Bernoulli **30**(2), 2024, 1029–1054
https://doi.org/10.3150/23-BEJ1623

A recursive distributional equation for the stable tree

NICHOLAS CHEE[1,a], FRANZ REMBART[2,b] and MATTHIAS WINKEL[3,c]

[1,a]*niccheekn@gmail.com*
[2]*Columbia University, Department of Statistics, 1255 Amsterdam Avenue, New York, NY 10027, USA,* [b]*franz.rembart@gmail.com*
[3]*University of Oxford, Department of Statistics, 24–29 St Giles', Oxford OX1 3LB, UK,* [c]*winkel@stats.ox.ac.uk*

We provide a new characterisation of Duquesne and Le Gall's α-stable tree, $\alpha \in (1,2]$, as the solution of a recursive distributional equation (RDE) of the form $\mathcal{T} \overset{d}{=} g(\xi, \mathcal{T}_i, i \geq 0)$, where g is a concatenation operator, $\xi = (\xi_i, i \geq 0)$ a sequence of scaling factors, $\mathcal{T}_i$, $i \geq 0$, and $\mathcal{T}$ are i.i.d. trees independent of ξ. This generalises the characterisation of the Brownian Continuum Random Tree proved by Albenque and Goldschmidt, based on self-similarity observed by Aldous. By relating to previous results on a rather different class of RDE, we explore the present RDE and obtain for a large class of similar RDEs that the fixpoint is unique (up to multiplication by a constant) and attractive.

Keywords: Gromov–Hausdorff distance; recursive distributional equation; $\mathbb{R}$-tree; stable tree

1. Introduction

$\mathbb{R}$-trees constitute a class of loop-free length spaces which frequently arise as scaling limits of discrete trees. Significant attention turned to random $\mathbb{R}$-trees following Aldous's introduction of the Brownian Continuum Random Tree (BCRT) [4] and Evans's initiation of Gromov–Hausdorff topologies in probability theory [17]. The BCRT is generalised by Duquesne and Le Gall's α-stable trees [15], which represent the genealogies of continuous-state branching processes with branching mechanism $\psi(\lambda) = \lambda^\alpha, \alpha \in (1,2]$. When $\alpha = 2$, we recover the BCRT. α-stable trees emerge in scaling limits of numerous discrete tree structures, e.g. Bienaymé–Galton–Watson trees [4] and their vertex-cut trees [14] and conditioned stable Lévy forests [11]. Particular aspects of α-stable trees include invariance under random re-rooting [18], decompositions along the diameter [16], an embedding property [13], and links to beta coalescents [7]. We emphasise a crucial self-similarity property of α-stable trees: decomposing an α-stable tree above a certain height or at appropriate nodes results in the connected components after decomposition forming rescaled independent copies of the original tree. This observation was first formalised by Miermont [20], building upon Bertoin's self-similar fragmentation theory [8].

In this paper, we express the self-similarity of the α-stable tree by a new *recursive distributional equation* (RDE) in the setting of Aldous and Bandyopadhyay's survey paper [6]. Given a random variable $\mathcal{T}$ taking values in a Polish metric space $(\mathbb{T}, d)$, an RDE is a stochastic equation of the form

$$\mathcal{T} \overset{d}{=} g(\xi, \mathcal{T}_i, i \geq 0) \quad \text{ on } \mathbb{T},$$

where $(\mathcal{T}_i, i \geq 0)$ are i.i.d. and distributed as $\mathcal{T}$, g is a measurable mapping, and ξ is independent of $(\mathcal{T}_i, i \geq 0)$. RDEs are pertinent in various contexts with recursive structures, ranging from Bienaymé–Galton–Watson branching processes [6] to Quicksort algorithms [25].

RDEs have been employed in the recursive construction of the BCRT by Albenque and Goldschmidt [3], recursively concatenating three rescaled trees at a single point. Broutin and Sulzbach [10] extended this to further recursive combinatorial structures and weighted $\mathbb{R}$-trees under a finite concatenation

1350-7265 This research was funded, in whole or in part, by [UKRI].

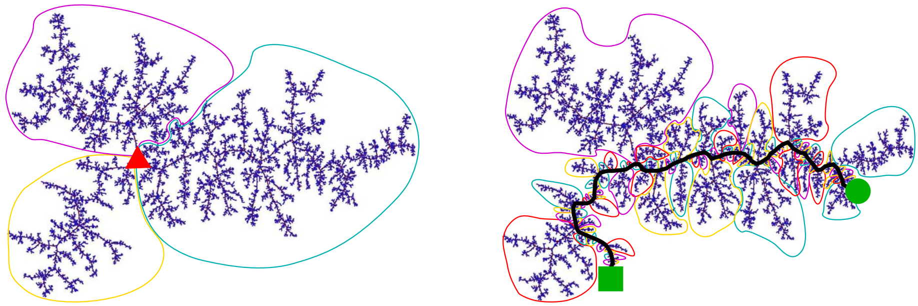

Figure 1. RDEs derived from the decomposition of the BCRT (simulation courtesy of Igor Kortchemski) around a branchpoint (the red triangle) into three parts and along the spine from the root (green square) to a random leaf (green circle) into "infinitely many" parts.

operation. Rembart and Winkel [24] did similarly with $\mathbb{R}$-trees under a different operation that concatenates a countable (possibly infinite) number of rescaled trees to a branch/spine. See Figure 1.

In this paper, we consider as g the operation that concatenates at a single point a countable number of $\mathbb{R}$-trees $\mathcal{T}_i \overset{d}{=} \mathcal{T}$, rescaled by $\xi_i \geq 0$, $i \geq 0$, respectively, seeking to obtain a version of $\mathcal{T}$. Theorem 3.5 shows that the law of the α-stable tree is a fixpoint solution of an RDE of this type. This is illustrated in Figure 2. Our primary argument appeals to Marchal's random growth algorithm [19], which provides a recursive method of constructing α-stable trees as scaling limits. To explore the uniqueness of this solution (up to rescaling distances by a constant) we require certain finite height moments. In the absence of this condition, further solutions can be obtained, e.g., by decorating the α-stable tree with massless branches, see Remark 3.6.

Let us explore our approach to uniqueness and attraction in the context of the literature. While our results resemble [3] and [10] for the (binary) BCRT and other finitely branching structures, our results extend finite concatenation operations to handle trees such as the α-stable trees, whose branch points are of countably infinite multiplicity. Extending their uniqueness and attraction results is not straightforward using the methods of [3,10]. On the other hand, [24] presents an RDE for which the law of the α-stable tree is a unique and attractive fixpoint, but the concatenation approach of em-

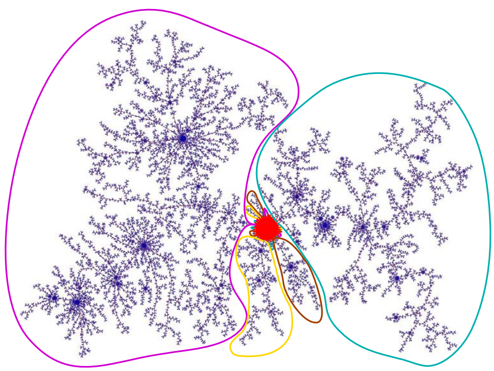

Figure 2. RDEs derived from the decomposition of a stable tree (simulation courtesy of Igor Kortchemski) around a branchpoint (the red star) into "infinitely many" parts.

ploying strings of beads (weighted intervals) and bead-splitting processes of [23] is different. Our RDEs only require countably infinite weight sequences such as Poisson–Dirichlet sequences and give a less technical recursive construction of α-stable trees that elucidates how mass partitions in α-stable trees relate to urn models and partition-valued processes. This approach allows us to work under the Gromov–Hausdorff–Prokhorov topology hence responding to a suggestion in [3, Section 4.2], who use Gromov–Prokhorov.

Specifically, we prove the self-similarity property of α-stable trees decomposed at a branch point solely via the recursive nature of Marchal's algorithm, without need for Miermont's fragmentation tree theory [20]. To prove our uniqueness and attraction result, Theorem 4.2, we establish a connection between the two types of RDE, which effectively breaks down the proofs here into a one-dimensional martingale argument, the uniqueness and attraction of the RDE of [24] and a tightness argument that again builds on [24] by constructing an auxiliary dominating CRT.

The structure of this paper is as follows. In Section 2, we state background results on $\mathbb{R}$-trees and α-stable trees, and we collect some of the tools we use, chiefly the setup of Marchal's algorithm and RDEs. Section 3 establishes an RDE for the law of the α-stable tree and indicates other fixpoint solutions to the same RDE. In Section 4, we obtain the uniqueness and attraction properties of the RDE solution up to multiplicative constants. The latter arguments are in a general setup where g is the single-point concatenation operation, but the distribution of ξ is just subject to non-degeneracy assumptions.

2. Preliminaries

2.1. $\mathbb{R}$-trees and topologies on sets of (marked or weighted) $\mathbb{R}$-trees

Definition 2.1 ($\mathbb{R}$-tree). A metric space $(\mathcal{T},d)$ is an $\mathbb{R}$-*tree* if for every $a,b\in\mathcal{T}$,

(i) there exists a unique isometry $f_{a,b}\colon [0,d(a,b)]\to\mathcal{T}$ such that $f_{a,b}(0)=a$ and $f_{a,b}(d(a,b))=b$. In this case, let $[\![a,b]\!]$ denote the image $f_{a,b}([0,d(a,b)])$,

(ii) if $h\colon [0,1]\to\mathcal{T}$ is a continuous injective map with $h(0)=a$ and $h(1)=b$, then $h([0,1])=[\![a,b]\!]$, i.e. the only non self-intersecting path from a to b is $[\![a,b]\!]$.

A *rooted* $\mathbb{R}$-tree $(\mathcal{T},d,\rho)$ is an $\mathbb{R}$-tree $(\mathcal{T},d)$ with a distinguished vertex $\rho\in\mathcal{T}$ called the *root*. The *degree* of a vertex $a\in\mathcal{T}$ is the number of connected components of $\mathcal{T}\setminus\{a\}$. A *leaf* is a vertex $a\in\mathcal{T}\setminus\{\rho\}$ with degree one. We denote the set of leaves in $\mathcal{T}$ by $\mathcal{L}(\mathcal{T})$. We say that $a\in\mathcal{T}\setminus\{\rho\}$ is a *branch point* if its degree is at least three. Finally, for any $a\in\mathcal{T}$, we define the *height* of a as $d(\rho,a)$, and the *height* of $\mathcal{T}$ as $\mathrm{ht}(\mathcal{T}):=\sup_{a\in\mathcal{T}}d(\rho,a)$.

We will want to specify a marked point on a compact rooted $\mathbb{R}$-tree. Two marked compact rooted $\mathbb{R}$-trees $(\mathcal{T},d,\rho,x)$ and $(\mathcal{T}',d',\rho',x')$ are GH^{m}-*equivalent* if there is an isometry $f\colon\mathcal{T}\to\mathcal{T}'$ such that $f(\rho)=\rho'$ and $f(x)=x'$. We denote the set of equivalence classes of marked compact rooted $\mathbb{R}$-trees by $\mathbb{T}_{\mathrm{m}}$. For two marked compact rooted $\mathbb{R}$-trees, the *marked Gromov–Hausdorff distance* is defined as

$$d_{\mathrm{GH}}^{\mathrm{m}}((\mathcal{T},d,\rho,x),(\mathcal{T}',d',\rho',x')):=\inf_{\phi,\phi'}\big(\delta_{\mathrm{H}}(\phi(\mathcal{T}),\phi'(\mathcal{T}'))\vee\delta(\phi(\rho),\phi'(\rho'))\vee\delta(\phi(x),\phi'(x'))\big),$$

where the infimum is taken over all metric spaces (X,δ) and all isometric embeddings $\phi\colon\mathcal{T}\to X$ and $\phi'\colon\mathcal{T}'\to X$. The marked Gromov–Hausdorff distance only depends on the GH^{m}-equivalence classes of $(\mathcal{T},d,\rho,x)$ and induces a metric on $\mathbb{T}_{\mathrm{m}}$, which we also denote by $d_{\mathrm{GH}}^{\mathrm{m}}$.

Suppose now that (X,δ) is a complete metric space, then (X,δ,μ) is a *metric measure space* if (X,δ) is equipped with a Borel probability measure μ. A *weighted* $\mathbb{R}$-tree is a compact rooted $\mathbb{R}$-tree $(\mathcal{T},d,\rho)$

equipped with a Borel probability measure μ, referred to as the *mass measure*. We will often write $\mathcal{T}$ for a weighted $\mathbb{R}$-tree, the distance, the root and the mass measure being implicit. We write

$$d_{\mathrm{GHP}}((\mathcal{T},d,\rho,\mu),(\mathcal{T}',d',\rho',\mu')) := \inf_{\phi,\phi'} \left(\delta_{\mathrm{H}}(\phi(\mathcal{T}),\phi'(\mathcal{T}')) \vee \delta(\phi(\rho),\phi'(\rho')) \vee \delta_{\mathrm{P}}(\phi_*\mu,\phi'_*\mu')\right),$$

for the Gromov–Hausdorff–Prokhorov distance between weighted $\mathbb{R}$-trees $(\mathcal{T},d,\rho,\mu)$, $(\mathcal{T}',d',\rho',\mu')$, where the infimum is taken over all metric spaces (X,δ) and all isometric embeddings $\phi\colon \mathcal{T} \to X$ and $\phi'\colon \mathcal{T}' \to X$, δ_{P} denotes the Prokhorov metric, and $\phi_*\mu$, $\phi'_*\mu'$ are push-forwards of μ,μ'.

Two weighted $\mathbb{R}$-trees $(\mathcal{T},d,\rho,\mu)$ and $(\mathcal{T}',d',\rho',\mu')$ are considered GHP-*equivalent* if there is an isometry $f\colon (\mathcal{T},d,\rho,\mu) \to (\mathcal{T}',d',\rho',\mu')$ such that $f(\rho)=\rho'$ and μ' is the push-forward of μ under f. Denote the set of equivalence classes of weighted $\mathbb{R}$-trees by $\mathbb{T}_{\mathrm{w}}$. The Gromov–Hausdorff–Prokhorov distance naturally induces a metric on $\mathbb{T}_{\mathrm{w}}$.

Proposition 2.2 (e.g., Proposition 9(ii) of [21] and Theorem 2.7 of [1].). *The spaces* $\left(\mathbb{T}_{\mathrm{m}}, d_{\mathrm{GH}}^{\mathrm{m}}\right)$ *and* $(\mathbb{T}_{\mathrm{w}}, d_{\mathrm{GHP}})$ *are Polish.*

In [4], Aldous originally built his theory of *continuum trees*, in $\ell_1(\mathbb{N})$. Indeed, some of our arguments will benefit from specific representatives in $\ell_1(\mathbb{U})$, where $\mathbb{U}$ is the countable set of integer words. In any case, [4, Theorem 3] connects Aldous's $\ell_1(\mathbb{N})$ embedding and the above setup of weighted $\mathbb{R}$-trees. So, we call a weighted $\mathbb{R}$-tree $(\mathcal{T},d,\rho,\mu)$ a *continuum tree* if the Borel probability measure μ satisfies the following properties.

(i) $\mu(\mathcal{L}(\mathcal{T}))=1$, that is, μ is supported by the leaves of $\mathcal{T}$.
(ii) μ is non-atomic, that is, if $a \in \mathcal{L}(\mathcal{T})$, then $\mu(\{a\})=0$.
(iii) For every $a \in \mathcal{T} \setminus \mathcal{L}(\mathcal{T})$, we have $\mu(\mathcal{T}(a))>0$, where $\mathcal{T}(a) := \{\sigma \in \mathcal{T}\colon a \in [\![\rho,\sigma]\!]\}$ is the subtree above a in $\mathcal{T}$.

A *Continuum Random Tree* (CRT) is a random variable taking values in a space (of GHP-equivalence classes) of continuum trees. Note that conditions (i)–(ii) above imply that $\mathcal{L}(\mathcal{T})$ is uncountable for any CRT $\mathcal{T}$. It is not obvious how to determine the distribution of a CRT $(\mathcal{T},d,\rho,\mu)$ simply by this definition. To do this, Aldous introduced the notion of *reduced trees*. Let $m \geq 1$. A *uniform sample* of m points according to the measure μ is a vector $(V_1,\ldots,V_m)$ such that $V_i \sim \mu$, $i=1,\ldots,m$, are i.i.d.. The associated *m-th reduced subtree* of $(\mathcal{T},d,\rho,\mu)$ is the subtree of $\mathcal{T}$ spanned by $V_1,\ldots,V_m$ and ρ, i.e. $\bigcup_{1\leq j\leq m}[\![\rho,V_j]\!]$.

The distribution of the m-th reduced subtree is fully specified by its *tree shape* when regarded as a discrete, graph-theoretic, rooted tree with m labelled leaves, and by its *edge lengths*. The consistent system of m-th reduced subtree distributions, $m \geq 1$, may be regarded as a system of finite-dimensional distributions of a CRT [3]. It is well-known that they determine the distribution of a CRT on $\mathbb{T}_{\mathrm{w}}$.

We now turn to Marchal's algorithm which leads to the definition of a special class of continuum random trees, the α-stable trees with parameter $\alpha \in (1,2]$.

2.2. Marchal's algorithm and α-stable trees

Definition 2.3 (Marchal's algorithm). Given $\alpha \in (1,2]$, recursively construct a sequence $(\mathbf{T}_\alpha(n))_{n\geq 1}$ valued in the set of leaf-labelled discrete trees, with $\mathbf{T}_\alpha(n)$ having n leaves and a root, as follows.

(I) Initialise $\mathbf{T}_\alpha(1)$ as the unique tree consisting of a single edge connecting a root vertex and a leaf.
(II) For $n \geq 1$, given $\mathbf{T}_\alpha(n)$, assign weight $\alpha-1$ to each edge and weight $d-1-\alpha$ to each branch point of degree $d \geq 3$. Choose an edge or a branch point with probability proportional to its weight.

(III) (a) If an edge was selected, split the chosen edge into two edges at its midpoint by a new middle vertex. To this new vertex, attach a new edge carrying the $(n+1)$-st leaf.
(b) If a branch point was selected, attach a new edge carrying the $(n+1)$-st leaf to this vertex.
(IV) Denote the resulting tree by $\mathbf{T}_\alpha(n+1)$ and repeat from (II) with $n \mapsto n+1$.

Define the measure $W(\cdot)$ which assigns the total weight to sub-structures in Marchal's algorithm. It is easy to see that, regardless of tree shape, for all $n \geq 1$, the total weight of the tree is $W(\mathbf{T}_\alpha(n)) = n\alpha - 1$. The distribution of the shape of the trees in Marchal's algorithm was given in [19, Theorem 1]:

Proposition 2.4. *Suppose* $\mathbf{t}$ *is a given leaf-labelled tree with n leaves and a root, where* $n \geq 2$, *then the tree shape of* $\mathbf{T}_\alpha(n)$ *has distribution*

$$\mathbb{P}(\mathbf{T}_\alpha(n) = \mathbf{t}) = \frac{\prod_{v \in \mathbf{t}} p_{\deg(v)}}{\prod_{i=1}^{n-1}(i\alpha - 1)},$$

where $p_1 = 1$, $p_2 = 0$, *and* $p_k = \left|\prod_{i=1}^{k-2}(\alpha - i)\right|$ *for* $k \geq 3$.

We state a result by Curien and Haas [13, Theorem 5(iii)], strengthening [19, Theorem 2], allowing us to regard the *scaling limit* of the sequence of trees generated by Marchal's algorithm in $(\mathbb{T}_{\mathrm{w}}, d_{\mathrm{GHP}})$.

Proposition 2.5. *For* $\alpha \in (1,2]$, *let* $\beta := 1 - 1/\alpha \in (0, 1/2]$. *Let* μ_n *denote the empirical mass measure on the leaves of* $T_\alpha(n)$, d_n *be the graph distance on* $T_\alpha(n)$, *and* ρ_n *be the root. Then*

$$\left(T_\alpha(n), \frac{d_n}{\alpha n^\beta}, \rho_n, \mu_n\right) \xrightarrow{a.s.} (\mathcal{T}_\alpha, d_\alpha, \rho_\alpha, \mu_\alpha) \qquad \text{as } n \to \infty,$$

in the Gromov–Hausdorff–Prokhorov topology, for some CRT $(\mathcal{T}_\alpha, d_\alpha, \rho_\alpha, \mu_\alpha)$.

Definition 2.6 (α-stable tree). We call $(\mathcal{T}_\alpha, d_\alpha, \rho_\alpha, \mu_\alpha)$ the *α-stable tree*, $\alpha \in (1,2]$.

We highlight that each element in the sequence of trees produced by Marchal's algorithm is a Bienaymé–Galton–Watson tree conditioned on a fixed number of leaves, whose offspring distributions lie in the domain of attraction of an α-stable law. This can be seen from Proposition 2.4, see also [19, Section 2.3] and [15, Theorems 3.2.1 and 3.3.3]. Indeed, the convergence stated in Proposition 2.5 holds as a distributional scaling limt for any family of such conditioned Bienaymé–Galton–Watson trees.

It is useful to parametrize the α-stable tree by an index $\beta := 1 - 1/\alpha \in (0, 1/2]$. We often rescale trees: distances by c^β and masses by c, as in

$$\left(\mathcal{T}_\alpha, c^\beta d_\alpha, \rho_\alpha, c\mu_\alpha\right).$$

When $\alpha = 2$, no weight is ever given to a vertex of $\mathbf{T}_2(n)$, $n \geq 1$, in the second step of Marchal's algorithm. In the scaling limit, this coheres with the fact that $\mathcal{T}_2$ is binary almost surely.

The tree $(\mathcal{T}_\alpha, d_\alpha, \rho_\alpha, \mu_\alpha)$ induces a distribution ς_α on $\mathbb{T}_{\mathrm{w}}$. We call the distribution ς_α the *law of the α-stable tree*. Similarly, we will consider the distribution $\varsigma_\alpha^{\mathrm{m}}$ of $(\mathcal{T}_\alpha, d_\alpha, \rho_\alpha, x_\alpha)$ on $\mathbb{T}_{\mathrm{m}}$ when $x_\alpha \sim \mu_\alpha$ is a marked element of $\mathcal{T}_\alpha$ sampled from μ_α, which we call the *law of the marked α-stable tree*.

In [13], Curien and Haas exploit the recursive nature of Marchal's algorithm property to obtain a randomly rescaled α'-stable tree from an α-stable tree by pruning, where $1 < \alpha < \alpha' \leq 2$. They identified sub-constructions within Marchal's algorithm with parameter α that evolve as a time-changed Marchal algorithm with parameter α'. We use a similar approach in Section 3 to find a recursive distributional equation where the law of the α-stable tree is a solution.

2.3. Mittag–Leffler distributions and Chinese restaurant processes

Given $\beta > 0$ and $\theta > -\beta$, a random variable L valued in $[0,\infty)$ has a *generalised Mittag–Leffler distribution* with parameters (β,θ), denoted by $L \sim \mathrm{ML}(\beta,\theta)$, if it has p-th moment

$$\mathbb{E}[L^p] = \frac{\Gamma(\theta+1)\Gamma(\theta/\beta+1+p)}{\Gamma(\theta/\beta+1)\Gamma(\theta+\beta p+1)}, \qquad p \geq 1. \tag{1}$$

Indeed, the moments (1) uniquely characterise a distribution, see e.g. [22]. It was shown in [2, Lemma 11] that α times the distance between two points sampled from μ_α in an α-stable tree $\mathcal{T}_\alpha$ is $\mathrm{ML}(\beta,\beta)$-distributed, where $\beta = 1 - 1/\alpha$. By invariance under random re-rooting [18, Theorem 11], this is also the distribution of α times the distance between the root and point sampled from μ_α.

Definition 2.7 (Chinese restaurant process). Given $\beta \in [0,1]$ and $\theta > -\beta$, the two-parameter *Chinese restaurant process* with a (β,θ) seating plan, denoted by $\mathrm{CRP}(\beta,\theta)$, proceeds as follows. Label customers by $n \geq 1$. Seat customer 1 at the first table. For $n \geq 1$, let K_n denote the number of tables occupied after customer n has been seated and let $N_j(n)$ denote the number of customers seated at the j-th table for $j \in \{1,\dots,K_n\}$. At the next arrival, conditional on $(N_1(n),\dots,N_j(n))$, customer $n+1$

- sits at the j-th table with probability $(N_j(n)-\beta)/(n+\theta)$ for $j \in \{1,\dots,K_n\}$,
- opens the (K_n+1)-st table with the complementary probability $(\theta + K_n\beta)/(n+\theta)$.

The CRP satisfies the following limit theorem, cf. [22, Theorems 3.2 and 3.8].

Proposition 2.8. *Consider a two-parameter Chinese restaurant process with parameters $\beta \in (0,1)$ and $\theta > -\beta$, denoted by $\mathrm{CRP}(\beta,\theta)$. Then the number of tables K_n at time n satisfies*

$$n^{-\beta} K_n \xrightarrow{a.s.} K_\infty \qquad \text{as } n \to \infty,$$

where $K_\infty \sim \mathrm{ML}(\beta,\theta)$. Furthermore, relative table sizes have almost sure limits

$$\left(\frac{N_1(n)}{n}, \frac{N_2(n)}{n}, \dots, \frac{N_{K_n}(n)}{n}, 0, 0, \dots\right) \xrightarrow{a.s.} \left(W_1, \overline{W}_1 W_2, \overline{W}_1 \overline{W}_2 W_3, \dots\right) \qquad \text{as } n \to \infty,$$

where $W_j \sim \mathrm{Beta}(1-\beta, \theta + j\beta)$, $j \geq 1$, are independent and $\overline{W}_j := 1 - W_j$ for all $j \geq 1$.

The distribution of the vector $(P_1,P_2,P_3,\dots) := \left(W_1, \overline{W}_1 W_2, \overline{W}_1 \overline{W}_2 W_3, \dots\right)$ as defined in Proposition 2.8 is the *Griffiths–Engen–McCloskey distribution* with parameters (β,θ), denoted by $\mathrm{GEM}(\beta,\theta)$. Ordering $(P_i, i \geq 1)$ in decreasing order yields the Poisson–Dirichlet distribution with parameters (β,θ), in short $\mathrm{PD}(\beta,\theta)$, i.e. $\left(P_i^{\downarrow}, i \geq 1\right) := (P_i, i \geq 1)^{\downarrow} \sim \mathrm{PD}(\beta,\theta)$.

2.4. Recursive distributional equations (RDEs)

We review RDEs in their full generality, as presented in [6, Section 2.1]. Denote our underlying probability space by $(\Omega, \mathcal{F}, \mathbb{P})$. Given two measurable spaces $(\mathbb{S}, \mathcal{F}_{\mathbb{S}})$ and $(\Theta, \mathcal{F}_{\Theta})$, consider

$$\Theta^* := \Theta \times \bigcup_{0 \le m \le \infty} \mathbb{S}^m, \tag{2}$$

where the union is disjoint over $\mathbb{S}^m$, the space of $\mathbb{S}$-valued sequences of lengths $0 \le m \le \infty$, and where $\mathbb{S}^0 := \{\Delta\}$ is the singleton set and $\mathbb{S}^\infty$ is constructed as a typical sequence space.

Equip Θ^* with the product sigma-algebra. Let $g \colon \Theta^* \to \mathbb{S}$ be a measurable map, and define random variables $(\mathcal{S}_i, i \ge 0) \in \mathbb{S}^\infty$, $(\xi, N) \in \Theta \times \overline{\mathbb{N}} := \Theta \times \{0, 1, \dots; \infty\}$ as follows.

(i) $(\xi, N) \sim \nu$, where ν is a probability measure on $\Theta \times \overline{\mathbb{N}}$.
(ii) $\mathcal{S}_i \sim \eta$, $i \ge 0$, i.i.d., where η is a probability measure on $\mathbb{S}$.
(iii) (ξ, N) and $(\mathcal{S}_i, i \ge 0)$ are independent.

Denote by $\mathcal{P}(\mathbb{S})$ the set of probability measures on $(\mathbb{S}, \mathcal{F}_{\mathbb{S}})$. Given the distribution ν on $\Theta \times \overline{\mathbb{N}}$, we obtain a mapping

$$\Phi \colon \mathcal{P}(\mathbb{S}) \to \mathcal{P}(\mathbb{S}), \qquad \eta \mapsto \Phi(\eta), \tag{3}$$

where $\Phi(\eta)$ is the distribution of $\mathcal{S} := g(\xi, \mathcal{S}_i, 0 \le i \le^* N)$, and where the notation $\le^* N$ means $\le N$ for $N < \infty$ and $< \infty$ for $N = \infty$. This lends itself to a fixpoint perspective on RDEs, where we wish to find a distribution of $\mathcal{S}$ such that

$$\eta = \Phi(\eta) \iff \mathcal{S} \overset{d}{=} g(\xi, \mathcal{S}_i, 0 \le i \le^* N) \quad \text{on } \mathbb{S}. \tag{4}$$

2.5. An independence criterion

To end the Preliminaries section, we introduce an elementary lemma, which will help us verify certain required independences. We leave its proof to the reader.

Lemma 2.9. *Let T be an a.s. finite stopping time with respect to a filtration $(\mathcal{F}_n)_{n \ge 1}$. Suppose that X is a non-negative and bounded random variable satisfying, for each $n \ge 1$,*

$$\mathbb{E}[X \mid \mathcal{F}_T] = \mathbb{E}[X \mid \mathcal{F}_m] \quad a.s.,$$

for all $m \ge n$ on $\{T = n\}$. Then $\mathbb{E}[X|\mathcal{F}_T] = \mathbb{E}[X|\mathcal{F}_\infty]$ a.s. where $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \ge 1)$.

3. An RDE for $\mathbb{R}$-trees from Marchal's algorithm

In this section, fix $\alpha \in (1,2]$ and let $\beta = 1 - 1/\alpha \in (0, 1/2]$. Unless ambiguity arises, we suppress α hereafter. Note that, in Marchal's algorithm, $\mathbf{T}(2)$ is deterministic, comprising a Y-shape with a root and two leaves that we denote by A_0, A_1 and A_2 and an internal vertex denoted by V_2. Denote the edges by $e_0 := [\![A_0, V_2]\!]$, $e_1 := [\![A_1, V_2]\!]$ and $e_2 := [\![A_2, V_2]\!]$. The following paragraph, inspired by the proof of [13, Proposition 10], outlines the argument in this section.

The independent choice at each step of Marchal's algorithm entails that we have independent subconstructions of Marchal's algorithm with parameter α evolving along each edge of $\mathbf{T}(2)$. This yields three independent copies of $\mathcal{T}_\alpha$, denoted by τ_0, τ_1 and τ_2, subject to rescaling depending on the eventual proportion of mass distributed to each tree. For $\alpha \in (1,2)$, the internal vertex V_2 will give rise to a further countably infinite and independent collection of copies of $\mathcal{T}_\alpha$ a.s.. Denote this infinite collection by $(\tau_i, i \geq 3)$, which is independent of τ_0, τ_1 and τ_2. We will rescale and concatenate our collection $(\tau_i, i \geq 0)$ of independent copies of $\mathcal{T}_\alpha$ at V_2 to get a copy of $\mathcal{T}_\alpha$. Denote the collection of scaling factors in the limit by $\xi = (\xi_i, i \geq 0)$ and the concatenation operator by g. We obtain an RDE

$$\mathcal{T}_\alpha \stackrel{d}{=} g\,(\xi, \tau_i, i \geq 0)$$

in the form (4). To be rigorous, we need to address the following questions.

1. What is the distribution of the limiting scaling factors $\xi = (\xi_i, i \geq 0)$?
2. Are the random variables $(\tau_i, i \geq 0)$ independent of ξ, as well as of each other?
3. How do we construct the concatenation operation in a measurable way?

3.1. The scaling factors $\xi = (\xi_i, i \geq 0)$ and subtrees $(\tau_i, i \geq 0)$

For $i \in \{0,1,2\}$ and $n \geq 0$, define $\tau_i^{(n)}$ as the subtree of $\mathbf{T}(n+2)$ cut at V_2 containing the edge e_i. For example, we have $\tau_i^{(0)} = e_i$ for each $i \in \{0,1,2\}$. Let $\mathcal{K}_n$ denote the set of edges incident to V_2 in $\mathbf{T}(n+2)$ excluding $\{e_i, i = 0,1,2\}$, and set $K_n = |\mathcal{K}_n|$. For $\mathcal{K}_n \neq \varnothing$, $\mathcal{K}_n = \{e_j, j = 3, \ldots, K_n + 2\}$, ordered according to least leaf labels. Define $\sigma^{(n)}$ as the remaining components of $\mathbf{T}(n+2)$ cut at V_2 excluding $\bigcup_{i=0}^{2} \tau_i^{(n)}$. If $\mathcal{K}_n = \varnothing$, then $\sigma^{(n)} = \varnothing$. Otherwise, $\sigma^{(n)} = \bigcup_{j=3}^{K_n+2} \tau_j^{(n)}$ is a union of subtrees $\{\tau_j^{(n)}, j = 3, \ldots, K_n + 2\}$ growing along their respective edges in $\mathcal{K}_n$. We illustrate this in Figure 3.

Regard V_2 as a (weightless) root from the perspective of each element of $\{\tau_i^{(n)}, i = 1, \ldots, K_n + 2\}$ and as a marked leaf of $\tau_0^{(n)}$. For each $i = 1, \ldots, K_n + 2$, mark the first leaf created in $\tau_i^{(n)}$ by Marchal's algorithm, that is, the other endpoint of e_i which is not V_2. We treat A_0 as the root of $\tau_0^{(n)}$.

Denote the number of leaves of $\mathbf{T}(n+2)$ in $\tau_i^{(n)}$ by $N_i(n)$ for all $i = 0, 1, \ldots, K_n + 2$, and define its inverse $N_i^{-1}(n) := \inf\{k \geq 0 \colon N_i(k) = n\}$ as the first time k at which $\tau_i^{(k)}$ has n leaves, with the convention $\inf \varnothing = \infty$.

Recall that $W(\cdot)$ measures the total weight of a given sub-structure, e.g., for each $i \in \{0,1,2\}$, $W(\tau_i^{(0)}) = \alpha - 1$. The following result shows that the weight of a particular subtree only depends on the number of leaves it has, and not on its shape.

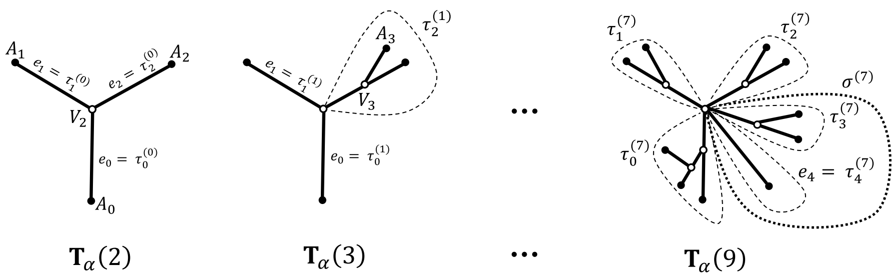

Figure 3. Illustration of Marchal's random growth algorithm and notation employed

Lemma 3.1. *Regardless of its shape, the total weight of the i-th subtree is* $W(\tau_i^{(n)}) = \alpha N_i(n) - 1$ *for* $i = 0, 1, \ldots, K_n + 2$ *and* $n \geq 0$.

Proof. This follows simply by induction applied to each subtree. □

The following result gives the limiting weight partition of subtrees and proves that the subtrees essentially evolve as independent copies at the first branch point.

Proposition 3.2.

(i) *For* $\alpha \in (1,2)$, $K_n \longrightarrow \infty$ *as* $n \longrightarrow \infty$ *a.s.. When* $\alpha = 2$, $K_n = 0$ *a.s. for all* $n \geq 0$.

(ii) *The relative weight split in* $\mathbf{T}_\alpha(n)$ *has an almost sure limit as* $n \longrightarrow \infty$ *given by*

$$\left(\frac{W\left(\tau_0^{(n)}\right)}{(n+2)\alpha - 1}, \frac{W\left(\tau_1^{(n)}\right)}{(n+2)\alpha - 1}, \frac{W\left(\tau_2^{(n)}\right)}{(n+2)\alpha - 1}, \frac{W\left(\sigma^{(n)}\right) + W(\{V_2\})}{(n+2)\alpha - 1} \right) \xrightarrow{a.s.} (X_0, X_1, X_2, X_3), \quad (5)$$

where $(X_0, X_1, X_2, X_3) \sim \mathrm{Dir}(\beta, \beta, \beta, 1 - 2\beta)$, *noting that if* $\alpha = 2$, *no weight is distributed to* X_3. *For* $\alpha \in (1,2)$, *within the last co-ordinate, denote the limiting weight proportion of the subtree* τ_{i+2} *by* P_i *for* $i \geq 1$. *Then,* $(P_i, i \geq 1) \sim \mathrm{GEM}(1 - \beta, 1 - 2\beta)$. *In particular, the subtrees* τ_i, $i \geq 3$, *have a relative weight partition that follows a* $\mathrm{PD}(1 - \beta, 1 - 2\beta)$ *distribution, when ranked in decreasing order.*

(iii) *Let* $\alpha \in (1,2)$, *for* $n \geq 1$ *and* $i \geq 0$, *we have* $\tau_i^{(N_i^{-1}(n))} \overset{d}{=} \mathbf{T}(n)$. *That is, at transition times in which a leaf is added into the i-th subtree, it evolves as Marchal's algorithm with parameter* α *with initial edge* e_i. *The sigma-field generated by*

$$\left(\tau_i^{(N_i^{-1}(n))}, n \geq 1 \right)_{i \geq 0}$$

is independent of the sigma-field generated by $(N_i(n), n \geq 1, i \geq 0)$. *Consequently, the limiting CRTs* τ_i, $i \geq 0$, *are independent. Furthermore,* $(\tau_i, i \geq 0)$ *is independent of* $(N_i(n), n \geq 1, i \geq 0)$. *An analogous result holds when* $\alpha = 2$.

(iv) *For* $\alpha \in (1,2)$, *the random variables* (X_0, X_1, X_2, X_3) *and* $(P_i, i \geq 1)$ *are independent, and independent of* $(\tau_i, i \geq 0)$. *This fully specifies their joint distribution.*

Proof of (i). We prove the result for $\alpha \in (1,2)$ with the $\alpha = 2$ case similarly argued. From Lemma 3.1, conditional on an edge or branch point in $\tau_i^{(n)}$ being selected in the next step of Marchal's algorithm, we increase the weight in $\tau_i^{(n)}$ by α. It is easy to check that this holds for $\sigma^{(n)}$ with one weighted copy of V_2 included. Hence,

$$\left(W\left(\tau_0^{(n)}\right), W\left(\tau_1^{(n)}\right), W\left(\tau_2^{(n)}\right), W\left(\sigma^{(n)}\right) + W(\{V_2\}) \right) \quad (6)$$

evolves precisely as a Pólya urn scheme with 4 colours and weight vector $\vec{\gamma} = (\alpha-1, \alpha-1, \alpha-1, 2-\alpha)$ which increases by α at each update for the chosen colour.

Next, we focus on the subtrees within $\sigma^{(n)}$. The above implies that $W\left(\sigma^{(n)}\right) + W(\{V_2\}) \to \infty$ as $n \to \infty$ a.s.. So, a.s., we observe infinitely many leaves being added to $\left(\sigma^{(n)}, n \geq 1\right)$. We may then condition on the times where a leaf is added to $\left(\sigma^{(n)}, n \geq 1\right)$, say $(q_i, i \geq 1)$, where $1 \leq q_1 < q_2 < \cdots < q_n < q_{n+1} < \cdots$ is an infinite sequence a.s.. Conditional on the preceding event, the first leaf added creates $\tau_3^{(q_1)}$. At each $r \in \{q_n, \dots, q_{n+1} - 1\}$, we have n leaves (not including V_2) with K_{q_n} subtrees whose union is $\sigma^{(r)}$. For $j = 3, \dots, K_{q_n} + 2$, $\tau_j^{(r)}$ has $N_j(q_n)$ leaves (not including V_2), and so has total weight $\alpha N_j(q_n) - 1$, by Lemma 3.1. Thus, as the total weight of V_2 is $2 + K_{q_n} - \alpha$, the total weight of $\sigma^{(r)}$ and $\{V_2\}$ is $\alpha n + (2 - \alpha)$. At the next arrival time q_{n+1}, we add a leaf to $\tau_j^{(q_{n+1}-1)}$ with probability $\left(\alpha N_j(q_n) - 1\right) / (\alpha n + 2 - \alpha)$ and we create a new sub-tree with probability $\left(2 + K_{q_n} - \alpha\right) / (\alpha n + 2 - \alpha)$. Regarding the leaves (excluding V_2) as customers and each subtree as a table, this models a CRP$(1 - \beta, 1 - 2\beta)$. From Proposition 2.8, $K_n \to \infty$ as $n \to \infty$ almost surely. □

Proof of (ii). We continue in the setting introduced in the proof of (i). For the Pólya urn, (5) is well-known to hold [9]. Now let $\alpha \in (1,2)$ and recall that $q_1 < \infty$ almost surely, so we may assume $n \geq q_1$. From Proposition 2.8, we can identify the almost sure limiting proportion of leaves split within subtrees of $\sigma^{(n)}$ as GEM$(1 - \beta, 1 - 2\beta)$ holding along the increasing subsequence $(q_i, i \geq 1)$. That is,

$$\left(\frac{N_3(q_n)}{n}, \dots, \frac{N_{K_{q_n}+2}(q_n)}{n}, 0, 0, \dots\right) \xrightarrow{a.s.} (P_i, i \geq 1) \qquad \text{as } n \to \infty,$$

where $(P_i, i \geq 1) \sim$ GEM$(1 - \beta, 1 - 2\beta)$. Write $N_\sigma(n)$ as the number of leaves in $\sigma^{(n)}$ excluding V_2. Noting that $N_\sigma(n) > 0$ for $n \geq q_1$, we may rephrase the above as

$$\left(\frac{N_3(n)}{N_\sigma(n)}, \dots, \frac{N_{K_n+2}(n)}{N_\sigma(n)}, 0, 0, \dots\right) \xrightarrow{a.s.} (P_i, i \geq 1) \qquad \text{as } n \to \infty. \tag{7}$$

Using the relation $W\left(\sigma^{(n)}\right) + W(\{V_2\}) = \alpha N_\sigma(n) + 2 - \alpha$, and the aggregation property of the Dirichlet distribution applied to (6), we get that

$$\frac{N_\sigma(n)}{n} \xrightarrow{a.s.} X_3 \qquad \text{as } n \to \infty, \tag{8}$$

where $X_3 \sim$ Beta$(1 - 2\beta, 3\beta)$. Applying the algebra of a.s. convergence for all $j = 3, \dots, K_n + 2$ and $n \geq q_1$, and using the fact that $W\left(\tau_j^{(n)}\right) = \alpha N_j(n) - 1$, the results above imply that, jointly in j,

$$\frac{W\left(\tau_j^{(n)}\right)}{(n+2)\alpha - 1} = \frac{\frac{N_j(n)}{n} - \frac{1}{\alpha n}}{\frac{n+2}{n} - \frac{1}{\alpha n}} \xrightarrow{a.s.} X_3 P_{j-2} \qquad \text{as } n \to \infty,$$

where $X_3 \sim$ Beta$(1 - 2\beta, 3\beta)$ and $(P_i, i \geq 1) \sim$ GEM$(1 - \beta, 1 - 2\beta)$. Thus, we have obtained the almost sure limiting weight partition for the subtrees $(\tau_j^{(n)}, j \geq 0)$. □

Proof of (iii). From (i), $K_n \to \infty$ a.s. as $n \to \infty$. In particular, for all $i \geq 0$ and $n \geq 1$, $N_i^{-1}(n) < \infty$ a.s.. We assume this holds henceforth. It suffices to show the independence of the sigma-fields generated by

$$(N_i(n), n \geq 1, i \geq 0) \qquad \text{and} \qquad \left(\tau_i^{(N_i^{-1}(n))}, n \geq 1\right)_{0 \leq i \leq m+2},$$

respectively, where $m \geq 0$ is arbitrary but fixed.

Since $N_i^{-1}(n) < \infty$ a.s., the distributional identity $\tau_i^{(N_i^{-1}(n))} \overset{d}{=} \mathbf{T}(n)$ in the i-th subtree for $0 \leq i \leq m+2$ holds by virtue of Marchal's algorithm.

Let $M > 1$ be arbitrary, but fixed, and denote the natural filtration of $(N_i(n), i \geq 0)_{n\geq 1}$ by $(\mathcal{F}_n)_{n\geq 1}$. Note that for any fixed $n \geq 1$, $(N_i(n), i \geq 0)$ is almost surely a vector with finitely many non-trivial entries. Define $T := \max_{i=0,\dots,m+2} N_i^{-1}(M)$, which is a stopping time with respect to $(\mathcal{F}_n)_{n\geq 1}$. By assumption, $T < \infty$ a.s.. Conditional on $\mathcal{F}_T$ (which is the same as conditioning on relative weights on subtrees until time T), we have factorisation of tree shape probabilities into tree shape probabilities for the respective subtrees cut at V_2. In particular, given $\mathcal{F}_T$, the tree shapes of $\big(\tau_i^{(N_i^{-1}(n))}, 1 \leq n \leq M\big)_{0\leq i\leq m+2}$ are independent. Furthermore, on the event $\{T = t\}$, conditioning on the sigma-field generated at a later time $k \geq t$ does not affect the tree shapes under consideration. Hence, the hypotheses in Lemma 2.9 are fulfilled. Let $\mathbf{t}_i^{(n)}$ be some given leaf-labelled trees with n leaves and a root. Then,

$$\begin{aligned}
&\mathbb{P}\left(\tau_i^{(N_i^{-1}(n))} = \mathbf{t}_i^{(n)}, 1 \leq n \leq M, 0 \leq i \leq m+2 \,\middle|\, \mathcal{F}_\infty\right) \\
&= \mathbb{P}\left(\tau_i^{(N_i^{-1}(n))} = \mathbf{t}_i^{(n)}, 1 \leq n \leq M, 0 \leq i \leq m+2 \,\middle|\, \mathcal{F}_T\right) = \mathbb{P}\left(\tau_i^{(N_i^{-1}(M))} = \mathbf{t}_i^{(M)}, 0 \leq i \leq m+2 \,\middle|\, \mathcal{F}_T\right) \quad (9)\\
&= \prod_{i=0}^{m+2} \mathbb{P}\left(\tau_i^{(N_i^{-1}(M))} = \mathbf{t}_i^{(M)} \,\middle|\, \mathcal{F}_T\right) = \prod_{i=0}^{m+2} \mathbb{P}\left(\mathbf{T}(M) = \mathbf{t}_i^{(M)} \,\middle|\, \mathcal{F}_T\right) = \prod_{i=0}^{m+2} \mathbb{P}\left(\mathbf{T}(M) = \mathbf{t}_i^{(M)}\right), \quad (10)
\end{aligned}$$

where (9) holds by Proposition 2.4 and since $\tau_i^{(N_i^{-1}(M))}$ determines $\tau_i^{(N_i^{-1}(n))}$ for all $1 \leq n \leq M$. (10) follows since there is no dependence on $\mathcal{F}_\infty$ in the final expression and we are conditioning over an almost surely finite number of discrete random variables. Furthermore, this implies that the sigma-field of $\big(\tau_i^{(N_i^{-1}(n))}, 1 \leq n \leq M\big)_{0\leq i\leq m+2}$ is independent of $\mathcal{F}_\infty$. Letting $M \to \infty$, and recalling $m \geq 0$ is arbitrary, the claimed independence of the sigma-fields follows.

As the collection $(\tau_i, i \geq 0)$ is measurably constructed from $\big(\tau_i^{(N_i^{-1}(n))}, n \geq 1\big)_{i\geq 0}$, it is independent of $(N_i(n), n \geq 1, i \geq 0)$. Dropping the conditioning in (10), we get that $\big(\tau_i^{(N_i^{-1}(n))}, 1 \leq n \leq M\big), 0 \leq i \leq m+2$, are independent. Thus, in the limit as $M \to \infty$, $\big(\tau_i^{(N_i^{-1}(n))}, n \geq 1\big)$, $0 \leq i \leq m+2$, are independent. As τ_i is measurably constructed from $\big(\tau_i^{(N_i^{-1}(n))}, n \geq 1\big)$ for each $0 \leq i \leq m+2$, $\big(\tau_i, 0 \leq i \leq m+2\big)$ are independent. Let $m \to \infty$ to conclude that $(\tau_i, i \geq 0)$ are independent. □

Proof of (iv). Recall that $N_\sigma(n)$ denotes the number of leaves (excluding V_2) in $\sigma^{(n)}$, and define $N_\sigma^{-1}(n) := \inf\{k \geq 0 \colon N_\sigma(k) = n\}$, which is a.s. finite by (i) since $\sigma^{(n)}$ has at least K_n leaves. Let $(\mathcal{F}_n)_{n\geq 1}$ denote the natural filtration of $\big(N_0^{-1}(n), N_1^{-1}(n), N_2^{-1}(n), N_\sigma^{-1}(n)\big)_{n\geq 1}$. Conditional on a leaf being added to $\sigma^{(n)}$ at time $n+1$, the process of adding leaves to each subtree within $\sigma^{(n)}$ is modelled by a CRP $(1-\beta, 1-2\beta)$ and does not depend on the times at which the leaf is added. By a similar argument to (iii) and applying Lemma 2.9, for all non-negative integers $l_j^{(n)}$,

$$\mathbb{P}\left(N_j\left(N_\sigma^{-1}(n)\right) = l_j^{(n)}, 1 \leq n \leq M, 3 \leq j \leq m \,\middle|\, \mathcal{F}_\infty\right) = \mathbb{P}\left(N_j\left(N_\sigma^{-1}(n)\right) = l_j^{(n)}, 1 \leq n \leq M, 3 \leq j \leq m\right).$$

This implies that the sigma-field generated by $\big(N_j\big(N_\sigma^{-1}(n)\big), 1 \leq n \leq M\big)_{3\leq j\leq m}$ is independent of $\mathcal{F}_\infty$ for all $M \geq 1$. Let $M \to \infty$ to conclude that the sigma-field generated by $\big(N_j\big(N_\sigma^{-1}(n)\big), n \geq 1\big)_{3\leq j\leq m}$ is independent of $\mathcal{F}_\infty$. We may rewrite equations (7) and (8) to get

$$\frac{N_{i+2}\left(N_\sigma^{-1}(n)\right)}{n} \xrightarrow{a.s.} P_i \quad \text{and} \quad \frac{n}{N_\sigma^{-1}(n)} \xrightarrow{a.s.} X_3 \quad \text{as } n \to \infty.$$

So, $(P_i, i \geq 1)$ are measurable with respect to the sigma-field generated by $\left(N_j\left(N_\sigma^{-1}(n)\right), n \geq 1\right)_{j \geq 3}$. Also, X_3 is $\mathcal{F}_\infty$-measurable, similarly for X_0, X_1 and X_2. The desired result follows. □

We finally obtain the main result regarding the self-similarity of Marchal's algorithm, which proves the self-similarity property of limiting α-stable tree when decomposed at the "first" branch point.

Theorem 3.3. *For any $\alpha \in (1,2]$, the limiting trees $(\tau_i, i \geq 0)$ in Marchal's algorithm are independent. Furthermore, they are independent of their scaling factors. For each subtree $\tau_i^{(n)}$, let $d_i^{(n)}$ denote the graph distance and $\mu_i^{(n)}$ the empirical mass measure on its leaves. For each $i \geq 0$, we have*

$$\left(\tau_i^{(n)}, \frac{d_i^{(n)}}{\alpha n^\beta}, \mu_i^{(n)}\right) \xrightarrow{a.s.} \left(\tau_i, \xi_i^\beta d_i, \xi_i \mu_i\right) \qquad \text{as } n \to \infty,$$

in the Gromov–Hausdorff–Prokhorov topology, where $(\tau_i, d_i, \mu_i), i \geq 1$, are i.i.d. with

$$(\tau_i, d_i, \mu_i) \stackrel{d}{=} (\mathcal{T}_\alpha, d_\alpha, \mu_\alpha), \quad i \geq 0,$$

and $\xi_i = X_i$ for $i \in \{0,1,2\}$ and, for $\alpha \in (1,2)$, $\xi_{j+2} = X_3 P_j$ for $j \geq 1$. For $\alpha = 2$, $(\xi_0, \xi_1, \xi_2) \sim \mathrm{Dir}(1/2, 1/2, 1/2)$. Otherwise, for $\alpha \in (1,2)$, (X_0, X_1, X_2, X_3) and $(P_i, i \geq 1)^\downarrow$ are independent with $(X_0, X_1, X_2, X_3) \sim \mathrm{Dir}(\beta, \beta, \beta, 1-2\beta)$ and $(P_i, i \geq 1)^\downarrow \sim \mathrm{PD}(1-\beta, 1-2\beta)$.

Proof. The almost sure convergence in the rescaled subtrees arises by applying Proposition 2.5 and Proposition 3.2(iii). The independence between the limiting subtrees comes immediately from Proposition 3.2(iii). The arguments in Lemma 3.1 and Proposition 3.2(ii) show that the limiting proportion of weights is measurably constructed from $\left(N_i(n), n \geq 1, i \geq 0\right)$. Hence, by Proposition 3.2(iii), the limiting subtrees are independent of their scaling factors. The representation of $(\xi_i, i \geq 0)$ and hence its distribution are a consequence of Proposition 3.2(ii) and Proposition 3.2(iv). □

The results of Theorem 3.3 agree with similar decompositions of the BCRT at a branch point in Aldous [5, Theorem 2], Albenque and Goldschmidt [3, Section 1.4], and Croydon and Hambly [12, Lemma 6], where the branch point is uniquely determined by a uniformly chosen point according to the mass measure within each of the three subtrees. We point out that Albenque and Goldschmidt deal with an unrooted BCRT, while Croydon and Hambly's construction uses a doubly-marked rooted BCRT. Our construction thus far does not require a notion of a mass measure (even though we have chosen to include the mass measure in our statements), but rather a single marked point in each subtree.

3.2. Formal specification of the concatenation operation

After verifying that the subtrees $(\tau_i, i \geq 0)$ are rescaled versions of $\mathcal{T}_\alpha$ in the limit with the required independences, the next step is to show that the concatenation operation induced by Marchal's algorithm

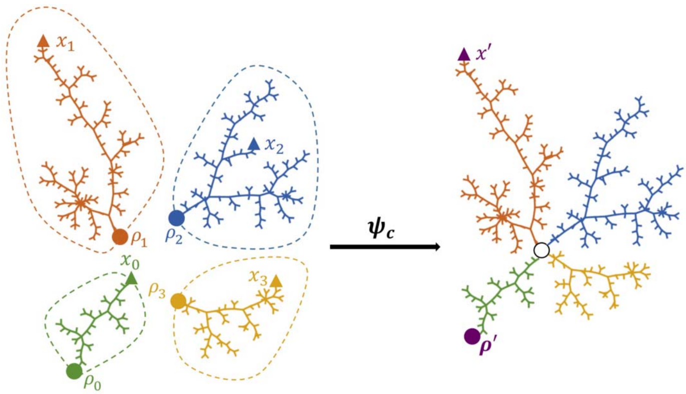

Figure 4. Construction of concatenated tree from 4 marked trees, rescaling is not shown.

is well-defined and measurable as an operation on $\mathbb{T}_\mathrm{m}$. We adapt the general setup and terminology from [24]. Let

$$\Xi := \left\{(x_0, x_1, x_2, x_3 p_j, j \geq 1): \; x_0, x_1, x_2, x_3 \geq 0, \sum_{i=0}^{3} x_i = 1, \, p_1 \geq p_2 \geq \cdots \geq 0, \sum_{j=1}^{\infty} p_j = 1\right\}.$$

For notational convenience, we write $\xi_i = \begin{cases} x_i & \text{if } i \in \{0,1,2\}, \\ x_3 p_{i-2} & \text{otherwise.} \end{cases}$

Set $\Xi^* := \Xi \times \mathbb{T}_\mathrm{m}^\infty$, as in (2), which is a Polish space since it is a product of Polish spaces. We formally define our concatenation operator. Let $\xi \in \Xi$ and let $(\tau_i, d_i, \rho_i, x_i)$ be representatives of GH^m-equivalence classes in $\mathbb{T}_\mathrm{m}$ for $i \geq 0$. Define the *concatenated tree* (τ', d', ρ', x') as follows.

1. Let $\widetilde{\tau}' := \coprod_{i \geq 0} \tau_i$ be the disjoint union of trees. Let $\sim_c$ be the equivalence relation on $\widetilde{\tau}'$ in which $\rho_i \sim_c x_0$ for all $i \geq 1$. Define $\tau' := \widetilde{\tau}' / \sim_c$. Write ψ_c for the canonical projection from $\widetilde{\tau}'$ onto τ'.
2. Define d' as the metric induced on τ' under ψ_c by the metric $\widetilde{d'}$ on $\widetilde{\tau}'$ such that

$$\widetilde{d'}(u,v) = \begin{cases} \xi_i^\beta d_i(u,v) & \text{if } u,v \in \tau_i, i \geq 0, \\ \xi_0^\beta d_0(u,x_0) + \xi_j^\beta d_j(\rho_j, v) & \text{if } u \in \tau_0 \text{ and } v \in \tau_j, j \neq 0, \\ \xi_i^\beta d_i(u,\rho_i) + \xi_0^\beta d_0(x_0, v) & \text{if } u \in \tau_i \text{ and } v \in \tau_0, i \neq 0, \\ \xi_i^\beta d_i(u,\rho_i) + \xi_j^\beta d_j(\rho_j, v) & \text{if } u \in \tau_i \text{ and } v \in \tau_j, i,j \neq 0. \end{cases} \tag{11}$$

3. Retain $x' = \psi_c(x_1)$ as our marked point in τ' and set $\rho' = \psi_c(\rho_0)$ as the root of τ'.

We illustrate this construction in Figure 4.

By virtue of this construction, the GH^m-equivalence class of (τ', d', ρ', x') only depends on the GH^m-equivalence classes of $(\tau_i, d_i, \rho_i, x_i)$ for $i \geq 0$. Thus, it makes sense to define $C_\beta \subseteq \Xi^*$ as the set of elements $\kappa = (\xi, \tau_i, i \geq 0) \in \Xi^*$ such that the concatenated tree (τ', d', ρ', x') formed by any equivalence class representatives of $((\tau_i, d_i, \rho_i, x_i), i \geq 0)$ is compact. Equip $\mathbb{T}_\mathrm{m}$ and Ξ^* with their respective Borel sigma-algebras, $\mathcal{B}(\mathbb{T}_\mathrm{m})$ and $\mathcal{B}(\Xi^*)$. The *concatenation operator* $g_\beta \colon \Xi^* \to \mathbb{T}_\mathrm{m}$ is,

$$g_\beta(\kappa) = \begin{cases} (\tau', d', \rho', x') & \text{if } \kappa \in C_\beta, \\ (\{x'\}, 0, x', x') & \text{otherwise,} \end{cases} \tag{12}$$

where $(\{x'\}, 0, x', x')$ denotes the equivalence class of a trivial one-point rooted tree.

Proposition 3.4. *The map $g_\beta\colon \Xi^* \to \mathbb{T}_\mathrm{m}$ is $\mathcal{B}(\Xi^*)$-measurable.*

Proof. The proof can be adapted from [24, Proposition 3.2]. □

3.3. First main result: The RDE satisfied by the stable tree

We now deduce our main theorem in this section.

Theorem 3.5. *The marked α-stable tree $(\mathcal{T}_\alpha, d_\alpha, \mu_\alpha, x_\alpha)$ with $x_\alpha \sim \mu_\alpha$ satisfies the RDE*

$$\mathcal{T}_\alpha \overset{d}{=} g_\beta\,(\xi, \mathcal{T}_i, i \geq 0) \tag{13}$$

on $\mathbb{T}_\mathrm{m}$, where $(\mathcal{T}_i, i \geq 0)$ is a sequence of independent copies of $\mathcal{T}_\alpha$, independent of the sequence $\xi = \big(X_0, X_1, X_2, X_3 P_j, j \geq 1\big) \in \Xi$, and where the following holds.

- *If $\alpha = 2$, then $\xi_{j+2} = X_3 P_j = 0$ almost surely for all $j \geq 1$ and $(X_0, X_1, X_2) \sim \mathrm{Dir}\,(1/2, 1/2, 1/2)$.*
- *If $\alpha \in (1,2)$, then $(X_0, X_1, X_2, X_3) \sim \mathrm{Dir}\,(\beta, \beta, \beta, 1-2\beta)$ and $(P_j, j \geq 1) \sim \mathrm{PD}(1-\beta, 1-2\beta)$, where (X_0, X_1, X_2, X_3) and $(P_j, j \geq 1)$ are independent.*

In other words, the law of the marked α-stable tree $\varsigma_\alpha^\mathrm{m}$ satisfies the fixpoint equation $\eta = \Phi_\beta\,(\eta)$ on $\mathcal{P}(\mathbb{T}_\mathrm{m})$, where $\Phi_\beta\colon \mathcal{P}\,(\mathbb{T}_\mathrm{m}) \to \mathcal{P}\,(\mathbb{T}_\mathrm{m})$ is the mapping on $\mathcal{P}(\mathbb{T}_\mathrm{m})$ induced by (13), and where we recall that $\mathcal{P}(\mathbb{T}_\mathrm{m})$ denotes the set of Borel probability measures on $\mathbb{T}_\mathrm{m}$.

Proof. Recall that for the subtrees involved in the recursive application of Marchal's algorithm, we regarded V_2 as a root and marked the first leaf in the i-th subtree for each $i \geq 1$. We regarded A_0 as the root for the overall tree, and V_2 as a marked leaf for the 0-th subtree. Thus, our construction using Marchal's algorithm agrees with the concatenation operator g_β acting on the subtrees. Theorem 3.3 gives the required independences and the distribution of $\xi = (\xi_i, i \geq 0)$. Proposition 3.4 ascertains the measurability of g_β. □

In general, the marked α-stable tree is not the only fixpoint of (13). If the metrics d_i of (the representatives of) $(\tau_i, d_i, \rho_i, x_i) \in \mathbb{T}_\mathrm{m}$ in (11) were multiplied by some constant $c > 0$, then the concatenated tree will also have its metric d' multiplied by c. Furthermore, if the original concatenated tree were a marked compact rooted $\mathbb{R}$-tree, then so would the concatenated tree with metric multiplied by c. Thus, since $(\mathcal{T}_\alpha, d_\alpha, \rho_\alpha, x_\alpha)$ is a distributional fixpoint of (13), so is $(\mathcal{T}_\alpha, c d_\alpha, \rho_\alpha, x_\alpha)$ for any $c > 0$.

Remark 3.6. There also exist solutions to RDE (13) with infinite $1/\beta$-th height moment. This can be shown by grafting mass-less length-y branches onto a stable tree with intensity proportional to $y^{-1-1/\beta} dy \mu(dx)$, see e.g. [10] and [3] for such constructions in the context of related RDEs with finite concatenation operations – the arguments there are not affected by the change of setting here. We will establish uniqueness of the solution to (13) up to multiplication of distances by a constant, under suitable constraints on height moments.

4. Uniqueness and attraction for a general RDE on $\mathbb{T}_\mathbf{m}$

4.1. Recursive tree frameworks (RTF) and recursive tree processes (RTP)

In a *recursive tree framework*, the approach of Section 2.4 is extended recursively to $\mathcal{S}_i, i \geq 1$, and beyond. To this end, we will work with the Ulam–Harris-indexation of integer words

$$\mathbb{U} := \bigcup_{n\geq 0} \mathbb{N}^n, \qquad \text{where } \mathbb{N} := \{0,1,2,\dots\}.$$

Consider a sequence of i.i.d. $\Theta \times \overline{\mathbb{N}}$-valued random variables $(\xi_\mathbf{u}, N_\mathbf{u}), \mathbf{u} \in \mathbb{U}$. Furthermore, suppose that there are random variables $\tau_\mathbf{u}, \mathbf{u} \in \mathbb{U}$, possibly on an extended probability space, as follows.

(i) For all $\mathbf{u} \in \mathbb{U}$,

$$\tau_\mathbf{u} = g\left(\xi_\mathbf{u}, \tau_{\mathbf{u}j}, 1 \leq j \leq^* N_\mathbf{u}\right) \qquad \text{a.s..} \tag{14}$$

(ii) The variables $(\tau_\mathbf{u}, \mathbf{u} \in \mathbb{N}^n)$ are i.i.d. with some distribution η_n, $n \geq 1$.
(iii) The variables $(\tau_\mathbf{u}, \mathbf{u} \in \mathbb{N}^n)$ are independent of the variables $(\xi_\mathbf{u}, N_\mathbf{u}, \mathbf{u} \in \bigcup_{k=0}^n \mathbb{N}^k)$.

In this setup, we may define a *recursive tree framework* as follows.

Definition 4.1 (Recursive tree framework). A pair $((\xi_\mathbf{u}, N_\mathbf{u}, \mathbf{u} \in \mathbb{U}), g)$ is called a *recursive tree framework* if $(\xi_\mathbf{u}, N_\mathbf{u}, \mathbf{u} \in \mathbb{U})$ is an i.i.d. family of $\Theta \times \overline{\mathbb{N}}$-valued random variables $(\xi_\mathbf{u}, N_\mathbf{u}) \sim \nu, \mathbf{u} \in \mathbb{U}$, and $g : \Theta^* \to \mathbb{T}$ is a measurable map.

Enriching an RTF with the random variables $\tau_\mathbf{u}, \mathbf{u} \in \mathbb{U}$ gives us a so-called *recursive tree process* (RTP). Sometimes, RTPs are only considered up to generation n, that is, only for $\tau_\mathbf{u}, \mathbf{u} \in \bigcup_{k=0}^n \mathbb{N}^k$. We then speak of an *RTP of depth n*. Finite-depth RTPs can always be defined for any distribution η_n of $\tau_\mathbf{u}, \mathbf{u} \in \mathbb{N}^n$, by use of (14) with i.i.d. $\xi_\mathbf{u}$, $\mathbf{u} \in \bigcup_{k=0}^{n-1} \mathbb{N}^k$, also independent of the i.i.d. family $\left(\tau_\mathbf{u}, \mathbf{u} \in \mathbb{N}^n\right)$.

4.2. Second main result: Uniqueness and attraction for the new RDE

We now turn to the uniqueness and attraction of the fixpoints in (13). By Theorem 3.5 and Remark 3.6, uniqueness will only hold up to multiplication by a constant and under additional moment conditions on tree heights in the specific setting of Theorem 3.5. As our setup works for more general $\xi \in \Xi$, we will broaden our scope and establish a uniqueness theorem for the RDE (13) in a setting that keeps the same concatenation operation g_β of Section 3.2, but allows a more general distribution of ξ.

It will be useful to work in the a recursive tree framework, as defined in Section 4.1. Let us consider an i.i.d. family of sequences of scaling factors $(\xi_{\mathbf{u}i}, i \geq 0), \mathbf{u} \in \mathbb{U}$, with some distribution ν on Ξ, where we recall the Ulam–Harris notation $\mathbb{U} = \bigcup_{n\geq 0} \mathbb{N}^n$. Then $(((\xi_{\mathbf{u}i}, i \geq 0), \mathbf{u} \in \mathbb{U}), g_\beta)$ is a recursive tree framework.

For $n \geq 1$ and $\eta \in \mathcal{P}(\mathbb{T}_\mathrm{m})$, we would like to study the distribution $\Phi_\beta^n(\eta)$ of $\mathcal{T}_n := \tau_\varnothing^{(n)}$, where

$$\tau_\mathbf{u}^{(n)} := g_\beta\left((\xi_{\mathbf{u}i}, i \geq 0), \left(\tau_{\mathbf{u}i}^{(n)}, i \geq 0\right)\right), \quad \mathbf{u} \in \mathbb{N}^k, \quad k = n-1, \dots, 0, \tag{15}$$

for $\tau_{\mathbf{u}i}^{(n)} \sim \eta, i \geq 0$, $\mathbf{u} \in \mathbb{N}^{n-1}$, i.i.d.. We also set $\mathcal{T}_0 := \tau_\varnothing^{(0)} \sim \eta$. Note that this setup induces a recursive tree process of depth n, for any $n \geq 0$ and $\eta \in \mathcal{P}(\mathbb{T}_\mathrm{m})$.

Furthermore, let $\mathcal{P}_\infty(\mathbb{T}_m) \subset \mathcal{P}(\mathbb{T}_m)$ be defined as

$$\mathcal{P}_\infty(\mathbb{T}_m) := \{\eta \in \mathcal{P}(\mathbb{T}_m) : \mathbb{E}[\mathrm{ht}(\mathcal{T})^p] < \infty \text{ for all } p > 0 \text{ where } (\mathcal{T}, d, \rho, x) \sim \eta\}.$$

Our main result in this section is as follows.

Theorem 4.2. *For any Ξ-valued random variable $\xi = (\xi_i, i \geq 0)$ such that $\mathbb{P}(\xi_0 + \xi_1 < 1) = 1$ and $\mathbb{P}(\xi_0 > 0, \xi_1 > 0) = 1$, choose $\beta \in (0,1)$ such that $\mathbb{E}[\xi_0^\beta + \xi_1^\beta] = 1$. Then, for any $\eta \in \mathcal{P}_\infty(\mathbb{T}_m)$ with $h := \mathbb{E}[d(\rho, x)]$ for $(\mathcal{T}, d, \rho, x) \sim \eta$,*

$$\Phi_\beta^n(\eta) \to \eta_h^* \text{ weakly as } n \to \infty,$$

where η_h^* is the unique fixpoint of Φ_β in $\mathcal{P}_\infty(\mathbb{T}_m)$ with $\mathbb{E}[d^*(\rho^*, x^*)] = h$ for $(\mathcal{T}^*, d^*, \rho^*, x^*) \sim \eta_h^*$.

Note that the function $f : [0,1] \to (0, \infty)$, $\beta \mapsto \mathbb{E}[\xi_0^\beta + \xi_1^\beta]$ is continuous with $f(0) = 2$ and $f(1) < 1$ when $\mathbb{P}(\xi_0 + \xi_1 < 1) = 1$. Hence, there is always some $\beta \in (0,1)$ such that $f(\beta) = 1$ in the situation of Theorem 4.2.

The uniqueness and attraction of the marked α-stable tree in (13) is a direct consequence of Theorem 4.2.

Corollary 4.3. *Let $\alpha \in (1,2]$ and $\beta = 1 - 1/\alpha$. Furthermore, let $\xi = (X_0, X_1, X_2, X_3 P_j, j \geq 1)$ for independent $(X_0, X_1, X_2, X_3) \sim \mathrm{Dir}(\beta, \beta, \beta, 1 - 2\beta)$ and $(P_j, j \geq 1) \sim \mathrm{PD}(1 - \beta, 1 - 2\beta)$. Then the law $\varsigma_\alpha^{\mathrm{m}}$ of the marked α-stable tree is the unique fixpoint of Φ_β on $\mathcal{P}_\infty(\mathbb{T}_m)$ with $\mathbb{E}[d(\rho, x)] = \alpha\Gamma(\beta)/\Gamma(2\beta)$ for $(\mathcal{T}, d, \rho, x) \sim \eta$, $\eta \in \mathcal{P}_\infty(\mathbb{T}_m)$. Furthermore, for any $\eta \in \mathcal{P}_\infty(\mathbb{T}_m)$ with $\mathbb{E}[d(\rho, x)] = h$ for $(\mathcal{T}, d, \rho, x) \sim \eta$, we have*

$$\Phi_\beta^n(\eta) \to \varsigma_{\alpha,h}^{\mathrm{m}} \text{ weakly as } n \to \infty,$$

where $\varsigma_{\alpha,h}^{\mathrm{m}}$ denotes the distribution of the marked α-stable tree with distances scaled by the factor $h\Gamma(2\beta)/(\alpha\Gamma(\beta))$.

Proof. Apply Theorem 4.2 with the specific distribution for ξ, and $\beta = 1 - 1/\alpha \in (0, 1/2]$. Furthermore, recall from Theorem 3.5 that the marked α-stable tree is a fixpoint of the resulting RDE, is well-known to have height moments of all orders (e.g. from its construction via Theorem 4.6), and from Section 2.3 that the distance between the root and a uniformly sampled leaf of the α-stable tree has distribution $\mathrm{ML}(\beta, \beta)$ scaled by α, which has mean $\alpha\Gamma(\beta)/\Gamma(2\beta)$ by (1). □

To prove Theorem 4.2, we first focus on the case when η is supported on the space of probability measures on *trivial* trees, that is, single branch trees with a root and exactly one leaf (which is marked). We further require that the length of such a tree has moments of orders $p > 0$. Specifically, we consider

$$\mathbb{T}_m^{\mathrm{tr}} := \{(\mathcal{T}, d, 0, y) \in \mathbb{T}_m : \mathcal{T} = [\![0, y]\!], y > 0\}.$$

For most of the proof, we will work in the special case of $\mathbb{T}_m^{\mathrm{tr}}$-valued initial distributions:

Assumption (A). $\eta \in \mathcal{P}_\infty\left(\mathbb{T}_m^{\mathrm{tr}}\right) := \left\{\eta \in \mathcal{P}\left(\mathbb{T}_m^{\mathrm{tr}}\right) : \mathbb{E}[(\mathrm{ht}(\mathcal{T}))^p] < \infty \text{ for all } p > 0 \text{ where } \mathcal{T} \sim \eta\right\}$.

Under Assumption (A), we will show the convergence of scaling factors on the spine from the root to the marked point in the RDE (Section 4.3), the convergence of subtrees spanned by leaves up to recursion depth k (Section 4.4), the CRT limit as $k \to \infty$ (Section 4.6) and establish that the RDE is attractive, pulling threads together via a tightness argument (Section 4.7). We finally strengthen this to lift Assumption (A) and complete the proof of Theorem 4.2.

For the remainder of this section, we write $(\mathcal{T}_n, n \geq 0)$ for the sequence of trees constructed around (15) from i.i.d. $\tau_{\mathbf{u}}^{(n)} \sim \eta$, $\mathbf{u} \in \mathbb{N}^n$, $n \geq 0$, for some $\eta \in \mathcal{P}(\mathbb{T}_{\mathrm{m}})$. We write $Y_{\mathbf{u}} := \mathrm{ht}(\tau_{\mathbf{u}}^{(n)})$, $\mathbf{u} \in \mathbb{N}^n$, $n \geq 0$.

4.3. Scaling factors on the spine from the root to the marked point in the RDE

We first study an $\mathcal{L}^p$-bounded martingale in the general setting of Theorem 4.2, which arises from the scaling factors that are relevant for the spine from the root to the marked point under $\Phi_\beta^n(\eta)$, $n \geq 0$.

Lemma 4.4. *Let ξ be a Ξ-valued random variable with $\mathbb{P}(\xi_0 > 0, \xi_1 > 0) = 1$. Let $\beta \in (0,1]$ such that $\mathbb{E}[\xi_0^\beta + \xi_1^\beta] = 1$, let $(\xi_{\mathbf{u}j}, j \geq 0), \mathbf{u} \in \mathbb{U}$, be i.i.d. with the same distribution as ξ, and define $\overline{\xi}_\varnothing := 1$ and*

$$\overline{\xi}_{\mathbf{u}} := \xi_{u_1}\xi_{u_1u_2}\cdots\xi_{u_1\dots u_n}, \quad \mathbf{u} = u_1\dots u_n \in \mathbb{N}^n, \quad n \geq 1. \tag{16}$$

Then the process

$$L_n = \sum_{\mathbf{u}\in\{0,1\}^n} \overline{\xi}_{\mathbf{u}}^\beta, \quad n \geq 0, \tag{17}$$

is a mean-1 martingale that converges a.s. and in $\mathcal{L}^p$ for all $p \geq 1$.

Proof. It is straightforward to show that $(L_n, n \geq 0)$ is a martingale, so we focus on the $\mathcal{L}^p$-boundedness. Indeed, for $p = 1$, we have for all $n \geq 1$,

$$\mathbb{E}[L_n] = \sum_{\mathbf{u}\in\{0,1\}^n} \mathbb{E}\left[\xi_{u_1}^\beta\right]\cdots\mathbb{E}\left[\xi_{u_1\dots u_n}^\beta\right] = \left(\mathbb{E}\left[\xi_0^\beta\right] + \mathbb{E}\left[\xi_1^\beta\right]\right)^n = 1.$$

Inductively, if for all $j \leq p-1$ and $n \geq 1$, we have $\mathbb{E}[L_n^j] \leq f(j)$, then for all $n \geq 1$,

$$\begin{aligned}
\mathbb{E}\left[L_n^p\right] &= \sum_{\mathbf{u}^{(1)},\dots,\mathbf{u}^{(p)}\in\{0,1\}^n} \mathbb{E}\left[\overline{\xi}_{\mathbf{u}^{(1)}}^\beta\cdots\overline{\xi}_{\mathbf{u}^{(p)}}^\beta\right] \\
&= \sum_{\mathbf{v}\in\{0,1\}^n}\mathbb{E}\left[\overline{\xi}_{\mathbf{v}}^{p\beta}\right] + \sum_{k=0}^{n-1}\sum_{\mathbf{v}\in\{0,1\}^k}\mathbb{E}\left[\overline{\xi}_{\mathbf{v}}^{p\beta}\right]\sum_{j=1}^{p-1}\binom{p}{j}\mathbb{E}\left[\xi_0^{j\beta}\xi_1^{(p-j)\beta}\right] \\
&\quad\times\mathbb{E}\left[\sum_{\substack{\mathbf{w}^{(1)},\dots,\mathbf{w}^{(j)}\\ \in\{0,1\}^{n-k-1}}}\left(\overline{\xi}_{\mathbf{w}^{(1)}}^\beta\cdots\overline{\xi}_{\mathbf{w}^{(j)}}^\beta\right)\right]\mathbb{E}\left[\sum_{\substack{\mathbf{w}^{(j+1)},\dots,\mathbf{w}^{(p)}\\ \in\{0,1\}^{n-k-1}}}\left(\overline{\xi}_{\mathbf{w}^{(j+1)}}^\beta\cdots\overline{\xi}_{\mathbf{w}^{(p)}}^\beta\right)\right].
\end{aligned}$$

Specifically, we split the sum over $\mathbf{u}^{(1)},\dots,\mathbf{u}^{(p)}$ according to the number k of initial entries that are common to all $\mathbf{u}^{(1)},\dots,\mathbf{u}^{(p)}$ and according to the number j of entries in the $(k+1)$-st place of $\mathbf{u}^{(1)},\dots,\mathbf{u}^{(p)}$ that equal 0. For each k and j, there are $\binom{p}{j}$ ways to choose which j they are. By symmetry, the contribution is the same as if they are $1,\dots,j$, so that we write the sum as a sum over

$$\mathbf{u}^{(1)} = \mathbf{v}0\mathbf{w}^{(1)},\dots,\mathbf{u}^{(j)} = \mathbf{v}0\mathbf{w}^{(j)},\mathbf{u}^{(j+1)} = \mathbf{v}1\mathbf{w}^{(j+1)},\dots,\mathbf{u}^{(p)} = \mathbf{v}1\mathbf{w}^{(p)}.$$

By the induction hypothesis, we can further bound $\mathbb{E}[L_n^p]$ above by

$$\sum_{k=0}^{n}\sum_{\mathbf{v}\in\{0,1\}^k}\mathbb{E}\left[\overline{\xi}_{\mathbf{v}}^{p\beta}\right]\sum_{j=1}^{p-1}\binom{p}{j}f(j)f(p-j)\leq\left(\sum_{j=1}^{p-1}\binom{p}{j}f(j)f(p-j)\right)\frac{1}{1-\mathbb{E}\left[\xi_0^{p\beta}+\xi_1^{p\beta}\right]}=:f(p),$$

and this is finite. An application of the Martingale Convergence Theorem completes the proof. □

4.4. Convergence of subtrees spanned by leaves up to depth k

In the following, we associate a notion of *depth* with every leaf of the trees $\mathcal{T}_n := \tau_\varnothing^{(n)}$, $n \geq 1$, of (15), so as to give a meaning to "the subtree of $\mathcal{T}_n$ spanned by the leaves up to depth k". Let us do this under Assumption (A). Recall that $\mathcal{T}_n$, for $n \geq 1$, is built recursively by scaling and concatenating trivial trees $\tau_{\mathbf{u}}^{(n)}$, $\mathbf{u} \in \mathbb{N}^n$, in such a way that the unique leaf of $\tau_{\mathbf{u}}^{(n)}$ gives rise to a leaf (if $u_1,\ldots,u_n \in \mathbb{N}\setminus\{0\}$) or branch point (otherwise) of $\mathcal{T}_n$, which we denote by $\Sigma_{n,\mathbf{u}} \in \mathcal{T}_n$, $\mathbf{u} \in \mathbb{N}^n$. We say $\Sigma_{n,\mathbf{u}}$ has depth $k \in \{1,\ldots,n\}$ if $\mathbf{u} = u_1\cdots u_n \in \mathbb{N}^n$, with $u_k \neq 1$ and $u_{k+1} = \ldots = u_n = 1$. We also say that the marked leaf $\Sigma_{n,11\cdots 1}$ of $\mathcal{T}_n$ has depth 0. Then the subtree of $\mathcal{T}_n$ spanned by the leaves up to depth k has the same tree shape as $\mathcal{T}_k$, for all $n \geq k$, with edge lengths, whose convergence will establish the following proposition.

Proposition 4.5. *Suppose Assumption (A) holds and $\mathcal{T}_n := \tau_\varnothing^{(n)}$ in the setting of (15), $n \geq 0$. Let $k \in \mathbb{N}$. For $n \geq k$, let $\mathcal{T}_n^k$ be the subtree of $\mathcal{T}_n$ spanned by the root and the leaves up to depth k. We consider $\Sigma_{n,11\cdots 1}$ as the respective marked point. Then there is an increasing sequence $(\mathcal{T}^k, k \geq 0)$ of marked trees such that, for all $k \geq 0$,*

$$\mathcal{T}_n^k \to \mathcal{T}^k \text{ in probability as } n \to \infty$$

in the marked Gromov–Hausdorff topology.

Proof. For $k = 0$, $\mathcal{T}_n^0$ is a trivial one-branch tree with a root and a marked leaf, and total length

$$\widetilde{L}_n^0 = \sum_{\mathbf{u}\in\{0,1\}^n}\overline{\xi}_{\mathbf{u}}^{\beta}Y_{\mathbf{u}}.$$

Recall the martingale $(L_n, n \geq 0)$ from (17) and denote its limit by L_∞. Let $m := \mathbb{E}[Y_\varnothing]$, and note that,

$$\begin{aligned}\mathbb{E}\left[\left(\widetilde{L}_n^0 - mL_n\right)^2\right] &= \mathbb{E}\left[\left(\sum_{\mathbf{u}\in\{0,1\}^n}\overline{\xi}_{\mathbf{u}}^{\beta}(Y_{\mathbf{u}}-m)\right)^2\right]\\ &= \sum_{\mathbf{u}\in\{0,1\}^n}\sum_{\mathbf{v}\in\{0,1\}^n}\mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{\beta}\overline{\xi}_{\mathbf{v}}^{\beta}\right]\mathbb{E}\left[(Y_{\mathbf{u}}-m)(Y_{\mathbf{v}}-m)\right]\\ &= \sum_{\mathbf{u}\in\{0,1\}^n}\mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{2\beta}\right]\mathbb{E}\left[(Y_{\mathbf{u}}-m)^2\right] = \mathrm{Var}(Y_\varnothing)\left(\mathbb{E}\left[\xi_0^{2\beta}+\xi_1^{2\beta}\right]\right)^n \to 0\end{aligned}$$

as $n \to \infty$, where we used the facts that $Y_{\mathbf{u}}$ and $Y_{\mathbf{v}}$ are independent for $\mathbf{u} \neq \mathbf{v}$, and $\mathbb{E}[\xi_0^{2\beta}+\xi_1^{2\beta}] < 1$ as $0 < \xi_0, \xi_1 < 1$ a.s.. Therefore, $\widetilde{L}_n^0 \to \widetilde{L}_\infty^0 := m \cdot L_\infty$ in $\mathcal{L}^2$ and almost surely as $n \to \infty$.

Under Assumption (A), the $Y_{\mathbf{u}}$ also have finite p-th moment for all $p \geq 3$ and splitting p-fold sums as in the proof of Lemma 4.4, it is straightforward to strengthen this convergence to $\mathcal{L}^p$-convergence.

Now, let $k \geq 1$, and note that the shapes of $\mathcal{T}_n^k$ and $\mathcal{T}_k$ coincide for all $n \geq k$. Let $\widetilde{L}_{n,\mathbf{u}}^k, \mathbf{u} \in \mathbb{N}^k$, denote the lengths of the edges of $\mathcal{T}_n^k$ using obvious notation, i.e.

$$\widetilde{L}_{n,\mathbf{u}}^k := \sum_{\mathbf{v} \in \{0,1\}^{n-k}} \overline{\xi}_{\mathbf{uv}}^{\beta} Y_{\mathbf{uv}}, \quad \mathbf{u} \in \mathbb{N}^k.$$

Furthermore, let $\mathcal{T}^k$ have the same shape and the same marked leaf as $\mathcal{T}_k$ with edge lengths $\widetilde{L}_{\infty,\mathbf{u}}^k, \mathbf{u} \in \mathbb{N}^k$, given by

$$\widetilde{L}_{\infty,\mathbf{u}}^k = \lim_{t \to \infty} \sum_{\mathbf{v} \in \{0,1\}^t} \overline{\xi}_{\mathbf{uv}}^{\beta} Y_{\mathbf{uv}}, \quad \mathbf{u} \in \mathbb{N}^k,$$

which exists a.s. as a $\overline{\xi}_{\mathbf{u}}^{\beta}$-scaled copy of $\widetilde{L}_{\infty}^0$, independent for $\mathbf{u} \in \mathbb{N}^k$.

Hence, for each $k \geq 0$, the differences $\left|\widetilde{L}_{n,\mathbf{u}}^k - \widetilde{L}_{\infty,\mathbf{u}}^k\right|$, $\mathbf{u} \in \mathbb{N}^k$, are $\overline{\xi}_{\mathbf{u}}^{\beta}$-scaled independent copies of $|\widetilde{L}_n^0 - \widetilde{L}_{\infty}^0|$. Therefore, for $p \geq 1/\beta$, as every leaf of $\mathcal{T}_n^k$ or $\mathcal{T}^k$ is at most 2^k edges from the root and from another leaf,

$$\begin{aligned}\mathbb{E}\left[\left(d_{\mathrm{GH}}^{\mathrm{m}}\left(\mathcal{T}_n^k, \mathcal{T}^k\right)\right)^p\right] &\leq 2^{pk}\mathbb{E}\left[\max_{\mathbf{u} \in \mathbb{N}^k}\left|\widetilde{L}_{n,\mathbf{u}}^k - \widetilde{L}_{\infty,\mathbf{u}}^k\right|^p\right] \leq 2^{pk} \sum_{\mathbf{u} \in \mathbb{N}^k} \mathbb{E}\left[\left|\widetilde{L}_{n,\mathbf{u}}^k - \widetilde{L}_{\infty,\mathbf{u}}^k\right|^p\right] \\ &= 2^{pk} \sum_{\mathbf{u} \in \mathbb{N}^k} \mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{p\beta}\right] \mathbb{E}\left[\left|\widetilde{L}_n^0 - \widetilde{L}_{\infty}^0\right|^p\right].\end{aligned}$$

Since $\sum_{\mathbf{u} \in \mathbb{N}^k} \mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{p\beta}\right] < \infty$ for $p \geq 1/\beta$ and $\widetilde{L}_n^0 \to \widetilde{L}_{\infty}^0$ in $\mathcal{L}^p$ as $n \to \infty$, we conclude that, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(d_{\mathrm{GH}}^{\mathrm{m}}\left(\mathcal{T}_n^k, \mathcal{T}^k\right) > \epsilon\right) \leq \lim_{n \to \infty} \epsilon^{-p}\mathbb{E}\left[\left(d_{\mathrm{GH}}^{\mathrm{m}}\left(\mathcal{T}_n^k, \mathcal{T}^k\right)\right)^p\right] = 0.$$

Hence, $\mathcal{T}_n^k \to \mathcal{T}^k$ in probability in the marked Gromov–Hausdorff topology as $n \to \infty$. □

4.5. An RDE on $\mathbb{T}$ and associated constructions in $\mathbb{T}_{\mathrm{w}}$ of [24]

In [24], two of us established a recursive construction method for CRTs by successively replacing the atoms of a random *string of beads*, that is, a random interval $[0, L]$ for some $L > 0$ equipped with a random discrete probability measure μ, with scaled independent copies of itself. More general versions of the CRT construction using so-called *generalised strings* were established to capture multifurcating self-similar CRTs. Let us briefly recap this construction here, and refer to [24] for more details.

Strings of beads can be represented in the form $([0,\ell], (x_i)_{i \in I}, (q_i)_{i \in I})$, where $\ell > 0$ denotes the length of the interval, and $x_i \in [0,\ell]$, $i \in I$, are distinct and describe the locations of the atoms with respective masses $q_i \geq 0$, $i \in I$, of a discrete measure $\mu = \sum_{i \in I} q_i \delta_{x_i}$ on $[0,\ell]$, where $\sum_{i \in I} q_i = 1$, where I is some countable index set. In this representation, the concept of a string of beads is naturally generalised by allowing for non-distinct x_i's. We call such $([0,\ell], (x_i)_{i \in I}, (q_i)_{i \in I})$ a *generalised string*. More generally, we write $(\mathcal{T}, (x_i)_{i \in I}, (q_i)_{i \in I})$ for an $\mathbb{R}$-tree $(\mathcal{T}, d)$, $x_i \in \mathcal{T}$, $i \in I$, and $q_i \geq 0$ with $\sum_{i \in I} q_i = 1$. The following theorem is a (slightly simplified) version of the main result in [24].

Theorem 4.6. *Let $\beta \in (0,\infty)$ and $p > 1/\beta$. Let $\zeta = (\check{\mathcal{T}}_0, (\check{X}_i^{(0)})_{i\in I_0}, (\check{Q}_i^{(0)})_{i\in I_0})$ be a random generalised string with length $L > 0$ such that $\mathbb{E}[L^p] < \infty$, and atom masses $0 \le \check{Q}_i^{(0)} < 1$ a.s. for all $i \in I_0$ and such that $\sum_{i\in I_0} \check{Q}_i^{(0)} = 1$ a.s.. For $n \ge 0$, the $(n+1)$st $\mathbb{R}$-tree with a measure represented by atom locations and masses,*

$$\left(\check{\mathcal{T}}_{n+1}, \left(\check{X}_i^{(n+1)}\right)_{i\in I_{n+1}}, \left(\check{Q}_i^{(n+1)}\right)_{i\in I_{n+1}}\right),$$

conditionally given $(\check{\mathcal{T}}_n, (\check{X}_i^{(n)})_{i\in I_n}, (\check{Q}_i^{(n)})_{i\in I_n})$, is obtained by attaching to each $\check{X}_i^{(n)} \in \check{\mathcal{T}}_n$ an independent isometric copy of ζ with metric rescaled by $(\check{Q}_i^{(n)})^\beta$ and atom masses rescaled by $\check{Q}_i^{(n)}$, $i \in I_n$.

Let $\check{\mu}_n = \sum_{i\in I_n} \check{Q}_i^{(n)} \delta_{\check{X}_i^{(n)}}, n \ge 0$. Then there exists a random weighted $\mathbb{R}$-tree $(\check{\mathcal{T}}, \check{\mu})$ such that

$$\lim_{n\to\infty} \left(\check{\mathcal{T}}_n, \check{\mu}_n\right) = \left(\check{\mathcal{T}}, \check{\mu}\right) \qquad a.s.$$

in the Gromov–Hausdorff–Prokhorov topology in $\mathbb{T}_{\mathrm{w}}$. Furthermore, $\mathbb{E}[\mathrm{ht}(\check{\mathcal{T}})^p] < \infty$ for all $p < p^* := \sup\{p \ge 1 : \mathbb{E}[L^p] < \infty\}$.

The convergence in Theorem 4.6 holds in particular in the Gromov–Hausdorff sense when we omit mass measures. In fact, this construction is naturally carried out in the Banach space $\ell_1(\mathbb{U})$, $\mathbb{U} := \bigcup_{n\ge 0} \mathbb{N}^n$, which is a variant of Aldous's $\ell_1(\mathbb{N})$ since $\mathbb{U}$ is countable. So embedded, the convergence holds with respect to the Hausdorff metric (or a Hausdorff–Prokhorov metric) for compact subsets (equipped with a probability measure) of $\ell_1(\mathbb{U})$, as a consequence of the arguments of [24]. In particular, the α-stable tree was characterised as the limit in the case of a *β-generalised string* for $\beta = 1 - 1/\alpha \in (0, 1/2]$, that is, a generalised string of the form $\big([0,L], (X_i)_{i\ge 1}, (P_i)_{i\ge 1}\big)$ where, for $(Q_m, m \ge 1) \sim \mathrm{PD}(\beta,\beta)$ independent of i.i.d. $(R_j^{(m)}, j \ge 1) \sim \mathrm{PD}(1-\beta, -\beta)$, $m \ge 1$, the atom sizes are given via

$$(P_i, i \ge 1) = \left(Q_m R_j^{(m)}, j \ge 1, m \ge 1\right)^{\downarrow},$$

and the atom locations are defined via i.i.d. Unif $([0,1])$-variables $(U_m, m \ge 1)$ and

$$L := \lim_{m\to\infty} m\Gamma(1-\beta) Q_m^\beta, \quad X_i = LU_m \text{ if } P_i = Q_m R_j^{(m)}, \qquad i,j \ge 1.$$

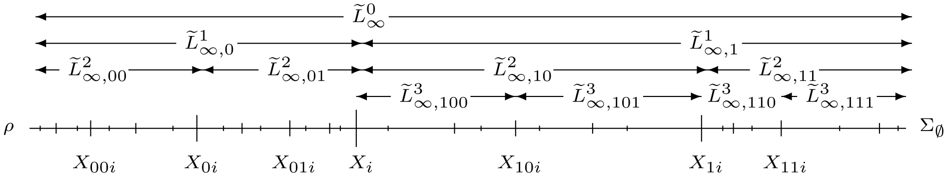

Figure 5. The dyadic structure of limiting branch lengths $\widetilde{L}^k_{\infty,\mathbf{u}} = \widetilde{L}^{k+1}_{\infty,\mathbf{u}0} + \widetilde{L}^{k+1}_{\infty,\mathbf{u}1}$, ordered in lexicographical order; atom positions $X_{\mathbf{u}i}$, not depending on $i \ge 2$, are between fragments $\widetilde{L}^n_{\infty,\mathbf{u}0}$ and $\widetilde{L}^n_{\infty,\mathbf{u}1}$, $\mathbf{u} \in \{0,1\}^n$, $n \ge 0$.

4.6. The CRT limit of $\mathcal{T}^k$ as $k \to \infty$

Next, we want to prove the convergence of the limits $\mathcal{T}^k$ of Proposition 4.5 as $k \to \infty$. We first identify a suitable candidate for the limit. We employ the recursive construction method for CRTs as described in Section 4.5. Let us define a generalised string that will capture subtree masses on the spine $\mathcal{T}^0$ of length $\widetilde{L}^0_\infty$. Viewed as a subset of $\mathcal{T}^k$, this spine consists of 2^k spinal edges. In the notation of the proof of Proposition 4.5, they have lengths $\widetilde{L}^k_{\infty,\mathbf{u}}$, $\mathbf{u} \in \{0,1\}^k$, and their order from the root to the marked leaf is the lexicographical order (see Figure 5), that is

$$\mathbf{v} = v_1 \dots v_k \prec u_1 \dots u_k = \mathbf{u} \iff \exists t \in \{1,\dots,k\} \text{ such that } \forall j < t\colon v_j = u_j \text{ and } v_t < u_t.$$

In view of the concatenation operation g_β, subtrees and subtree masses are naturally parametrised by $\mathbb{U}^* := \bigcup_{k\geq 0}\{0,1\}^k \times \{2,3,\dots\}$. We set

$$\zeta = \left(\left[0, \widetilde{L}^0_\infty\right], (X_\mathbf{u})_{\mathbf{u}\in\mathbb{U}^*}, (Q_\mathbf{u})_{\mathbf{u}\in\mathbb{U}^*}\right) \tag{18}$$

where $(Q_\mathbf{u})_{\mathbf{u}\in\mathbb{U}^*}$ and $(X_\mathbf{u})_{\mathbf{u}\in\mathbb{U}^*}$ are defined by recursively splitting mass 1 and length $\widetilde{L}^0_\infty$, as follows.

- Let $Q_i := \xi_i$, $i \geq 2$, and, for $\mathbf{u} = u_1 \dots u_n \in \mathbb{U}^*$, define masses

$$Q_\mathbf{u} := \xi_{u_1}\xi_{u_1u_2}\cdots\xi_{u_1u_2\dots u_n} = \overline{\xi}_{u_1u_2\dots u_n}.$$

  Note that $0 \leq Q_\mathbf{u} < 1$ a.s. for all $\mathbf{u} \in \mathbb{U}^*$, $\sum_{\mathbf{u}\in\mathbb{U}^*} Q_\mathbf{u} = 1$ a.s., and $\mathbb{E}\left[\sum_{\mathbf{u}\in\mathbb{U}^*} Q_\mathbf{u}^{p\beta}\right] < 1$ for $p > 1/\beta$.
- Define the locations $(X_\mathbf{u})_{\mathbf{u}\in\mathbb{U}^*}$ of the atoms with respective masses $(Q_\mathbf{u})_{\mathbf{u}\in\mathbb{U}^*}$ by $X_i = \widetilde{L}^1_{\infty,0}$, $i \geq 2$, and, for general $\mathbf{u} = u_1 \dots u_n \in \mathbb{U}^*$, the following sum of up to $2^n - 1$ edge lengths

$$X_\mathbf{u} = \sum_{\substack{v_1\dots v_n \in \{0,1\}^n\colon v_1\dots v_n \prec u_1\dots u_n \\ v_1\dots v_{n-1}v_n \neq u_1\dots u_{n-1}1}} \widetilde{L}^n_{\infty,v_1\dots v_n}. \tag{19}$$

Noting in particular that this specifies $X_{u_1\dots u_{n-1}i} = X_\mathbf{u}$ for all $i \geq 2$ and each $\mathbf{u} = u_1 \dots u_n \in \mathbb{U}^*$, the scaled lengths and dyadic splits to depth $k = 3$ are illustrated in Figure 5.

We now apply the recursive construction as outlined in Theorem 4.6 to the generalised string ζ, which results in an $\mathbb{R}$-tree $\mathcal{T}$, whose distribution we denote by η^*.

Proposition 4.7. *Let $\beta \in (0,1]$, and $p > 1/\beta$. Consider the generalised string ζ given by (18). Consider a sequence of random weighted $\mathbb{R}$-trees obtained inductively by the construction of Theorem 4.6, which we denote here by $(\mathcal{T}^*_n, \mu^*_n)$, $n \geq 0$. Then $(\mathcal{T}^*_n, \mu^*_n) \to (\mathcal{T}, \mu)$ a.s. in the Gromov–Hausdorff–Prokhorov topology for some random weighted compact $\mathbb{R}$-tree $(\mathcal{T}, \mu)$ with $\mathbb{E}[\mathrm{ht}(\mathcal{T})^p] < \infty$ for all $p > 0$.*

It will be convenient to refer to the two endpoints of the rescaled generalised strings attached in the construction of $\mathcal{T}$ as $\rho_\mathbf{u}$ and $\Sigma_\mathbf{u}$, $\mathbf{u} \in \mathbb{U}$. Specifically, we denote by $\rho = \rho_\varnothing$ and $\Sigma_\varnothing$ the endpoints of $\mathcal{T}^*_0 \subset \mathcal{T}$. The first step of the construction performed in Proposition 4.7 consists of attaching to $\mathcal{T}^*_0$ rescaled independent copies of the generalised string ζ to $\mathcal{T}^*_0$, one for each $u \in \mathbb{U}^*$. In particular, we have $\Sigma_\mathbf{u} \in \mathcal{T}^*_1$ for all $\mathbf{u} \in \mathbb{U}^*$. The way we specified ζ in terms of $(\xi_\mathbf{u})_{\mathbf{u}\in\mathbb{U}^*}$, removing Σ_0 from $(\mathcal{T}^*_1, \mu^*_1)$ or $(\mathcal{T}, \mu)$ yields a mass split $(\xi_i, i \geq 0)$. For any $\mathbf{u} \in \mathbb{U}$, we have $\Sigma_\mathbf{u} \in \mathcal{T}^*_n \setminus \mathcal{T}^*_{n-1}$ if and only if $\mathbf{u} = \mathbf{u}^{(1)}i_1 \dots \mathbf{u}^{(n)}i_n\mathbf{u}^{(n+1)}$ for some $\mathbf{u}^{(j)} \in \bigcup_{k\geq 0}\{0,1\}^k$, $1 \leq j \leq n+1$, and $i_j \in \{2,3,\dots\}$, $1 \leq j \leq n$.

We will now couple the vectors $(\xi_{\mathbf{v}i}, i \geq 0)$, $\mathbf{v} \in \mathbb{U}$, of the construction of $(\mathcal{T}_n, n \geq 0)$ and $(\mathcal{T}^k, k \geq 0)$ in the setting of Proposition 4.5, and the generalised strings $\zeta_\mathbf{u}$, $\mathbf{u} \in \mathbb{U}$, in the construction of Proposition

4.7. Specifically, the recursive nature of both constructions ensures that we can achieve that removing $\rho_{\mathbf{u}} = \rho_{\mathbf{u}0}$, $\Sigma_{\mathbf{u}0} = \rho_{\mathbf{u}1}$ and $\Sigma_{\mathbf{u}1} = \Sigma_{\mathbf{u}}$ from $\mathcal{T}$, or indeed from $\mathcal{T}_n^*$ for n sufficiently large, yields a relative mass split $(\xi_{\mathbf{u}i}, i \geq 0)$ for the components adjacent to $\Sigma_{\mathbf{u}0}$, and these splits are i.i.d., $\mathbf{u} \in \mathbb{U}$.

We can represent $\mathcal{T}$ and $\mathcal{T}_n$, $n \geq 0$, in $\ell_1(\mathbb{U})$ in such a way that the convergences of Proposition 4.5 hold for the Hausdorff metric on compact subsets of $\ell_1(\mathbb{U})$. The arguments in the proof of Proposition 4.5 ensure that we have further a.s. convergence of $\Sigma_{n,\mathbf{u}}$ to limits that we denote by $\Sigma_{\mathbf{u}}$, for all $\mathbf{u} \in \mathbb{U}$. Then the trees $\mathcal{T}^k$ are spanned by $\Sigma_{\mathbf{u}}$, $\mathbf{u} \in \bigcup_{0 \leq j \leq k} \mathbb{N}^j$, while $\mathcal{T}_n^*$ is spanned by $\Sigma_{\mathbf{u}}$, $\mathbf{u} = \mathbf{u}^{(1)} v_1 \mathbf{u}^{(2)} v_2 \cdots \mathbf{u}^{(n)} v_n \mathbf{u}^{(n+1)}$, $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(n+1)} \in \bigcup_{m \geq 0} \{0,1\}^m$, $v_1, \ldots, v_n \in \mathbb{N}$.

Lemma 4.8. *Let $(\mathcal{T}^k, k \geq 0)$ be the sequence of trees from Proposition 4.5, and let $(\mathcal{T}_n^*, n \geq 0)$ be the sequence of trees from Proposition 4.7 with $\mathcal{T}_n^* \to \mathcal{T}$ a.s. as $n \to \infty$. Then $\mathcal{T}^k \to \mathcal{T}$ a.s. as $k \to \infty$ in the marked Gromov–Hausdorff topology.*

Proof. Since the sequence of trees $(\mathcal{T}^k, k \geq 0)$ is increasing and embedded in $\mathcal{T}$ with the same marked point, it remains to show that the almost sure limit of $\mathcal{T}^k$ is the whole of $\mathcal{T}$.

Let $(\mathcal{T}_{\mathbf{u}j}, j \geq 2)$, $\mathbf{u} \in \bigcup_{t=0}^\infty \mathbb{N}^k \times \{0,1\}^t$, denote the connected components of $\mathcal{T} \setminus \mathcal{T}^k$, $k \geq 0$, where we write $(\mathcal{T}_{u_1 \ldots u_n j}, j \geq 2)$ for the subtrees of $\mathcal{T} \setminus \mathcal{T}^k$ rooted at the edge of $\mathcal{T}_k$ of length $\widetilde{L}^k_{\infty, u_1 \ldots u_k}$, $n \geq k$, using notation from Proposition 4.5. By the recursive construction, each $\mathcal{T}_{\mathbf{u}j}$ is a $\overline{\xi}_{\mathbf{u}j}$-scaled independent copy of $\mathcal{T}$, and we obtain for $k \geq 0$ and $p > 1/\beta$,

$$
\begin{aligned}
\mathbb{E}\left[\left(d_{\mathrm{GH}}^{\mathrm{m}}\left(\mathcal{T}^k, \mathcal{T}\right)\right)^p\right] &\leq \mathbb{E}\left[\left(\max_{\mathbf{u} \in \bigcup_{t=0}^\infty \mathbb{N}^k \times \{0,1\}^t, j \geq 2} \mathrm{ht}\left(\mathcal{T}_{\mathbf{u}j}\right)\right)^p\right] \\
&\leq \mathbb{E}[(\mathrm{ht}(\mathcal{T}))^p] \sum_{\mathbf{u} \in \bigcup_{t=0}^\infty \mathbb{N}^k \times \{0,1\}^t, j \geq 2} \mathbb{E}\left[\overline{\xi}_{\mathbf{u}j}^{p\beta}\right] \leq \mathbb{E}[(\mathrm{ht}(\mathcal{T}))^p] \sum_{\mathbf{u} \in \bigcup_{t=0}^\infty \mathbb{N}^k \times \{0,1\}^t} \mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{p\beta}\right] \\
&\leq \mathbb{E}[(\mathrm{ht}(\mathcal{T}))^p] \left(\mathbb{E}\left[\sum_{j \geq 0} \xi_j^{p\beta}\right]\right)^k \mathbb{E}\left[\sum_{t=0}^\infty \left(\xi_0^{p\beta} + \xi_1^{p\beta}\right)^t\right] \to 0 \qquad \text{as } k \to \infty
\end{aligned}
$$

since $\mathbb{E}[\mathrm{ht}(\mathcal{T})^p] < \infty$ and $\mathbb{E}\left[\sum_{j \geq 0} \xi_j^{p\beta}\right] < 1$. Hence, for any $\epsilon > 0$ and $p > 1/\beta$,

$$
\mathbb{P}\left(d_{\mathrm{GH}}^{\mathrm{m}}\left(\mathcal{T}^k, \mathcal{T}\right) > \epsilon\right) \leq \epsilon^{-p} \mathbb{E}\left[\left(d_{\mathrm{GH}}^{\mathrm{m}}\left(\mathcal{T}^k, \mathcal{T}\right)\right)^p\right] \to 0
$$

as $k \to \infty$. Therefore, due to the embedding of $(\mathcal{T}^k, k \geq 0)$ into $\mathcal{T}$, $\mathcal{T}^k \to \mathcal{T}$ a.s. as $k \to \infty$. □

4.7. Attraction of the RDE and the proof of Theorem 4.2

Theorem 4.2 claims a general convergence to a unique fixpoint. Under Assumption (A) of initial distributions concentrated on trivial trees, Proposition 4.5 and Lemma 4.8 establish a two-step convergence as first $n \to \infty$ for subtrees of $\mathcal{T}_n$ spanned by leaves up to depth $k \geq 1$ and then $k \to \infty$. To complete the proof of Theorem 4.2, we need a tightness result, and we need to lift Assumption (A). For the former, we show that the supremum of the height moments of $\mathcal{T}_n$ is finite, employing the recursive construction of CRTs for a generalised string defined in a similar manner as in the discussion before Proposition 4.7.

Lemma 4.9. *Under Assumption (A), the sequence of trees $(\mathcal{T}_n, n \geq 0)$ satisfies*

$$\mathbb{E}\left[\sup_{n\geq 0} \mathrm{ht}(\mathcal{T}_n)^p\right] < \infty \text{ for all } p > 0. \tag{20}$$

Proof. The idea of the proof is to construct a CRT $\widehat{\mathcal{T}}$ whose height dominates $\mathrm{ht}(\mathcal{T}_n)$ for all $n \geq 0$. Indeed, we apply the recursive construction of CRTs (cf. the construction of $\mathcal{T}$ in Section 4.6) to the generalised string $\widehat{\zeta}$ obtained by modifying the definition of ζ in (18). Specifically, note that we can write the locations on the string in (19) as

$$X_{\mathbf{u}} = \lim_{t\to\infty} \left\{ \sum_{\substack{v_1\ldots v_k \in \{0,1\}^k :\ v_1\ldots v_k \prec u_1\ldots u_k \\ v_1\ldots v_{k-1} v_k \neq u_1\ldots u_{k-1} 1}} \ \sum_{v_{k+1}\ldots v_{k+t} \in \{0,1\}^t} \overline{\xi}^{\beta}_{v_1\ldots v_{k+t}} Y_{v_1\ldots v_{k+t}} \right\}.$$

To define $\widehat{X}_{\mathbf{u}}$, we replace $\lim_{t\to\infty}$ by $\sup_{t\geq 0}$. In the same way, we define the length of the string as $\widehat{L}^0_\infty := \sup_{t\geq 0} \sum_{\mathbf{u}\in\{0,1\}^t} \overline{\xi}^{\beta}_{\mathbf{u}} Y_{\mathbf{u}}$. If $Y_{\mathbf{u}} = m$ a.s., we have $\mathbb{E}[(\widehat{L}^0_\infty)^p] < \infty$ by Lemma 4.4 and Doob's L^p-inequality. In the general case, we split $Y_{\mathbf{u}} = m + (Y_{\mathbf{u}} - m)$ and argue as in the proof of Proposition 4.5 (and the proof of Lemma 4.4) to see that $\mathbb{E}[(\widehat{L}^0_\infty)^p] < \infty$ for all $p > 0$.

This ensures that each atom on $\widehat{\zeta}$ is placed at the furthest position away from the left end point which appears in the course of the construction of $\mathcal{T}_n, n \geq 0$. As a consequence, using $\widehat{\zeta}$ instead of ζ in the recursive construction, coupled by being derived derived from the same $(\xi_{\mathbf{u}}, \mathbf{u} \in \mathbb{U})$ and $(Y_{\mathbf{u}}, \mathbf{u} \in \mathbb{U})$, all distances from leaves and branch points to the root will be larger than in any of the trees $\mathcal{T}_n, n \geq 0$.

Applying Theorem 4.6 to the generalised string $\widehat{\zeta}$, we obtain a CRT $\widehat{\mathcal{T}}$ which has finite height moments of all orders. By the underlying coupling, $\mathrm{ht}(\mathcal{T}_n) \leq \mathrm{ht}(\widehat{\mathcal{T}})$ for all $n \geq 0$, i.e., the claim follows. ☐

Corollary 4.10. *Consider the sequences of trees $(\mathcal{T}_n, n \geq 0)$ and $(\mathcal{T}^k_n, n \geq k)$, $k \geq 0$, where we recall that, for $n \geq k$, $\mathcal{T}^k_n$ is the subtree of $\mathcal{T}_n$ spanned by the root and the leaves up to depth k. Then, for any $\epsilon > 0$,*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d^{\mathrm{m}}_{\mathrm{GH}}\left(\mathcal{T}^k_n, \mathcal{T}_n\right) > \epsilon\right) = 0. \tag{21}$$

Proof. Let $\mathcal{T}^k_{n,\mathbf{u}} \setminus \{\rho^k_{n,\mathbf{u}}\}, \mathbf{u} \in \bigcup_{t=0}^{n-k-1} \mathbb{N}^k \times \{0,1\}^t \times \{2,3,\ldots\}$, denote the subtrees of $\mathcal{T}_n \setminus \mathcal{T}^k_n$, $n \geq k+1$:

$$\mathcal{T}_n \setminus \mathcal{T}^k_n = \bigcup_{\mathbf{u}\in\bigcup_{t=0}^{n-k-1} \mathbb{N}^k\times\{0,1\}^t\times\{2,3,\ldots\}} \mathcal{T}^k_{n,\mathbf{u}} \setminus \{\rho^k_{n,\mathbf{u}}\}.$$

Then, for any $\epsilon > 0$ and $p > 1/\beta$,

$$\begin{aligned}\mathbb{P}\left(d^{\mathrm{m}}_{\mathrm{GH}}\left(\mathcal{T}^k_n, \mathcal{T}_n\right) > \epsilon\right) &\leq \epsilon^{-p}\mathbb{E}\left[\max_{\mathbf{u}\in\bigcup_{t=0}^{n-k} \mathbb{N}^k\times\{0,1\}^t\times\{2,3,\ldots\}} \mathrm{ht}\left(\mathcal{T}^k_{n,\mathbf{u}}\right)^p\right] \\ &\leq \epsilon^{-p} \sum_{\mathbf{u}\in\bigcup_{t=0}^{n-k-1} \mathbb{N}^k\times\{0,1\}^t\times\{2,3,\ldots\}} \mathbb{E}\left[\overline{\xi}^{p\beta}_{\mathbf{u}}\right] \mathbb{E}\left[\mathrm{ht}\left(\mathcal{T}_{n-|\mathbf{u}|}\right)^p\right].\end{aligned}$$

By Lemma 4.9, it remains to show that

$$\lim_{k\to\infty} \limsup_{n\to\infty} \sum_{\mathbf{u}\in\bigcup_{t=0}^{n-k-1} \mathbb{N}^k\times\{0,1\}^t\times\{2,3,\ldots\}} \mathbb{E}\left[\overline{\xi}^{p\beta}_{\mathbf{u}}\right] = 0, \tag{22}$$

First, note that the left-hand side of (22) is bounded above by

$$\lim_{k\to\infty}\sup_{n\geq k+1}\sum_{\mathbf{u}\in\mathbb{N}^k}\sum_{t=0}^{n-k-1}\sum_{\mathbf{v}\in\{0,1\}^t\times\{2,3,\dots\}}\mathbb{E}\left[\overline{\xi}_{\mathbf{uv}}^{p\beta}\right], \tag{23}$$

where we also slightly rewrote the expression. By the fact that $(\xi_{\mathbf{u}j}, j\geq 0), \mathbf{u}\in\mathbb{U}$, are i.i.d., we have

$$\mathbb{E}\left[\overline{\xi}_{\mathbf{uv}}^{p\beta}\right]=\mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{p\beta}\right]\mathbb{E}\left[\overline{\xi}_{\mathbf{v}}^{p\beta}\right]\leq\mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{p\beta}\right]\mathbb{E}\left[\overline{\xi}_{\mathbf{v}}\right],$$

where we used $\overline{\xi}_{\mathbf{v}}<1$ a.s. and $p\beta>1$ in the last inequality.

Furthermore, as $\sum_{j\geq 0}\xi_{\mathbf{v}j}=1$,

$$\sum_{t=0}^{n-k-1}\sum_{\mathbf{v}\in\{0,1\}^t\times\{2,3,\dots\}}\mathbb{E}\left[\overline{\xi}_{\mathbf{v}}\right]\leq\sum_{t=0}^{n-k-1}(\mathbb{E}[\xi_0+\xi_1])^t\leq\sum_{t=0}^{\infty}(\mathbb{E}[\xi_0+\xi_1])^t=(1-\mathbb{E}[\xi_0+\xi_1])^{-1}$$

where we also used the i.i.d. property of the $(\xi_{\mathbf{v}j}, j\geq 0)$, $\mathbf{v}\in\bigcup_{t=0}^{n-k-1}\{0,1\}^t$, and $\mathbb{E}[\xi_0+\xi_1]<1$. Hence, (23) can be further bounded above by

$$(1-\mathbb{E}[\xi_0+\xi_1])^{-1}\lim_{k\to\infty}\sum_{\mathbf{u}\in\mathbb{N}^k}\mathbb{E}\left[\overline{\xi}_{\mathbf{u}}^{p\beta}\right]=(1-\mathbb{E}[\xi_0+\xi_1])^{-1}\lim_{k\to\infty}\left(\mathbb{E}\left[\sum_{i\geq 0}\xi_i^{p\beta}\right]\right)^k. \tag{24}$$

As $p\beta>1$ and $0\leq\xi_i<1$ a.s. for all $i\geq 0$, $\mathbb{E}\left[\sum_{i\geq 0}\xi_i^{p\beta}\right]<1$, and we conclude that (24) is 0. □

We are now ready to prove our final result.

Corollary 4.11. *Under Assumption (A), let $(\mathcal{T}_n, n\geq 0)$ be as above, and let $\mathcal{T}$ be the tree from Proposition 4.7. We have the convergence*

$$\mathcal{T}_n\to\mathcal{T}\text{ in probability as } n\to\infty$$

in the marked Gromov–Hausdorff topology.

Proof. Let $\epsilon>0$, and use the triangle inequality twice to get, for $n\in\mathbb{N}$ and $k\leq n$,

$$\mathbb{P}(d_{\mathrm{GH}}^{\mathrm{m}}(\mathcal{T}_n,\mathcal{T})>3\epsilon)\leq\mathbb{P}(d_{\mathrm{GH}}^{\mathrm{m}}(\mathcal{T}_n,\mathcal{T}_n^k)>\epsilon)+\mathbb{P}(d_{\mathrm{GH}}^{\mathrm{m}}(\mathcal{T}_n^k,\mathcal{T}^k)>\epsilon)+\mathbb{P}(d_{\mathrm{GH}}^{\mathrm{m}}(\mathcal{T}^k,\mathcal{T})>\epsilon).$$

All three terms converge to 0 as $n\to\infty$, and then $k\to\infty$, cf. Proposition 4.5, Lemma 4.8 and Corollary 4.10. □

Theorem 4.2 is now a direct consequence of Corollary 4.11.

Proof of Theorem 4.2. Let $\eta\in\mathcal{P}_\infty(\mathbb{T}_m)$ be a general distribution of a marked $\mathbb{R}$-tree. For a marked $\mathbb{R}$-tree $(\mathcal{T}_0, d_0, \rho_0, x_0)\sim\eta$, we define the induced distribution $\eta^\circ\in\mathcal{P}_\infty(\mathbb{T}_{\mathrm{m}}^{\mathrm{tr}})$ as the distribution of $[\![\rho_0, x_0]\!]$. We construct coupled sequences $(\mathcal{T}_n, n\geq 0)$ and $(\mathcal{T}_n^\circ, n\geq 0)$ from the same recursive tree framework $((\xi_{\mathbf{u}i}, i\geq 0), \mathbf{u}\in\mathbb{U})$ and from coupled systems of i.i.d. η- and η°-distributed trees, according to (15), with $\mathcal{T}_0\sim\eta$ and $\mathcal{T}_0^\circ=[\![\rho_0, x_0]\!]\sim\eta^\circ$. Then $\mathcal{T}_0\setminus\mathcal{T}_0^\circ$ consists of subtrees of heights bounded by $\mathrm{ht}(\mathcal{T}_0)$. By construction, $\mathcal{T}_n \setminus \mathcal{T}_n^\circ$ consists of subtrees of heights bounded by the maximum of $\overline{\xi}_{\mathbf{u}}$-scaled independent copies of $\mathrm{ht}(\mathcal{T}_0)$. Hence,

$$\mathbb{E}\left[\left(d_{\mathrm{GH}}^{\mathrm{m}}(\mathcal{T}_n, \mathcal{T}_n^\circ)\right)^p\right] \le \mathbb{E}\left[(\mathrm{ht}(\mathcal{T}_0))^p\right]\left(\mathbb{E}\left[\sum_{j\ge 0} \xi_j^{p\beta}\right]\right)^n \to 0,$$

as $n \to \infty$. By Corollary 4.11, we have $\mathcal{T}_n^\circ \to \mathcal{T}$ and hence $\mathcal{T}_n \to \mathcal{T}$ in probability as $n \to \infty$ in the marked Gromov–Hausdorff topology. Uniqueness follows from the attraction property. □

Rights retention

Acknowledgements

We'd like to thank the referee for valuable suggestions that improved the legibility of the paper.

References

[1] Abraham, R., Delmas, J.-F. and Hoscheit, P. (2013). A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.* **18** 14. MR3035742 https://doi.org/10.1214/EJP.v18-2116

[2] Addario-Berry, L., Dieuleveut, D. and Goldschmidt, C. (2019). Inverting the cut-tree transform. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 1349–1376. MR4010938 https://doi.org/10.1214/18-aihp921

[3] Albenque, M. and Goldschmidt, C. (2015). The Brownian continuum random tree as the unique solution to a fixed point equation. *Electron. Commun. Probab.* **20** 61. MR3399812 https://doi.org/10.1214/ECP.v20-4250

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

[5] Aldous, D. (1994). Recursive self-similarity for random trees, random triangulations and Brownian excursion. *Ann. Probab.* **22** 527–545. MR1288122

[6] Aldous, D.J. and Bandyopadhyay, A. (2005). A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* **15** 1047–1110. MR2134098 https://doi.org/10.1214/105051605000000142

[7] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2007). Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35** 1835–1887. MR2349577 https://doi.org/10.1214/009117906000001114

[8] Bertoin, J. (2002). Self-similar fragmentations. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 319–340. MR1899456 https://doi.org/10.1016/S0246-0203(00)01073-6

[9] Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. MR0362614

[10] Broutin, N. and Sulzbach, H. (2021). Self-similar real trees defined as fixed points and their geometric properties. *Electron. J. Probab.* **26** 88. MR4278599 https://doi.org/10.1214/21-ejp647

[11] Chaumont, L. and Pardo, J.C. (2009). On the genealogy of conditioned stable Lévy forests. *ALEA Lat. Am. J. Probab. Math. Stat.* **6** 261–279. MR2534486

[12] Croydon, D. and Hambly, B. (2008). Self-similarity and spectral asymptotics for the continuum random tree. *Stochastic Process. Appl.* **118** 730–754. MR2411518 https://doi.org/10.1016/j.spa.2007.06.005

[13] Curien, N. and Haas, B. (2013). The stable trees are nested. *Probab. Theory Related Fields* **157** 847–883. MR3129805 https://doi.org/10.1007/s00440-012-0472-x

[14] Dieuleveut, D. (2015). The vertex-cut-tree of Galton-Watson trees converging to a stable tree. *Ann. Appl. Probab.* **25** 2215–2262. MR3349006 https://doi.org/10.1214/14-AAP1047

[15] Duquesne, T. and Le Gall, J.-F. (2002). Random trees, Lévy processes and spatial branching processes. *Astérisque* **281**.

[16] Duquesne, T. and Wang, M. (2017). Decomposition of Lévy trees along their diameter. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 539–593. MR3634265 https://doi.org/10.1214/15-AIHP725

[17] Evans, S.N. (2008). *Probability and Real Trees*. *Lecture Notes in Math.* **1920**. Berlin: Springer. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. MR2351587 https://doi.org/10.1007/978-3-540-74798-7

[18] Haas, B., Pitman, J. and Winkel, M. (2009). Spinal partitions and invariance under re-rooting of continuum random trees. *Ann. Probab.* **37** 1381–1411. MR2546748 https://doi.org/10.1214/08-AOP434

[19] Marchal, P. (2008). A note on the fragmentation of a stable tree. In *Fifth Colloquium on Mathematics and Computer Science*. *Discrete Math. Theor. Comput. Sci. Proc., AI* 489–499. Nancy: Assoc. Discrete Math. Theor. Comput. Sci. MR2508809

[20] Miermont, G. (2005). Self-similar fragmentations derived from the stable tree. II. Splitting at nodes. *Probab. Theory Related Fields* **131** 341–375. MR2123249 https://doi.org/10.1007/s00440-004-0373-8

[21] Miermont, G. (2009). Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)* **42** 725–781. MR2571957 https://doi.org/10.24033/asens.2108

[22] Pitman, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Berlin: Springer. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. MR2245368

[23] Pitman, J. and Winkel, M. (2015). Regenerative tree growth: Markovian embedding of fragmenters, bifurcators, and bead splitting processes. *Ann. Probab.* **43** 2611–2646. MR3395470 https://doi.org/10.1214/14-AOP945

[24] Rembart, F. and Winkel, M. (2018). Recursive construction of continuum random trees. *Ann. Probab.* **46** 2715–2748. MR3846837 https://doi.org/10.1214/17-AOP1237

[25] Rösler, U. and Rüschendorf, L. (2001). The contraction method for recursive algorithms. *Algorithmica* **29** 3–33.

Received July 2022 and revised April 2023