(X) \mathbf{i}]|}{|\mathcal{R}[\mathbf{i}_{j-1}(X)]|}$  and  $q_i = \frac{|\Psi(\mathcal{R}[\mathbf{i}_{j-1}(X) \mathbf{i}])|}{|\Psi(\mathcal{R}[\mathbf{i}_{j-1}(X)])|}$ . This ensures that (20) is satisfied for the appropriate choice of  $\mu$  given in the lemma.  $\square$

*Proof of Lemma 17.* For  $\mu > 0$ , we have

$$\mathbb{E} \left[ \exp \left( \frac{1}{2}(Z + \mu) \right) \right] = e^{\frac{1}{2}\mu} \sum_{i=1}^3 p_i^{\frac{1}{2}} \cdot q_i^{\frac{1}{2}}.$$

For each term, we have  $p_i^{\frac{1}{2}} \cdot q_i^{\frac{1}{2}} \leq \frac{p_i + q_i}{2}$ . Moreover, let  $a$  be such that  $|p_a - q_a| \geq \delta$ . We have

$$p_a^{\frac{1}{2}} \cdot q_a^{\frac{1}{2}} = \frac{p_a + q_a}{2} - \frac{1}{2} \left( p_a^{\frac{1}{2}} - q_a^{\frac{1}{2}} \right)^2 \leq \frac{p_a + q_a}{2} - \frac{1}{2} \left( \frac{\delta}{2} \right)^2,$$

since  $\sqrt{y} - \sqrt{x} \geq \frac{1}{2}(y - x)$  for  $0 \leq x \leq y \leq 1$ . Hence, we have

$$\mathbb{E} \left[ \exp \left( \frac{1}{2}(Z + \mu) \right) \right] = e^{\frac{1}{2}\mu} \sum_{i=1}^3 p_i^{\frac{1}{2}} \cdot q_i^{\frac{1}{2}} \leq e^{\frac{1}{2}\mu} \left( -\frac{\delta^2}{8} + \sum_{i=1}^3 \frac{p_i + q_i}{2} \right) = e^{\frac{1}{2}\mu} \left( 1 - \frac{\delta^2}{8} \right).$$

This proves our claim, by taking  $\mu > 0$  small enough.  $\square$

Given Propositions 13 and 16, it is now easy to conclude the proof of Proposition 3.

*Proof of Proposition 3.* We know that almost surely, for  $k$  large enough, the conclusions of Lemma 11 and Proposition 13 hold. We fix  $k$  for which it is the case. Let  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$  and consider some  $\mathbf{i} \in U^{k, \Psi}$ , meaning that  $|\Psi(\mathcal{R}[\mathbf{i}])| > e^{-\eta k} |\mathcal{R}[\mathbf{i}]|$ . Then

- either  $\mathbf{i} \in U^k$  as defined by Lemma 11, meaning that  $\mathbf{i}$  has less than  $\beta k$  good scales,
- or Item 2 of Proposition 13 holds, as Item 1 is prohibited since  $|\Psi(\mathcal{R}[\mathbf{i}])| > e^{-\eta k} |\mathcal{R}[\mathbf{i}]|$ , so the region  $\mathbf{i}$  must have at least  $\lfloor \frac{\beta}{2} k \rfloor$  scales that are  $\delta$ -mismatches, hence also weak  $\delta$ -mismatches. This entails, using our initial assumption on  $\mathbf{i}$ , that  $\mathbf{i} \in V^{k, \Psi}$ .

Therefore, on the event that we considered, we have  $U^{k, \Psi} \subset U^k \cup V^{k, \Psi}$ , so

$$\left| \bigcup_{\mathbf{i} \in U^{k, \Psi}} \mathcal{R}[\mathbf{i}] \right| < e^{-\eta k} + e^{-\gamma k} \leq e^{-\zeta k},$$

where  $\gamma$  and  $\eta$  are given respectively by Lemma 11 and Proposition 16. This concludes the proof by taking  $0 < \zeta < \min(\gamma, \eta)$ .  $\square$

## 4 Proofs of the main results

### 4.1 Hölder homeomorphisms

We start with the proof of Theorem 2, which follows quite straightforwardly from Proposition 3.

*Proof of Theorem 2.* Let  $(\mathcal{T}, (\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3})$  be a decomposed Brownian tree and let  $\mathcal{T}'$  be an independent Brownian tree. Let also  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$  be a homeomorphism. We will find a constant  $\zeta > 0$  such that almost surely, for  $k$  large enough, we can find  $\mathbf{i} \in \mathbb{T}_3^k$  such that

$$\text{diam}(\Psi(\mathcal{R}[\mathbf{i}]))^{1-\zeta} \leq \text{diam}(\mathcal{R}[\mathbf{i}]). \quad (22)$$

Since the maximal diameter over all the regions of level  $k$  tends to 0 as  $k \rightarrow +\infty$ , this will entail that  $\Psi^{-1}$  cannot be  $(1 - \frac{\zeta}{2})$ -Hölder. Since the problem is symmetric in  $\mathcal{T}$  and  $\mathcal{T}'$ , this also shows that  $\Psi$  cannot be  $(1 - \frac{\zeta}{2})$ -Hölder either.

Let  $k \geq 0$ . On the one hand, we know that almost surely, for  $k$  large enough, the conclusions of Proposition 3 and Lemma 9 hold. On the other hand, we recall that by Proposition 6, conditionally on their masses, the regions  $(\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3^k}$  are independent Brownian trees with those respective masses. Moreover, there exists a constant  $u > 0$  such that the Brownian tree  $\mathcal{T}$  of mass 1 satisfies  $\mathbb{P}(\text{diam}(\mathcal{T}) \leq x) \leq \exp(-ux^{-2})$  for all  $x > 0$  (this can e.g. be deduced from the explicit distribution of the maximum [18]). By union bound and Borel–Cantelli, there is a constant  $c > 0$  such that almost surely, we have for  $k$  large enough

$$\min_{\mathbf{i} \in \mathbb{T}_3^k} \left\{ \frac{\text{diam}(\mathcal{R}[\mathbf{i}])}{\sqrt{|\mathcal{R}[\mathbf{i}]|}} \right\} \geq \frac{c}{\sqrt{k}}.$$

Combining this with the upper bound of Lemma 9 (which ensures that  $|\mathcal{R}[\mathbf{i}]|$  behaves roughly as a decreasing exponential in  $k$ ), this implies that for any  $\varepsilon > 0$ , almost surely for  $k$  large enough and  $\mathbf{i} \in \mathbb{T}_3^k$ , we have

$$\text{diam}(\mathcal{R}[\mathbf{i}]) \geq c \sqrt{\frac{|\mathcal{R}[\mathbf{i}]|}{k}} \geq |\mathcal{R}[\mathbf{i}]|^{\frac{1}{2-\varepsilon}}. \quad (23)$$

Controlling the diameter of the regions  $\Psi(\mathcal{R}[\mathbf{i}])$  of  $\mathcal{T}'$  cannot be done in the same way, as we do not have a priori estimates on the shape of  $\Psi(\mathcal{R}[\mathbf{i}])$ . Therefore, we will use the definition of  $\mathcal{T}'$  via the Brownian excursion, and the fact the excursion is Hölder. More precisely, we recall from Section 2.2 that  $\mathcal{T}'$  is built as a quotient of  $[0, 1]$  using a Brownian excursion  $\mathbf{e}' = (e'_t)_{0 \leq t \leq 1}$ . We denote by  $p_{\mathbf{e}'}$  the canonical projection from  $[0, 1]$  to  $\mathcal{T}'$ . Now, for any  $\mathbf{i} \in \mathbb{T}_3$ , the region  $\mathcal{R}[\mathbf{i}]$  is delimited by at most 2 points. Therefore, the same is true for the region  $\Psi(\mathcal{R}[\mathbf{i}])$  of  $\mathcal{T}'$ . This implies that  $\Psi(\mathcal{R}[\mathbf{i}])$  is of the form  $p_{\mathbf{e}'}(I_1 \cup I_2 \cup I_3)$ , where  $I_1 \cup I_2 \cup I_3$  is the union of three sub-intervals of  $[0, 1]$ . By connectedness of  $\Psi(\mathcal{R}[\mathbf{i}])$ , we have

$$\text{diam}(\Psi(\mathcal{R}[\mathbf{i}])) \leq \sum_{j=1}^3 \text{diam}(p_{\mathbf{e}'}(I_j)).$$

Now let  $\varepsilon > 0$ . We know that if  $|s - t|$  is small enough, then  $|e'_s - e'_t| \leq \frac{1}{3}|s - t|^{\frac{1}{2}-\varepsilon}$ . On the other hand, we know that  $|\Psi(\mathcal{R}[\mathbf{i}])|$  goes a.s. to 0 as the depth of  $\mathbf{i}$  goes to  $+\infty$ , so almost surely, for  $k$  large enough and  $\mathbf{i}$  of depth  $k$ , we can write

$$\text{diam}(\Psi(\mathcal{R}[\mathbf{i}])) \leq \sum_{j=1}^3 \text{diam}(p_{\mathbf{e}'}(I_j)) \leq \sum_{j=1}^3 \frac{1}{3}|I_j|^{\frac{1}{2}-\varepsilon} \leq |\Psi(\mathcal{R}[\mathbf{i}])|^{\frac{1}{2}-\varepsilon}, \quad (24)$$

since  $p_{\mathbf{e}'}$  is measure-preserving.



We can finally put things together. Using the conclusion of Proposition 3, almost surely for  $k$  large enough, there exists  $\mathbf{i} \in \mathbb{T}_3^k$  such that

$$|\Psi(\mathcal{R}[\mathbf{i}])| \leq e^{-\eta k} |\mathcal{R}[\mathbf{i}]| \leq |\mathcal{R}[\mathbf{i}]|^{1+\eta/C}, \quad (25)$$

where the second inequality comes from Lemma 9. For this  $\mathbf{i}$ , we have

$$\text{diam}(\Psi(\mathcal{R}[\mathbf{i}])) \stackrel{(24)}{\leq} |\Psi(\mathcal{R}[\mathbf{i}])|^{\frac{1}{2}-\varepsilon} \stackrel{(25)}{\leq} |\mathcal{R}[\mathbf{i}]|^{(\frac{1}{2}-\varepsilon)(1+\frac{\eta}{C})} \stackrel{(23)}{\leq} \text{diam}(\mathcal{R}[\mathbf{i}])^{(\frac{1}{2}-\varepsilon)(2-\varepsilon)(1+\frac{\eta}{C})}.$$

From there, we just need to take  $\varepsilon > 0$  small enough to conclude. This proves the theorem and we can take  $\varepsilon_2 < \frac{\eta}{2C}$ .  $\square$

## 4.2 Maximum agreement subtree bound

In this section, we prove Theorem 1 from two intermediate results. The first one, Corollary 18, is a direct corollary of Proposition 3. The second one, Lemma 19, roughly says that the square root upper bound holds simultaneously in all regions of  $T_n$  and  $T'_n$  (its proof relies on the same ideas as the classic square root upper bound). We will only state this lemma here, and prove it in the next subsection.

**Corollary 18.** *Let  $(\mathcal{T}, (\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3})$  be a decomposed Brownian tree and let  $\mathcal{T}'$  be an independent Brownian tree. There exists a constant  $\rho > 0$  such that with probability  $1 - o_k(1)$ , for any homeomorphism  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$ , we have*

$$\sum_{\mathbf{i} \in \mathbb{T}_3^k} \sqrt{|\mathcal{R}[\mathbf{i}]| \cdot |\Psi(\mathcal{R}[\mathbf{i}])|} \leq \exp(-\rho \cdot k).$$

*Proof.* Let  $\xi, \eta > 0$  be given by Proposition 3, and let  $\Psi : \mathcal{T} \rightarrow \mathcal{T}'$  be a homeomorphism. For  $U^{k, \Psi}$  defined as in Proposition 3 and on the event of probability  $1 - o_k(1)$  on which its conclusion holds, we can write for  $k$  large enough

$$\begin{aligned} \sum_{\mathbf{i} \in \mathbb{T}_3^k} \sqrt{|\mathcal{R}[\mathbf{i}]| \cdot |\Psi(\mathcal{R}[\mathbf{i}])|} &= \sum_{\mathbf{i} \in \mathbb{T}_3^k \setminus U^{k, \Psi}} \sqrt{|\mathcal{R}[\mathbf{i}]| \cdot |\Psi(\mathcal{R}[\mathbf{i}])|} + \sum_{\mathbf{i} \in U^{k, \Psi}} \sqrt{|\mathcal{R}[\mathbf{i}]| \cdot |\Psi(\mathcal{R}[\mathbf{i}])|} \\ &\leq \sum_{\mathbf{i} \in \mathbb{T}_3^k \setminus U^{k, \Psi}} |\mathcal{R}[\mathbf{i}]| \cdot \sqrt{\frac{|\Psi(\mathcal{R}[\mathbf{i}])|}{|\mathcal{R}[\mathbf{i}]|}} + \sqrt{\sum_{\mathbf{i} \in U^{k, \Psi}} |\mathcal{R}[\mathbf{i}]|} \cdot \sqrt{\sum_{\mathbf{i} \in U^{k, \Psi}} |\Psi(\mathcal{R}[\mathbf{i}])|} \\ &\leq 1 \cdot e^{-\eta k/2} + e^{-\xi k/2} \cdot 1 \\ &< e^{-\rho k}, \end{aligned}$$

where  $\rho$  is chosen so that  $\rho < \frac{1}{2} \min(\eta, \xi)$ , the first inequality follows from the Cauchy–Schwarz inequality, and the second from Proposition 3.  $\square$

**Lemma 19.** *Let  $\varepsilon > 0$ . With high probability as  $n \rightarrow \infty$ , for any two regions  $R \subset T_n$  and  $R' \subset T'_n$ , we have*

$$\text{MAST}(R, R') \leq 4e\sqrt{2} \cdot \left( n^\varepsilon \vee \sqrt{\frac{(\#R) \cdot (\#R')}{n}} \right). \quad (26)$$

We can now prove Theorem 1.

*Proof of Theorem 1.* We let  $\theta = \min\left(\frac{1}{2C}, \frac{1}{4 \log 3}\right)$ , where  $C > 0$  is given by Lemma 9 and we take  $k = \lfloor \theta \log n \rfloor$ . Note that in particular, the number of regions of the recursive decomposition of

$\mathcal{T}$  at scale  $k$  is  $3^k \leq n^{\theta \log 3} \leq n^{\frac{1}{4}}$ . We now let  $\varepsilon = \frac{\theta \rho}{4}$ , where  $\rho$  is given by Corollary 18. We assume that  $T'_n$  is coupled with an  $n$ -pointed Brownian tree  $(\mathcal{T}', (X'_j)_{1 \leq j \leq n})$  in the way described in Section 2.4. We also assume that  $(\mathcal{T}', (X'_j))$  is independent from  $\mathcal{T}, (\mathcal{R}[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3^k}$  and  $T_n$ .

Suppose that for some  $S \subset \{1, \dots, n\}$ , we have  $T_n|_S = T'_n|_S = t$ . Then we claim that there exists a homeomorphism  $\Psi$  from  $\mathcal{T}$  to  $\mathcal{T}'$  such that  $\Psi(X_j) = X'_j$  for all  $j \in S$ . Indeed, up to reparametrization of the edges, there exists a unique embedding  $\varphi$  of  $t$  into  $\mathcal{T}$  that sends the leaf labelled  $j$  to  $X_j$  for any label  $j$  appearing in  $t$ , and a unique embedding  $\varphi'$  of  $t$  into  $\mathcal{T}'$  that sends similarly  $j$  to  $X'_j$ . For every edge  $e = (x_e, y_e)$  of  $t$ , let  $\mathcal{T}_e$  be the set of points  $x \in \mathcal{T}$  such that the closest point of  $\varphi(t)$  to  $x$  belongs to  $\varphi(e)$ . We define similarly the region  $\mathcal{T}'_e \subset \mathcal{T}'$ . For every  $e$ , the regions  $\mathcal{T}_e$  and  $\mathcal{T}'_e$  are compact real trees where branching points are dense and all have degree 3, so they have the same topology by [8, Theorem 1] (see also [12])<sup>3</sup>. Therefore, there exists a homeomorphism  $\Psi_e : \mathcal{T}_e \rightarrow \mathcal{T}'_e$  such that  $\Psi_e(\varphi(x_e)) = \varphi'(x_e)$  and  $\Psi_e(\varphi(y_e)) = \varphi'(y_e)$ . The homeomorphism  $\Psi$  is obtained by patching the  $\Psi_e$  together. Finally, for any  $\mathbf{i} \in \mathbb{T}_3^k$ , we denote by  $R[\mathbf{i}]$  the smallest region of  $T_n$  that contains all the labels  $j \in \{1, \dots, n\}$  such that  $X_j \in \mathcal{R}[\mathbf{i}]$ . We define similarly the regions  $R'[\mathbf{i}]$  of  $T'_n$  using  $\mathcal{T}'$  and  $(X'_j)$ .

For every  $\mathbf{i} \in \mathbb{T}_3^k$ , note that by definition of  $R[\mathbf{i}], R'[\mathbf{i}]$  and by the fact that  $\Psi(X_j) = X'_j$  for  $j \in S$ , we have  $S \cap R[\mathbf{i}] = S \cap R'[\mathbf{i}]$ . In particular, this subset induces the same subtree in  $R[\mathbf{i}]$  and in  $R'[\mathbf{i}]$ . Therefore, we can write

$$\#S = \sum_{\mathbf{i} \in \mathbb{T}_3^k} \#(S \cap R[\mathbf{i}]) \leq \sum_{\mathbf{i} \in \mathbb{T}_3^k} \text{MAST}(R[\mathbf{i}], R'[\mathbf{i}]).$$

On the other hand, Lemma 19 ensures that with probability  $1 - o_n(1)$ , for all  $\mathbf{i} \in \mathbb{T}_3^k$  we have

$$\begin{aligned} \text{MAST}(R[\mathbf{i}], R'[\mathbf{i}]) &\leq 24e \cdot \left( n^\varepsilon \vee \sqrt{\frac{(\#R[\mathbf{i}])(\#R'[\mathbf{i}])}{n}} \right) \\ &\leq 24e \left( n^\varepsilon + \sqrt{\frac{(\#R[\mathbf{i}])(\#R'[\mathbf{i}])}{n}} \mathbb{1}_{\{\#R[\mathbf{i}] \geq n^\varepsilon \text{ and } \#R'[\mathbf{i}] \geq n^\varepsilon\}} \right). \end{aligned}$$

We now use Lemma 5: with probability  $1 - o_n(1)$ , for any  $\mathbf{i} \in \mathbb{T}_3^k$  such that  $\#R[\mathbf{i}], \#R'[\mathbf{i}] \geq n^\varepsilon$ , we have  $\#R[\mathbf{i}] \leq n^{1+\varepsilon} \cdot |\mathcal{R}[\mathbf{i}]|$  and  $\#R'[\mathbf{i}] \leq n^{1+\varepsilon} \cdot |\Psi(\mathcal{R}[\mathbf{i}])|$ , so

$$\sqrt{\frac{(\#R[\mathbf{i}])(\#R'[\mathbf{i}])}{n}} \leq n^{\frac{1}{2}+\varepsilon} \cdot \sqrt{|\mathcal{R}[\mathbf{i}]| \cdot |\Psi(\mathcal{R}[\mathbf{i}])|}.$$

Finally putting everything together, we get

$$\begin{aligned} \#S &\leq 24e \cdot \left( 3^k \cdot n^\varepsilon + n^{\frac{1}{2}+\varepsilon} \cdot \sum_{\mathbf{i} \in \mathbb{T}_3^k} \sqrt{|\mathcal{R}[\mathbf{i}]| \cdot |\Psi(\mathcal{R}[\mathbf{i}])|} \right) \\ &\stackrel{\text{Cor. 18}}{=} O(n^{\frac{1}{4}+\varepsilon}) + O(n^{\frac{1}{2}+\varepsilon-\rho\theta}) \\ &\stackrel{\text{choice of } \varepsilon}{<} n^{\frac{1}{2}-\frac{\rho\theta}{2}}, \end{aligned}$$

with probability  $1 - o_n(1)$ . □

<sup>3</sup>The results in [8] are only stated for unpointed Brownian trees, but the proofs extend straightforwardly to bipointed trees.

### 4.3 The refined square root bound

This section is devoted to proving Lemma 19. Before getting to the proof, we will first state and prove two other intermediate results, Lemma 20 and Lemma 21.

First, Lemma 20, stated below, is a consequence of [10, Lemma 4.1], see also [7, Lemma 4.1, Proposition 4.2], but expressed in a slightly different context. We provide a quick proof for completeness, adapted from the same references. We say that a random labelled tree is *exchangeable* if its distribution is invariant under uniform random permutation of the labels on the leaves.

**Lemma 20.** *Suppose that  $S_m$  and  $S'_m$  are independent exchangeable random variables on  $\mathcal{B}_m$  with  $m$  leaves (with possibly different distribution). Then for any  $s \geq 1$ ,*

$$\mathbb{P}(\text{MAST}(S_m, S'_m) \geq s) \leq \binom{m}{s} \frac{2^{s-2s}}{s!}.$$

In particular

$$\mathbb{P}(\text{MAST}(S_m, S'_m) \geq 2e\sqrt{m}) \leq \exp(-2e(\log 2)\sqrt{m} + o(\sqrt{m}))$$

as  $m \rightarrow \infty$ .

*Proof.* We use a first moment method to write

$$\begin{aligned} \mathbb{P}(\text{MAST}(S_m, S'_m) \geq s) &\leq \mathbb{E} \left[ \sum_{\substack{A \subset \{1, \dots, m\} \\ \#A=s}} \mathbb{1}_{S_m|_A = S'_m|_A} \right] \\ &= \binom{m}{s} \mathbb{P}((S_m)|_{\{1, \dots, s\}} = (S'_m)|_{\{1, \dots, s\}}), \end{aligned}$$

where the last equality follows from the exchangeability of the leaf-labels. Now, we can bound the probability appearing on the right-hand-side of the last display by

$$\mathbb{P}((S_m)|_{\{1, \dots, s\}} = (S'_m)|_{\{1, \dots, s\}}) \leq \sup_{t \in \mathcal{B}_s} \mathbb{P}((S_m)|_{\{1, \dots, s\}} = t).$$

For  $t, t' \in \mathcal{B}_s$  we write  $t \sim t'$  if  $t'$  can be obtained from  $t$  by relabeling its leaves. Note that this defines an equivalence relation on  $\mathcal{B}_s$ . Then for any random variable  $T_s$  on  $\mathcal{B}_s$  that satisfy the exchangeability property, we have

$$\mathbb{P}(T_s = t) = \mathbb{P}(T_s \sim t) \cdot \mathbb{P}(T_s = t \mid T_s \sim t) \leq \mathbb{P}(T_s = t \mid T_s \sim t) = \frac{1}{\#\{t' \mid t' \sim t\}}.$$

Since the distribution of  $(S_m)|_{\{1, \dots, s\}}$  satisfies the exchangeability condition, we just need to check that the number appearing at the denominator on the right-hand-side of the inequality is bounded from below by  $\frac{s!}{2^{s-2s}}$  for any  $t \in \mathcal{B}_s$ . For that, it suffices to prove that the number of graph automorphisms of any tree  $t \in \mathcal{B}_s$  is bounded below by  $2^{s-2s}$ . This follows from the fact that an automorphism of a tree is determined by the image of one leaf and a cyclic ordering of the edges around each node. This proves the first claim of the lemma, and the second follows by the Stirling formula.  $\square$

In the proof of Lemma 19, the above result will be used to control the size of the MAST of two regions  $R \subset T_n$  and  $R' \subset T_n$  in terms of their number of common labels. The goal of the next result is to bound the number of these common labels in terms of  $m = \#R$  and  $m' = \#R'$ .

**Lemma 21.** *Let  $\varepsilon > 0$ . Let  $S$  and  $S'$  be two independent uniform random subsets of  $\{1, \dots, n\}$  of respective sizes  $m, m'$ . Then we have the following bound:*

$$\mathbb{P}\left(\#(S \cap S') \geq 8 \cdot \left(n^\varepsilon \vee \frac{mm'}{n}\right)\right) \leq 2 \exp\left(-\frac{2n^\varepsilon}{3}\right).$$

*Proof.* We write  $\#(S \cap S')$  as a sum of indicators  $\#(S \cap S') = \sum_{i=1}^n \mathbb{1}_{i \in S \cap S'}$ . We also denote by  $(\mathcal{H}_i)_{0 \leq i \leq n}$  the filtration generated by the sequence  $(\mathbb{1}_{i \in S}, \mathbb{1}_{i \in S'})_{1 \leq i \leq n}$ . We can check that

$$\mathbb{P}(i+1 \in S \cap S' \mid \mathcal{H}_i) = \frac{m - \sum_{j=1}^i \mathbb{1}_{j \in S}}{n-i} \cdot \frac{m' - \sum_{j=1}^i \mathbb{1}_{j \in S'}}{n-i} \leq \frac{mm'}{(n-i)^2}.$$

Therefore, we have the following stochastic domination

$$\sum_{i=1}^{\frac{n}{2}} \mathbb{1}_{i \in S \cap S'} \preceq_{\text{st}} \text{Bin}\left(\frac{n}{2}, 1 \wedge \frac{4mm'}{n^2}\right) \preceq_{\text{st}} \text{Bin}\left(\frac{n}{2}, 1 \wedge 4 \cdot \left(n^{\varepsilon-1} \vee \frac{mm'}{n^2}\right)\right).$$

By shuffling the indices, the same domination holds for  $\sum_{i=1}^{\frac{n}{2}} \mathbb{1}_{i \in S \cap S'}$ , so we obtain

$$\begin{aligned} & \mathbb{P}\left(\#(S \cap S') \geq 8 \cdot \left(n^{\varepsilon} \vee \frac{mm'}{n}\right)\right) \\ & \leq \mathbb{P}\left(\sum_{i=1}^{\frac{n}{2}} \mathbb{1}_{i \in S \cap S'} \geq 4 \cdot \left(n^{\varepsilon} \vee \frac{mm'}{n}\right)\right) + \mathbb{P}\left(\sum_{i=\frac{n}{2}+1}^n \mathbb{1}_{i \in S \cap S'} \geq 4 \cdot \left(n^{\varepsilon} \vee \frac{mm'}{n}\right)\right) \\ & \leq 2 \exp\left(-\frac{2n^{\varepsilon}}{3}\right), \end{aligned}$$

where for the last inequality we use that for a binomial random variable  $X$  with expectation  $\mu$ , we have  $\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp(-\frac{\varepsilon\mu}{3})$ .  $\square$

We can finally prove Lemma 19.

*Proof of Lemma 19.* We will actually prove a much stronger statement, which is that the estimate of Lemma 19 holds even if we condition on the shape of  $T_n, T'_n$ .

More precisely, let  $\sigma, \sigma'$  be two independent uniform random permutations of  $\{1, \dots, n\}$ , defined on the same probability space and independent of  $T_n, T'_n$ . We denote by  $T_n^{\sigma}$  (resp.  $(T'_n)^{\sigma'}$ ) the tree obtained from  $T_n$  (resp.  $T'_n$ ) by replacing each label  $j$  by the label  $\sigma(j)$  (resp.  $\sigma'(j)$ ). By exchangeability of the model, the couple  $(T_n^{\sigma}, (T'_n)^{\sigma'})$  has the same distribution as  $(T_n, T'_n)$  so we can prove the lemma for the former, and it will hold for the latter as well. For any region  $R$  (resp.  $R'$ ) of  $T_n$  (resp.  $T'_n$ ), we denote by  $R^{\sigma}$  (resp.  $(R')^{\sigma'}$ ) the corresponding region in  $T_n^{\sigma}$  (resp.  $(T'_n)^{\sigma'}$ ). That is, the label  $j$  is in  $R^{\sigma}$  if and only if  $\sigma^{-1}(j)$  is in  $R$ .

Note that if one of our regions is empty or consists of a single leaf, the result is obvious, so we may focus on regions delimited by 1 or 2 nodes. Let  $E_n$  be the event that there exist two regions  $R \subset T_n$  and  $R' \subset T'_n$  delimited by at most two nodes such that (26) fails for the regions  $R^{\sigma}$  and  $(R')^{\sigma'}$ . We want to show that  $\mathbb{P}(E_n) \rightarrow 0$  as  $n \rightarrow +\infty$ . We will actually show that this is true even if we condition on  $T_n, T'_n$ , that is

$$\max_{t, t' \in \mathcal{B}_n} \mathbb{P}(E_n \mid (T_n, T'_n) = (t, t')) \xrightarrow{n \rightarrow +\infty} 0. \quad (27)$$

For this, we fix  $t, t' \in \mathcal{B}_n$ . From now on, we condition on  $(T_n, T'_n) = (t, t')$ . Let  $r, r'$  be two regions of  $t, t'$  delimited by at most 2 nodes each. Since the number of nodes in those two trees is fixed and equal to  $n-2$ , there are  $O(n^4)$  such pairs  $(r, r')$ . We denote by  $m, m'$  the respective number of leaves of  $r$  and  $r'$ . We denote by  $\text{Lab}(\sigma, r)$  (resp.  $\text{Lab}(\sigma', r')$ ) the set of labels contained in the region  $r^{\sigma}$  (resp.  $(r')^{\sigma'}$ ) of  $T_n^{\sigma}$  (resp.  $(T'_n)^{\sigma'}$ ). Then  $\text{Lab}(\sigma, r)$  and  $\text{Lab}(\sigma', r')$  are two independent random uniform subsets of  $\{1, \dots, n\}$  of respective size  $m$  and  $m'$ .

Any common induced subtree to  $r^{\sigma}$  and  $(r')^{\sigma'}$  can only use leaf-labels that are common to those two regions, so denoting  $L = \text{Lab}(\sigma, r) \cap \text{Lab}(\sigma', r')$  we have

$$\text{MAST}(r^{\sigma}, (r')^{\sigma'}) = \text{MAST}(r^{\sigma}|_L, (r')^{\sigma'}|_L).$$

Moreover, conditionally on  $(T_n, T'_n, L)$ , the random labelled trees  $r^\sigma|_L$  and  $(r')^{\sigma'}|_L$  are independent (maybe with different distributions) and have exchangeable leaf-labels taken from  $L$ .

Piecing things together, we get

$$\begin{aligned} & \mathbb{P} \left( \text{MAST}(r^\sigma, (r')^{\sigma'}) \geq 4e\sqrt{2} \left( n^\varepsilon \vee \sqrt{\frac{mm'}{n}} \right) \mid (T_n, T'_n) = (t, t') \right) \\ & \leq \mathbb{P} \left( \#L \geq 8 \cdot \left( \frac{mm'}{n} \vee n^{2\varepsilon} \right) \mid (T_n, T'_n) = (t, t') \right) \\ & + \mathbb{P} \left( \text{MAST}(r^\sigma|_L, (r')^{\sigma'}|_L) \geq 4e\sqrt{2} \cdot \sqrt{n^{2\varepsilon} \vee \frac{mm'}{n}} \text{ and } \#L \leq 8 \cdot \left( \frac{mm'}{n} \vee n^{2\varepsilon} \right) \mid (T_n, T'_n) = (t, t') \right) \\ & \leq 2 \exp \left( -\frac{2n^{2\varepsilon}}{3} \right) + \exp \left( -4e\sqrt{2}(\log 2)n^\varepsilon + o(n^\varepsilon) \right). \end{aligned}$$

Where we have bounded the first term using Lemma 21, and the second by Lemma 20 (which applies by exchangeability) applied to  $8 \left( \frac{mm'}{n} \wedge n^{2\varepsilon} \right)$ . Moreover, those bounds do not depend on  $(t, t')$  and on  $(r, r')$ , so we can sum over the  $O(n^4)$  choices of  $(r, r')$  to obtain (27).  $\square$

## 5 Discussion of the results

### 5.1 Explicit constants

The goal of this paragraph is to obtain a quantitative lower bound for the constants appearing in Theorems 1 and 2. We do not try to optimize the computations.

**Choice of  $C$  in Lemma 9.** In (6) in the proof of Lemma 9, a quick computation shows that  $\mathbb{E} \left[ e^{-\lambda \log(W_1)} \right] = \frac{1}{1-2\lambda}$  so that we are looking for  $\lambda < \frac{1}{2}$  and  $C > 0$  such that  $\frac{e^{-\lambda C}}{1-2\lambda} \leq \frac{1}{4}$ . Set  $\lambda = \frac{1}{2} - \frac{1}{C}$ , assuming  $C > 2$ . We now want  $\frac{C}{2} e^{1-\frac{C}{2}} \leq \frac{1}{4}$ , so we can take  $C = 7.5$ .

**Choice of  $\alpha, \beta, \gamma$  in Lemma 11.** To get a quantitative bound on the probability  $p$  appearing in (10) we can use a union bound and the fact that  $W_1, W_2, W_3$  all have distribution  $\text{Beta}(\frac{1}{2}, 1)$  with density  $\frac{1}{2\sqrt{x}} \cdot \mathbb{1}_{(0,1]}(x)$ . We get

$$\mathbb{P}(W_1, W_2, W_3 \geq \alpha) \geq 1 - 3\mathbb{P}(W_1 < \alpha) = 1 - 3\sqrt{\alpha}.$$

We decide to take  $\alpha = 10^{-6}$ . We then have  $p \geq \frac{1}{9} \cdot (1 - 3\sqrt{\alpha}) = \frac{0.997}{9}$  and we need  $\beta < \frac{p}{2}$ , so we take  $\beta = \frac{1}{20}$ . Finally, to conclude the proof in a quantitative way, we use Hoeffding's inequality and we find  $\gamma = (p - 2\beta)^2 \approx 10^{-4}$  (its exact value won't be needed for the computations below).

**Choice of  $\delta$  in Proposition 13.** We recall from (14) that the constant  $K$  appearing in Lemma 14 is given by  $K = \frac{6}{\alpha} \cdot e^{C+2} \cdot 2^{\frac{2C+4}{\beta}}$ . Now it follows from (17) in the proof of Proposition 13 that it is sufficient to choose  $\delta$  such that  $6 \cdot \left( \frac{2\delta K}{\alpha\sqrt{\alpha}} \right)^{\frac{\beta}{2}} < 1$ . This is equivalent to

$$\delta < 6^{-\frac{2}{\beta}} \cdot \frac{\alpha^{\frac{3}{2}}}{2K} = \frac{6^{-1-\frac{2}{\beta}}}{2} \cdot e^{-(C+2)} \cdot \alpha^{\frac{5}{2}} \cdot 2^{-\frac{2C+4}{\beta}} = \frac{6^{-41}}{2} \cdot e^{-9.5} \cdot 10^{-15} \cdot 2^{-19 \cdot 20} \approx 1.89 \cdot 10^{-166},$$

so we can take  $\delta = 10^{-166}$ . In particular, considering only  $\beta k$  good scales is the reason why our final value is very small.

**Choice of  $\mu, \eta$  in Lemma 17.** Following the proof of Lemma 17, we choose  $\mu$  so that

$$e^{\frac{1}{2}\mu} \left(1 - \frac{\delta^2}{8}\right) \leq 1 \quad \text{i.e.} \quad \mu \leq -2 \log \left(1 - \frac{\delta^2}{8}\right).$$

For example, we may choose  $\mu = \frac{\delta^2}{10} = 10^{-333}$ . Moreover, we have defined  $\eta$  as  $\frac{\mu\beta}{8} = \frac{\mu}{160} > 10^{-336}$ .

**Choice of  $\xi$  of Proposition 3.** We only need  $\xi < \min(\gamma, \eta)$ , so we can take  $\xi = 10^{-336}$ .

**Choice of  $\rho$  in Corollary 18 and  $\theta, \varepsilon_1$  in Theorem 1, and  $\varepsilon_2$  in Theorem 2.** In the proof of Corollary 18, the constant  $\rho$  has to be chosen so that  $\rho < \frac{1}{2} \min(\eta, \xi) = \frac{\xi}{2}$ . In the proof of Theorem 1, we can take  $\theta = \frac{1}{2C} = \frac{1}{15}$ , and finally  $\varepsilon_1 < \frac{\theta}{2}$ , so we can take  $\varepsilon_1 = 10^{-338}$ . Similarly  $\varepsilon_2$  has to be chosen such that  $\varepsilon_2 < \frac{\eta}{2C} = \frac{\eta}{15}$ , so we can take  $\varepsilon_2 = 10^{-338}$ .

## 5.2 Remarks and open questions

**The expected maximum agreement subtree.** In all the arguments of the paper, the estimates that are stated with probability  $1 - o_n(1)$  actually hold with probability  $1 - O(n^{-a})$  for some (small)  $a > 0$  (or with probability  $1 - O(e^{-ak})$ , which is equivalent since we take  $k$  of order  $\log n$ ). Hence, by Theorem 1 and Lemma 20, we can write

$$\mathbb{P} \left( \text{MAST}(T_n, T'_n) \geq n^{1/2-\varepsilon_1} \right) \leq n^{-a}, \quad \text{and} \quad \mathbb{P} \left( \text{MAST}(T_n, T'_n) \geq Cn^{1/2} \right) \leq e^{-c\sqrt{n}}$$

for some constants  $C, c > 0$ . Therefore, since  $\text{MAST}(T_n, T'_n)$  is bounded by  $n$ , we can write

$$\mathbb{E} [\text{MAST}(T_n, T'_n)] \leq n^{1/2-\varepsilon_1} + n^{-a} \times C\sqrt{n} + e^{-c\sqrt{n}} \times n = O \left( n^{1/2-\min(\varepsilon_1, a)} \right),$$

so the *expected* MAST is also much less than  $\sqrt{n}$ .

**Other models of random trees.** Another natural random tree model where the Maximum Agreement Subtree has been investigated [10, 7] is the Yule–Harding model  $Y_n$ , i.e. the model where the binary tree  $Y_n$  is obtained from  $Y_{n-1}$  by choosing a leaf uniformly at random and splitting it into one node and two leaves. The best known lower and upper bounds for this model are given by [7] and are respectively of order  $n^{0.344}$  and  $\sqrt{n}$ . It seems likely to us that adaptations of the ideas developed in the present paper could be used to prove Theorem 1 for the Yule–Harding model.

Beyond the binary case, another natural question would be to try to estimate the MAST between more general Galton–Watson trees. We believe that similar results could be obtained provided the tail of the offspring distribution is light enough (with the technical difficulty that the coupling between the discrete model and the Brownian tree would not be as simple). On the other hand, when the tail is very heavy, the MAST should become larger because of star-shaped subtrees.

**Optimal regularity for homeomorphisms between continuous random structures.** It seems natural to introduce the following quantity

$$\gamma_+ := \inf \{ \gamma \geq 0 \mid \text{there a.s. exists no homeomorphism } \Psi : \mathcal{T} \rightarrow \mathcal{T}' \text{ that is } \gamma\text{-H\"older} \},$$

which, in some sense, captures how metrically different two independent realizations of the Brownian tree are. Theorem 2 ensures that  $\gamma_+ \leq 1 - 10^{-338}$ , and as it was pointed out in the introduction, Aldous’s construction in [5] amounts to constructing a  $(5 - 2\sqrt{6})$ -H\"older homeomorphism

between  $\mathcal{T}$  and  $\mathcal{T}'$  so  $\gamma_+ \geq 5 - 2\sqrt{6} \approx 0.1010$ . It would be an interesting direction of research to find tighter bounds on  $\gamma_+$  or a good heuristic as to what the value of  $\gamma_+$  may be.

The question of finding the optimal regularity for homeomorphisms between independent copies of random metric spaces can be asked for many other models. It is for example natural to ask whether an analog of Theorem 2 could be proved for some models with the topology of the plane such as the Brownian map [22, 24], or more generally Liouville Quantum Gravity metrics [17].

## A Construction of a Hölder homeomorphism

**Theorem 22.** *Let  $\mathcal{T}$  and  $\mathcal{T}'$  be two independent copies of the Brownian tree. There almost surely exists a homeomorphism  $\Psi$  from  $\mathcal{T}$  to  $\mathcal{T}'$  such that  $\Psi$  and  $\Psi^{-1}$  are both  $\gamma$ -Hölder, for any  $\gamma < 5 - 2\sqrt{6}$ .*

*Proof.* Consider the Aldous decomposition of the tree  $\mathcal{T}$ , as described in Section 2.5, with the associated collection  $(x[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3 \setminus \{\emptyset\}}$  of points and the collection  $(b[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3}$  of branching points. Let also  $(\mathcal{T}', (x'[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3 \setminus \{\emptyset\}}, (b'[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3})$  be an independent copy of  $(\mathcal{T}, (x[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3 \setminus \{\emptyset\}}, (b[\mathbf{i}])_{\mathbf{i} \in \mathbb{T}_3})$ , so that in particular  $\mathcal{T}$  and  $\mathcal{T}'$  are Brownian trees. We construct a map  $\Psi : \{b[\mathbf{i}] \mid \mathbf{i} \in \mathbb{T}_3\} \rightarrow \{b'[\mathbf{i}] \mid \mathbf{i} \in \mathbb{T}_3\}$  by setting  $\Psi(b[\mathbf{i}]) = b'[\mathbf{i}]$ . We are going to show that for any  $\gamma < 5 - 2\sqrt{6}$ , this map is almost surely  $\gamma$ -Hölder. Then, by density of the set of branching points that we consider, we can extend  $\Psi$  in a unique way into a  $\gamma$ -Hölder map from  $\mathcal{T}$  to  $\mathcal{T}'$ . By symmetry, the same can be done with  $\Psi^{-1}$ , so the map  $\Psi$  is indeed a homeomorphism, with  $\Psi$  and  $\Psi^{-1}$  that are  $\gamma$ -Hölder, almost surely. Since this holds almost surely for a countable sequence of values of  $\gamma$  tending to  $5 - 2\sqrt{6}$ , the claim of the theorem holds.

From now on, we fix  $\gamma < 5 - 2\sqrt{6}$  and introduce, for any  $k \geq 1$ ,

$$S_k := \max_{\substack{\mathbf{i}, \mathbf{j} \in \mathbb{T}_3 \\ |\mathbf{i}|, |\mathbf{j}| \leq k}} \frac{d_{\mathcal{T}'}(b'[\mathbf{i}], b'[\mathbf{j}])}{(d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{j}]))^\gamma} \quad \text{and} \quad S := \sup_{\mathbf{i}, \mathbf{j} \in \mathbb{T}_3} \frac{d_{\mathcal{T}'}(b'[\mathbf{i}], b'[\mathbf{j}])}{(d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{j}]))^\gamma} = \sup_{k \geq 1} S_k.$$

Proving that  $\Psi$  is a.s.  $\gamma$ -Hölder continuous on  $\{b[\mathbf{i}] \mid \mathbf{i} \in \mathbb{T}_3\}$  reduces to proving that  $S$  is almost surely finite. We prove that claim in two steps:

- First, we obtain a control on the distances between "neighbor branching points" of level  $k$  that will hold true simultaneously for all pairs of such points. Such points appear as pairs of the form  $(x[\mathbf{i}1], x[\mathbf{i}2])$  in the decomposition described in Section 2.5. This estimate is the content of the next lemma.
- Then, we use this estimate to bound  $S_k - S_{k-1}$ , so that  $S_k$  increases to a finite limit.

**Lemma 23.** *Almost surely, for  $k$  large enough, for any  $\mathbf{i} \in \mathbb{T}_3^k$ , we have*

$$d_{\mathcal{T}'}(x'[\mathbf{i}1], x'[\mathbf{i}2]) \leq k^{-2} \cdot d_{\mathcal{T}}(x[\mathbf{i}1], x[\mathbf{i}2])^\gamma. \quad (28)$$

*Proof of Lemma 23.* Recall Proposition 6, which states that for any  $\mathbf{i} \in \mathbb{T}_3^k$ , conditionally on  $\mathcal{R}[\mathbf{i}]$ , the region  $(\mathcal{R}[\mathbf{i}], x[\mathbf{i}1], x[\mathbf{i}2])$  has the distribution of a bi-pointed Brownian tree of mass  $|\mathcal{R}[\mathbf{i}]|$ . Therefore, for any  $\mathbf{i} \in \mathbb{T}_3$ , conditionally on  $|\mathcal{R}[\mathbf{i}]|$ , the distance  $d_{\mathcal{T}}(x[\mathbf{i}1], x[\mathbf{i}2])$ , is distributed as

$$d_{\mathcal{T}}(x[\mathbf{i}1], x[\mathbf{i}2]) \stackrel{d}{=} |\mathcal{R}[\mathbf{i}]|^{\frac{1}{2}} \cdot Y,$$

where  $Y$  has the Rayleigh distribution with density  $4xe^{-2x^2} dx$  on  $[0, \infty)$ , see for example [21, Theorem 2.11]. The same is of course true for analogous quantities in  $\mathcal{T}'$ .



For any  $\gamma \in (0, 1)$  and  $\mathbf{i} \in \mathbb{T}_3^k$  we write

$$\begin{aligned} & \mathbb{P} \left( d_{\mathcal{T}'}(x'[\mathbf{i}1], x'[\mathbf{i}2]) > k^{-2} \cdot d_{\mathcal{T}}(x[\mathbf{i}1], x[\mathbf{i}2])^\gamma \right) \\ &= \mathbb{P} \left( |\mathcal{R}'[\mathbf{i}]|^{\frac{1}{2}} \cdot Y' > k^{-2} \cdot \left( |\mathcal{R}[\mathbf{i}]|^{\frac{1}{2}} \cdot Y \right)^\gamma \right) \\ &= \mathbb{P} \left( \frac{|\mathcal{R}'[\mathbf{i}]|^{\frac{1}{2}} \cdot Y'}{\left( |\mathcal{R}[\mathbf{i}]|^{\frac{1}{2}} \cdot Y \right)^\gamma} > k^{-2} \right) \\ &\leq k^{2\lambda} \cdot \mathbb{E} \left[ |\mathcal{R}'[\mathbf{i}]|^{\frac{\lambda}{2}} \right] \cdot \mathbb{E} \left[ |\mathcal{R}[\mathbf{i}]|^{-\frac{\lambda\gamma}{2}} \right] \cdot \mathbb{E} \left[ (Y')^\lambda \right] \cdot \mathbb{E} \left[ Y^{-\gamma\lambda} \right], \end{aligned}$$

where in the last line we raised everything to some power  $\lambda < \frac{2}{\gamma}$  and used the Markov inequality. The condition  $\lambda < \frac{2}{\gamma}$  ensures that the quantity  $C(\gamma, \lambda) := \mathbb{E} \left[ (Y')^\lambda \right] \cdot \mathbb{E} \left[ Y^{-\gamma\lambda} \right]$  is finite. On the other hand, it follows from Proposition 6 that  $|\mathcal{R}[\mathbf{i}]|$  is the product of  $k$  i.i.d. Beta(1/2, 1)-random variables. Since the  $p$ -th moment of such a Beta variable is given by  $\frac{1}{2^{p+1}}$  for  $p > -\frac{1}{2}$ , we get

$$\mathbb{E} \left[ |\mathcal{R}'[\mathbf{i}]|^{\frac{\lambda}{2}} \right] \mathbb{E} \left[ |\mathcal{R}[\mathbf{i}]|^{-\frac{\lambda\gamma}{2}} \right] = \left( \frac{1}{(1+\lambda)(1-\gamma\lambda)} \right)^k.$$

The expression  $(1+\lambda)(1-\gamma\lambda)$  appearing in the denominator is maximized in  $\lambda$  at  $\lambda = \frac{1-\gamma}{2\gamma} < \frac{2}{\gamma}$  and the value of the maximum is  $1 + \frac{(1-\gamma)^2}{4\gamma}$ . This value decreases in  $\gamma$  and attains 3 at  $5 - 2\sqrt{6}$ . Hence, for  $\gamma < 5 - 2\sqrt{6}$  we can write a union bound over all  $\mathbf{i} \in \mathbb{T}_3^k$  to get

$$\mathbb{P} \left( \exists \mathbf{i} \in \mathbb{T}_3^k : d_{\mathcal{T}'}(x'[\mathbf{i}1], x'[\mathbf{i}2]) > k^{-2} \cdot d_{\mathcal{T}}(x[\mathbf{i}1], x[\mathbf{i}2])^\gamma \right) \leq k^{\frac{1-\gamma}{\gamma}} \cdot \left( \frac{3}{1 + \frac{(1-\gamma)^2}{4\gamma}} \right)^k \cdot C(\gamma, \frac{1-\gamma}{2\gamma}),$$

which is summable in  $k \geq 1$ . This ensures using the Borel-Cantelli lemma that our lemma holds true.  $\square$

From now on, we let  $K_0$  be the first time for which (28) holds for all  $k \geq K_0$ , and fix  $k \geq K_0$ . Consider two indices in  $\mathbb{T}_3$ , say  $\mathbf{i}$  and  $\mathbf{j}$ , such that  $|\mathbf{i}|, |\mathbf{j}| \leq k$ , and consider the path  $p_{\mathbf{i} \rightarrow \mathbf{j}}$  in the tree  $\mathcal{T}$  going from  $b[\mathbf{i}]$  to  $b[\mathbf{j}]$ . We denote by  $\mathbf{l}$  the index such that  $b[\mathbf{l}]$  is the first point of  $\{b[\mathbf{i}] \mid \mathbf{i} \in \mathbb{T}_3, |\mathbf{i}| \leq k-1\}$  visited by  $p_{\mathbf{i} \rightarrow \mathbf{j}}$  (note that this index can possibly be  $\mathbf{i}$  itself, if  $|\mathbf{i}| \leq k-1$ ). Similarly we let  $\mathbf{m}$  be so that  $b[\mathbf{m}]$  is the last such point. We can then write

$$d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{j}]) = d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{l}]) + d_{\mathcal{T}}(b[\mathbf{l}], b[\mathbf{m}]) + d_{\mathcal{T}}(b[\mathbf{m}], b[\mathbf{j}]).$$

Now, either  $b[\mathbf{i}] = b[\mathbf{l}]$ , in which case their distance is 0, or  $b[\mathbf{i}] \neq b[\mathbf{l}]$  and in that case this pair of branching points can be written as  $\{b[\mathbf{i}], b[\mathbf{l}]\} = \{x[\mathbf{u}1], x[\mathbf{u}2]\}$  for some  $\mathbf{u} \in \mathbb{T}_3$  with  $|\mathbf{u}| = k+1$ . Indeed, two consecutive branching points of depth at most  $k$  along  $p_{\mathbf{i} \rightarrow \mathbf{j}}$  are always of this form, and it always holds that one has depth exactly  $k$  and the other at most  $k-1$ , so  $b[\mathbf{l}]$  is either  $b[\mathbf{i}]$  or the second branching point of  $p_{\mathbf{i} \rightarrow \mathbf{j}}$ . The same considerations hold in  $\mathcal{T}'$  so in any case, by Lemma 23, we have

$$d_{\mathcal{T}'}(b'[\mathbf{i}], b'[\mathbf{l}]) \leq (k+1)^{-2} \cdot d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{l}])^\gamma,$$

and the analogous inequality is true for  $d_{\mathcal{T}'}(b'[\mathbf{m}], b'[\mathbf{j}])$ .

Therefore, for  $k \geq K_0$ , we have

$$\begin{aligned}
& d_{\mathcal{T}'}(b'[\mathbf{i}], b'[\mathbf{j}]) \\
&= d_{\mathcal{T}'}(b'[\mathbf{i}], b'[\mathbf{1}]) + d_{\mathcal{T}'}(b'[\mathbf{1}], b'[\mathbf{m}]) + d_{\mathcal{T}'}(b'[\mathbf{m}], b'[\mathbf{j}]) \\
&\leq (k+1)^{-2} \cdot d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{1}])^\gamma + S_{k-1} \cdot d_{\mathcal{T}}(b[\mathbf{1}], b[\mathbf{m}])^\gamma + (k+1)^{-2} \cdot d_{\mathcal{T}}(b[\mathbf{m}], b[\mathbf{j}])^\gamma \\
&\leq \left( (k+1)^{-\frac{2}{1-\gamma}} + S_{k-1}^{\frac{1}{1-\gamma}} + (k+1)^{-\frac{2}{1-\gamma}} \right)^{1-\gamma} \cdot (d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{1}]) + d_{\mathcal{T}}(b[\mathbf{1}], b[\mathbf{m}]) + d_{\mathcal{T}}(b[\mathbf{m}], b[\mathbf{j}]))^\gamma \\
&\leq \left( S_{k-1}^{\frac{1}{1-\gamma}} + 2(k+1)^{-\frac{2}{1-\gamma}} \right)^{1-\gamma} \cdot d_{\mathcal{T}}(b[\mathbf{i}], b[\mathbf{j}])^\gamma,
\end{aligned}$$

where we used the Hölder inequality with  $p = \frac{1}{1-\gamma}$  and  $q = \frac{1}{\gamma}$ . This ensures that for  $k \geq K_0$ , we have

$$S_k \leq \left( S_{k-1}^{\frac{1}{1-\gamma}} + 2(k+1)^{-\frac{2}{1-\gamma}} \right)^{1-\gamma} \leq S_{k-1} + \frac{2}{(k+1)^2},$$

and so  $\sup_{k \geq 1} S_k < \infty$ . This concludes the proof.  $\square$

## References

- [1] L. ADDARIO-BERRY, N. BROUTIN, AND C. GOLDSCHMIDT, *Critical random graphs: limiting constructions and distributional properties*, Electron. J. Probab., 15 (2010), pp. 741–775. Id/No 25.
- [2] D. ALDOUS, *The Continuum Random Tree. I*, The Annals of Probability, 19 (1991), pp. 1 – 28.
- [3] ———, *The Continuum Random Tree III*, The Annals of Probability, 21 (1993), pp. 248 – 289.
- [4] D. ALDOUS, *Recursive self-similarity for random trees, random triangulations and Brownian excursion*, Ann. Probab., 22 (1994), pp. 527–545.
- [5] D. J. ALDOUS, *On the largest common subtree of random leaf-labeled binary trees*, SIAM J. Discrete Math., 36 (2022), pp. 299–314.
- [6] F. BASSINO, M. BOUVEL, M. DRMOTA, V. FÉRAY, L. GERIN, M. MAZOUN, AND A. PIÉROT, *Linear-sized independent sets in random cographs and increasing subsequences in separable permutations*, Combinatorial Theory, 2 (2022).
- [7] D. I. BERNSTEIN, L. S. T. HO, C. LONG, M. STEEL, K. S. JOHN, AND S. SULLIVANT, *Bounds on the expected size of the maximum agreement subtree*, SIAM J. Discrete Math., 29 (2015), pp. 2065–2074.
- [8] M. BONK AND H. TRAN, *The continuum self-similar tree*, in Fractal Geometry and Stochastics VI, U. Freiberg, B. Hambly, M. Hinz, and S. Winter, eds., Cham, 2021, Springer International Publishing, pp. 143–189.
- [9] J. BORGA, W. DA SILVA, AND E. GWYNNE, *Power-law bounds for increasing subsequences in brownian separable permutations and homogeneous sets in brownian cographs*, (2023).
- [10] D. BRYANT, A. MCKENZIE, AND M. STEEL, *The size of a maximum agreement subtree for random binary trees*, in Bioconsensus. DIMACS working group meetings on bioconsensus, October 25–26, 2000 and October 2–5, 2001, DIMACS Center, Providence, RI: American Mathematical Society (AMS), 2003, pp. 55–65.
- [11] D. BURAGO, Y. BURAGO, AND S. IVANOV, *A course in metric geometry*, vol. 33 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001.
- [12] D. CROYDON AND B. HAMBLY, *Self-similarity and spectral asymptotics for the continuum random tree*, Stochastic Processes and their Applications, 118 (2008), pp. 730–754.

- [13] D. DUFRESNE, *Algebraic properties of beta and gamma distributions, and applications*, Adv. Appl. Math., 20 (1998), pp. 285–299.
- [14] C. R. FINDEN AND A. D. GORDON, *Obtaining common pruned trees*, Journal of Classification, 2 (1985), pp. 255–276.
- [15] A. D. GORDON, *On the assessment and comparison of classifications*, Analyse de Données et Informatique, (1980), p. 149–160.
- [16] M. GROMOV, *Carnot-Carathéodory spaces seen from within*, Birkhäuser Basel, Basel, 1996, pp. 79–323.
- [17] E. GWYNNE AND J. MILLER, *Existence and uniqueness of the Liouville quantum gravity metric for  $\gamma \in (0, 2)$* , Inventiones Mathematicae, 223 (2021), pp. 213–333.
- [18] D. P. KENNEDY, *The distribution of the maximum Brownian excursion*, J. Appl. Probability, 13 (1976), pp. 371–376.
- [19] A. KHEZELI, *An improved lower bound on the largest common subtree of random leaf-labeled binary trees*, arXiv:2208.05148 (2022).
- [20] E. KUBICKA, G. KUBICKI, AND F. MCMORRIS, *On agreement subtrees of two binary trees*, Congressus Numerantium, (1992), pp. 217–217.
- [21] J.-F. LE GALL, *Random trees and applications*, Probab. Surv., 2 (2005), pp. 245–311.
- [22] ———, *Uniqueness and universality of the Brownian map*, Ann. Probab., 41 (2013), pp. 2880–2960.
- [23] A. MARKIN, *On the extremal maximum agreement subtree problem*, Discrete Applied Mathematics, 285 (2020), pp. 612–620.
- [24] G. MIERMONT, *The Brownian map is the scaling limit of uniform random plane quadrangulations*, Acta Math., 210 (2013), pp. 319–401.
- [25] P. MISRA AND S. SULLIVANT, *Bounds on the expected size of the maximum agreement subtree for a given tree shape*, SIAM J. Discrete Math., 33 (2019), pp. 2316–2325.
- [26] J. PITMAN, *Combinatorial stochastic processes*, vol. 1875 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.
- [27] B. PITTEL, *Expected number of induced subtrees shared by two independent copies of a random tree*, SIAM Journal on Discrete Mathematics, 37 (2023), pp. 1–16.
- [28] J.-L. RÉMY, *Un procédé itératif de dénombrement d’arbres binaires et son application à leur génération aléatoire*, RAIRO. Informatique théorique, 19 (1985), pp. 179–195.
- [29] M. STEEL AND T. WARNOW, *Kaikoura tree theorems: Computing the maximum agreement subtree*, Information Processing Letters, 48 (1993), pp. 77–82.