Electronic Journal of Probability

Electron. J. Probab. **25** (2020), article no. 131, 1–28.
ISSN: 1083-6489 https://doi.org/10.1214/20-EJP522

Exchangeable hierarchies and mass-structure of weighted real trees*

Noah Forman†

Abstract

Rooted, weighted continuum random trees are used to describe limits of sequences of random discrete trees. Formally, they are random quadruples $(\mathcal{T}, d, r, p)$, where $(\mathcal{T}, d)$ is a tree-like metric space, $r \in \mathcal{T}$ is a distinguished *root*, and p is a probability measure on this space. Intuitively, these trees have a combinatorial "underlying branching structure" implied by their topology but otherwise independent of the metric d. We explore various ways of making this rigorous, using the weight p to do so without losing the fractal complexity possible in continuum trees. We introduce a notion of mass-structural equivalence and show that two rooted, weighted $\mathbb{R}$-trees are equivalent in this sense if and only if the discrete hierarchies derived by i.i.d. sampling from their weights, in a manner analogous to Kingman's paintbox, have the same distribution. We introduce a family of trees, called "interval partition trees" that serve as representatives of mass-structure equivalence classes, and which naturally represent the laws of the aforementioned hierarchies.

Keywords: real tree; continuum random tree; exchangeability; hierarchy; interval partition.
MSC2020 subject classifications: 60B05; 60G09; 60C05.
Submitted to EJP on March 17, 2018, final version accepted on September 6, 2020.
Supersedes arXiv:1801.02700.

1 Introduction

This paper explores three related ideas: a notion of "mass-structural equivalence" between rooted, weighted real trees; a family of such trees, called "interval partition (IP) trees," in which the metric is, in a sense, specified by the weight and underlying branching structure; and continuum random tree representations of exchangeable random hierarchies on $\mathbb{N}$.

Definition 1.1. *A* real tree *(*$\mathbb{R}$-tree*) is a complete, separable metric space* $(\mathcal{T}, d)$ *with the property that: (i) for each* $x, y \in \mathcal{T}$*, there is a unique non-self-intersecting path in* $\mathcal{T}$

*Research supported in part by EPSRC grant EP/K029797/1 and NSF grant DMS-1444084.
†McMaster University, Canada. E-mail: noah.forman@gmail.com

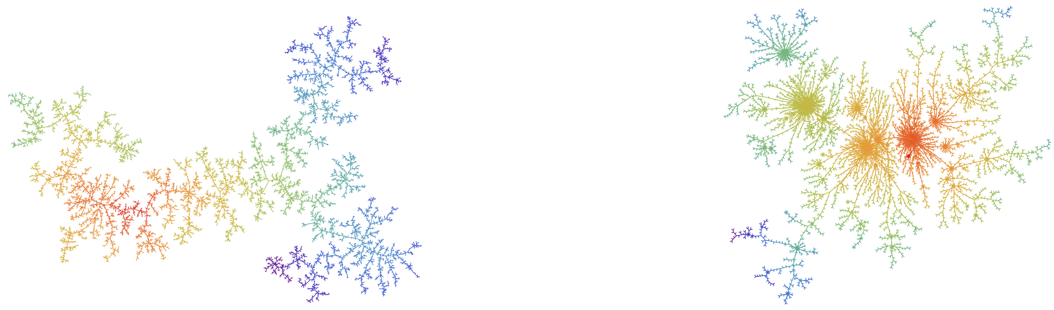

Figure 1: Left: BCRT. Right: Stable(1.1) CRT. Simulations by I. Kortchemski.

from x to y, called a segment $[[x,y]]_{\mathcal{T}}$*, and (ii) each segment is isometric to a closed real interval.*

A rooted, weighted $\mathbb{R}$-tree *is a quadruple $(\mathcal{T},d,r,p)$, where $(\mathcal{T},d)$ is a $\mathbb{R}$-tree, $r \in \mathcal{T}$ is a distinguished vertex called the* root*, and p is a probability distribution on $\mathcal{T}$ with respect to the Borel σ-algebra generated by d.*

We call two rooted, weighted $\mathbb{R}$-trees isomorphic *if there exists a root- and weight-preserving isometry between them.*

$\mathbb{R}$-trees have long been studied by topologists, e.g. [9, 34]. Random $\mathbb{R}$-trees, called *continuum random trees* (*CRTs*), were first studied by Aldous [1, 2]. In particular, Aldous introduced the Brownian CRT (BCRT), which arises as a scaling limit of various families of random discrete trees, including critical Galton-Watson trees conditioned on total progeny. The BCRT is a random fractal in the sense that, if we decompose it around a suitably chosen random branch point, then the components are each distributed as scaled copies of a BCRT and are conditionally independent given their sizes. Since Aldous's work, other authors have introduced similarly complex CRTs, such as the Stable CRTs [11, 10]. See Figure 1. For a survey of the field, see [12, 23].

Authors often take "CRT" to refer only to random $\mathbb{R}$-trees that share certain properties with the BCRT. We use this term more generally to refer to random $\mathbb{R}$-trees, with the only additional assumption being boundedness.

In a rooted $\mathbb{R}$-tree $(\mathcal{T},d,r)$, a point $x \in \mathcal{T}$ is a *branch point* if $\mathcal{T} \setminus \{x\}$ has at least three connected components or a *leaf* if $\mathcal{T} \setminus \{x\}$ has only one connected component. The complement of the set of leaves is the *skeleton* of the tree. The *fringe subtree* of $(\mathcal{T},d,r)$ rooted at x is

$$F_{\mathcal{T}}(x) := \{y \in \mathcal{T} : x \in [[r,y]]_{\mathcal{T}}\}. \tag{1.1}$$

The "between-ness" relations among the branch points, leaves, and root comprise complex "underlying combinatorial tree structures" within the Brownian and Stable CRTs; e.g. Figure 1 appears to depict two vastly complex combinatorial trees. We are interested in formalizations of the notion of underlying tree structure that depend upon the topology of the tree but not upon the metrization. We will see in Section 6.2 that some more direct approaches to this lose sight of the complexity of the Brownian and other self-similar CRTs. In this paper, we instead formalize the "interaction between tree structure and mass" in rooted, weighted $\mathbb{R}$-trees in order to retain a rich picture of the Brownian and other CRTs.

Definition 1.2. *Consider a rooted, weighted $\mathbb{R}$-tree $(\mathcal{T},d,r,p)$. The* subtree spanned by (the support of) p *is*

$$\text{span}(p) := \bigcup_{x \in \text{support}(p)} [[r,x]]_{\mathcal{T}}. \tag{1.2}$$

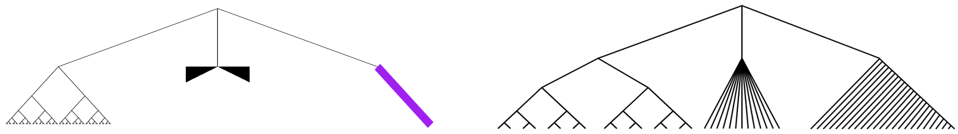

Figure 2: Left: IP tree representation of the hierarchy in (1.7), as in Theorem 1.8(ii). The root is at the top. The black wedge shape in the middle represents an atom of p. The heavy, shaded line represents continuous mass on the skeleton. Right: combinatorial tree representation of a finite version of this hierarchy.

The special points *of* $(\mathcal{T}, d, r, p)$ *are:*

(a) the locations of atoms of p*,*

(b) the branch points of $\mathrm{span}(p)$*, and*

(c) the isolated leaves *of* $\mathrm{span}(p)$*, by which we mean leaves of* $\mathrm{span}(p)$ *that are not limit points of the branch points of* $\mathrm{span}(p)$*.*

Definition 1.3. *Let* $\mathscr{S}_i$ *denote the set of special points of a rooted, weighted* ℝ*-tree* $(\mathcal{T}_i, d_i, r_i, p_i)$ *for* $i = 1, 2$*. A* mass-structural isomophism *between these trees is a bijection* $\phi\colon \mathscr{S}_1 \to \mathscr{S}_2$ *with the following properties.*

(i) Mass preserving*. For every* $x \in \mathscr{S}_1$, $p_1\big([[r_1, x]]_{\mathcal{T}_1}\big) = p_2\big([[r_2, \phi(x)]]_{\mathcal{T}_2}\big)$, $p_1\{x\} = p_2\{\phi(x)\}$*, and* $p_1\big(F_{\mathcal{T}_1}(x)\big) = p_2\big(F_{\mathcal{T}_2}(\phi(x))\big)$*.*

(ii) Structure preserving*. For* $x, y \in \mathscr{S}_1$ *we have* $x \in [[r_1, y]]_{\mathcal{T}_1}$ *if and only if* $\phi(x) \in [[r_2, \phi(y)]]_{\mathcal{T}_2}$*.*

We say that two rooted, weighted ℝ*-trees are* mass-structurally equivalent *if there exists a mass-structural isomorphism from one to the other. It is straightforward to confirm that this is an equivalence relation.*

Definition 1.4. *A rooted, weighted* ℝ*-tree* $(\mathcal{T}, d, r, p)$ *is an* interval partition tree *(*IP tree*) if it possesses the following properties.*

Spanning *Every leaf of* $\mathcal{T}$ *is in the support of* p*, i.e.* $\mathcal{T} = \mathrm{span}(p)$*.*

Spacing *For* $x \in \mathcal{T}$*, if* x *is either a branch point or lies in the support of* p *then*

$$d(r, x) + p(F_{\mathcal{T}}(x)) = 1. \tag{1.3}$$

See Figures 2 and 3 for examples IP trees. Note that, as a consequence of (1.3), each leaf either is an atom of p or sits at unit distance from the root. Relatedly, not all IP trees are locally compact: in each of the IP trees depicted in Figure 3, each branch point is a limit point of other branch points, and each branch extends to a different leaf at unit distance from the root.

Theorem 1.5. *Each mass-structural equivalence class of rooted, weighted* ℝ*-trees contains exactly one isomorphism class of IP trees.*

In light of this theorem, the isomorphism classes of IP trees can be taken as representatives of the mass-structural equivalence classes. We could refer to the isomorphism class of IP trees that are mass-structurally equivalent to a given rooted, weighted ℝ-tree as the *mass-structure* of that tree. As one application of this, we show in Section 6.1 that a Brownian CRT (BCRT) is specified, up to isomorphism, by its mass-structure; i.e. the map from rooted, weighted ℝ-trees to mass-structures has a BCRT-a.s. left inverse.

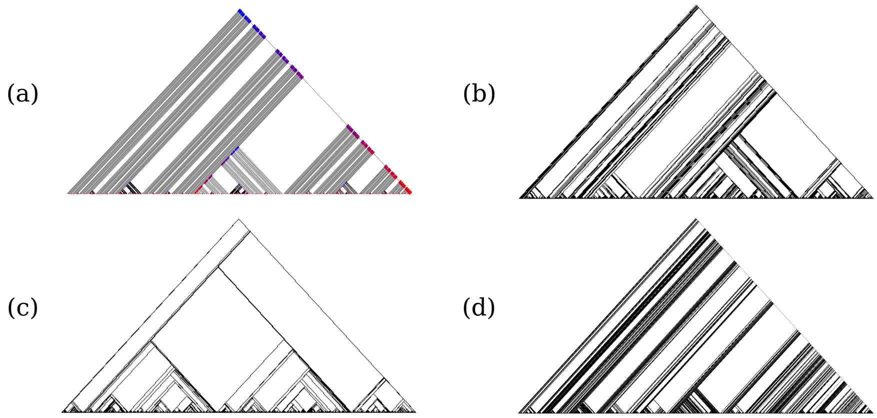

Figure 3: Simulations of IP trees defined in Section 2.2, represented as in Figure 2. Leaves from different branches appear to touch on a horizontal bottom line; this is not intended. Length along this line in (b,c,d) equals continuous mass in the leaf set; in (a) mass is supported in the shaded bits of the skeleton. (a) Fat Cantor IP tree. (b) Brownian IP tree. (c) $(.2, .2)$-IP tree. (d) $(.8, .5)$-IP tree.

Definition 1.6. *(i) A* hierarchy *on a finite set S is a collection $\mathcal{H}$ of subsets of S such that*

(a) if $A, B \in \mathcal{H}$ then $A \cap B$ equals either A or B or $\varnothing$, and

(b) $S \in \mathcal{H}$, $\varnothing \in \mathcal{H}$, and $\{s\} \in \mathcal{H}$ for all $s \in S$.

The trivial hierarchy on S *is*

$$\Xi(S) := \{S, \varnothing\} \cup \{\{a\} : a \in S\}. \tag{1.4}$$

(ii) A hierarchy on $\mathbb{N}$ *is a sequence $(\mathcal{H}_n, n \geq 1)$, with each $\mathcal{H}_n$ a hierarchy on $[n] := \{1, 2, \ldots, n\}$, with the consistency condition that*

$$\mathcal{H}_n = \mathcal{H}_{n+1}\big|_{[n]} := \big\{A \cap [n] : A \in \mathcal{H}_{n+1}\big\} \qquad \text{for } n \geq 1.$$

(iii) Permutations act on hierarchies by relabeling the contents of constituent sets: if $\mathcal{H}$ is a hierarchy on $[n]$ and π a permutation of $[n]$, then

$$\pi(\mathcal{H}) := \{\{\pi(j) : j \in A\} : A \in \mathcal{H}\}.$$

A random hierarchy $\mathcal{H}$ on $[n]$ is exchangeable *if*

$$\pi(\mathcal{H}) \stackrel{d}{=} \mathcal{H} \qquad \text{for every permutation } \pi \text{ of } [n].$$

A hierarchy $(\mathcal{H}_n, n \geq 1)$ on $\mathbb{N}$ is said to be exchangeable if every $\mathcal{H}_n$ is exchangeable. Exchangeable hierarchies on $\mathbb{N}$ were studied in [14].

(iv) A random hierarchy $(\mathcal{H}_n, n \geq 1)$ on $\mathbb{N}$ is independently generated *if for every N and every vector $(A_1, \ldots, A_k)$ of disjoint subsets of $[N]$, the restrictions of $\mathcal{H}_N$ to these subsets, $\big(\mathcal{H}_N|_{A_1}, \ldots, \mathcal{H}_N|_{A_k}\big)$, are jointly independent. We write* e.i.g. *to abbreviate "exchangeable and independently generated."*

The convention in (i) that $\varnothing$ must belong to every hierarchy is intended to avoid unwanted distinctions between two otherwise equal hierarchies, one of which contains

$\varnothing$ while the other does not. We could equivalently require that $\varnothing$ be excluded from every hierarchy. The setup in (ii) serves a similar purpose of avoiding unwanted distinctions: we could instead simply consider the definition in (i) with infinite sets S, but then two hierarchies might be distinct, despite all of their finite restrictions being equal. We want to avoid drawing such distinctions. The method in (ii) of representing an infinite combinatorial object by a projective system has often been used to study limits of exchangeable combinatorial structures; see e.g. [17], [27, Chapter 2.2].

Think of e.i.g. as an analogue in the setting of combinatorial structures to i.i.d. sequences. By way of analogy to de Finetti's Theorem for exchageable sequences, in [14, Theorem 2] it was shown that exchangeable laws of hierarchies on $\mathbb{N}$ can be represented as convex combinations of e.i.g. laws.

A hierarchy on a finite set S can be represented as a tree rooted at S, with the non-empty blocks of the hierarchy as the nodes and singleton blocks as the leaves. It can be constructed by partitioning S, then iteratively partitioning the resulting blocks until only singletons remain. The collection of subsets obtained in the course of this process, plus the empty set, forms the hierarchy.

Now, consider a rooted, weighted $\mathbb{R}$-tree $(\mathcal{T}, d, r, p)$. Let $(x_i,\ i \geq 1)$ be an i.i.d. random sequence with law p. Set

$$\mathcal{H}_n := \big\{\{i \in [n] \colon x_i \in F_{\mathcal{T}}(x)\} \colon x \in \mathcal{T}\big\} \cup \Xi([n]) \qquad \text{for } n \geq 1. \tag{1.5}$$

We say that $(\mathcal{H}_n,\ n \geq 1)$ is *derived by sampling from* $(\mathcal{T}, d, r, p)$. Let $\Theta(\mathcal{T}, d, r, p)$ denote the law of $(\mathcal{H}_n,\ n \geq 1)$. This is an e.i.g. law. If two rooted, weighted $\mathbb{R}$-trees are isomorphic, then Θ maps them to the same law. If $\mathscr{T}$ is an isomorphism class of such trees, we write $\Theta(\mathscr{T})$ to denote the unique e.i.g. law that appears in the image of the class under Θ.

Theorem 1.7. *Two rooted, weighted $\mathbb{R}$-trees are mass-structurally equivalent if and only if they have the same image under Θ.*

For a hierarchy $(\mathcal{H}_n,\ n \geq 1)$, we denote the associated tail σ-algebra by

$$\text{tail}(\mathcal{H}_n) := \bigcap_{j \geq 1} \sigma\left(\mathcal{H}_k|_{\{j,j+1,\ldots,k\}},\ k \geq j\right). \tag{1.6}$$

We resolve [14, Conjecture 1] and strengthen Theorem 5, which was the main result of that paper, as follows.

Theorem 1.8. *(i) For $(\mathcal{H}_n,\ n \geq 1)$ an exchangeable random hierarchy on $\mathbb{N}$, there exists an a.s. unique, $\text{tail}(\mathcal{H}_n)$-measurable random isomorphism class of IP trees, $\mathscr{T}$, such that $\Theta(\mathscr{T})$ is a regular conditional distribution (r.c.d.) for $(\mathcal{H}_n,\ n \geq 1)$ given $\text{tail}(\mathcal{H}_n)$.*

(ii) The map Θ is a bijection from the set of isomorphism classes of IP trees to the set of e.i.g. laws of hierarchies on $\mathbb{N}$.

This theorem is a hierarchies analogue to Kingman's paintbox theorem [22], which describes exchangeable partitions of $\mathbb{N}$, or to de Finetti's theorem for exchangeable sequences of random variables [21]. We recall [14, Example 1].

Example 1.9. Informally, the following is a hierarchy on the interval $[0, 3)$:

$$\mathscr{H} := \{[0,1), [1,2), [2,3)\} \cup \left\{\bigcup\nolimits_{n\geq 1} \left\{ \left[\frac{j}{2^n}, \frac{j+1}{2^n}\right) : 0 \leq j \leq 2^n - 1 \right\}\right\}$$
$$\cup \{[x,3) \colon 2 < x < 3\} \cup \{\{x\} \colon x \in [0,3)\} \cup \{[0,3), \varnothing\};$$

c.f. [14, Definition 5]. Let $(s_i,\ i \geq 1)$ be an i.i.d. sequence of Uniform$[0,3)$ random variables, and define an e.i.g. hierarchy on $\mathbb{N}$ by

$$\mathcal{H}_n := \{\{i \in [n] \colon s_i \in B\} \colon B \in \mathscr{H}\} \cup \Xi([n]) \qquad \text{for } n \geq 1. \tag{1.7}$$

See Figure 2 for combinatorial and IP tree representations of this hierarchy.

In [14], the authors pose the "Naïve conjecture" that exchangeable hierarchies are characterized by a mixture of the three behaviors exhibited in Example 1.9: *macroscopic splitting*, *broom-like explosion*, and *comb-like erosion*. This is formalized in Conjecture 2 of that paper, which is verified by Theorem 1.8 above and the following.

Theorem 1.10. *For $(\mathcal{T}, d, r, p)$ an IP tree, p can be decomposed uniquely as $p^a + p^s + p^l$, with p^a purely atomic, p^s the restriction of length measure to a subset of the skeleton of $\mathcal{T}$, and p^l a diffuse measure on the leaf set of $\mathcal{T}$.*

By length measure we mean the measure supported on the skeleton of $\mathcal{T}$ that assigns mass $d(x,y)$ to each segment $[[x,y]]_{\mathcal{T}}$. The only non-trivial assertion in this theorem is that the diffuse component of p on the skeleton is a restriction of length measure. We formulate the theorem in this way, splitting p into three components, to highlight the connection to the Naïve conjecture. Explosions, erosion, and macroscopic splitting correspond to p^a, p^s, and branch points, respectively, with p^l corresponding to the singletons that are eventually isolated by repeated splitting.

1.1 Applications and related literature

The IP tree representation of a BCRT, mentioned after Theorem 1.5, has been applied in my work on a 1999 conjecture of Aldous [3] on the existence of a continuum analogue to a natural Markov chain on cladograms. My collaborators and I have constructed two representations of this continuum analogue in [15]. The first representation – the one intended by Aldous – is a path-continuous $\mathbb{R}$-tree-valued Markov process, stationary with the law of the BCRT. However, we find that this representation is not strongly Markovian at exceptional times when branch points collide. To obtain a strongly Markovian version of the process, we use an IP tree representation. Informally, this works because IP trees do a better job than the BCRT or other well-studied CRTs at keeping space between branch points.

A parallel effort by Löhr, Mytnick, and Winter [24] has proven existence of a continuum analogue to Aldous's Markov chain on a space of "algebraic measure trees." The algebraic measure trees, introduced in [25] concurrently with and independently of the present work, serve a similar purpose to IP trees in representing mass-structure, but the authors take a more algebraic approach. There is an obvious distinction in that the objects of the present work are all rooted trees, whereas algebraic trees are unrooted. However, one could naturally define rooted algebraic trees or, going the other way, accept a convention of rooting an unrooted weighted tree at its centroid in order to obtain an IP tree representation. It is an open question to determine whether equivalence of algebraic measure tree representations is the same as mass-structural equivalence.

The nested Chinese restaurant process (NCRP) [6] is a Markov chain $(\mathcal{H}_n,\ n \geq 1)$ where each $\mathcal{H}_n$ is an exchangeable hierarchy on $[n]$, and these are projectively consistent, comprising an exchangeable hierarchy on $\mathbb{N}$. It is applied as a Bayesian non-parametric nested topic model, in which a collection of documents is automatically clustered into topics, subtopics, and sub-subtopics, etc.. Rather than being given a fixed hierarchy of topics, such models infer natural clusters from the set of documents they are given. In the nested hierarchical Dirichlet process [26], documents are classed as convex combinations of subtopics, again with a topic tree arising from the NCRP.

Exchangeable hierarchies also relate to fragmentation and coagulation processes [5], in which sets break down or aggregate over time. Hierarchies differ from these processes in that they do not give an account of the *times* at which sets join or break apart; they only describe which sets eventually arise in such a process. Hierarchies relate to other phylogenetic models, as well, such as phylogenetic trees [33]. A more complete catalog of references related to exchangeable hierarchies can be read from [14].

Theorem 1.10, with our interpretation as it relates to explosions, erosion, and macroscopic splitting, is a hierarchies analogue to Bertoin's result that the same three behaviors characterize self-similar fragmentations [4]. In light of Theorems 1.8 and 1.10, IP trees may be understood as recipes for combining and interspersing these behaviors. This insight may be of value where exchangeable hierarchies are used in applications. For example, new nested topic models could be constructed that would allow all three of these dynamics in the infinite limit, rather than solely the iterative infinitary macroscopic splitting that appears in the limit in the nested Chinese restaurant process [6].

In Section 2 we introduce a general "bead-crushing" construction of IP trees and the related notion of strings of beads from [29]. Section 3 recounts relevant background from [14] relating hierarchies to CRTs, then connects this material to IP trees. The main mathematical work of the paper is done in Section 4, with proofs of two key propositions building towards the main results, all of which are then proved in Section 5. In Section 6 we offer final thoughts and open questions, including discussion of the Brownian IP tree.

2 Interval partition trees

We will construct IP trees as subsets of the following space.

Definition 2.1. *Let ℓ_1 denote the Banach space of absolutely summable sequences of reals under the norm $\|(x_i,\, i \geq 1)\| = \sum_i |x_i|$. We write $\ell_1(x,y) := \|y-x\|$. Let $(\mathbf{e}_j,\, j \geq 1)$ be the coordinate vectors, $\mathbf{e}_1 = (1,0,0,\ldots)$, $\mathbf{e}_2 = (0,1,0,\ldots)$, etc.. For $m \geq 1$ let $\pi_m \colon \ell_1 \to$ span$(\mathbf{e}_1,\ldots,\mathbf{e}_m)$ denote the projection onto the first m coordinates, and let π_0 send everything to $(0,0,\ldots)$, which we denote by 0. Let cl denote the topological closure map on subsets of ℓ_1.*

Definition 2.2. *Following Aldous [2], for $x \in \ell_1$ let $[[0,x]]_\ell$ denote the path that proceeds from 0 to x along successive directions:*

$$[[0,x]]_\ell := \{x\} \cup \bigcup_{m\geq 0} \{t\pi_m(x) + (1-t)\pi_{m+1}(x) : t \in [0,1]\}. \tag{2.1}$$

For $x, y \in \ell_1$ with all non-negative coordinates,

$$[[0,x]]_\ell \cap [[0,y]]_\ell = [[0,z]]_\ell$$

for some $z \in \ell_1$, possibly equal to zero. We define

$$(x \wedge y)_\ell := z, \qquad [[x,y]]_\ell := \big([[0,x]]_\ell \cup [[0,y]]_\ell \setminus [[0,z]]_\ell\big) \cup \{z\}. \tag{2.2}$$

For example, if $x = 2\mathbf{e}_1 + \mathbf{e}_3$ then $[[0,x]]_\ell$ is a union of two segments parallel to the first and third coordinate axes, $\mathbf{e}_1[0,2] \cup (2\mathbf{e}_1 + \mathbf{e}_3[0,1])$. Generally, if x has only finitely many non-zero coordinates then the last of these segments terminates at x, and the singleton $\{x\}$ on the right hand side in (2.1) is redundant.

Definition 2.3. *We call a probability measure q with compact support $K \subseteq [0,1]$* uniformized *if $q[0,x) = x$ for every $x \in K$. Let $F \colon \mathbb{R} \to [0,1]$ be a cumulative distribution function for a probability measure μ on $\mathbb{R}$. The* uniformization *of μ is the measure q on $[0,1]$ specified by $q[0,x] = \inf($range$(F) \cap [x,1])$.*

Note that the uniformization q of a measure μ is uniformized, and that if μ is uniformized then $q = \mu$. Gnedin introduced uniformized measures in the context of the following bijection.

Lemma 2.4 (Gnedin [17], Section 3)**.** *The map $q \mapsto [0,1) \setminus \mathrm{support}(q)$ is a bijection from the uniformized probability measures to the open subsets of $(0,1)$.*

We represent a generic open set $U \subseteq (0,1)$ as a disjoint finite or countably infinite union, $\bigcup_i (a_i, b_i)$. Given U, let $q^a = \sum_i (b_i - a_i)\delta_{a_i}$ and let q^d denote the restriction of Lebesgue measure to $[0,1)\setminus U$. Then $q = q^a + q^d$ is uniformized and $U = [0,1)\setminus \mathrm{support}(q)$.

Example 2.5 (Fat Cantor measure)**.** Let $A_0 := [0,1]$. Let $A_1 := A_0 \setminus (3/8, 5/8)$. We carry on recursively, as follows. For $n \geq 1$, A_n comprises 2^n disjoint closed intervals of the same length. We form A_{n+1} by removing an open interval of length 4^{-n-1} from the middle of each component of A_n. This sequence decreases to a *fat Cantor set* $A_\infty = \bigcap_{n\geq 1} A_n$, also called a Smith-Volterra-Cantor set, with Lebesgue measure $1/2$; see [16, p. 89].

The fat Cantor set is closed. We define the *fat Cantor measure* to be the unique uniformized probability measure supported on A_∞. This equals the restriction of Lebesgue measure to A_∞, plus the sum of one atom at the left end of each interval removed in the construction, each having mass equal to the length of the removed interval.

Lemma 2.6. *A probability measure q on $\mathbb{R}$ is uniformized if and only if $([0,L], d, 0, q)$ is an IP tree, where d is the Euclidean metric and L is the maximum of the compact support of q.*

Proof. The Spanning property follows from our definition of L. The Spacing property is then equivalent to the uniformization property. □

This lemma characterizes all IP trees that lack branch points.

2.1 The bead-crushing construction of IP trees

Bead-crushing constructions were introduced in [29] and generalized in [32]; we discuss their "strings of beads" and their construction in Sections 2.2 and 6.1, respectively. Bead-crushing builds upon Aldous's line-breaking construction [2]. We have adapted bead-crushing to yield IP trees. See Figure 4.

Ingredients. A finite or infinite sequence of uniformized probability measures, $(q_n, n \in [1,N])$ or $(q_n, n \geq 1)$, with $L_n := \max(\mathrm{support}(q_n))$ for each n.

Initial step. $(\mathcal{T}_0, p_0) := (\{0\}, \delta_0)$, where 0 denotes the origin in ℓ_1.

Recursive step. Assume $(\mathcal{T}_n, \ell_1, 0, p_n)$ is a rooted, weighted $\mathbb{R}$-tree embedded in the first n coordinates in ℓ_1 and p_n is not purely diffuse. Choose an atom $m_n\delta_{x_n}$ of p_n and choose $a_n \in (0, m_n]$, perhaps at random. Set

$$\begin{aligned} \phi_n(z) &:= x_n + \big(p_n(F_{\mathcal{T}_n}(x_n)) - a_n\big)\mathbf{e}_{n+1} + z a_n \mathbf{e}_{n+1} \quad \text{for } z \in [0, L_{n+1}], \\ \mathcal{T}_{n+1} &:= \mathcal{T}_n \cup [[x_n, \phi_n(L_{n+1})]]_\ell, \\ p_{n+1} &:= p_n - a_n\delta_{x_n} + a_n\phi_n(q_{n+1}), \end{aligned} \tag{2.3}$$

where $\phi_n(q_{n+1})$ denotes the pushforward of the measure.

If p_{n+1} is purely diffuse or $n+1 = N$ then the construction terminates with $(\mathcal{T}, p) = (\mathcal{T}_{n+1}, p_{n+1})$. If this never arises then the recursive step repeats ad infinitum and we proceed to the following step.

Take the limit. Let $\mathcal{T} := \mathrm{cl}(\bigcup_{n\geq 1} \mathcal{T}_n)$. For every $k > n \geq 1$ we have $p_n = \pi_n(p_k)$, where π_n is as in Definition 2.1. Thus, by the Daniell-Kolmogorov extension theorem, there exists a measure p on $[0,1]^{\mathbb{N}}$ such that $\pi_n(p) = p_n$ for every $n \geq 1$. Moreover, since $(\mathcal{T}, \ell_1)$ is complete, p is supported on $\mathcal{T}$.

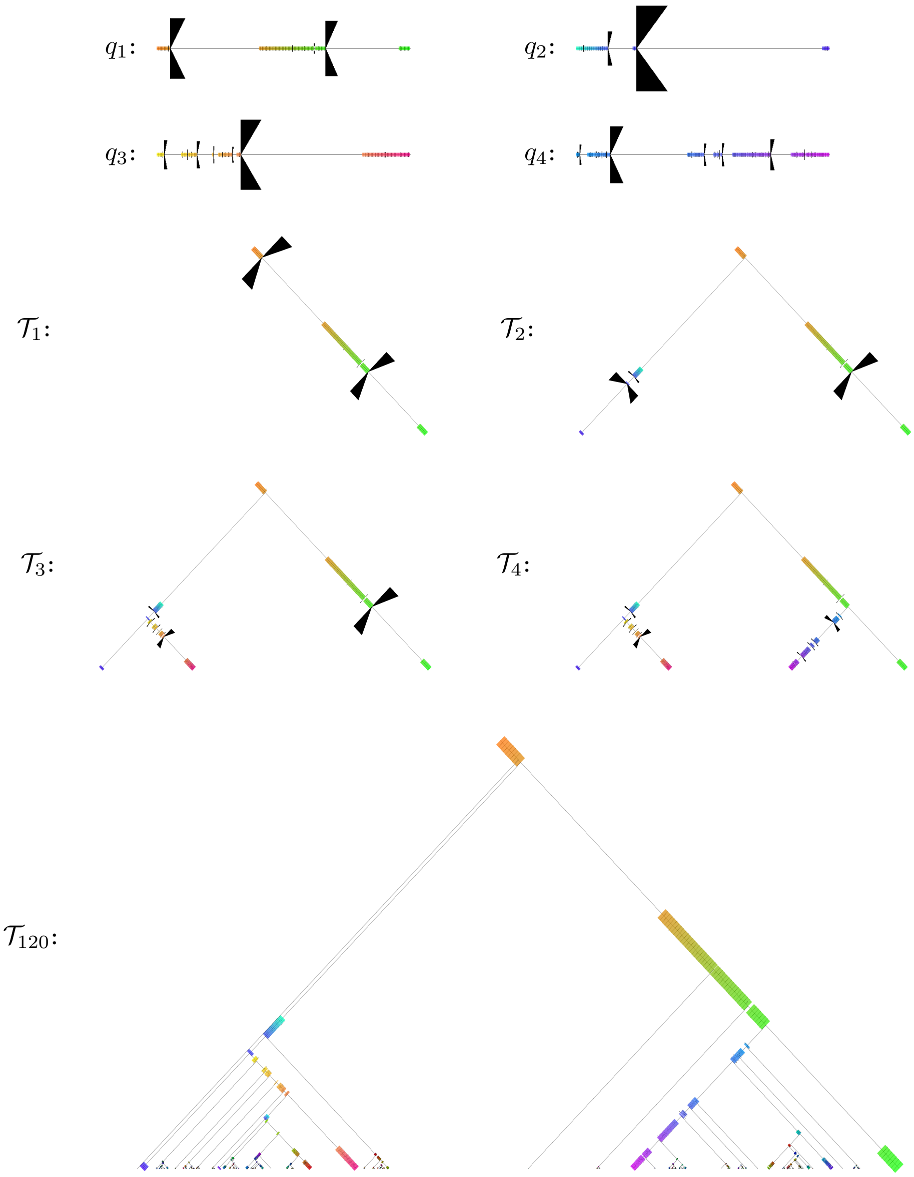

Figure 4: A color illustration of the bead crushing construction described in Section 2.1. In each tree image, the root is at the top and leaves are along a line at the bottom. Heavy, shaded lines mark subsets of the skeleton on which q_j or p_j equals length measure (in image of $\mathcal{T}_{120}$, we make these thinner to avoid branches appearing to overlap). Black wedge shapes, many of which are barely visible, represent atoms, or "beads," of q_j or p_j.

Remark 2.7. This construction can be reframed so that all of its inputs are given in advance, rather than some during each recursive step. In particular, we can adapt [14, equation (31)] so that each x_n is a function of p_n and a number $u_n \in (0,1)$, which need not depend on $(\mathcal{T}_n, p_n)$. Likewise, we could define $a_n := m_n b_n$, where $b_n \in (0,1]$ does not depend on $(\mathcal{T}_n, p_n)$.

Remark 2.8. In the infinite recursive construction, the measures p_n can be shown to converge to p in the first Wasserstein metric (defined, e.g., in [31]), though we will not use this.

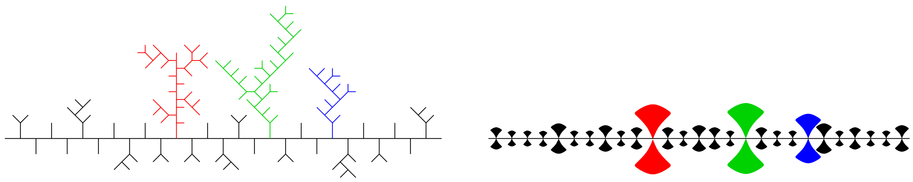

Figure 5: A string of beads in a discrete tree.

Proposition 2.9. *For any finite or infinite sequence, $(q_n,\ n \in [1,N])$ or $(q_n,\ n \geq 1)$, of uniformized probability measures, and any choices of (x_n, a_n), $n \geq 0$, in the course of this construction, the resulting quadruple $(\mathcal{T}, \ell_1, 0, p)$ is an IP tree.*

We prove this in Section 2.3, after giving some examples of the construction.

Theorem 2.10. *Every IP tree can be isomorphically embedded in ℓ_1 by the above bead-crushing construction.*

We prove this in Section 5. It is easily seen that any separable $\mathbb{R}$-tree can be isometrically embedded in ℓ_1, by constructing in the n^{th} step a tree isometric to one spanned by the first n points in a countable, dense sequence in the tree. The main assertion of Theorem 2.10 is rather that this can always be accomplished via the construction described above.

2.2 Example IP trees, strings of beads, the Brownian IP tree

Definition 2.11. *The* simple bead-crushing construction of IP trees *is a randomization of the construction in Section 2.1 in which: (i) the measures $(q_n,\ n \geq 1)$ are i.i.d. picks from some law on uniformized probability measures, with not all $q_n = \delta_0$; (ii) at each step, $m_n\delta_{x_n}$ is a size-biased pick from among the atoms of p_n; and (iii) at each step, $a_n = m_n$.*

This variant of the construction always yields a random IP tree with only binary branch points and a purely diffuse weight measure. If we carry out this construction with each q_n (deterministically) equal to the Fat Cantor measure of Example 2.5, then we get a binary branching IP tree with length measure interspersed among the branch points in such a way that the support of the measure does not include any non-trivial segments, as depicted in Figure 3(a).

Let $(\mathcal{T}, d, r, p)$ be a rooted, weighted $\mathbb{R}$-tree, and fix $x \in \mathcal{T}$. Consider the decomposition of $\mathcal{T}$ into the path $[[r, x]]$, called a *spine*, and the collection of subtrees, called *bushes*, branching out from the branch points along the spine, with perhaps a final bush rooted at x, if x is not a leaf. The bushes are totally ordered by increasing distance from the root. We may project p down onto the spine, replacing the mass distribution over each bush with an atom at the root of the bush. The resulting measure is called a *string of beads*, with the spine being the string and the atoms of the projection of p comprising the beads; see Figure 5. Spinal decompositions have a long history dating at least to 1981 [20]. Strings of beads were introduced in [29].

Example 2.12. The *two-parameter Poisson-Dirichlet distributions* [30], denoted by $\texttt{PoiDir}(\alpha, \theta)$ with $\alpha \in [0, 1)$ and $\theta > -\alpha$, are probability distributions on the Kingman simplex: the set of non-increasing infinite sequences of real numbers that sum to 1. These distributions arise in many settings and applications. Fix $\alpha \in (0, 1)$. Let $(U_i,\ i \geq 1)$ be i.i.d. $\texttt{Uniform}[0, 1]$, and let $(P_i,\ i \geq 1)$ be independent of this sequence with $\texttt{PoiDir}(\alpha, \alpha)$ distribution. We define

$$L := \lim_{n \to \infty} n(P_n)^\alpha \Gamma(1 - \alpha) \quad \text{and} \quad \mu := \sum_{i \geq 1} P_i \delta_{U_i L}. \tag{2.4}$$

The quantity L, called the *α-diversity* or sometimes the *local time*, is known to be a.s. positive and finite, with a known probability distribution; see [28, eqn. 83] or [29, eqn. 6]. The measure μ is an *(α,α)-string of beads* [29].

In Section 6.1, we describe the bead crushing construction of [29], which differs from that in Section 2.1. In particular, plugging i.i.d. $\left(\frac{1}{2},\frac{1}{2}\right)$-strings of beads into the former construction yields a Brownian CRT.

Definition 2.13. *Fix $\alpha \in (0,1)$. The* (α,α)-IP tree *is the IP tree arising from the simple bead crushing construction of Definition 2.11, with each q_n, $n \geq 1$, being the uniformization of an (α,α)-string of beads. In the case $\alpha = \frac{1}{2}$, we call it a* Brownian IP tree. *See Figure 3.*

This construction can be carried out with the full two-parameter family of (α,θ)-strings, with $\theta \geq 0$, introduced in [29]. We discuss the connection between the Brownian CRT and the Brownian IP tree in Section 6.1.

2.3 Proof of Propostion 2.9

We begin with a lemma.

Lemma 2.14. *For any sequence $(q_n,\ n \geq 1)$ of uniformized probability measures, and any choices of (x_n, a_n), $n \geq 0$, in the course of the bead-crushing construction of Section 2.1, the resulting trees $(\mathcal{T}_n, \ell_1, 0, p_n)$, $n \geq 0$, are IP trees.*

Proof. It is easily seen that these trees possess the Spanning property, so we need only check the Spacing property stated in equation (1.3). This holds by construction for $n = 0$. Assume for induction that it holds from some $n \geq 0$. The reader may check that for $y \in \mathcal{T}_n$, we get

$$p_{n+1}(F_{\mathcal{T}_{n+1}}(y)) = p_n(F_{\mathcal{T}_n}(y)), \tag{2.5}$$

regardless of the position of y relative to the point x_n of insertion of the new branch. Thus, $(\mathcal{T}_{n+1}, \ell_1, 0, p_{n+1})$ satisfies (1.3) at all branch points of $\mathcal{T}_n$ and all points in the closed support of p_n. It remains to check (1.3) at points $y \in \mathcal{T}_{n+1} \setminus \mathcal{T}_n$ in the closed support of p_{n+1}, including x_n. By definition of p_{n+1}, each such y equals $\phi_n(z)$ for some z in the closed support of q_{n+1}. Thus,

$$\begin{aligned} p_{n+1}(F_{\mathcal{T}_{n+1}}(y)) &= p_{n+1}\big(F_{\mathcal{T}_{n+1}}(\phi_n(z))\big) = a_n q_{n+1}([z, L_{n+1}]) = a_n(1-z) \\ &= \big(1 - \|x_n\| - p(F_{\mathcal{T}}(x_n))\big) - a_n(z-1) = 1 - \|\phi_n(z)\|, \end{aligned}$$

where the second equality results from the definition of p_{n+1}, the third from the uniformized property of q_{n+1} at z, the fourth from the Spacing property of p_n at x_n, and the last from the definition of ϕ_n. We conclude that $(\mathcal{T}_{n+1}, \ell_1, 0, p_{n+1})$ possesses the Spacing property, as needed for our induction. □

The lemma immediately proves Proposition 2.9 in the case that the construction terminates after finitely many steps. We now resolve the infinitely recursive case, showing that the limiting tree $(\mathcal{T}, \ell_1, 0, p)$ is an IP tree.

Spacing. By construction and the definition of p via projective consistency,

$$p(F_{\mathcal{T}}(y)) = p_n(\pi_n(F_{\mathcal{T}}(y))) = p_n(F_{\mathcal{T}_n}(y)) \qquad \text{for } n \geq 0,\ y \in \mathcal{T}_n. \tag{2.6}$$

Consider y in the closed support of p. We will abbreviate $y_n := \pi_n(y)$. For each $n \geq 1$, y_n lies in the closed support of p_n. Therefore,

$$\begin{aligned} p\left(F_{\mathcal{T}}(y)\right) &= p\left(\bigcap\nolimits_{n\geq 1} F_{\mathcal{T}}(y_n)\right) = \lim_{n\to\infty} p(F_{\mathcal{T}}(y_n)) \\ &= \lim_{n\to\infty} p_n(F_{\mathcal{T}_n}(y_n)) = \lim_{n\to\infty} 1 - \|y_n\| = 1 - \|y\|, \end{aligned} \tag{2.7}$$

where the first and last equalities follow from the convergence $y_n \to y$ along the segment $[[0,y]]_\ell$, the second follows from the countable additivity of p and the nesting $F_\mathcal{T}(y_n) \supseteq F_\mathcal{T}(y_N)$ for $n \leq N$, the third from (2.6), and the fourth from the Spacing property of the trees $(\mathcal{T}_n, p_n)$ proved in Lemma 2.14.

Spanning. Let y be a leaf of $\mathcal{T}$. As before, let $y_n := \pi_n(y)$. Then for all $n \geq 1$, either y_n is a leaf in $\mathcal{T}_n$ or it lies on an atom of p_n, which then arises as an attachment point for a new branch later in the construction. By the Spanning property of $(\mathcal{T}_n, p_n)$, y_n is in the closed support of p_n regardless. This condition is sufficient to apply the argument (2.7). In particular, if $\|y\| < 1$ then $p\{y\} = p(F_\mathcal{T}(y)) = 1 - \|y\| > 0$, so y is in the closed support of p.

Now, suppose $\|y\| = 1$ and fix $\epsilon > 0$. We will show the ϵ-ball about y has positive p-measure. Take N sufficiently large so that $\|y_N\| > 1 - \epsilon/4$ and let z denote the point on $[[0,y_N]]_\ell$ at distance $\epsilon/4$ from y_N. Since $\|z\| > 1 - \epsilon/2$ and no point in $\mathcal{T}$ lies farther than one unit from the origin,

$$\|x - y\| \leq \|x - z\| + \|y_N - z\| + \|y - y_N\| < \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} \quad \text{for } x \in F_\mathcal{T}(z).$$

Moreover, by the Spanning property of $\mathcal{T}_N$, $p(F_\mathcal{T}(z)) = p_N(F_{\mathcal{T}_N}(z)) > 0$. In other words, the ϵ-ball about y has positive measure under p. $\square$

2.4 Metrization and measurability of spaces of IP trees

IP trees $(\mathcal{T}, \ell_1, 0, p)$ constructed by bead crushing have the property

$$\mathcal{T} = \bigcup_{\text{leaves } x \in \mathcal{T}} [[0,x]]_\mathcal{T} = \bigcup_{x \in \text{support}(p)} [[0,x]]_\ell. \tag{2.8}$$

In particular, $\mathcal{T}$ is a function of p. We metrize the set of IP trees $(\mathcal{T}, \ell_1, 0, p)$ that satisfy (2.8) via the Prokhorov metric on their weight measures:

$$d_P(p,q) = \inf\left\{\epsilon > 0 \colon \forall A \in \mathcal{B},\ p(A^\epsilon) + \epsilon \geq q(A) \text{ and } q(A^\epsilon) + \epsilon \geq p(A)\right\}, \tag{2.9}$$

where $\mathcal{B}$ is the Borel σ-algebra on ℓ_1 and $A^\epsilon = \{x \colon \exists y \in A \text{ s.t. } \|x - y\| < \epsilon\}$. We equip this space of IP trees with the resulting Borel σ-algebra.

Likewise we metrize the space of isomorphism classes of IP trees with the pointed Gromov-Prokhorov metric of [18], which is the infimum of Prokhorov distances between the two weight measures over all isometric embeddings of the two trees into a common space with a common root. Again, we equip this space with the resulting Borel σ-algebra.

We must confirm that the pointed Gromov-Prokhorov metric is positive-definite between isometry classes of IP trees; it finds zero distance between trees if and only if there is a measure- and root-preserving isometry between the supports of the measures unioned with their roots. Indeed, in the setting of $\mathbb{R}$-trees with the Spanning property, if such a map between the roots and supports exists then it can be extended to a measure- and root-preserving isometry between the entire trees. This observation relates to the representation of $\mathbb{R}$-trees as ultrametric spaces and [19, Theorem 4]. We sketch its proof.

Proof sketch. Suppose $(\mathcal{T}_i, d_i, r_i, p_i)$, $i = 1,2$, are rooted, weighted $\mathbb{R}$-trees with the Spanning property with $\phi \colon \{r_1\} \cup \text{support}(p_1) \to \{r_2\} \cup \text{support}(p_2)$ a weight-preserving isometry. Let $(t_i^1, i \geq 1)$ denote i.i.d. samples from p_1 and $t_i^2 := \phi(t_i^1)$, $i \geq 1$, so that the latter are i.i.d. samples from p_2. The distance from the root to the branch point separating two points in an $\mathbb{R}$-tree is a function of the pairwise distances between the root and the two points: $d(r,(x \wedge y)) = (d(r,x) + d(r,y) - d(x,y))/2$. Thus, since $\{r_1\} \cup \{t_j^1,\, 1 \leq j \leq n\}$

is isometric to $\{r_2\} \cup \{t_j^2,\, 1 \le j \le n\}$ for any n, it follows that the subtrees spanned by these points are also isometric. Let ϕ_n denote this isometry. This gives rise to a sequence of consistent isometries between two sequences of growing subtrees. Since every point in the skeletons of $\mathcal{T}_1$ and $\mathcal{T}_2$ eventually falls into these spanned subtrees, the skeletons are isometric. And since the skeletons are dense and both trees are complete, the trees are isometric. □

Under this σ-algebra, the following map is measurable: the map from an (isomorphism class of) IP tree(s) to the law of the infinite matrix of pairwise distances between the root and a sequence of i.i.d. samples from the weight measure [18, Definition 2.8]. The hierarchy generated by n samples can be obtained as a function of that matrix of distances, so that the resulting law on hierarchies is also a measurable function of the IP tree.

3 IP tree representation of an exchangeable hierarchy

The proofs of our main results build upon the study of CRT representations of exchangeable hiearchies in [14]. We recall the relevant background and connect it to IP trees in this section.

Definition 3.1. *A hierarchy $\mathcal{H}$ on a finite set S is partially ordered under the subset relation. For $x, y \in S$, their* most recent common ancestor (MRCA) *is*

$$(x \wedge y) := \bigcap_{G \in \mathcal{H}:\ x,y \in G} G. \tag{3.1}$$

If $(\mathcal{H}_n,\, n \ge 1)$ is hierarchy on $\mathbb{N}$, then we define the MRCA of i and j in this hierarchy to be

$$(i \wedge j) := \bigcup_{n \ge \max\{i,j\}} (i \wedge j)_n, \tag{3.2}$$

where $(i \wedge j)_n$ denotes the MRCA of i and j in $\mathcal{H}_n$.

MRCAs in hierarchies on $\mathbb{N}$ are projectively consistent [14, Proposition 1]:

$$(i \wedge j)_n = (i \wedge j)_N \cap [n] = (i \wedge j) \cap [n] \qquad \text{for } i, j \le n \le N. \tag{3.3}$$

When constructing a $\mathbb{R}$-tree representation of a hierarchy, we find it convenient to work with a hierarchy on $\mathbb{Z}$. Our strategy is to construct the tree by a line-breaking procedure, with the endpoints of successively added branch corresponding to -1, -2, …, in the hierarchy. Then, the positive indices correspond to samples t_1, t_2, …, in the tree, which, by exchangeability, specify a driving measure on the tree.

Let $(\mathcal{H}'_n, n \ge 1)$ be an exchangeable hierarchy on $\mathbb{N}$ and let $b : \mathbb{N} \to \mathbb{Z}$ denote the bijection that sends odd numbers to sequential non-positive numbers and evens to sequential positive numbers. For $n \ge 1$ set

$$\mathcal{H}_n := \left\{ \{b(k) : k \in A\} : A \in \mathcal{H}'_{2n+1} \right\}. \tag{3.4}$$

Then $\mathcal{H}_n$ is a hierarchy on $[\pm n] := \{-n, \ldots, 0, \ldots, n\}$ and $\mathcal{H}_{n+1}\big|_{[\pm n]} = \mathcal{H}_n$ for every $n \ge 1$. Definition 3.1 extends to this context without modification.

Proposition 3.2 ([17] Theorem 11, [8] Theorem 5, [14] Proposition 2)**.** *Let $(\mathcal{H}_n,\, n \ge 1)$ be an exchangeable hierarchy on $\mathbb{Z}$.*

(i) For $i, j \in \mathbb{Z}$, the following limit exists almost surely:

$$X_j^i := 1 - \lim_{n \to \infty} \frac{\#((i \wedge j) \cap [\pm n])}{2n}. \tag{3.5}$$

(ii) For bijections $\sigma\colon \mathbb{Z}\to\mathbb{Z}$ with finitely many non-fixed points,

$$\left(X^i_j;\ i,j\in\mathbb{Z},\ i\neq j\right) \stackrel{d}{=} \left(X^{\sigma(i)}_{\sigma(j)};\ i,j\in\mathbb{Z},\ i\neq j\right). \tag{3.6}$$

In particular, for $i\in\mathbb{Z}$, the family $(X^i_j,\ j\in\mathbb{Z}\setminus\{i\})$ is exchangeable.

(iii) For $i,j,k\in\mathbb{N}$, the following events are almost surely equal:

$$\{X^i_j\le X^i_k\} = \{(i\wedge k)\subseteq(i\wedge j)\} = \{k\in(i\wedge j)\}. \tag{3.7}$$

Recall the notation of Definitions 2.1 and 2.2 for a standard basis $(\mathbf{e}_n,\ n\ge 1)$, projection maps $(\pi_n,\ n\ge 1)$, and segments $[[0,x]]_\ell$ in ℓ_1. We adopt the convention that for $k<0$, $[k]:=\{k,k+1,\ldots,-1\}$.

Definition 3.3. *For all $j\in\mathbb{Z}$, set $t^0_j=0$. Iteratively, for each $k\le 0$, we define*

$$\begin{aligned} t^{k-1}_j &:= t^k_j + \mathbf{e}_{|k-1|}\left(X^{k-1}_j - \left\|t^k_j\right\|\right)_+ \quad \text{for } j\in\mathbb{Z}\setminus[k-1],\\ \mathcal{T}_k &:= \mathrm{cl}\left(\bigcup\nolimits_{j\ge 1} [\![0,t^k_j]\!]_\ell\right), \end{aligned} \tag{3.8}$$

where $(a)_+ := \max\{a,0\}$. We treat 0 as the root of each of the trees.

It is easily checked that

$$\left\|t^k_j\right\| = \max_{m\in[k]} X^m_j \quad \text{and} \quad \pi_{|i|}\left(t^k_j\right) = t^i_j \qquad \text{for } k<i\le 0,\ j\notin[k].$$

Thus, for any j, the sequence $(\left\|t^k_j\right\|,\ k<0)$ is monotone increasing as k decreases. When this magnitude holds constant from k to $k-1$, so does t^k_j; and when the magnitude increases, it does so because $t^{k-1}_j - t^k_j$ is a vector in the $\mathbf{e}_{|k-1|}$ direction. In this latter case, we say that t^{k-1}_j is "pushed out" from t^k_j.

Proposition 3.4 (Lemma 1 and Propositions 4, 5, and 6 of [14])**.**

(i) Line-breaking property of $\mathcal{T}$. *For $k\le -1$ and $j\in\mathbb{Z}\setminus[k-1]$, if $t^{k-1}_j\neq t^k_j$ then $t^k_j = t^k_{k-1}$. Informally, all samples that are "pushed out" in passing from t^k_j to t^{k-1}_j come from the same spot on $\mathcal{T}_k$, namely t^k_{k-1}. Moreover, regardless of whether $t^{k-1}_j = t^k_j$,*

$$\left(X^{k-1}_j - \left\|t^k_j\right\|\right)_+ = \left(X^{k-1}_j - \left\|t^k_{k-1}\right\|\right)_+. \tag{3.9}$$

(ii) For each $j\ge 1$, $(t^k_j,\ k<0)$ converges a.s. in ℓ_1. Call the limit t_j.

(iii) The family $(t_j,\ j\ge 1)$ is exchangeable and has a driving measure p supported on a subset of

$$\mathcal{T} := \mathrm{cl}\left(\bigcup\nolimits_{k<0}\mathcal{T}_k\right).$$

Likewise, for every $k<0$, the family $(t^k_j, j\ge 1)$ is exchangeable and has a driving measure p_k.

(iv) For distinct $u,v\in\mathbb{N}$,

$$(u\wedge v)_{\mathcal{H}}\cap\mathbb{N} = \{j\in\mathbb{N}\colon t_j\in F_{\mathcal{T}}((t_u\wedge t_v)_\ell)\}. \tag{3.10}$$

Theorem 3.5 (Theorem 5 and its proof in [14])**.** *Let $(\mathcal{H}'_n,\ n\ge 1)$ denote a hierarchy on $\mathbb{N}$, and let $(\mathcal{T},p)$ be as in Proposition 3.4 above, following from the construction in Definition 3.3. Then the random law $\Theta(\mathcal{T},\ell_1,0,p)$ is a r.c.d. for $(\mathcal{H}'_n,\ n\ge 1)$ on $\mathrm{tail}(\mathcal{H}'_n)$.*

To this description, we add the following.

Proposition 3.6. *In the setting of Theorem 3.5, $(\mathcal{T},\ell_1,0,p)$ is a random IP tree arising from a bead crushing construction as in Section 2.1, with the caveat that at some steps k, $(\mathcal{T}_{k-1},p_{k-1})=(\mathcal{T}_k,p_k)$.*

The formulation of bead crushing in Section 2.1 does not allow this possibility of the tree going unchanged in one of the steps. We refer to this variant of bead crushing as *bead crushing with pauses*. Of course, trees arising from the construction with pauses are still IP trees.

Proof of Proposition 3.6. We restate (2.3) from the recursive step in the bead-crushing construction, for use in the present setting:

$$\begin{aligned} \phi_k(z) &:= t^k_{k-1} + \big(p_k\big(F_{\mathcal{T}_k}\big(t^k_{k-1}\big)\big) - a_k\big)\mathbf{e}_{|k-1|} + z a_k \mathbf{e}_{|k-1|}, \quad z\in[0,L_{k-1}],\\ \mathcal{T}_{k-1} &:= \mathcal{T}_k \cup \big[\!\big[t^k_{k-1},\phi_k(L_{k-1})\big]\!\big]_\ell,\\ p_{k-1} &:= p_k - a_k\delta_{t^k_{k-1}} + a_k\phi_k(q_{k-1}), \end{aligned} \tag{3.11}$$

where $L_{k-1}=\max(\text{support}(q_{k-1}))$ and $a_k\in\big(0,p_k\big\{t^k_{k-1}\big\}\big]$. We will prove that, at each step in the iterative construction of $(\mathcal{T},p)$, if $(\mathcal{T}_{k-1},p_{k-1})\neq(\mathcal{T}_k,p_k)$ then there exists a uniformized law q_{k-1} and a mass $a_k\in(0,m_k]$, where $m_k:=p_k\{t^k_{k-1}\}>0$, such that $(\mathcal{T}_{k-1},p_{k-1})$ is obtained from $(\mathcal{T}_k,p_k)$ as in (3.11).

Base step: $k=0$. By definition, $t^0_{-1}=0$ and $t^{-1}_j=X^{-1}_j\mathbf{e}_1$ for each $j\neq-1$. We set $a_0:=m_0=1$. Then, following (3.11), $\phi_{-1}(z)=z\mathbf{e}_1$ for $z\in[0,1]$. Let q_{-1} denote the driving measure of the sequence $(X^{-1}_j,\ j\geq1)$.

$$\begin{aligned} 1-X^{-1}_j &= \lim_{n\to\infty}\frac{\#((-1\wedge j)\cap[\pm n])}{2n}\\ &= \lim_{n\to\infty}\frac{\#\{i\in[\pm n]\colon X^{-1}_i\geq X^{-1}_j\}}{2n} = q_{-1}[X^{-1}_j,1], \end{aligned} \tag{3.12}$$

where the first equation follows from (3.5), the second from (3.7), and the last from the definition of q_{-1}. Since the X^{-1}_j are dense in the closed support of q_{-1}, we find that q_{-1} is uniformized, in the sense of Definition 2.3. Since $t^{-1}_j=\phi_{-1}(X^{-1}_j)$ and p_{-1} is the driving measure of the (t^{-1}_j), we conclude that $p_{-1}=\phi_{-1}(q_{-1})$, consistent with the last line of (3.11). Thus $(\mathcal{T}_1,\ell_1,0,p_1)$ is an IP tree arising from a single step of a bead crushing construction.

Inductive step. Fix $k<0$ and assume that $(\mathcal{T}_k,\ell_1,0,p_k)$ is an IP tree arising from $|k|$ steps of the bead crushing construction with pauses. Let

$$S:=\big\{j\in\mathbb{Z}\setminus[k-1]\colon((k-1)\wedge j)_{\mathcal{H}}\cap[k]=\emptyset\big\}.$$

Informally, S is the set of indices that remain in a block with $k-1$ in the hierarchy until after $k-1$ has branched away from all of the indices $k,k+1,\ldots,-1$. By (3.7) and the definition of the (t^i_j),

$$S=\big\{j\in\mathbb{Z}\setminus[k-1]\colon X^{k-1}_j>\max_{i\in[k]}X^i_j\big\}=\big\{j\in\mathbb{Z}\setminus[k-1]\colon t^{k-1}_j\neq t^k_j\big\}.$$

Thus, $S=\emptyset$ if and only if $(\mathcal{T}_{k-1},p_{k-1})=(\mathcal{T}_k,p_k)$, in which case we have nothing to prove. So assume $S\neq\emptyset$.

The family $(\mathbf{1}\{j\in S\},\ j\in\mathbb{Z}\setminus[k-1])$ is exchangeable, and $S\neq\emptyset$ means that not all entries are zero, so by de Finetti's theorem,

$$a_k:=\lim_{n\to\infty}\frac{\#(S\cap[\pm n])}{2n}>0.$$

By Proposition 3.4(i), every index $j \in S$ satisfies $t_j^k = t_{k-1}^k$. Thus, a_k is bounded above by $m_k = p_k\{t_{k-1}^k\}$.

Now, for $j \in S$, let

$$Y_j := 1 - \frac{1 - X_j^{k-1}}{a_k} = 1 - \lim_{n\to\infty} \frac{\#\big(((k-1)\wedge j) \cap [\pm n]\big)}{\#(S \cap [\pm n])}. \tag{3.13}$$

Here, the rightmost formula follows by plugging in the definitions of a_k and X_j^{k-1} and canceling out factors of $2n$. Let f denote the unique increasing bijection from $\mathbb{N}$ to $S \cap \mathbb{N}$. The sequence $(Y_{f(j)},\, j \in \mathbb{N})$ is exchangeable; let q_{k-1} denote its driving measure. By an argument similar to that in (3.12), $Y_j = 1 - q_{k-1}[Y_j, 1]$ for each $j \in S$. Since the $(Y_j,\, j \in S)$ are dense in the closed support of q_{k-1}, we conclude that q_{k-1} is uniformized.

Now, consider the map ϕ_k as defined in (3.11). Note that

$$\big\|\phi_k(Y_j)\big\| = \big\|t_{k-1}^k\big\| + p_k(F_{\mathcal{T}}(t_{k-1}^k)) + a_k\big(Y_j - 1\big) = X_j^{k-1},$$

where the second equality follows by appealing to the Spacing property of $(\mathcal{T}_k, p_k)$ at t_{k-1}^k and plugging in the definition of Y_j. Thus, for $j \in S$, $\phi_k(Y_j)$ is a point embedded in the first $|k|+1$ coordinates in ℓ_1 whose projection onto the first $|k|$ coordinates is t_{k-1}^k, and with $|k|+1^{\text{st}}$ coordinate equal to $X_j^{k-1} - \big\|t_{k-1}^k\big\|$. We conclude that $\phi_k(Y_j) = t_j^{k-1}$. Since p_{k-1} is the driving measure for the sequence $(t_j^{k-1},\, j \geq 1)$, we find that it satisfies the third formula in (3.11). Therefore, $(\mathcal{T}_{k-1}, \ell_1, 0, p_{k-1})$ is an IP tree arising from $|k|+1$ steps of a bead-crushing construction with pauses, which completes our induction. $\square$

4 Two key propositions

To prove our theorems we require two more major intermediate steps. Let $(\mathcal{S}, d, r, q)$ be a rooted, weighted $\mathbb{R}$-tree, let $(s_i,\, i \in \mathbb{Z})$ denote i.i.d. samples from q, and let $(\mathcal{H}_n,\, n \geq 1)$ denote the hierarchy on $\mathbb{Z}$ derived from $(\mathcal{S}, d, r)$ via these samples. In other words, modulo our choice to label with $\mathbb{Z}$ rather than $\mathbb{N}$, $(\mathcal{H}_n)$ is exchangeable and independently generated (e.i.g.) with law $\Theta(\mathcal{S}, d, r, q)$. Let $(\mathcal{T}, \ell_1, 0, p)$ and $(t_j,\, j \geq 1)$ denote the random IP tree and samples that arise from applying the construction of Section 3 to $(\mathcal{H}_n)$.

Proposition 4.1. *For every rooted, weighted $\mathbb{R}$-tree, there is a deterministic bead-crushing construction, as in Section 2.1, that yields an IP tree that: (i) is mass-structurally equivalent to $(\mathcal{S}, d, r, q)$ and (ii) has the same image under Θ as $(\mathcal{S}, d, r, q)$. In particular, the law of the random IP tree $(\mathcal{T}, \ell_1, 0, p)$ is supported on the set of IP trees with these two properties.*

Proposition 4.2. *If two IP trees are mass-structurally equivalent then they are isomorphic.*

We prove these propositions in Sections 4.1 and 4.2, respectively. To do so, we require two lemmas. Extending the notation $(y \wedge z)_\ell$ of Definition 2.2, for $y, z \in \mathcal{S}$, let $(y \wedge z)_{\mathcal{S}}$ denote the unique point in the intersection $[[r, y]] \cap [[r, z]] \cap [[y, z]]$. This equals the branch point that separates y, z, and r, except in the degenerate circumstance that all three lie on a common segment, in which case $(y \wedge z)_{\mathcal{S}}$ equals whichever of y, z, or r lies between the other two.

Lemma 4.3. *It is a.s. the case that for every $j \in \mathbb{N}$ and $\epsilon > 0$, there is some $i \neq j$ for which $d((s_i \wedge s_j)_{\mathcal{S}}, s_j) < \epsilon$.*

Proof. Fix $j \in \mathbb{N}$ and $\epsilon > 0$. Recall from Definition 1.1 that we require $\mathbb{R}$-trees to be separable and thus second countable. Thus, there exists a countable collection $\mathcal{A}$ of open sets of diameter at most ϵ that cover $\mathcal{S}$. It is a.s. the case that for every $U \in \mathcal{A}$, if $q(U) = 0$ then $\{i\colon s_i \in U\} = \emptyset$. Consequently, the ϵ-ball about s_j a.s. has positive q-measure. Therefore there is a.s. some other sample s_i with $d(s_i, s_j) < \epsilon$. Finally, $d((s_i \wedge s_j)_{\mathcal{S}}, s_j) < d(s_i, s_j) < \epsilon$. $\square$

We define

$$\begin{aligned} I^{\mathcal{S}}(a) &:= \{j \in \mathbb{Z}\colon a \in [[r, s_j]]_{\mathcal{S}}\} \quad \text{for } a \in \mathcal{S} \\ \text{and} \quad I^{\mathcal{T}}(b) &:= \{j \in \mathbb{N}\colon b \in [[0, t_j]]_{\ell}\} \quad \text{for } b \in \mathcal{T}. \end{aligned} \tag{4.1}$$

Lemma 4.4. *For $i, j, u, v \in \mathbb{N}$ with $i \neq j$ and $u \neq v$, up to null events,*

$$I^{\mathcal{S}}((s_u \wedge s_v)_{\mathcal{S}}) \cap \mathbb{N} = (u \wedge v) \cap \mathbb{N} = I^{\mathcal{T}}((t_u \wedge t_v)_{\ell}), \tag{4.2}$$

$$I^{\mathcal{S}}(s_u) \cap \mathbb{N} = \mathbb{N} \cap \bigcap_{k \in \mathbb{Z} \setminus \{u\}} (u \wedge k) = I^{\mathcal{T}}(t_u), \tag{4.3}$$

$$\big\{(s_u \wedge s_v)_{\mathcal{S}} = (s_i \wedge s_j)_{\mathcal{S}}\big\} = \big\{(t_u \wedge t_v)_{\ell} = (t_i \wedge t_j)_{\ell}\big\}, \tag{4.4}$$

$$\text{and} \quad \{s_u = s_v\} = \{t_u = t_v\}. \tag{4.5}$$

Proof. (4.2): Note that for $u, v \in \mathbb{N}$ distinct and $n > u, v$,

$$\begin{aligned} (u \wedge v)_n = \bigcap_{A \in \mathcal{H}_n\colon u, v \in A} A &= \bigcap_{x \in \mathcal{S}\colon s_u, s_v \in F_{\mathcal{S}}(x)} \big([\pm n] \cap I^{\mathcal{S}}(x)\big) \\ &= [\pm n] \cap I^{\mathcal{S}}((s_u \wedge s_v)_{\mathcal{S}}), \end{aligned}$$

where the first equation is Definition 3.1 of the MRCA, the second follows from the definition of $\mathcal{H}_n$ via the samples (s_j), and the last follows because every fringe subtree containing both s_u and s_v must contain the branch point $(s_u \wedge s_v)_{\mathcal{S}}$. This proves the first equation in (4.2). The second has already been established in Proposition 3.4(iv).

(4.3): By Lemma 4.3,

$$F_{\mathcal{S}}(s_u) = \bigcap_{k \in \mathbb{Z} \setminus \{u\}} F_{\mathcal{S}}((s_u \wedge s_k)_{\mathcal{S}});$$

$$\text{thus,} \quad I^{\mathcal{S}}(s_u) \cap \mathbb{N} = \mathbb{N} \cap \bigcap_{k \in \mathbb{Z} \setminus \{u\}} I^{\mathcal{S}}((s_u \wedge s_k)_{\mathcal{S}}) = \mathbb{N} \cap \bigcap_{k \in \mathbb{Z} \setminus \{u\}} (u \wedge k)_{\mathcal{H}},$$

with the last equation following from (4.2). By Proposition 3.4(iii), the $(t_i,\ i \geq 1)$ have p as their driving measure, so the same argument via Lemma 4.3 applies to $I^{\mathcal{T}}(t_u)$, thus proving (4.3).

(4.4): Note that $(s_i \wedge s_j)_{\mathcal{S}} = (s_u \wedge s_v)_{\mathcal{S}}$ if and only if both $i, j \in I^{\mathcal{S}}((s_u \wedge s_v)_{\mathcal{S}})$ and $u, v \in I^{\mathcal{S}}((s_i \wedge s_j)_{\mathcal{S}})$. The corresponding claim holds for samples in $\mathcal{T}$. Thus, (4.4) follows from (4.2).

(4.5): Note that $s_u = s_v$ if and only if both $v \in I^{\mathcal{S}}(s_u)$ and $u \in I^{\mathcal{S}}(s_v)$. The corresponding claim holds for t_u and t_v. Thus, (4.5) follows from (4.3). $\square$

4.1 Proof of Proposition 4.1

We know from Theorem 3.5 that $\Theta(\mathcal{T}, \ell_1, 0, p) = \Theta(\mathcal{S}, d, r, q)$ a.s.. Thus, it suffices to show that these two trees are a.s. mass-structurally equivalent. First, we will define a function ϕ mapping the special points of $\mathcal{S}$, in the sense of Definition 1.2, to those of $\mathcal{T}$, and we will show that it is a bijection. Then we will show that ϕ is mass and structure preserving.

Recall that, by Proposition 3.6, $(\mathcal{T}, \ell_1, 0, p)$ is an IP tree. In particular, it possesses the Spanning property, $\mathcal{T} = \mathrm{span}(p)$.

Definition of a bijection, ϕ Recall from Definition 1.2 that there are three types of special points: locations of atoms, branch points of the subtree spanned by the measure, and isolated leaves of said subtree. Therefore, we define a bijection ϕ in these three cases.

(a) If y is the location of an atom of q then there is a.s. some i for which $s_i = y$. We define $\phi(y) := t_i$. By (4.5), it is a.s. the case that $t_j = t_i$ if and only if $s_j = s_i$, for $j \geq 1$, so this is well-defined. Moreover, by the law of large numbers and Proposition 3.4(iii),

$$q\{s_i\} = \lim_{n\to\infty} \frac{\#\{j \in [n]\colon s_j = s_i\}}{n} = \lim_{n\to\infty} \frac{\#\{j \in [n]\colon t_j = t_i\}}{n} = p\{t_i\}.$$

By the preceding argument, ϕ is injective from atoms of q to those of p. The same argument in reverse shows that it bijects these sets of atoms.

(b) If x is a branch point of $\mathrm{span}(q)$, in the sense of Definition 1.2, then there is a.s. some pair $i, j \in \mathbb{N}$ for which $x = (s_i \wedge s_j)_{\mathcal{S}}$ with $x \neq s_i$ and $x \neq s_j$. In particular, $s_i \notin [[r, s_j]]_{\mathcal{S}}$ and vice versa. By (4.3), $t_i \notin [[r, t_j]]_{\mathcal{S}}$ and vice versa, so $(t_i \wedge t_j)_\ell$ is a branch point of $\mathcal{T} = \mathrm{span}(p)$. And by (4.4), the vertex $(t_i \wedge t_j)_\ell$ is a.s. the same across all pairs i, j for which $x = (s_i \wedge s_j)_{\mathcal{S}}$. We define $\phi(x) := (t_i \wedge t_j)_\ell$. By this same argument in reverse, starting with a branch point of $\mathcal{T}$, we see that ϕ bijects the branch points of $\mathrm{span}(q)$ with those of $\mathcal{T}$.

In the special case that q has an atom located at the branch point x, this agrees with our previous definition of ϕ for atoms. In this case, there exist samples $s_u = s_v = x$ with $u \neq v$. Then $(s_i \wedge s_j)_{\mathcal{S}} = x = s_u = (s_u \wedge s_v)_{\mathcal{S}}$. By (4.4) this means $(s_i \wedge s_j)_{\mathcal{S}} = (t_u \wedge t_v)_\ell$, and by (4.5), $t_u = t_v$. Then we conclude $\phi(x) = (t_i \wedge t_j)_\ell = t_u$.

(c) Now suppose $z \in \mathcal{S}$ is an isolated leaf of $\mathrm{span}(q)$, in the sense that there is a non-trivial segment $[[x, z]]_{\mathcal{S}} \subseteq [[r, z]]_{\mathcal{S}}$ that contains no branch points of $\mathrm{span}(q)$ and every such segment has positive mass under q. Recall the map $I^{\mathcal{S}}$ defined in (4.1). Consider

$$J := \left\{ i \geq 1 \;\middle|\; \forall j \in I^{\mathcal{S}}(s_i),\ z \in F_{\mathcal{S}}(s_j) \right\}. \tag{4.6}$$

This is the set of indices of all samples that lie on a branch with the properties mentioned above. The samples $(s_i,\, i \in J)$ all lie along $[[r, z]]_{\mathcal{S}}$, and they are totally ordered, up to equality, along this segment. Since z is in the closed support of q, it is the unique limit point of this set at maximal distance from r. By (4.3), the samples $(t_i,\, i \in J)$ are correspondingly totally ordered along a segment. As $\mathcal{T}$ is bounded and complete under ℓ_1, these samples also have a unique limit point $z' \in \mathcal{T}$ at maximal distance from 0. We define $\phi(z) := z'$.

To show that this is a bijection between the sets of isolated leaves, we consider properties of the set J. This set a.s. satisfies:

(i) $\forall i \in J,\ \mathbb{N} \cap I^{\mathcal{S}}(s_i) \subseteq J$ and

(ii) $\forall i, j \in J,\ j \in I^{\mathcal{S}}(s_i)$ and/or $i \in I^{\mathcal{S}}(s_j)$.

Condition (i) asserts, roughly, that J comprises indices of all samples that fall into some fringe subtree $B \subseteq \mathcal{S}$. Condition (ii) asserts that these samples are totally ordered, up to equality, along a branch going away from r. I.e. the support of q on B is contained within a single segment aligned with r. By its definition, J is maximal with these two properties. If we view (4.6) as a map sending z to J, then this is a bijection from isolated leaves of $\mathrm{span}(q)$ to maximal sets of indices J that satisfy properties (i) and (ii) above. Likewise,

$$z' \mapsto \left\{ i \geq 1 \;\middle|\; \forall j \in I^{\mathcal{T}}(t_i),\ z' \in F_{\mathcal{T}}(t_j) \right\}.$$

is a bijection from isolated leaves of $\mathcal{T}$ to maximal sets J satisfying:

(i') $\forall i \in J,\ I^{\mathcal{T}}(t_i) \subseteq J$ and

(ii') $\forall i, j \in J,\ j \in I^{\mathcal{T}}(t_i)$ and/or $i \in I^{\mathcal{T}}(t_j)$.

Finally, by (4.3), conditions (i) and (ii) are equivalent to (i') and (ii'). Therefore, ϕ bijects the isolated leaves of span(q) with those of $\mathcal{T}$.

In the special case that q has an atom at z, this again agrees with our previous definition of ϕ for atoms. In this case, there exists some i with $s_i = z$. Since z is a leaf of span(q), $i \in J$ and s_i is the least upper bound of samples $(s_j,\ j \in J)$. By (4.3), t_i is then the least upper bound of samples $(t_j,\ j \in J)$. Thus, $\phi(z) = t_i$, as desired.

Mass preserving We have already established that $q\{x\} = p\{\phi(x)\}$ for all points $x \in \mathcal{S}$ at which q has atoms, and that ϕ bijects the locations of atoms of q with those of p.

Recall the maps $I^{\mathcal{S}}$ and $I^{\mathcal{T}}$ defined in (4.1). For $j \geq 1$, it is a.s. the case that

$$\begin{aligned} q([[r, s_j]]_{\mathcal{S}}) &= \lim_{n\to\infty} \frac{\#\{i \in [n] \colon s_i \in [[r, s_j]]_{\mathcal{S}}\}}{n} \\ &= \lim_{n\to\infty} \frac{\#\{i \in [n] \colon j \in \mathcal{I}^{\mathcal{S}}(i)\}}{n} = \lim_{n\to\infty} \frac{\#\{i \in [n] \colon j \in \mathcal{I}^{\mathcal{T}}(i)\}}{n} \\ &= \lim_{n\to\infty} \frac{\#\{i \in [n] \colon t_i \in [[r, t_j]]_{\ell}\}}{n} = p([[0, t_j]]_{\ell}), \end{aligned}$$

with the first and last equations a consequence of q and p being driving measures for the (s_i) and (t_i), respectively; the second and fourth following from the definition of fringe subtrees; and the third following from (4.3). An analogous derivation, making use of (4.2) in place of (4.3), shows that $q([[r, (s_i \wedge s_j)_{\mathcal{S}}]]_{\mathcal{S}}) = p([[0, (t_i \wedge t_j)_{\ell}]]_{\ell})$. This proves that $q([[r, x]]_{\mathcal{S}}) = p([[0, \phi(x)]]_{\ell})$ when x is the location of an atom of q or a branch point of span(q). Finally, the map $x \mapsto q([[r, x]]_{\mathcal{S}})$ is continuous at points x that are neither branch points nor locations of atoms of q, and correspondingly for p. Thus, by passing through a limit with samples converging to an isolated leaf, the result also holds when x is an isolated leaf of span(q).

If z is an isolated leaf of q at which there is no atom, then $q\left(F_{\mathcal{S}}(z)\right) = 0 = p\left(F_{\mathcal{T}}(\phi(z))\right)$. Finally, for y a branch point of span(q) or the location of an atom of q, we can write $y = (s_i \wedge s_j)_{\mathcal{S}}$ for some $1 \leq i < j$. Then, by (4.2),

$$\begin{aligned} q\left(F_{\mathcal{S}}((s_i \wedge s_j)_{\mathcal{S}})\right) &= \lim_{n\to\infty} n^{-1} \#\left(I^{\mathcal{S}}((s_i \wedge s_j)_{\mathcal{S}}) \cap [n]\right) \\ &= \lim_{n\to\infty} n^{-1} \#\left(I^{\mathcal{T}}((t_i \wedge t_j)_{\ell}) \cap [n]\right) = p\left(F_{\mathcal{T}}((t_i \wedge t_j)_{\ell})\right), \end{aligned}$$

as desired.

Structure preserving We must confirm that structure is preserved, in the sense of Definition 1.3(ii), between any two special points in $\mathcal{S}$. Again, we approach this case-by-case for the different types of special points.

For branch points y_1 and y_2 of span(q), we have $y_1 = (s_i \wedge s_j)_{\mathcal{S}}$ and $y_2 = (s_u \wedge s_v)_{\mathcal{S}}$ for some $i, j, u, v \in \mathbb{N}$. Then by (4.2) and the definition of $(a \wedge b)_{\mathcal{S}}$,

$$\begin{aligned} (s_i \wedge s_j)_{\mathcal{S}} \in [[r, (s_u \wedge s_v)_{\mathcal{S}}]]_{\mathcal{S}} \ &\Leftrightarrow\ u, v \in I^{\mathcal{S}}((s_i \wedge s_j)_{\mathcal{S}}) \\ &\Leftrightarrow\ u, v \in I^{\mathcal{T}}((t_i \wedge t_j)_{\ell}) \\ &\Leftrightarrow\ (t_i \wedge t_j)_{\ell} \in [[0, (t_u \wedge t_v)_{\ell}]]_{\ell}. \end{aligned}$$

The same argument shows that ϕ preserves structure between two locations of atoms x_1, x_2, or between a branch point and an atom, by taking $s_i = s_j = x_1$ for some pair $i \neq j$ so that $(s_i \wedge s_j)_{\mathcal{S}} = x_1$, and correspondingly for x_2.

If z_1 and z_2 are both isolated leaves of span(q) then $z_1 \notin [[r,z_2]]_{\mathcal{S}}$ and $z_2 \notin [[r,z_1]]_{\mathcal{S}}$, since both are leaves of the same tree, and likewise for $\phi(z_1)$ and $\phi(z_2)$. Thus, structure is preserved here as well.

Finally, suppose that z is an isolated leaf of span(q) with $q\{z\} = 0$ and x is either the location of an atom of q or a branch point of span(q). In either case, $x = (s_i \wedge s_j)_{\mathcal{S}}$ for some distinct $i,j \in \mathbb{N}$. We cannot have $z \in [[r,x]]_{\mathcal{S}}$, nor can we have $\phi(z) \in [[0,\phi(x)]]_\ell$, since z and $\phi(z)$ are leaves and do not equal x or $\phi(x)$, respectively. Let J be as in (4.6). Then

$$\begin{aligned}(s_i \wedge s_j)_{\mathcal{S}} \in [[r,z]]_{\mathcal{S}} \;&\Leftrightarrow\; I^{\mathcal{S}}((s_i \wedge s_j)_{\mathcal{S}}) \cap J \neq \emptyset \\ &\Leftrightarrow\; I^{\mathcal{T}}((t_i \wedge t_j)_\ell) \cap J \neq \emptyset \;\Leftrightarrow\; (t_i \wedge t_j)_\ell \in [[0,\phi(z)]]_\ell.\end{aligned}$$

Thus, ϕ preserves structure between isolated leaves of span(q) and other special points. □

4.2 Proof of Proposition 4.2

Let $(\mathcal{T}_i, d_i, r_i, p_i)$ for $i = 1,2$ be a pair of IP trees, with special point sets $\mathscr{S}_1$ and $\mathscr{S}_2$ and $\phi\colon \mathscr{S}_1 \to \mathscr{S}_2$ a mass-structural isomorphism. We begin with a pair of observations.

First, the roots r_1 and r_2 need not be special points. However, for $x \in \mathscr{S}_1$,

$$d_2(r_2,\phi(x)) = 1 - p_2\left(F_{\mathcal{T}_2}(\phi(x))\right) = 1 - p_1\left(F_{\mathcal{T}_1}(x)\right) = d_1(r_1,x), \tag{4.7}$$

by the Spacing properties of the two IP trees and the mass preserving property of ϕ. Taking $x = r_1$ or $\phi(x) = r_2$ shows that r_1 is a special point if and only if r_2 is, in which case $\phi(r_1) = \phi(r_2)$. If they are not special points, then we define $\phi(r_1) := r_2$.

Second, since $\mathscr{S}_1$ and $\mathscr{S}_2$ contain all branch points of the two trees, it follows from the structure preserving property that $\phi((x \wedge y)_{\mathcal{T}_1}) = (\phi(x),\phi(y))_{\mathcal{T}_2}$ for every $x,y \in \mathscr{S}_1$. Thus,

$$\begin{aligned}d_1(x,y) &= d_1\big(x,(x\wedge y)_{\mathcal{T}_1}\big) + d_1\big((x\wedge y)_{\mathcal{T}_1},y\big)\\ &= 2p\big(F_{\mathcal{T}_1}((x\wedge y)_{\mathcal{T}_1})\big) - p\big(F_{\mathcal{T}_1}(x)\big) - p\big(F_{\mathcal{T}_1}(y)\big)\\ &= 2p\big(F_{\mathcal{T}_2}(\phi((x\wedge y)_{\mathcal{T}_1}))\big) - p\big(F_{\mathcal{T}_2}(\phi(x))\big) - p\big(F_{\mathcal{T}_2}(\phi(y))\big)\\ &= d_2\big(\phi(x),\phi((x\wedge y)_{\mathcal{T}_1})\big) + d_2\big(\phi((x\wedge y)_{\mathcal{T}_1}),\phi(y)\big) = d_2(\phi(x),\phi(y)),\end{aligned}$$

where the second and fourth lines follow from the Spacing properties of $\mathcal{T}_1$ and $\mathcal{T}_2$ and the third is an application of the mass preserving property of ϕ. In other words, ϕ is an isometry from $(\mathscr{S}_1 \cup \{r_1\}, d_1)$ to $(\mathscr{S}_2 \cup \{r_2\}, d_2)$.

We must show that the IP trees $(\mathcal{T}_i, d_i, r_i, p_i)$ for $i = 1,2$ are isomorphic. First, we will define a map $\psi\colon \mathcal{T}_1 \to \mathcal{T}_2$ that preserves distance from the root; then, we show that ψ is an isometry; and finally we prove that ψ is measure-preserving.

Definition of ψ We extend ϕ to define $\psi\colon \mathcal{T}_1 \to \mathcal{T}_2$ by two mechanisms, which we call overshooting and approximation. Consider $z \in \mathcal{T}_1 \setminus \mathscr{S}_1$.

Case 1 (overshooting): $F_{\mathcal{T}_1}(z) \cap \mathscr{S}_1 \neq \emptyset$. Consider $x \in F_{\mathcal{T}_1}(z) \cap \mathscr{S}_1$. Define $\psi(z)$ to be the point along $[[r_2,\phi(x)]]_{\mathcal{T}_2}$ at distance $d_1(r_1,z)$ from r_2. This definition does not depend on our choice of x: if $x_1,x_2 \in F_{\mathcal{T}_1}(z) \cap \mathscr{S}_1$ then $x^* := (x_1 \wedge x_2)_{\mathcal{T}} \in F_{\mathcal{T}_1}(z) \cap \mathscr{S}_1$ as well. In that case, $d_1(r_1,z) < d_1(r_1,x^*) = d_2(r_2,\phi(x^*))$, and by the structure preserving property of ϕ,

$$[[r_2,\phi(x^*)]]_{\mathcal{T}_2} = [[r_2,\phi(x_1)]]_{\mathcal{T}_2} \cap [[r_2,\phi(x_2)]]_{\mathcal{T}_2}.$$

Thus, the points along $[[r_2,\phi(x_i)]]_{\mathcal{T}_2}$ at distance $d_1(r_1,z)$ from r_2 are the same for $i = 1,2$, as both lie in $[[r_2,\phi(x^*)]]_{\mathcal{T}_2}$.

Case 2 (approximation): $F_{\mathcal{T}_1}(z) \cap \mathscr{S}_1 = \emptyset$. Then there is no branch point, nor any isolated leaf of $\mathrm{span}(p_1) = \mathcal{T}_1$ beyond z. Thus, z must be a leaf with a sequence of branch points $(x_i,\, i \geq 1)$ converging to it along $[[r_1, z]]_{\mathcal{T}_1}$. Moreover, since z is a leaf and not the location of an atom, $d_1(r_1, z) = 1$ by the Spacing property. Since ϕ is an isometry, the sequence $(\phi(x_i),\, i \geq 1)$ is a Cauchy sequence in d_2, so it has a limit z' with $d_2(r_2, z') = 1$. We define $\phi(z) := z'$. Again, this is well-defined. If $(y_i,\, i \geq 1)$ is another sequence of branch points converging to z, then so is $x_1, y_1, x_2, y_2, \ldots$, so the ϕ-images of these sequences must have the same limit.

Note that ψ preserves distance from the root, by definition. Moreover, if z is defined by approximation then

$$d_2(\phi(x), \phi(z)) = d_1(x, z) \qquad \text{for branch points } x \in [[r_1, z]]_{\mathcal{T}_1}. \tag{4.8}$$

Isometry It follows from Lemma 4.3 and the definition above that ψ is a surjection. The definition also implies that ψ preserves distance from the root. Thus, to show that it is an isometry, it suffices to show that it preserves structure, in the sense that $\psi(x) \in [[r_2, \psi(y)]]_{\mathcal{T}_2}$ if and only if $x \in [[r_1, y]]_{\mathcal{T}_1}$. We consider two cases in which $x \in [[r_1, y]]_{\mathcal{T}_1}$ and one in which $x \notin [[r_1, y]]_{\mathcal{T}_1}$.

Case A.I: $x \in [[r_1, y]]_{\mathcal{T}_1}$ and $y \in [[r_1, z]]_{\mathcal{T}_1}$ for some $z \in \mathscr{S}_1$. Then both $\psi(x)$ and $\psi(y)$ lie on $[[r_2, \psi(z)]]_{\mathcal{T}_2}$, at respective distances $d_1(r_1, x)$ and $d_1(r_1, y)$ from r_2. Since $d_1(r_1, x) \leq d_1(r_1, y)$, we get $\psi(x) \in [[r_2, \psi(y)]]_{\mathcal{T}_2}$, as desired.

Case A.II: $x \in [[r_1, y]]_{\mathcal{T}_1}$ and $\psi(y)$ is defined by approximation. This means that we can take $z \in [[r_1, y]]$ to be a branch point with $d_1(z, y) < d_1(x, y)/2$. Then z must belong to $F_{\mathcal{T}_1}(x)$, so by the definition of $\psi(x)$ by overshooting, $\psi(x) \in [[r_2, \psi(z)]]_{\mathcal{T}_2}$. Moreover,

$$d_2(\psi(x), \psi(z)) = d_1(x, z) > d_1(z, y) = d_2(\psi(z), \psi(y)),$$

with the first equation following from preservation of distance from the root, the inequality from our assumption that $d_1(z, y) < d_1(x, y)/2$, and the final equation from (4.8). The entire closed ball of radius $d_2(\psi(x), \psi(z))$ about $\psi(z)$ lies inside $F_{\mathcal{T}_2}(\psi(x))$. In particular, $\psi(y) \in F_{\mathcal{T}_2}(\psi(x))$, as desired.

Case B: $x \notin [[r_1, y]]_{\mathcal{T}_1}$ and $y \notin [[r_1, x]]_{\mathcal{T}_1}$. We take up the case in which $\psi(x)$ is defined by overshooting and $\psi(y)$ by approximation; the other cases can be addressed similarly. Let z be a special point in $F_{\mathcal{T}_1}(x)$ and z' a branch point in $[[r_1, y]]_{\mathcal{T}_1}$ with $d_1(z', y) < d_1((x \wedge y)_{\mathcal{T}_1}, y)/2$. Then $(z \wedge z')_{\mathcal{T}_1} = (x \wedge y)_{\mathcal{T}_1}$. Moreover, by the structure-preserving property of ϕ,

$$\psi((x \wedge y)_{\mathcal{T}_1}) = \phi((z \wedge z')_{\mathcal{T}_1}) = (\phi(z) \wedge \phi(z'))_{\mathcal{T}_2} = (\psi(z) \wedge \psi(z'))_{\mathcal{T}_2}.$$

By definition,

$$d_2(\psi(x), \psi(z)) = d_1(x, z) < d_1(z, (x \wedge y)_{\mathcal{T}_1}) = d_2\big(\psi(z), \psi((x \wedge y)_{\mathcal{T}_1})\big).$$

Thus, $\psi(x)$ is in the component of $F_{\mathcal{T}_2}(\psi((x \wedge y)_{\mathcal{T}_1})) \setminus \{\psi((x \wedge y)_{\mathcal{T}_1})\}$ that contains $\psi(z)$. Correspondingly,

$$d_2(\psi(y), \psi(z')) = d_1(y, z') < d_1(z', (x \wedge y)_{\mathcal{T}_1}) = d_2(z', \psi((z \wedge z')_{\mathcal{T}_1})).$$

Thus, $\psi(y)$ is in the component of $F_{\mathcal{T}_2}(\psi((x \wedge y)_{\mathcal{T}_1})) \setminus \{\psi((x \wedge y)_{\mathcal{T}_1})\}$ that contains $\psi(z')$. We conclude that $\psi(x) \notin [[r_2, \psi(y)]]_{\mathcal{T}_2}$ and vice versa, as desired.

Measure-preserving The fringe subtrees of $\mathcal{T}_1$ comprise a π-system that generates the Borel σ algebra on $\mathcal{T}_1$, and likewise for $\mathcal{T}_2$. Because ψ is a root-preserving isometry, for $x \in \mathcal{T}_1$, $\psi\big(F_{\mathcal{T}_1}(x)\big) = F_{\mathcal{T}_2}(\psi(x))$. Thus, by a monotone class argument, it suffices to show that for every $x \in \mathcal{T}_1$, $p_1\left(F_{\mathcal{T}_1}(x)\right) = p_2\left(F_{\mathcal{T}_2}(\psi(x))\right)$. We argue this in four cases.

Case 1: $x = r_1$. Then $p_1\left(F_{\mathcal{T}_1}(x)\right) = 1 = p_2\left(F_{\mathcal{T}_2}(\psi(x))\right)$.

Case 2: $x \in \mathscr{S}_1$. Then $\psi(x) = \phi(x)$, and the desired equality is exactly the Mass preserving property of ϕ.

Case 3: x is not special but is the limit of a sequence of special points $(x_i,\, i \geq 1)$ in $[[r_1, x]]_{\mathcal{T}_1} \cup F_{\mathcal{T}_1}(x)$. Then x is neither a branch point nor the location of an atom, and likewise for $\psi(x)$, so

$$p_1\left(F_{\mathcal{T}_1}(x)\right) = \lim_{i\to\infty} p_1\left(F_{\mathcal{T}_1}(x_i)\right) = \lim_{i\to\infty} p_2\left(F_{\mathcal{T}_2}(\psi(x_i))\right) = p_2\left(F_{\mathcal{T}_2}(\psi(x))\right).$$

Case 4: x is not special and is not a limit of special points. Then x cannot be a leaf. Let y and z be the points closest to x in $(\mathrm{cl}(\mathscr{S}_1) \cup \{r_1\}) \cap [[r_1, x]]_{\mathcal{T}_1}$ and $\mathrm{cl}(\mathscr{S}_1) \cap F_{\mathcal{T}_1}(x)$, respectively. The map $w \mapsto p_1([[r_1, w]]_{\mathcal{T}_1})$ is continuous except at locations of atoms of p_1, and correspondingly for p_2. By the Mass preserving property of ϕ, the isometry property of ψ, and this continuity,

$$\begin{aligned} M := p_1\big([[y,z]]_{\mathcal{T}_1} \setminus \{z\}\big) &= p_1([[r_1,z]]_{\mathcal{T}_1}) - p_1\{z\} - p_1([[r_1,y]]_{\mathcal{T}_1}) \\ &= p_2\big([[r_2,\psi(z)]]_{\mathcal{T}_2}\big) - p_2\{\psi(z)\} - p_2\big([[r_2,\psi(y)]]_{\mathcal{T}_2}\big) \\ &= p_2\big([[\psi(y),\psi(z)]]_{\mathcal{T}_2} \setminus \{\psi(z)\}\big) \end{aligned}$$

By the Spacing property, $d_1(y,z) = p_1\left(F_{\mathcal{T}_1}(y)\right) - p_1\left(F_{\mathcal{T}_1}(z)\right) \geq M$. Let v be the point in $[[y,z]]_{\mathcal{T}_1}$ at distance M from z. Then the Spacing property of $\mathcal{T}_1$ implies that p_1 is null on $[[y,v]]_{\mathcal{T}_1} \setminus \{y\}$ and equals length measure on $[[v,z]]_{\mathcal{T}_1} \setminus \{z\}$. Correspondingly, the Spacing property of $\mathcal{T}_2$ implies that p_2 is null on $[[\psi(y),\psi(v)]]_{\mathcal{T}_2}$ and equals length measure on $[[\psi(v),\psi(z)]]_{\mathcal{T}_2} \setminus \{\psi(z)\}$. In particular,

$$\begin{aligned} p_1(F_{\mathcal{T}_1}(x)) &= p_1([[x,z]]_{\mathcal{T}_1} \setminus \{z\}) + p_1(F_{\mathcal{T}_1}(z)) \\ &= \min\{d_1(x,z), M\} + p_1\left(F_{\mathcal{T}_1}(z)\right) \\ &= \min\big\{d_2(\psi(x),\psi(z)), M\big\} + p_2\big(F_{\mathcal{T}_2}(\psi(z))\big) = p_2\big(F_{\mathcal{T}_2}(\psi(x))\big). \end{aligned}$$ ☐

5 Proofs of theorems

Proof of Theorem 1.5. Consider a rooted, weighted ℝ-tree $(\mathcal{T}, d, r, p)$. By Proposition 4.1, it is mass-structurally equivalent to at least one IP tree. By Proposition 4.2, all such IP trees are isomorphic to each other. ☐

Proof of Theorem 2.10. Consider an IP tree $(\mathcal{T}, d, r, p)$. By Proposition 4.1, there exists an IP tree arising from a deterministic bead crushing construction that is mass-structurally equivalent to $(\mathcal{T}, d, r, p)$. By Proposition 4.2, the two IP trees are thus isomorphic. ☐

Proof of Theorem 1.10. By Theorem 2.10, it suffices to prove this theorem for IP trees that arise from the bead-crushing construction of Section 2.1. Consider such an IP tree $(\mathcal{T}, \ell_1, 0, p)$ constructed from a sequence of uniformized probability measures q_n, $n \geq 1$, with atom locations x_n and masses a_n, $n \geq 1$, chosen during the construction. We need only show that the restriction of the non-atomic component of p to the skeleton of $\mathcal{T}$ equals the restriction of the length measure to a subset of the skeleton.

In the "Take the limit" step in Section 2.1 we note that the sequence of measures $(p_n,\, n \geq 1)$ arising in the construction is projectively consistent, $p_n = \pi_n(p_{n+1})$ for $n \geq 1$, and we define the limiting measure p via the Daniell-Kolmogorov extension theorem. The skeleton of the tree contains only points in ℓ_1 with finitely many positive coordinates. Thus, the diffuse component of p on the skeleton, p^s, is the sum over n of the diffuse measure on the n^{th} branch added in the construction. To get the diffuse measure on each new branch, the construction takes the diffuse component of q_n – call it q_n^d – scales down both its total mass and the length of the segment supporting it by some factor $a_n \in (0,1)$, i.e. $q_n^d \mapsto a_n q_n^d(\,\cdot\,/a_n)$, and it transposes the measure from the line segment to the branch in $\mathcal{T}$. Now, the theorem follows from Lemma 2.4, which states that q_n^d is the restriction of Lebesgue measure to a subset of $[0,1]$. $\square$

Proposition 5.1. *Two IP trees $(\mathcal{T}_i, d_i, r_i, p_i)$, $i = 1, 2$, are isomorphic if and only if*

$$\Theta(\mathcal{T}_1, d_1, r_1, p_1) = \Theta(\mathcal{T}_2, d_2, r_2, p_2).$$

Proof. We have already mentioned, and it is easily seen, that isomorphic trees have the same image under Θ. Now, suppose that the two trees have the same image under Θ. Let $(\mathcal{H}_n,\, n \geq 1)$ be an exchangeable random hierarchy with law $\Theta(\mathcal{T}_1, d_1, r_1, p_1)$. Let $(\mathcal{T}, \ell_1, 0, p)$ denote the random IP tree representation of $(\mathcal{H}_n)$ obtained from the construction in Section 3. By Proposition 4.1, all three IP trees are a.s. mass-structurally equivalent. Then, by Proposition 4.2 they are a.s. isomorphic. In particular, the two deterministic trees must be isomorphic. $\square$

Proof of Theorem 1.7. First, suppose that $(\mathcal{T}_i, d_i, r_i, p_i)$, $i = 1, 2$, are two rooted, weighted $\mathbb{R}$-trees with the same image as each other under Θ. Then by the same argument as in the proof of Proposition 5.1, they must be mass-structurally equivalent to each other.

Now, suppose instead that the two rooted, weighted $\mathbb{R}$-trees are mass-structurally equivalent. By Proposition 4.1, each tree $(\mathcal{T}_i, d_i, r_i, p_i)$ is then mass-structurally equivalent to some IP tree $(\mathcal{S}_i, \ell_1, 0, q_i)$ for which $\Theta(\mathcal{T}_i, d_i, r_i, p_i) = \Theta(\mathcal{S}_i, \ell_1, 0, q_i)$. By the transitivity of mass-structural equivalence, the two IP trees are mass-structurally equivalent. By Proposition 4.2, that means the IP trees are isomorphic, so

$$\Theta(\mathcal{T}_1, d_1, r_1, p_1) = \Theta(\mathcal{S}_1, \ell_1, 0, q_1) = \Theta(\mathcal{S}_2, \ell_1, 0, q_2) = \Theta(\mathcal{T}_2, d_2, r_2, p_2). \qquad \square$$

Proof of Theorem 1.8. (i) Let $(\mathcal{H}_n,\, n \geq 1)$ be an exchangeable random hierarchy. By Theorem 3.5 and Proposition 3.6, there exists a random IP tree $(\mathcal{T}, \ell_1, 0, p)$ with the property that $\Theta(\mathcal{T}, \ell_1, 0, p)$ is a r.c.d. for $(\mathcal{H}_n,\, n \geq 1)$ given its tail σ-algebra, $\mathrm{tail}(\mathcal{H}_n)$. Since $\Theta(\mathcal{T}, \ell_1, 0, p)$ is $\mathrm{tail}(\mathcal{H}_n)$-measurable, it follows from Proposition 5.1 that the random isomorphism class $\mathscr{T}$ of $(\mathcal{T}, \ell_1, 0, p)$ is as well. Then $\Theta(\mathscr{T}) = \Theta(\mathcal{T}, \ell_1, 0, p)$ is a r.c.d. for $(\mathcal{H}_n)$.

(ii) Proposition 5.1 states that the map Θ from isomorphism classes of IP trees to e.i.g. hierarchy laws is injective. By Theorem 3.5 and Proposition 3.6, every e.i.g. law is the Θ image of an IP tree, so it is also surjective. $\square$

6 Complements

6.1 Connections to the Brownian CRT

Recall Definition 2.13 of the Brownian IP tree. The following result justifies that terminology.

Proposition 6.1. *It is possible to construct a Brownian CRT (BCRT) and a Brownian IP tree on a common probability space coupled so that they are a.s. mass-structurally equivalent.*

Proof. Following [29, 32], we can construct a BCRT via a bead crushing construction similar to that in Section 2.1. In fact, we will construct a coupled BCRT and Brownian IP tree.

Let $(q_n,\, n \geq 1)$ denote an i.i.d. sequence of $\left(\frac{1}{2}, \frac{1}{2}\right)$-strings of beads, as described in Example 2.12. For each n, denote by L_n the maximum of the support of q_n; this is a.s. finite. As in Section 2.1, we define $\mathcal{T}_0 := \{0\}$ and $p_0 := \delta_0$ and proceed recursively to construct a tree embedded in ℓ_1.

Assume $(\mathcal{T}_n, 0, \ell_1, p_n)$ is a rooted, weighted $\mathbb{R}$-tree embedded in the first n coordinates in ℓ_1, with p_n a purely atomic measure. Let X_n be a sample from p_n, so $p_n(X_n) =: M_n > 0$. I.e. $M_n\delta_{X_n}$ is a size-biased random atom of p_n. Set

$$\begin{aligned} \phi_n(z) &:= X_n + z\sqrt{M_n}\mathbf{e}_{n+1} \quad \text{for } z \in [0, L_{n+1}], \\ \mathcal{T}_{n+1} &:= \mathcal{T}_n \cup \phi_n[0, L_{n+1}] = \mathcal{T}_n \cup [[X_n, \phi_n(L_{n+1})]]_\ell, \\ p_{n+1} &:= p_n + M_n\left(-\delta_{X_n} + \phi_n\left(q_{n+1}\right)\right), \end{aligned} \tag{6.1}$$

where $\phi_n(q_{n+1})$ denotes the pushforward of the measure. As in Section 2.1, $p_n = \pi_n(p_N)$ for $N > n$, so again, by the Daniell-Kolmogorov extension theorem, there exists a measure p on ℓ_1 with $\pi_n(p) = p_n$ for $n \geq 1$. Setting $\mathcal{T} := \mathrm{cl}(\bigcup_{n\geq 1}\mathcal{T}_n)$, the tree $(\mathcal{T}, \ell_1, 0, p)$ is a BCRT [29].

We now construct a Brownian IP tree coupled with this BCRT. For $n \geq 1$, let q'_n be the uniformization of q_n, as in Definition 2.3. There is a natural bijection from atoms of q_n to those of q'_n – in fact, this bijection is a mass-structural isomorphism from $([0, L_n], d, 0, q_n)$ to $([0,1], d, 0, q'_n)$. We plug the measures $(q'_n,\, n \geq 1)$ into the bead-crushing construction of Section 2.1 to recursively construct trees $(\mathcal{T}'_n, \ell_1, 0, p'_n)$. We see inductively that at each step, this resulting IP tree is mass-structurally equivalent to $(\mathcal{T}_n, \ell_1, 0, p_n)$ from the other construction, and so to proceed to the next step we can crush an atom $X'_n\delta_{M_n}$ of p'_n that corresponds to the atom $X_n\delta_{M_n}$ that was crushed in the other construction. In particular, this choice of X'_n is a sample from p'_n. The resulting limiting tree $(\mathcal{T}', \ell_1, 0, p')$ is a Brownian IP tree, as in Definition 2.13.

Both p and p' are diffuse measures supported on the leaves of $\mathcal{T}$ and $\mathcal{T}'$, respectively. It follows from our inductive argument that there is a mass- and structure-preserving bijection from branch points of $\mathcal{T}$ to those of $\mathcal{T}'$. Thus, the two trees are mass-structurally equivalent. □

The next result requires additional notation. Consider a rooted, weighted $\mathbb{R}$-tree $(\mathcal{T}, d, r, p)$, where p does not assign any continuous mass to the skeleton of $\mathcal{T}$. For $y \in \mathcal{T}$, we define a purely atomic probability measure p^y on $[[r, y]]_\mathcal{T}$,

$$p^y\{x\} := p(B_\mathcal{T}(x, y)) \quad \text{where} \quad B_\mathcal{T}(x, y) := F_\mathcal{T}(x) \setminus \bigcup_{z \in [[x,y]]_\mathcal{T} \setminus \{x\}} F_\mathcal{T}(z)$$

for $x \in [[r, y]]_\mathcal{T}$. Note that $B_\mathcal{T}(y, y) = F_\mathcal{T}(y)$, and for $x \neq y$, $B_\mathcal{T}(x, y) = \{x\}$ if and only if x is not a branch point. In the language of Section 2.2, $B_\mathcal{T}(x, y)$ is the *bush* that branches off of the *spine* $[[r, y]]_\mathcal{T}$ at x.

Proposition 6.2. *Let $\mathcal{L}$ denote the BCRT probability distribution on a measurable space $(S, \mathcal{S})$ of rooted, weighted $\mathbb{R}$-trees. Let A denote the set of trees $(\mathcal{T}, d, r, p) \in S$ in which: branch points are dense, p does not assign any continuous mass to the skeleton of $\mathcal{T}$, and*

$$d(r, y) = \sqrt{\pi} \lim_{h\to 0} \sqrt{h}\#\{x \in [[r, y]]_\mathcal{T} \colon p^y\{x\} > h\} \qquad \text{for all } y \in \mathcal{T}. \tag{6.2}$$

Then $\mathcal{L}(A) = 1$ and any two trees in A are isomorphic if and only if they are mass-structurally equivalent.

Informally, this proposition states that the BCRT is a.s. uniquely specified, up to isomorphism, by its mass-structural equivalence class.

Proof. It follows from well-known properties of the BCRT and, for equation (6.2), results in [29], that $\mathcal{L}(A) = 1$.

Now, consider two trees $(\mathcal{T}_i, d_i, r_i, p_i) \in A$, $i = 1, 2$, and suppose they are mass-structurally equivalent via a mass-structural isomorphism ϕ from $\mathcal{T}_1$ to $\mathcal{T}_2$. Then for any two branch points $x, y \in \mathcal{T}_1$ with $x \in [[r_1, y]]_{\mathcal{T}_1}$, we get $p_1^y(x) = p_2^{\phi(y)}(\phi(x))$. Thus, for any branch point $y \in \mathcal{T}_1$,

$$\begin{aligned}
d_1(r_1, y) &= \sqrt{\pi} \lim_{h \to 0} \sqrt{h} \# \big\{ x \in [[r_1, y]]_{\mathcal{T}_1} : p_1^y\{x\} > h \big\} \\
&= \sqrt{\pi} \lim_{h \to 0} \sqrt{h} \# \big\{ x \in [[r_1, y]]_{\mathcal{T}_1} : p_2^{\phi(y)}\{\phi(x)\} > h \big\} \\
&= \sqrt{\pi} \lim_{h \to 0} \sqrt{h} \# \big\{ z \in [[r_2, \phi(y)]]_{\mathcal{T}_2} : p_2^{\phi(y)}\{z\} > h \big\} = d_2(r_2, \phi(y)).
\end{aligned}$$

Together with structure-preserving property of ϕ, this shows that ϕ is an isometry from the branch points of $\mathcal{T}_1$ to those of $\mathcal{T}_2$. As branch points are dense in both trees, we can extend ϕ uniquely to a root-preserving isometry $\psi \colon \mathcal{T}_1 \to \mathcal{T}_2$.

The mass-preserving property of ϕ and the root-preserving isometry property of ψ yield that for any branch point $x \in \mathcal{T}_1$ we have

$$p_1\big(F_{\mathcal{T}_1}(x)\big) = p_2\big(F_{\mathcal{T}_2}(\phi(x))\big) = p_2\big(\psi\big(F_{\mathcal{T}_1}(x)\big)\big).$$

The fringe subtrees rooted at branch points in $\mathcal{T}_1$ comprise a π-system that generates the Borel σ-algebra, and likewise for $\mathcal{T}_2$. Thus, by Dynkin's π-λ theorem, ψ is measure-preserving. We conclude that ψ is an isomorphism. □

6.2 Structural equivalence

In the introduction to this paper, we heuristically described mass-structural equivalence as equivalence of the interaction between mass and "underlying tree structure." Here, we present one notion of underlying structure as an object in itself. One approach to this would be to define structural equivalence as topological equivalence, i.e. equivalence up to homeomorphism. However, the coupled Brownian CRT and IP tree of Proposition 6.1 are not homeomorphic to each other. We present a weaker notion of structural equivalence.

Definition 6.3. *Consider a rooted* $\mathbb{R}$*-tree* $(\mathcal{T}, d, r)$*. A leaf* $x \in \mathcal{T}$ *is a* discrete leaf *if there exists some branch point* $y \in \mathcal{T}$ *(its* parent*) that separates* x *from all other branch points. These discrete leaves, along with the branch points and the root* r*, comprise the set of* structural points *of* $(\mathcal{T}, d, r)$*.*

Let $\mathscr{V}_i$ *denote the set of structural points of a tree* $(\mathcal{T}_i, d_i, r_i)$ *for* $i = 1, 2$*. A* structural isomophism *between these* $\mathbb{R}$*-trees is a bijection* $f \colon \mathscr{V}_1 \to \mathscr{V}_2$ *with the property that, for* $x, y \in \mathscr{V}_1$*, we have* $x \in [[r_1, y]]_{\mathcal{T}_1}$ *if and only if* $f(x) \in [[r_2, f(y)]]_{\mathcal{T}_2}$*.*

Two rooted $\mathbb{R}$*-trees are said to be* structurally equivalent *if there exists a structural isomorphism from one to the other. It is straightforward to confirm that this is an equivalence relation.*

The following example illustrates the subtle distinction between the discrete leaves defined here and the isolated leaves of Definition 1.2. We conjecture, and it should not be difficult to show, that replacing isolated leaves with discrete leaves of $\mathrm{span}(p)$ in Definition 1.2 would yield an equivalent notion of mass-structural equivalence, but we will not prove this.

Example 6.4. Let $(\mathcal{T}, \ell_1, 0, p)$ be a Brownian CRT embedded in ℓ_1 via the bead crushing construction discussed in the proof of Proposition 6.1. Let $x_1, x_2, \ldots$ be i.i.d. samples from p. For $n \geq 2$, let ϕ_n denote the linear transformation on ℓ_1 that sends each coordinate vector $\mathbf{e}_k$ to $\mathbf{e}_{nk}$ for $k \geq 1$. Then, define

$$\mathcal{T}_2 := \phi_2(\mathcal{T}) \cup \bigcup_{k\geq 1} \left(\phi_2(x_k) + [0, 2^{-k}]\mathbf{e}_{2k-1}\right),$$
$$\mathcal{T}_3 := \phi_3(\mathcal{T}) \cup \bigcup_{k\geq 1} \left(\phi_3(x_k) + \left([0, 2^{-k-1}]\mathbf{e}_{3k-2} \cup [0, 2^{-k-1}]\mathbf{e}_{3k-1}\right)\right), \quad \text{and}$$
$$\mathcal{T}_4 := \phi_4(\mathcal{T}) \cup \bigcup_{k\geq 1} \left(\phi_4(x_k) + \left([0, 2^{-k-1}]\mathbf{e}_{4k-3} \cup [0, 2^{-k-2}]\mathbf{e}_{4k-2} \cup [0, 2^{-k-2}]\mathbf{e}_{4k-1}\right)\right).$$

In other words, $\mathcal{T}_2$ is formed by isometrically re-embedding $\mathcal{T}$ into the even coordinates in ℓ_1, and then attaching new, macroscopic branches at each of the leaves x_k, $k \geq 1$; and $\mathcal{T}_3$ and $\mathcal{T}_4$ are correspondingly formed by attaching two or three new branches at each sampled leaf. For $n = 2, 3, 4$, let p_n denote the length measure on $\mathcal{T}_n \setminus \phi_n(\mathcal{T})$, and consider $(\mathcal{T}_n, \ell_1, 0, p_n)$ as a rooted, weighted $\mathbb{R}$-tree. Then $\mathcal{T}_2 = \text{span}(p_2)$ and the leaves $\phi(x_k) + 2^{-k}\mathbf{e}_{2k-1}$ are isolated leaves in the sense of Definition 1.2, and correspondingly for $\mathcal{T}_3$ and $\mathcal{T}_4$. However, the newly added leaves in $\mathcal{T}_2$, in particular, are not "discrete" in the sense of Definition 6.3, since leaves in a Brownian CRT do not have parent branch points but rather arise as limit points of branch points.

If we did not include isolated leaves, like those in $\mathcal{T}_2$, $\mathcal{T}_3$, and $\mathcal{T}_4$, as special points, but otherwise left Definitions 1.2 and 1.3 of special points and mass-structural equivalence as is, then $(\mathcal{T}_3, \ell_1, 0, p_3)$ and $(\mathcal{T}_4, \ell_1, 0, p_4)$ would be considered mass-structurally equivalent, and Theorems 1.5 and 1.7 would fail.

Now, define

$$\mathcal{T}_2' := \phi_2(\mathcal{T}) \cup \left(\phi_2((x_1 \wedge x_2)_{\mathcal{T}}) + [0, 1]\mathbf{e}_1\right) \quad \text{and}$$
$$\mathcal{T}_3' := \phi_2(\mathcal{T}) \cup \left(\phi_2((x_1 \wedge x_2)_{\mathcal{T}}) + ([0, 1]\mathbf{e}_1 \cup [0, 1]\mathbf{e}_3)\right).$$

Consider $(\mathcal{T}_2', \ell_1, 0, \phi_2(p))$ and $(\mathcal{T}_3', \ell_1, 0, \phi_2(p))$. The newly added branches do not belong to $\text{span}(\phi_2(p))$, so their endpoints are not isolated leaves, in the sense of Definition 1.2. But these endpoints are discrete leaves, in the sense of Definition 6.3. Consequently, the two trees are mass-structurally equivalent to each other and to $(\mathcal{T}, \ell_1, 0, p)$, but not structurally equivalent.

Structural equivalence may be an interesting notion of equivalence, but the "underlying structure" – i.e. structural equivalence class – as an object sacrifices much of what makes CRTs interesting. For example, Croyden and Hambly [7] showed that there is a homeomorphism class of $\mathbb{R}$-trees that a.s. contains the Brownian CRT; i.e. its underlying structure is deterministic. Without either distances or masses to indicate relative "sizes" of components in a decomposition of the Brownian CRT, the randomness and much of the interesting fractal structure are lost.

6.3 Directions for further study

(1) Introduce and study growth processes for exchangeable hierarchies on $\mathbb{N}$, in the spirit of the Chinese restaurant process for exchangeable partitions [27], for use in applications such as Bayesian non-parametric hierarchical clustering, e.g. for nested topic models [6, 26]. The three behaviors mentioned around the statement of Theorem 1.10 – macroscopic branching, broom-like explosion, and comb-like erosion – cannot all be distinguished in the discrete regime, but the insight that all three can appear in scaling limits may aid in defining models for finite exchangeable random hierarchies. Such models would also give rise to potentially interesting random IP trees, via Theorem 1.8.

(2) In connection with (1), do random IP trees arise as scaling limits of suitably metrized random discrete trees? Can we learn about the IP trees from this perspective?

This may tie back to the perspective in Theorem 1.7, of IP trees as corresponding to e.i.g. hierarchies on $\mathbb{N}$, and the latter being represented as projectively consistent sequences, as in Definition 1.6.

(3) Study the images of other rooted, weighted CRTs, for example those arising from bead-crushing constructions as in [29, 32] (including stable CRTs), under the map from a rooted, weighted $\mathbb{R}$-tree to an isomorphism class of mass-structurally equivalent IP trees.

(4) Relate IP trees and mass-structural equivalence to the algebraic measure trees of [25].

(5) Characterize mass-structural equivalence in terms of deformations or correspondences, in the sense described in [13]. How can a tree be stretched, pruned, contracted, or otherwise modified without changing its mass-structure?

(6) Study notions of structural equivalence of CRTs that do not depend on either mass or quantified distance, such as that in Definition 6.3 or equivalence up to homeomorphism, as in [7]. Look at a space of $\mathbb{R}$-tree structures. Consider random elements, metrize the space, etc..

References

[1] David Aldous, *The continuum random tree. I*, Ann. Probab. **19** (1991), no. 1, 1–28. MR-1085326

[2] David Aldous, *The continuum random tree. III*, Ann. Probab. **21** (1993), no. 1, 248–289. MR-1207226

[3] David Aldous, *From random walks on discrete tree space to diffusions on continuous tree space*, https://www.fields.utoronto.ca/programs/scientific/98-99/stochastic_processes/abstracts.html, March 1999, Seminar on Stochastic Processes, Fields Institute.

[4] Jean Bertoin, *Self-similar fragmentations*, Ann. Inst. H. Poincaré Probab. Statist. **38** (2002), no. 3, 319–340. MR-1899456

[5] Jean Bertoin, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, Cambridge, 2006. MR-2253162

[6] David M. Blei, Thomas L. Griffiths, and Michael I. Jordan, *The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies*, J. ACM **57** (2010), no. 2, Art. 7, 30. MR-2606082

[7] David Croydon and Ben Hambly, *Self-similarity and spectral asymptotics for the continuum random tree*, Stochastic Process. Appl. **118** (2008), no. 5, 730–754. MR-2411518

[8] Peter Donnelly and Paul Joyce, *Consistent ordered sampling distributions: characterization and convergence*, Adv. in Appl. Probab. **23** (1991), no. 2, 229–258. MR-1104078

[9] Andreas W. M. Dress, *Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces*, Adv. in Math. **53** (1984), no. 3, 321–402. MR-0753872

[10] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque (2002), no. 281, vi+147. MR-1954248

[11] Thomas Duquesne and Jean-François Le Gall, *Probabilistic and fractal aspects of Lévy trees*, Probab. Theory Related Fields **131** (2005), no. 4, 553–603. MR-2147221

[12] Steven N. Evans, *Probability and real trees*, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008, Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. MR-2351587

[13] Steven N. Evans, Jim Pitman, and Anita Winter, *Rayleigh processes, real trees, and root growth with re-grafting*, Probab. Theory Related Fields **134** (2006), no. 1, 81–126. MR-2221786

[14] Noah Forman, Chris Haulk, and Jim Pitman, *A representation of exchangeable hierarchies by sampling from random real trees*, Probab. Theory Related Fields **172** (2018), no. 1-2, 1–29. MR-3851828

[15] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Aldous diffusion I: a projective system of continuum k-tree evolutions*, arXiv:1809.07756 [math.PR], 37 pp., 2018. MR-4144083

[16] Bernard R. Gelbaum and John M. H. Olmsted, *Counterexamples in analysis*, Dover Publications, Inc., Mineola, NY, 2003, Corrected reprint of the second (1965) edition. MR-1996162

[17] Alexander V. Gnedin, *The representation of composition structures*, Ann. Probab. **25** (1997), no. 3, 1437–1450. MR-1457625

[18] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322. MR-2520129

[19] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Tree-valued resampling dynamics martingale problems and applications*, Probab. Theory Related Fields **155** (2013), no. 3-4, 789–838. MR-3034793

[20] André Joyal, *Une théorie combinatoire des séries formelles*, Adv. in Math. **42** (1981), no. 1, 1–82. MR-0633783

[21] Olav Kallenberg, *Probabilistic symmetries and invariance principles*, Probability and its Applications (New York), Springer, New York, 2005. MR-2161313

[22] J. F. C. Kingman, *The representation of partition structures*, J. London Math. Soc. (2) **18** (1978), no. 2, 374–380. MR-0509954

[23] Jean-François Le Gall, *Random trees and applications*, Probab. Surv. **2** (2005), 245–311. MR-2203728

[24] Wolfgang Löhr, Leonid Mytnik, and Anita Winter, *The Aldous chain on cladograms in the diffusion limit*, Ann. Probab. **48** (2020), no. 5, 2565–2590. MR-4152651

[25] Wolfgang Löhr and Anita Winter, *Spaces of algebraic measure trees and triangulations of the circle*, arXiv:1811.11734 [math.PR], November 2018.

[26] John Paisley, Chong Wang, David M. Blei, and Michael I. Jordan, *Nested hierarchical dirichlet processes*, IEEE Trans. Pattern Anal. Mach. Intell. **37** (2015), no. 2, 256–270.

[27] Jim Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, 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. MR-2245368

[28] Jim Pitman, *Poisson-Kingman partitions*, Statistics and science: a Festschrift for Terry Speed, IMS Lecture Notes Monogr. Ser., vol. 40, Inst. Math. Statist., Beachwood, OH, 2003, pp. 1–34. MR-2004330

[29] Jim Pitman and Matthias Winkel, *Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions*, Ann. Probab. **37** (2009), no. 5, 1999–2041. MR-2561439

[30] Jim Pitman and Marc Yor, *The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator*, Ann. Probab. **25** (1997), no. 2, 855–900. MR-1434129

[31] Svetlozar T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1991. MR-1105086

[32] Franz Rembart and Matthias Winkel, *Recursive construction of continuum random trees*, Ann. Probab. **46** (2018), no. 5, 2715–2748. MR-3846837

[33] Charles Semple and Mike Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and its Applications, vol. 24, Oxford University Press, Oxford, 2003. MR-2060009

[34] Jacques Tits, *A "theorem of Lie-Kolchin" for trees*, Contributions to Algebra (Collection of Papers Dedicated to Ellis Kolchin), Academic Press, 1977, pp. 377–388. MR-0578488

Acknowledgments. The author thanks Soumik Pal and Matthias Winkel for their support and suggestions, as well as the referees, whose comments led to improved presentation.